# *Sixteenth*
# *Biennial Scientific Report*

*March 2021 – March 2023*

# Contents

# Part I

# Overview – The Institute

# 1 Overview

## 1.1 Organization and Staff

**Directors and Departments:** The Max Planck Institute for Informatics was established in 1990, with Kurt Mehlhorn as founding director. Harald Ganzinger was appointed shortly afterwards; he passed away in 2004. Currently, the Institute has six directors (with appointment periods given in parentheses): Danupon Na Nongkai (2022–2048), Algorithms & Complexity (D1); Bernt Schiele (2010–2035), Computer Vision and Machine Learning (D2); Anja Feldmann (2018–2033), Internet Architecture (D3); Hans-Peter Seidel (1999–2026), Computer Graphics (D4); Gerhard Weikum (2003–2023), Databases and Information Systems (D5); and Christian Theobalt (2021–2043), heading the newly founded department on Visual Computing and Artificial Intelligence (D6). Thomas Lengauer (2001–2018) and Kurt Mehlhorn (1990–2019) have emeritus status; Kurt Mehlhorn continues as scientist in the department on Algorithms and Complexity (D1). In addition to the departments, the Institute has a permanent independent research group on Automation of Logic, headed by Christoph Weidenbach, and an independent research group on Network and Cloud Systems, headed by Yiting Xia. Yiting Xia successfully applied to the tenure-track openings of the Max Planck institutes in computer science, a joint recruitment campaign. Tenure track openings are an additional instrument for recruiting talent that has been installed recently. In total, the Institute currently has 140 scientists, out of which 96 are doctoral students[1] and 44 have a doctoral degree.

**Senior Researchers:** The Institute has five scientific ranks: director, senior researcher with tenure, senior researcher on tenure track, senior researcher, and researcher. Senior researchers with tenure and senior researchers roughly correspond to tenured associate professor and non-tenured assistant professor in the North American system. The process for appointing senior researchers is similar to faculty appointment procedures, and involves reference letters from international top researchers. The appointment committee comprises the directors of the Institute and a faculty member of Saarland University. In addition to the typically non-tenured senior researcher appointment process, we have established a tenure-track senior researcher career path. Together with the MPI for Security and Privacy and the MPI for Software Systems we invite applications in all areas of computer science. In the reporting period, the Institute has been home to 12 senior researchers, in addition to its directors.[2]

Senior researcher positions are not tenure-track, in general. We expect the vast majority of our senior researchers to become professors, ultimately full professors, or leading researchers

[1] Including nine students financed together with Saarland University.

[2] Martin Bromberger, Dengxin Dai, Tobias Fiebig, Oliver Gasser, Andreas Karrenbauer, Karol Myszkowski (tenured), Simon Razniewski, Rishiraj Saha Roy, Paul Swoboda, Christoph Weidenbach (tenured), Philip Wellnitz, Yiting Xia (tenure-track)

in industry. A strong indicator for the success of the model is the high number of faculty positions that our alumni have received at universities all over the world (see Section 1.5).

Figure 1.1 shows the organizational structure and the research areas of the Institute. Each department pursues a number of research areas, and each area has its coordinator(s). The coordinating scientists are senior researchers or postdoctoral researchers with strong potential for becoming senior researchers.

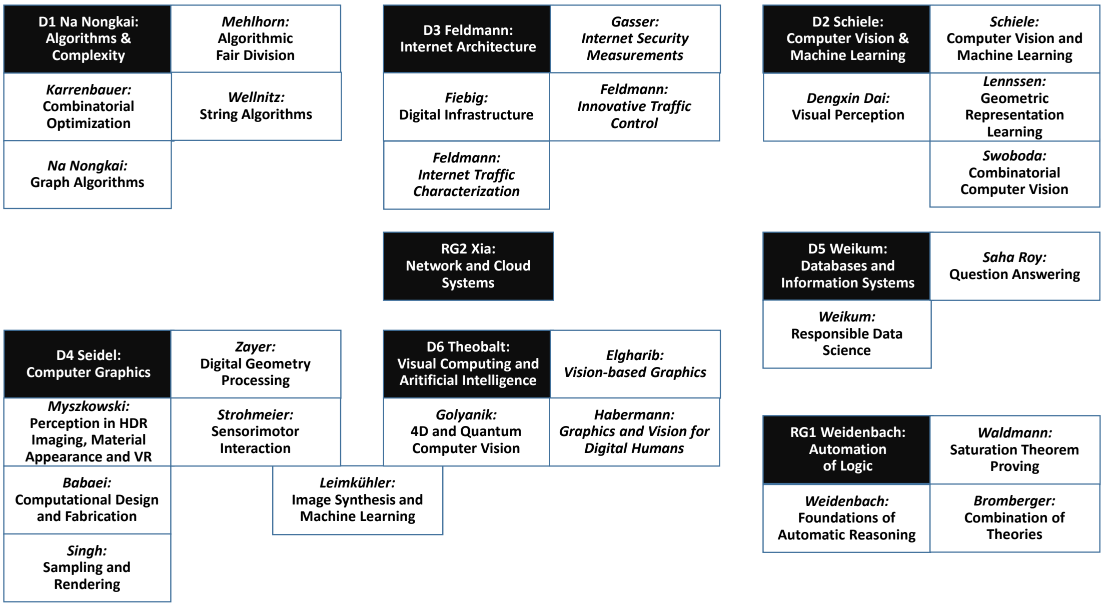



Figure 1.1: Research Areas of the Institute (as of March 2023)

**Doctoral Students:** As of March 2023, there are 96 doctoral students being supervised by members of the Institute, including 21 women (about 22%). 74 of the 96 students are non-German. The Institute does not grant degrees. In this regard, we closely collaborate with Saarland University. Members of the Institute teach courses at the university and supervise students at all levels. Upon appointment to senior researcher, the Computer Science Department of Saarland University decides whether to grant the senior researcher the right to supervise doctoral students on a case by case basis. Currently, this right has always been granted to all senior researchers of the Institute.

**Joint Administration and Technical Support:** The Institute shares the IT support group, administration, library, and facility management with the MPI for Software Systems. Both institutes currently count 54 full-time employees including 21 IT support staff in these shared areas.

## 1.2 Scientific Vision and Strategic Goals

Algorithms and their applications are and have always been the main focus of the Institute. They are the core of what makes computer systems useful and productive. They influence every aspect of our daily lives and are the basis for industrial change. Throughout the last decade, major parts of our research effort have focused on multimodal computing. The grand challenge is to understand, search, and organize large, distributed, noisy, incomplete, and diverse information in a robust, efficient, and intelligent manner. Our research ranges from foundations (algorithms and complexity, automation of logic) to a variety of multimodal domains (computer graphics and vision, geometric computation, intelligent information systems, adaptive networks). In recent years, research on foundations of machine learning, as well as the investigation of machine learning and artificial intelligence methods at the intersection to the aforementioned research domains, has become an important part of the research of our institute. The overarching mission of the Institute is to be one of the world's top players and strategic trend-setters on these topics.

Most of the major advances in computer science have come through the combination of new theoretical insights and application-oriented experimental validation, all driven by outstanding researchers. Our goal is, thus, to have impact through i) *publications*, ii) *software, services, and data resources* enabled by our research, and iii) *people* alike. In the following, Section 1.3 presents our achievements regarding the first two dimensions, and Section 1.5 discusses our performance history with respect to the third dimension.

## 1.3 Long-Term Achievements and Impact

Over the last twenty years, the Institute has pursued a number of high-risk high-gain endeavors, starting with foundational science and ultimately making great practical impact. In the following, we outline the highlights of the Institute's scientific achievements.

- Danupon Nanongkai's long-term vision is to develop techniques for designing efficient algorithms across computing paradigms and, along the way, achieve the following two goals simultaneously: (i) solutions for notorious long-standing open problems and (ii) efficient algorithms that can fully exploit the characteristics of modern computing devices and data. His group has and continues to make significant progress in this direction with, e.g., algorithms that resolved long-standing questions about shortest paths and graph connectivity that are theoretically efficient in many computing paradigms such as distributed, dynamic, and parallel settings. The former director Kurt Mehlhorn continues his work on fair division problems and algorithm engineering.
- Bernt Schiele's group has made strong contributions both to computer vision as well as machine learning. For example, the group has been contributing to multi-object tracking over the years and has won the Multi-Object Tracking Challenge twice. Another important direction has been person detection and multi-person pose estimation, both of which have lead to a wide range of follow-up work in the community. More recently, the group has made ground-breaking contributions to inherently interpretable machine learning methods.

- Anja Feldmann's research vision is to obtain insights from Internet measurements as foundation for shaping the evolution of the Internet by proposing optimizations and investigating alternative designs. As such her research group has and continues to address many challenges in *Internet measurement*, e.g., via a dead dive into the IoT backend ecosystem, as well as *innovative traffic control*, e.g., collaborative DDoS mitigation.
- The Computer Graphics Group stands out for its integrated view of *3D Image Analysis and Synthesis*. Hans-Peter Seidel has developed groundbreaking results on multiresolution modeling with special emphasis on new metaphors for editing and shape deformation, and novel data structures for high performance geometry processing. Karol Myszkowski has been and continues to be one of the pioneers and a driving force in perception-based graphics.
- The *Visual Computing and Artificial Intelligence Department* headed by Christian Theobalt made groundbreaking contributions to hard research problems at the intersection of computer graphics, computer vision and artificial intelligence, notably machine learning. The department has contributed pioneering methods for high-fidelity capture, modeling, rendering and simulation of virtual humans and general dynamic scenes, recently also on the basis of groundbreking new neuro-explicit algorithms. It is also did seminal work on neural scene representation and neural (inverse) rendering, generative modeling, robust real world perception an reconstruction, as well as foundational aspects of visual computing and machine learning.
- Christoph Weidenbach's group is on the next level of automated reasoning systems. While SAT and SMT solving has meanwhile found many places as a daily tool in standard processes in research and industry, we are working on lifting this success to more expressive logics. To this end we develop a new automatic reasoning framework called *SCL*: Clause Learning from Simple Models.
- Gerhard Weikum and his team pioneered the theme of *Knowledge Harvesting*: automatically building comprehensive knowledge bases from Internet contents. This work provided the blueprint for industrial-strength knowledge graphs that are key assets for search engines, question answering and text analytics (at Google, Microsoft, Amazon, Apple etc.). The MPI-INF team has also advanced use cases of large knowledge bases, most notably, for natural language understanding and for question answering. The ground-breaking nature and practical impact of the YAGO project has been recognized by the Influential Paper Award of the Artificial Intelligence Journal and the W3C Test-of-Time Award for the most influential papers in the WWW conference series.
- Yiting Xia's long-term goal is to re-architect cloud data centers with high-performance network infrastructures and networked systems. Her research group conducts fundamental research combining novel network hardware and software systems to re-design different layers of the network stack. Towards that, her group has built the first end-to-end testing framework for diverse data center network fabrics, proposed routing and transport protocols that challenge the conventional wisdom of TCP/IP, and optimized prominent cloud applications such as machine learning training and disaggregated storage.

**High-Risk Research:** When the above long-term projects started, they were far from the mainstream in their scientific communities; most were considered elusive and some even characterized as useless. Needless to say, not all of our bold endeavors have worked out. The sections about the six departments name some examples. Nevertheless, most of this work has resulted in novel insights (sometimes about what is, fundamentally or practically, non-viable) and often in influential publications.

## 1.4 Highlights 2021–2023 and New Research Directions

Danupon Nanongkai joined the Institute as a director for the Department of Algorithms and Complexity in August 2022. Until then, the department was led by the emeritus director, Kurt Mehlhorn, who continued his research as a researcher in the department. Christian Theobalt joined the Institute as a director for the department on Visual Computing and Artificial Intelligence in March 2021. The Institute is involved in the process of the Max Planck Society to establish departments in quantum computing.

The following are selected highlights from the scientific results that the Institute has achieved in the last two years.

- Danupon Nanongkai's group has been working on algorithms and complexity on graphs, numbers, geometric objects, and strings, as well as algorithmic game theory, and optimization. The group advanced the understanding of fast algorithms for many fundamental questions; results include fast shortest paths algorithms (best paper award at FOCS 2022), fast approximation algorithms for pattern matching and computing edit distance (STOC'22, FOCS'22, STOC'23), and fast algorithms and runtime (conditional) lower bounds for convolution, subset sums and Euclidean traveling salesperson (STOC'21, FOCS'21). The group also advanced the understanding of (i) the theory of fair division (e.g. simplified and stronger proofs for the existence of EFX allocations (EC'21, IJCAI'22, SODA'22, etc.) and (ii) learning algorithms (e.g. approximating the neural kernels (NeurIPS'21, NeurIPS'22) and learning-augmented online algorithms (AAAI'23). One of our works on optimization was also integrated into the production process of our industrial partner.
- In the reporting period, the group of Bernt Schiele has been working on various important areas in both computer vision as well as machine learning. For example, the group has contributed novel 3D scene understanding methods that not only unify 3D representations for a range of tasks such as detection, segmentation and classification, but also have won the Waymo Motion Prediction Challenge in 2022 and 2023. One of the most promising directions of the reporting period has been the investigation of explainability and inherently interpretable machine learning methods. We not only contributed an in-depth analysis of state-of-the-art post-hoc attribution methods, but also contributed inherently interpretable deep neural networks by including the goal of interpretability in the learning process itself. Another important direction was to learn without or with very little supervision. Here, a variety of powerful methods have been researched and proposed ranging from self-supervised, over few-shot-learning to label-efficient semi-supervised learning. Importantly, the above mentions just a few of

the many important directions pursued in the group both by Bernt Schiele as well as by the other senior researchers and research group leaders.

- Anja Feldmann's group has been working on understanding the capabilities and vulnerabilities of today's Internet. In recent work they have been highlighting, e.g., new capabilities to support the IoT eco-system (IMC 22), the ongoing evolution of the Internet topology (IMC 22), the cloudification of higher education infrastructure (PETS 22), and novel collaborative DDos mitigation techniques (CCS 22). This work has received an IETF Applied Networking Research award.
- In the Computer Graphics Group we achieved fundamental results on correlated sampling and on bridging the gap between Monte Carlo rendering and neural radiance fields (NeuRIPS'21, SIG'22, SIGAsia'22), and we substantially contributed towards improving the visual experience in VR and games (2x SIGAsia'21, SIG'22, SIGAsia'22, ACM TAP'23) and material appearance perception (SIG'21, TVC'21, SIGAsia'22). We achieved groundbreaking results on neural inverse design in computational design (SIG'22, NeuRIPS'22 (spotlight)), SIGAsia'22), and we significantly advanced tactile rendering (TEI'22, 2x CHI'23). We also continued our successful line of research on the mesh matrix formalism for geometric computing (CGF'21, EG'23).
- The Visual Computing an Artifical Intelligence department developed pioneering methods for human capture and modeling, such as pose estimation with explicit (differentiable) physics-modeling (SIGGRAPH'2021), egocentric motion capture (ECCV'22, CVPR'22), as well as monocular performance capture (DeepCap) (TPAMI'23). State-of-the-art methods for neural capture, modeling, editing, and (free-viewpoint) rendering of entire humans (e.g. (DDC) SIGGRAPH'21, (Neural Actor) SIGGRAPH Asia'21) and human faces (e.g. SIGGRAPH'21, CVPR'21) at highest fidelity were also presented. The team contributed groundbreaking works on neural and neuro-explicit modeling and rendering of general static (e.g. NeuS (NeurIPS'21)) and dynamic scenes (e.g. NR-NerF (CVPR'21), Phi-SfT (CVPR'22)) under illuminaton control (CVPR'22). Works on robust scene perception with neural analysis-by-synthesis (ECCV'22), and quantum computer vision (e.g. CVPR'21) opened new directions in the department. The works on 3D-aware high-quality generative modeling of shapes (ICLR'22, CVPR'22) and human motion (SIGGRAPH'22) were widely recognized.
- Christoph Weidenbach and his group have been working on arithmetic theory solving, higher-order formalization and mechanization, and SCL (Clause Learning from Simple models). In particular, we put SCL to practice. We showed that we can formalize, verify, and run so called *Supervisors* inside the SCL framework (TACAS'2022, VMCAI'2021, FroCos'2021). Supervisors are software that controls technical systems at a higher level of abstraction such as a lane change assistant in a car, or an electronic control unit for a combustion engine or an aircraft engine.
- Gerhard Weikum's group has further advanced its work on knowledge base construction, curation and application by spearheading new methods for capturing entities with quantities (such as running times of athletes or energy consumption of cars) and for cultural commonsense knowledge (such as food preferences or general habits in different cultures), and for conversational question answering over combinations of knowledge

graphs, tabular data and text corpora. These methods integrate new ways of leveraging pre-trained language models (such as GPT), but aim to avoid the risks and huge energy footprint of language models as sole source of (latent) knowledge. Results have appeared in top-tier venues like SIGIR, SIGMOD, EMNLP, WSDM and WWW. Also, a 350-page survey on knowledge bases has been published. In the ERC synergy grant imPACT, the focus has shifted from understanding privacy risks and their trade-offs with utility to enhancing the trustworthiness, explainability and controllability of recommenders and other machine-learning systems. The team has developed new methods for user-comprehensible explanations in recommenders, risk analysis for model deployment, and operationalizing fairness models in a practically viable manner. Results have appeared in premier venues like SIGIR, VLDB, NeurIPS, WWW and the Machine Learning Journal. Asia Biega's paper on fairness in rankings is one of the highest cited papers on this topic. Azin Ghazimatin's dissertation won the GI DBIS Dissertation Award.

– Yiting Xia's group has been working on performance optimization of network and cloud systems. The group pioneers on system support for optical data center networks, a high-performance and low-cost network fabric increasingly adopted by cloud providers. Their Hop-On Hop-Off (HOHO) routing algorithm is revolutionary in reduction of transmission latency in optical data center networks (APNet'22), and OpenOptics is the first end-to-end testing framework that makes optical data center networks accessible to academic researchers. Besides optical networks, the group has improved communication performance of machine learning training systems, through the EchelonFlow network abstraction for efficient collective flow scheduling (HotNets'22) and network-aware GPU sharing. In joint work with Meta, the group has also advanced data center network management with workflows and distributed database.

**New Research Directions:** We continue focusing on the grand challenge of exploring algorithms that understand, search, and organize large, distributed, noisy, incomplete, and diverse information in a robust, efficient, and intelligent manner. However, we expect a leap in research directions by the ongoing blending of virtual worlds with the real world. This trend involves several grand challenges on which we will embark. As digital contents and output from all kinds of sensors continue to explode, we need to move from data to understanding situations and anticipating user behavior. We need to support immersive interactions across all modalities, considering visual signals like facial expressions, gestures and body language, in conjunction with language, contextual knowledge and social interactions. All of the above entail profound methodical challenges to be able to handle, in the long run, tasks of high real world complexity, and to enable interaction with computing systems on a human-like level. To approach these challenges, we also intensify our investigation of new ways to unite data-driven machine learning-based concepts with explicitly designed methods. We further prepare for the resulting profound computing challenges resulting from this by investigating new algorithms that work across various models of computation, including quantum computing. Finally, with the rapid advance of machine learning and data-driven algorithmic decision-making, we need to better understand how to make computer behavior comprehensible. These challenges are the motivation for our foundational research on algorithms, visual computing, machine learning, internet architecture and knowledge discovery in the coming years.

## 1.5 Career Mentoring

The Institute has a strong track record on educating and mentoring young researchers, at both the doctoral student level and the postdoctoral level.

**Doctoral Student Training:** Since the Institute was established in 1990, a total number of 390 doctoral students have graduated. These include 73 women, and 179 non-Germans. A number of students have won prestigious national and international awards: 17 have been awarded the Otto Hahn Medal of the Max Planck Society. Andrey Rybalchenko (2007) and Fabian Suchanek (2012) have received the Otto Hahn Award for the best dissertations in the Chemistry, Physics, Technology Section of the Max Planck Society. This award includes a 5-year scholarship for an independent research group, and only one award per year is granted since 2007. In the reporting period Azin Ghazimatin won the DBIS Dissertation Award in 2023. Marc Habermann received the DAGM MVTec Dissertation Award 2022 and a Eurographics PhD Thesis Award 2022. Franziska Müller was awarded the Dr. Eduard Martin Prize 2021.

**Young Scientist Career Advancement:** A unique strength of the Institute is its successful fostering of young scientists. Our mentoring and career support efforts are most pronounced at the level of our senior researchers. Since the notion of senior researchers was explicitly introduced in 2007, a total of 85 young scientists have held such positions. 76 of them have meanwhile left the Institute. Out of these, 75 have accepted a tenured or tenure-track position offer at universities or university-like research organizations. We started tenure-track openings for senior researchers in 2019. We see this as an additional instrument for hiring talented researchers and, in particular, women. Several alumni (i. e., who graduated here or spent at least two years at the Institute and have since moved on) and current senior researchers of the Institute have won prestigious awards: Leibniz Prizes[3], ERC Grants[4], and other honors.

**Support for Women:** The percentage of women at our Institute is currently 22% for doctoral students and 23% at the postdoctoral level (including senior researchers). To increase the representation of women in our field, we have established the Lise Meitner Fellowship for outstanding female scientists at the postdoctoral level. So far, these two-year fellowships have been awarded to 17 women[5]. Out of the 15 recipients of the award who have meanwhile

[3] Susanne Albers 2008, Leif Kobbelt 2014, Thomas Neumann 2020, Peter Sanders 2012

[4] Susanne Albers 2016, Niko Beerenwinkel 2014, Jasmin Christian Blanchette 2016, 2022, Christoph Bock 2015, Karl Bringmann 2019, Andreas Bulling 2018, Parinya Chalermsook 2017, Piotr Didyk 2018, Jürgen Gall 2015, Bastian Goldlücke 2013, Matthias Hullin, 2018, Leif Kobbelt 2013, Christoph Lenzen 2016, 2020, 2023, Marcus Magnor 2010, Thomas Neumann 2016, Dániel Marx, 2017 Rafal Mantiuk 2016, Antti Oulasvirta 2014, Bodo Rosenhahn 2011, Thomas Sauerwald 2015, Pascal Schweitzer 2018, Jürgen Steimle 2016, 2023, Christian Theobalt 2013, 2017

[5] Anke van Zuylen 2010-2012, Anna Adamaszek 2012-2014, Zeynep Akata 2014-2016, Daria Stepanova 2015-2017, Hang Zhou 2015-2017, Shida Kunz (Beigpour) 2016-2018, Qianru Sun 2016-2018, Qiuhong Ke 2018-2020, Eunjin Oh 2018-2020, Sophie Tourret 2018-2020, Lingjie Liu 2019-2021, Erisa Terolli 2019-2021, Ana Serrano 2020-2021, Mengyu Chu 2020-2022, Jiangxin Dong 2020-2022, Roohani Sharma 2020-2022, Xinting Hu 2022-2024

left the Institute, twelve continued their career with a professor appointment. Our recently established senior researcher tenure-track career path constitutes an additional instrument for attracting women. Yiting Xia is the first tenure-track senior researcher at our Institute. A large number of female alumni of the Institute have become professors.[6]

## 1.6 Collaborations and Strategic Partnerships

An overriding goal that our Institute has been contributing to since its beginning in 1990, is to establish Saarbrücken as one of the world's premier sites in computer science. Indeed, over the last three decades the site as a whole has a unique track record, as exemplified by 7 ACM Fellows, 7 Leibniz Prizes (the highest scientific honor in Germany), and a total of 33 ERC Grants (at all levels).

**Saarland Informatics Campus:** Our most important partners are the Computer Science Department of Saarland University, the Helmholtz Center for Information Security (CISPA), the MPI for Software Systems and the German Research Center for Artificial Intelligence (DFKI). We have a long-standing tradition of teamwork and joint engagement in research, recruiting, and teaching. In the reporting period the senior researchers Mario Fritz, Dániel Marx, and Jilles Vreeken joined the CISPA faculty, and Karl Bringmann joined the Computer Science Department. There are numerous collaborations with faculty members from the university and with researchers from various institutes on campus, including colleagues from the Department for Computational Linguistics.

**The Research Center on Interactive Media, Smart Systems and Emerging Technologies (CYENS):** This center is located in Cyprus and was founded in 2017. It is funded within the framework of Horizon 2020. MPI is a strategic international partner of CYENS and represented on the Board by Hans-Peter Seidel.

**ERC Synergy Grant imPACT:** The ERC Synergy Grant has been awarded to Michael Backes (Helmholtz Center for Information Security), Peter Druschel (MPI for Software

[6] Anna Adamaszek (Univ. Copenhagen, Denmark), Zeynep Akata (Univ. Tübingen), Susanne Albers (TU Munich), Iris Antes (TU Munich), Oana Balalau (INRIA), Hannah Bast (Univ. Freiburg), Mengyu Chu (Peking University), Carola Dörr (CNRS, France), Jiangxin Dong (Nanjing U of Science and Technology), Panagiota Fatourou (Univ. of Crete), Lilia Georgieva (Heriot-Watt Univ., UK), Yulia Gryaditskaya (Univ. of Surrey, UK), Qiuhong Ke (Monash Univ., Australia), Katja Hose (Aalborg Univ., Denmark), Georgiana Ifrim (UC Dublin, Ireland), Mouna Kacimi (Univ. Bozen-Bolzano, Italy), Ruxandra Lasowski (Univ. of Applied Sciences Furtwangen), Lingjie Liu (Univ. of Pennsylvania, USA), Petra Mutzel (Univ. Bonn), Alice McHardy (TU Braunschweig), Nicole Megow (Univ. Bremen), Ndapa Nakashole (UC San Diego, USA), Marina Papatriantafilou (Chalmers Univ., Sweden), Ruzica Piskac (Yale Univ., USA), Nicoleta Preda (Univ. Versailles, France), Maya Ramanath (IIT Delhi, India), Anna Rohrbach (TU Darmstadt), Ana Serrano (Univ. Zaragoza, Spain), Ina Schäfer (TU Braunschweig), Renate Schmidt (Univ. Manchester, UK), Viorica Sofronie-Stokkermans (Univ. Koblenz-Landau), Qianru Sun (Singapore Management Univ.), Kavitha Telikepalli (Tata Institute, India), Erisa Terolli (Stevens College of Technology), Cara Tursun (Univ. of Groningen, Netherlands), Yafang Wang (Shandong Univ., China), Nicola Wolpert (Univ. of Applied Sciences Stuttgart), Shanshan Zhang (Nanjing Univ. of Science and Technology, China), Hang Zhou (École Polytechnique Paris, France), Anke van Zuylen (Cornell Univ., USA)

Systems), Rupak Majumdar (MPI for Software Systems) and Gerhard Weikum (MPI for Informatics) for joint research on the strategic research theme of “Privacy, Accountability, Compliance, and Trust for the Internet of Tomorrow”. The project has been funded with a total of 10 Million Euros over the timeframe 2015–2022.

**IMPRS on Trustworthy Computing (IMPRS-TRUST):** The International Max Planck Research School on Trustworthy Computing (IMPRS-TRUST) was established in 2020 as a joint program of the MPI for Informatics, the MPI for Software Systems, Saarland University, and TU Kaiserslautern. Currently, 97 doctoral students of MPI-INF, including 23 women (24%) and 66 non-Germans (68%), are part of the program.

**CS@Max Planck** Computer science centered research of the Max Planck Society has started a joint presence https://www.cis.mpg.de/. The institutes comprise the MPI for Informatics, the MPI for Software Systems, the MPI for Security and Privacy, the MPI for Molecular Genetics, and the MPI for Molecular Cell Biology and Genetics. The effort includes joint recruiting for tenure-track faculty positions and internships.

**The Max Planck Graduate Center for Computer and Information Science:** The Center is a highly selective doctoral program that grants admitted students full financial support to pursue doctoral research in the broad area of computer and information science, with faculty at the MPI for Informatics, the MPI for Software Systems, the MPI for Intelligent Systems, and top German universities.

**Saarbrücken Graduate School for Computer Science:** The school was established in 2007 and encompasses all doctoral training in computer science on campus. The school was largely modeled after the IMPRS-CS and adopted many of its elements. IMPRS-CS provides fellowships for doctoral students, within the structural framework of the Graduate School.

**VIA Research Center: Saarbrücken Center for Visual Computing, Interaction and Artificial Intelligence:** In 2022 the Max Planck Institute (MPI) for Informatics in Saarbrücken (Germany) and Google established a strategic partnership: the “Saarbrücken Research Center for Visual Computing, Interaction and Artificial Intelligence (VIA)” at the MPI for Informatics. The center investigates basic frontier research challenges in computer graphics, computer vision, and human machine interaction, at the intersection to artificial intelligence and machine learning. The VIA center is directed by Christian Theobalt and Bernt Schiele is a PI.

**ELLIS Unit SAM: Saarbrücken Artificial Intelligence & Machine Learning:** The ELLIS Unit SAM https://www.ellis-unit-sam.de has been founded in 2020 as part of the ELLIS Society which seeks to establish internationally visible, top-level research facilities in Europe in the area of machine learning and modern AI. The SAM Principal Investigators have agreed to jointly work on both the foundations for enhanced functionalities of Artificial Intelligence and Machine Learning (AIML) systems and the pressing needs for security, privacy, and trustworthiness that arise from the widespread use of Artificial Intelligence and Machine

Learning systems. Bernt Schiele is the scientific director of the unit, Christian Theobalt is a PI, and overall, the unit brings together nine PIs from the MPI for Informatics, MPI for Software Systems, UdS, and CISPA.

**ELIZA: Konrad Zuse School of Excellence in Learning and Intelligent Systems** The school is a graduate school in the field of artificial intelligence (AI) funded by the German Academic Exchange Service (DAAD). ELIZA's research and training activities focus on four main areas: the basics of machine learning (ML), including ML-driven fields like computer vision, Natural Language Processing (NLP), machine learning systems, applications in autonomous systems, as well as trans-disciplinary applications for machine learning in other scientific fields, from life sciences to physics. The ELLIS Unit SAM is one among the seven German partners, including Bernt Schiele (representative of the local ELIZA partner) and Christian Theobalt.

**Research Training Group: Neuroexplicit Models of Language, Vision, and Action** The research training group (in German 'Graduiertenkolleg') has been recently approved by the DFG (German Research Foundation) in 2023. It promotes early career researchers and is funded by the DFG for a period of up to nine years. The training group will lay the systematic foundation for so-called "neuroexplicit models". Such models attempt to combine the best aspects of previous approaches to create an AI that is safer, more reliable, and easier to interpret. In addition to the Max Planck Institute for Informatics (represented by Bernt Schiele and Christian Theobalt), partners include the Departments of Computer Science, Language Science and Technology and Mathematics at Saarland University, the German Research Center for Artificial Intelligence (DFKI), the Max Planck Institute for Software Systems, and the nearby CISPA Helmholtz Center for Information Security.

## 1.7 Results 2021–2023

**Publications, Software, Startups:** In the two-year time-frame 2021–2023, the Institute published more than 600 papers in peer-refereed conferences and journals. Many of these appeared in top-tier venues, with competitive conferences typically accepting only 10 to 20 percent of their submissions. Several publications won best paper awards or best student paper awards (at 3DV 22, CPM 22, DL-2022, FOCS-2022, IVA 21, IWOCA 22, PAM 23, SOFSEM 23).

Two startups that spun off from our research in 2014 and 2012, Captury and Logic4Business, respectively, are gaining traction in their respective markets. Two startups were founded with with seed-funding from the EXIST program of the German Ministry for Economy (BMWi): Ambiverse in 2016 and Oraclase in 2022.

**People:** In the two-year time-frame 2021–2023, 37 of our doctoral students graduated. These include 10 women. In the same time period, 12 of our researchers left the Institute to take a tenured or tenure-track faculty position [7]. These include 4 women.

[7] Mengyu Chu (Peking University, China), Yulia Gryaditskaya (University of Surrey, UK), Costas Iordanou (European University Cyprus, Cyprus), Petr Kellnhofer (TU Delft, Netherlands), Christoph Lenzen

**Awards:** Members of the Institute won prestigious awards. The following are the most prominent examples; full lists are in the respective sections of the departments. Anja Feldmann won the IETF/IRTF Applied Network Research Prize togehter with Daniel Wagner in 2022; in 2021, Harald Ganzinger was posthumously awarded with the CAV Award for pioneering contributions to the foundations of the theory and practice of satisfiability modulo theories (SMT); Hans-Peter Seidel became a member of the ACM SIGGRAPH Academy 2022; Gerhard Weikum was recognized with the Korad-Zuse-Medal 2021;

At the level of senior researchers, Vahid Babaei received the Herrmann-Neuhaus-Prize of the Max Planck Society 2022; Dengxin Dai won the German Pattern Recognition Award 2022; Shaoshuai Shi was recognized with the World Artificial Intelligence Conference Rising Star Award 2021;

At the student level, Corinna Coupette was awarded with the Caroline von Humboldt Prize; Ayush Tewari and Marc Habermann both won the Otto-Hahn-Medal in 2022 and 2023, respectively; Marc Habermann also received the DAGM MVTec Dissertation Award and the Eurographics PhD Award in 2022; Azin Ghazimatin was awarded the German GI-DBIS Dissertation Award 2023; Anna Kukleva received the Grace-Hopper Award of the University of Bonn in 2021; Jan Eric Lenssen won the TU Dormund Dissertation Award in 2022; Franziska Müller received the Dr. Eduard Martin Prize 2021; Karol Wegrzycki was recognized with the Prime Minister of Poland Award for his PhD thesis in 2022 and won the Witold Lipski Prize 2021 as well as the PCC Open Mind Prize 2022; Philip Wellnitz won the Dieter-Rampacher-Prize for the youngest doctoral student at the Max Planck Society 2021;

Further honors with considerable visibility include the following. Kurt Mehlhorn received a Honorary Doctorate Degree from Aalto University in 2023; Bernt Schiele and Thomas Lengauer were both appointed ACM Fellows in 2022; Christian Theobalt has been elected Fellow of the Eurographics Association in 2022; Yiting Xia has been included in the list of the top 10 Rising Stars in Computer Networking and Communications by the N2Women Association in 2021.

**Equal Opportunities Plan:** The equal opportunity plan is part of the MPI-INF policy to create a work environment that meets the diverse life situations and needs of all Institute members. For us, having a gender equality policy means establishing and shaping organizational and structural conditions within the Institute in order to optimize individual development opportunities for all our members. We have established continuous efforts to support young families: parent-child room, nursing room, reserved places in day care centers, babysitter agency, pacifier projects, and a high flexibility in home office work also after the pandemic. From the end of 2020 to December 31, 2022 we are allowed to call ourselves "Family-friendly Company". This certification was awarded to us by the state government in cooperation with the Chamber of Industry and Commerce.

We have been successful in attracting an outstanding woman as a director and as our

(CISPA, Germany), Lingjie Liu (University of Pennsylvania, USA), Habib Mostafaei (TU Eindhoven, Netherlands), Xingang Pan (Nanyang Technological University, Singapore), Ana Serrano (University of Zaragoza, Spain), Joachim Spoerhase (University of Sheffield, UK), Erisa Terolli (Stevens College of Technology, USA), Cara Tursun (University of Groningen, Netherlands), Andrew Yates (University of Amsterdam, Netherlands), Savvas Zannettou (TU Delft, Netherlands), Rhaleb Zayer (University of East Anglia, UK), Quan Zheng (Chinese Academy of Sciences, China) first tenure-track senior researcher. At the postdoc level, the Institute offers a distinguished fellowship for women, named Lise Meitner Scholarship, since 2013. The interest in this program has been strongly increasing in terms of both quantity and quality of applicants. In the last two years, two young women accepted the fellowship. The best postdocs are often candidates for becoming senior researchers after two years. Out of the 15 fellowship recipients who left the Institute, twelve continued with a professor appointment.

Attracting women to computer science needs to start already at the school level and constitutes a long-term effort. As a part of the BWINF, "Bundesweit Informatik Fördern", an organization mainly funded by the German government to support young talent in computer science, we have established a number of measures to inspire girls for computer science. In addition to three different competition formats aiming at different levels of skills and age, we are now organizing girl camps and have established a separate girl team participating in the informatics olympiad for pupils.

**Outreach and Visibility:** To increase the international visibility of Saarbrücken as a world-class CS hub, we have established an agreement for joint branding: almost all CS players on campus now use the label "Saarland Informatics Campus" as part of their official addresses. The label will also be used in Google Scholar profiles, academic rating sites, and other PR efforts (see http://sic.saarland/).

The Institute continues its role as a provider of for the BWINF, "Bundesweit Informatik Fördern", promoting young computer science talent in Germany.

# Part II

# Overview – The Research Units & Senior Researchers

# 2 D1: Algorithms and Complexity

## History

The Algorithms and Complexity group (D1) is currently led by Danupon Naongkai who started his position as a Scientific Director in August 2022. Danupon is the successor of Kurt Mehlhorn who established the group in 1990 as one of the two initial groups of the institute. Kurt moved to emeritus status on August 31, 2019, and has been an acting head of the department until Danupon joined the institute. Kurt has continued as a scientist till today.

The group is currently regrowing after it shrunk between 2019–2022. Meanwhile, the algorithm theory community on the campus has become stronger. Karl Bringmann became a Full Professor of Computer Science at Saarland University. Dániel Marx, Christoph Lenzen, and Sebastian Brandt became Faculties at CISPA, the Helmholtz Center for Information Security. The group hosts these researchers as long-term guests to foster collaborations within the campus. In general, the algorithms and complexity community is going strong on campus.

During the reporting period, the senior scientists and research area coordinators are Andreas Karrenbauer, Tomasz Kociumaka, Christoph Lenzen (moved to CISPA in July 2021) Kurt Mehlhorn, Adam Polak, Joachim Spoerhase (moved to the University of Sheffield in February 2023), and Philip Wellnitz. Some of our long-term guests also assisted in coordinating some research areas: Karl Bringmann and Evangelos Kipouridis from Saarland University and Sebastian Brandt, Christoph Lenzen and Dániel Marx from CISPA.

Section 27.1 lists the names of current and recent group members and the current positions of the group members that left during the report period. Our alumni continue to get very good positions, see Page 202; Christoph Lenzen moved to a faculty position at CISPA and Joachim Spoerhase moved to a faculty position at the University of Sheffield. Adam Polak joined the group before joining Bocconi University as a tenure-track assistant professor. Six group members completed their Ph.D., see Page 202. Group members received prestigious awards, see Page 203, for instance, Kurt Mehlhorn obtained an Honorary Doctorate Degree from Aalto University, Karol Węgrzycki received the Open Mind Prize in 2022, the Prime Minister of Poland Award for Ph.D. Thesis in 2022, and the Witold Lipski Prize in 2021, Philip Wellnitz received the Dieter-Rampacher-Prize for the youngest Ph.D. student of the Max Planck Society that graduated in 2021, Corinna Coupette received the Caroline von Humboldt Prize for outstanding female junior researchers in 2022, Roohani Sharma received the best paper award from SOFSEM 2023, Danupon Nanongkai received the best paper award from FOCS 2022, and Tomasz Kociumaka received the best paper awards from IWOCA 2022 and from CPM 2022. A former group member, Paolo Ferragina (now at the University of Pisa), recently received the 2023 ACM Paris Kanellakis Theory and Practice Award.

Some group members hold their own grants, see Page 203, for instance, Christoph held an ERC grant and Joachim Spoerhase holds a DFG grant.

We have published extensively and in excellent venues during the reporting period, for instance, 14 papers in FOCS and STOC, 12 papers in SODA, 5 papers in SOCG, and 5 papers in NEURIPS and ICML. We are also visible in, for instance, PODS, AAAI, IJCAI, KDD, ICLR, AI and Law, and Nature Communications.

## Vision and Research Strategy

The vision for D1 is to be a first-class algorithm and complexity group and a trendsetter in this field, and to have an impact on the research community and society through people, research results, techniques, software, and scientific leadership.

Most of our effort is on theoretical works while some group members also contributed to experimental works and software constructions. Our current research activities are organized into five areas.

- Algorithmic Game Theory,
- Algorithms and Complexity on Graphs,
- Algorithms and Complexity on Numbers and Geometric Objects,
- Algorithms on Strings, and
- Optimization.

Together, we span a large part of algorithmic research. The emphasis changes over the years as group members come and go. We hire postdocs mainly based on quality and less on thematic fit. There is considerable interaction and collaboration between the areas. The entire group meets twice a week to discuss science (Tuesday and Thursday noon seminar) and biweekly to discuss administrative matters. Various subgroups meet weekly and even daily. We try to create environments that encourage informal interactions as we believe that it is the best way to create synergies between different research fields.

We teach at all levels; see Page 200 for details. Frequently group members pair for lectures.

## Research Areas and Achievements

We discuss some of the main results obtained in the reporting period.

**Algorithmic game theory.** We worked mainly on fair division problems. In fair division, a set of items has to be allocated to a set of agents. The goal is to find a "fair" and "efficient" allocation. There are several notions of fairness. Envy-freeness up to any good (EFX) is arguably the most compelling fairness notion in the context of allocating indivisible goods. Indivisible goods are items that the agents would like to have as much as possible and they cannot be divided. In an EFX allocation, removing any good allocated to any agent A makes A's remaining bundle not interesting to other agents (no envy from other agents). The *existence* of EFX allocations is one of the most important problems in fair division.

Hannaneh Akrami, Bhaskar Ray Chaudhury, Kurt Mehlhorn, Pranabendu Misra, and their co-authors (EC 2021, arXiv 2022) simplified the proof (originally established by our group in the previous reporting period) that EFX-allocations exist for three agents and extended it to a larger class of valuations. They also showed that there exists an approximate EFX allocation with a sublinear number of unallocated goods by establishing an intriguing connection to a problem in zero-sum extremal combinatorics. They further studied a relaxation of EF$k$X, where $k$ goods can be removed to remove envy (IJCAI 2022), and the fair division of divisible chores where chores (or "bads") are items that agents dislike (SODA 2022).

Nash Social Welfare (NSW) is the oldest notion of fairness (Nash, 1950); the NSW of an allocation is the geometric mean of the utilities of the agents. An allocation maximizing NSW is Pareto-optimal, that is, there is no other allocation in which no agent is worse-off and some agent is strictly better-off. Finding an allocation maximizing NSW is NP-complete even for additive valuations.

Hannaneh Akrami, Bhaskar Ray Chaudhury, Kurt Mehlhorn, Golnoosh Shahkarami, and their collaborators (AAAI 2022, arXiv 2022) studied the case of 2-value additive valuations where each good is valued either 1 or $p/q$, for some fixed co-prime numbers $p, q \quad \mathbb{N}$ such that $1 \leq q < p$, and the value of a bundle is the sum of the values of the contained goods. They gave a complete characterization of polynomial-time tractability of NSW maximization that solely depends on the values of $q$.

**Algorithms and Complexity on Graphs.** Paths, cuts, and matchings are fundamental graph problems studied extensively for decades. In this reporting period, we made much progress on these problems. Karl Bringmann, Alejandro Cassis, Nick Fischer, Danupon Nanongkai and their collaborators developed techniques that resolved the long-standing negative-weight single-source shortest paths problem with an $\mathcal{O}(m \log^2(n) \log(nW) \log\log n)$-time algorithm, where $m$, $n$ and $W$ denote the number of edges, number of vertices, and the upper bound on the magnitude of the smallest negative-weight edge, respectively. One of the results received the best paper award at FOCS 2022 and was covered in the Quanta Magazine.

Another basic question about paths is constructing a data structure to report shortest-path distances called distance oracles. For sparse graphs and any constant $k \geq 1$, the state of the art is Thorup-Zwick distance oracle which achieves an approximation ratio (a "stretch") of $2k - 1$, preprocessing time $\mathcal{O}(kn^{1+1/k})$ and query time $\mathcal{O}(k)$.

Karl Bringmann, Nick Fischer, and their collaborators (STOC 2022, STOC 2023) addressed the following open question: *What is the best stretch $f(k)$ we can achieve if we insist on preprocessing time $\mathcal{O}(n^{1+1/k})$ and almost-constant query time $n^{o(1)}$?* They showed that $f(k) \geq k$ under the 3Sum Hypothesis, narrowing the gap to $f(k) \quad [k, 2k - 1]$.

For computing graph cuts and connectivity, we highlight two results in the distributed setting (the CONGEST model):

- Yonggang Jiang and his co-author (STOC 2023) achieved a vertex connectivity algorithm that works in sublinear time on distributed networks, and
- Christoph Lenzen and Hossein Vahidi (SIROCCO 2021) showed a tighter relationship between $(1 + \varepsilon)$-approximate shortest paths and approximate minimum directed spanning trees, implying improved distributed algorithms for the latter problem.

For a harder variant of *weighted or directed multicut problems*, Roohani Sharma, our guests from CISPA (Dániel Marx and Philipp Schepper) and their collaborators developed techniques to show that weighted multicut on trees and directed multicut with three terminal pairs are both FPT (WG 2022, SODA 2023).

For matching and its generalization to matroid problems, our main results are in the dynamic setting. We highlight two results:

- Peter Kiss (main affiliation: University of Warwick) and his co-authors achieved a dynamic $(1+\varepsilon)$-approximate matching algorithm with truly sublinear update time, making the first progress in a decade for one of the most central problems about dynamic matching, and
- Danupon Nanongkai, Ta-Wei Tu, and their collaborators introduced the model of dynamic matroid oracle and showed their applications in achieving fast graph algorithms.

Additionally, based on results on parameterized complexity for Perfect Matching and Generalized Matching problems on bounded tree-width graphs, Philip Wellnitz, our long-term guests from CISPA (Jacob Focke, Dániel Marx, Fionn Mc Inerney, and Philipp Schepper) and their collaborator studied the related vertex-selection problem about Counting "Generalized Dominating Sets" in Bounded-Treewidth Graphs (SODA 2023). They improve the time complexity significantly for many special cases such as the Perfect Code problem.

Other results include reducing the random bits needed for the Isolation Lemma for NP-complete problems (Karol Węgrzycki et al. STACS 2022) and some applied and experimental works such as analyzing graphs and hypergraphs (Corinna Coupette et al. KDD 2021, AAAI 2022, ICLR 2023) and computing graph hyperbolicity (André Nusser et al. ALENEX 2022).

**Algorithms and Complexity on Numbers and Geometric Objects.** Through the lens of fine-grained complexity, we improved and in some cases settled algorithmic questions about numbers and geometric objects. We highlight two results on numbers:

- Given integer vectors $A$ and $B$, the convolution problem is to compute $\sum_{i+j=k} A[i]B[j]$ for every $k$. This is a fundamental computational primitive that has been a vital component in computer algebra, signal processing, computer vision, and deep learning. In the reporting period, Karl Bringmann, Nick Fischer, and Vasileios Nakos showed a new algorithm and fine-grained reductions that imply essentially-optimal runtime for the sparse nonnegative variant of this problem (STOC 2021).
- A central problem in fine-grained complexity is the subset sum problem: given a set of $n$ integers and a target integer $t$, and the task is to find a subset of these numbers that sum to precisely $t$. Karol Węgrzycki and his collaborator (STOC 2021) improved over the four-decade-old time and space bounds for this problem. Adam Polak, Karol Węgrzycki, and a collaborator (ICALP 2021) also developed an improved pseudopolynomial-time algorithm.

Our highlights for geometric problems are as follows.

- The Euclidean traveling salesperson problem (TSP) has a well-known approximation scheme. In this reporting period, Sándor Kisfaludi-Bak and Karol Węgrzycki and their collaborator (FOCS 2021) developed a new algorithm and a reduction from the Gap-Exponential-Time Hypothesis (Gap-ETH) to achieve essentially tight runtime for this problem.
- Given simple polygons $P$ and $Q$, a basic problem is to determine the largest copy of $P$ that can be placed into $Q$. Despite a long line of research since the 1980s, it

remained open whether we can establish (a) hardness beyond quadratic time and (b) any superlinear bound for constant-sized $P$ or $Q$. André Nusser and Marvin Künnemann affirmatively answered these questions under the $k$-Sum Hypothesis in their SODA 2022 paper.

Other results include improved pseudopolynomial-time, approximation, and online algorithms for the knapsack problem by Karl Bringmann, Alejandro Cassis, Andreas Karrenbauer, Adam Polak, Karol Węgrzycki, and collaborators, computing the vertices of maximum level in an arrangement of lines, where Kurt Mehlhorn and his co-authors found that this textbook problem is not as easy as it looks, computing generalized convolution that unifies many convolution procedures under one general umbrella, various distance measures (for instance dynamic time warping, Fréchet and Hausdorff distances), problems on geometric intersection graphs, robot motion planning, and many foundational results of Fine-Grained Complexity Theory.

**Algorithms on Strings.** Two central topics in the area of computations on strings are pattern matching and similarity measures such as the edit distance. One of our highlights is on fast approximate pattern-matching: for the classic problem of finding the occurrences of a pattern $P$ in a long text $T$ where some edits are allowed, Tomasz Kociumaka, Philip Wellnitz, and their collaborator (FOCS 2022) developed an algorithm that breaks the long-standing bound from 1998.

Another highlight is computing the edit distance between two input strings (the number of edits to change one string to the other string). Karl Bringmann, Alejandro Cassis, Nick Fischer, Tomasz Kociumaka, Vasileios Nakos, and their collaborators (STOC 2022, FOCS 2022) developed adaptive approximate algorithms with sublinear time complexities, which are almost optimal in some regimes.

For computing the edit distance exactly, an optimal algorithm that is linear-time in the low-distance regime has been long known. Tomasz Kociumaka and collaborators (FOCS 2022, STOC 2023) showed that a similar complexity can be achieved for more general distance measures such as weighted edit distance and tree edit distance.

Other results include data structures that allow preprocessing or that can handle dynamic input, approximating various distances, approximating longest common subsequence, a compressibility measure, and a construction of compressed suffix arrays by Karl Bringmann, Alejandro Cassis, Nick Fischer, Tomasz Kociumaka, Karol Węgrzycki, and collaborators.

**Optimization.** Contributions by our researchers span from discrete optimization to continuous optimization and to machine learning applications. The *sample complexity* of a function-learning task is the smallest number of correctly labeled examples needed to successfully learn a target function. It can be characterized by the well-known VC-dimension but also the PB-dimension, NC-dimension, etc. Pieter Kleer, Hans U. Simon, and collaborators (NEURIPS 2022, JMLR 2023, etc.) studied these various dimensions including the relation between the primal and dual of various combinatorial dimensions for multi-valued function classes, the NC-teaching dimension in comparison to the teaching dimension induced by any collusion-free model, and the separation between NC-dimension and PB-dimension.

The Neural Tangent Kernel (NTK) characterizes the behavior of infinitely-wide neural networks trained under least squares loss by gradient descent. Amir Zandieh and collaborators (NEURIPS 2021) designed a near-input-sparsity time approximation algorithm for NTK via sketching algorithms to accelerate learning with NTK. His first algorithm can approximate the neural kernels with ReLU activation. His follow-up work (NEURIPS 2022) showed an extension to general activations.

An emerging subfield of algorithms and complexity is learning-augmented online algorithms. The goal is to improve online algorithms using predictions provided by machine learning techniques. Golnoosh Shahkarami and collaborators (SWAT 2022, AAAI 2023) explored this topic in the context of deadline-based speed-scaling scheduling and the Traveling Salesperson Problem (TSP) on the Line.

The slime mold Physarum polycephalum was experimentally demonstrated to be able to solve several algorithmic problems on graphs, including shortest path problem. A mathematical model defined via a system of differential equations for the dynamic behavior of the Physarum was shown to converge to the shortest path for all graphs and can solve positive linear programs. In this reporting period, Yuan Gao, Andreas Karrenbauer, Pavel Kolev, Kurt Mehlhorn, Golnoosh Shahkarami, and collaborators (Physical Review E 2021, TCS 2022, arXiv 2022) extended this research to study the influence of noise on the convergence behavior, a Physarum-inspired multi-commodity flow dynamics, and a Physarum-inspired Dynamics to Solve Semi-Definite Programs.

Other results include algorithms to determine car build sequence in assembly lines by Andreas Karrenbauer, Leonie Krull, Kurt Mehlhorn, Paolo Luigi Rinaldi, Anna Twelsiek, and a collaborator, where one of their results was integrated into the production process in a plant of an industrial partner, a computational approach for designing multi-modal electro-physiological sensors and for designing menu systems by Andreas Karrenbauer and collaborators, and clustering algorithms by Martin Herold, Evangelos Kipouridis, and Joachim Spoerhase.

# 3 D2: Computer Vision and Machine Learning

## Group Overview

The Computer Vision and Machine Learning group (D2) was established in 2010 with the appointment of Bernt Schiele. The group was called 'Computer Vision and Multimodal Computing' initially and, during the last reporting period, the group was renamed to become the Computer Vision and Machine Learning group. There were two main reasons for this: first, multimodal computing has become less of a research focus for the group over the years; and second, while machine learning always was an important component of our research, its importance has increased significantly both for our research in computer vision but also in general. Therefore, renaming the group is both underlining the importance of machine learning for our research in computer vision as well as clearly stating that machine learning is an integral part of our research agenda. This has also resulted in an increased presence of our group at top-tier machine learning venues.

During the reporting period, the group was home to six groups of senior researchers/research group leaders, of which three were fully financed by MPI, and three groups are affiliated and partially funded by MPI. The three senior researcher groups fully funded by MPI were the ones of Dengxin Dai, Jan Eric Lenssen, and Paul Swoboda, and the three affiliated research group leaders were Zeynep Akata (U Tübingen), Margret Keuper (U Siegen), and Gerard Pons-Moll (U Tübingen). Each of the six group leaders have their own PhD-students (and PostDocs) to conduct research in their respective area. Also, the group was home to six postdocs, and 28 PhD students. Among those 27.5% are female (two research group leaders, three postdocs, and six PhD students). Eight group members completed their PhD during the reporting period. Our researchers get very good offers for faculty positions in academia [1], postdoc positions in academia[2], and research positions in industry[3].

## Vision and Research Strategy

Understanding visual information and more generally multi-modal information is a fundamental problem in computer science. Scientific challenges cover the entire pipeline from uni-modal processing, over spatial and temporal fusion of multiple and divergent modalities to the complete description of large-scale multi-modal data. At the same time we observe a tremendous increase in both the quantity as well as the diversity of visual and multi-modal

[1] Jiangxin Dong (Nanjing U of Science and Technology), Li Jiang (Chinese U of Hong Kong), Margret Keuper (U Siegen), Gerard Pons-Moll (U of Tübingen, U of Luxembourg), Paul Swoboda (U of Mannheim).

[2] Stephan Alaniz (U of Tübingen), Apratim Bhattacharya (U of Tübingen), Yaoyao Liu (John Hopkins U), Yongqin Xian (ETH Zurich).

[3] Bharat Bhatnagar (Meta Reality Labs), Andrea Hornakova (Blindspot Solutions), Mohamed Omran (Qualcomm), Rakshith Shetty (Amazon), David Stutz (DeepMind), Yongqin Xian (Google).

information due to the increasing number of sensors embedded in a wide variety of digital devices and environments as well as due to the increasing storage of visual and multi-modal data (such as surveillance data, personal and multimedia databases, or simply the Internet). While storing and indexing large amounts of visual and multi-modal data has made tremendous progress, understanding of this data still lacks behind. Therefore the long-term goal of D2 is to make progress on how to process, structure, access, and truly understand visual and multi-modal data both for online use as well as for large-scale databases.

In the reporting period, the group focused on two main areas, namely computer vision and machine learning. In the area of computer vision we address some of the most fundamental problems of image and video understanding such as object class recognition, people detection and tracking, and scene understanding. In the area of machine learning, we are focusing, on the one hand, on problems at the intersection of computer vision and machine learning (such as semi-supervised learning or adversarial robustness), and on the other hand on more foundational problems in machine learning (such as interpretability of deep learning and continual learning).

## Research Areas and Achievements

In the following, we report highlights from our research largely following the sub-group structure of D2. As you will notice, there are substantial cooperations among the various sub-groups with many joint topics, cooperations and publications.

**Computer Vision and Machine Learning** *Investigator: Bernt Schiele*

Since the establishment of D2 at the Max Planck Institute for Informatics, the sub-group headed by Bernt Schiele has been working at the intersection of computer vision and machine learning. The range of topics in Bernt's subgroup ranges from computer vision topics such as 3D scene understanding and multi object tracking, over machine learning topics such as inherently interpretable models and semi-supervised learning, to topics at the intersection of machine learning and computer vision such as adversarial robustness. In the following we briefly summarize contributions to three specific topics, namely interpretability, 3D scene understanding and robustness. More details are given in section 28.4.

The motivation to aim for explainable and interpretable machine learning methods is obvious: despite the apparent success of deep learning for an incredibly wide range of tasks, it is still difficult to understand the 'decision making process' within deep neural networks and how information is aggregated and processed. Given how ubiquitously these models are employed in our everyday lives, however, it is of paramount importance to gain a better understanding of their inner workings. Especially in safety-critical tasks—such as autonomous driving, health care, or in the judicial system—one needs to ensure that decisions are made for the right reasons. In our work on Interpretable Machine Learning, we approached this problem setting from a variety of different angles. On the one hand, we performed an in-depth evaluation of the state-of-the art post-hoc attribution methods to understand their promises and shortcomings. On the other hand, we developed inherently interpretable deep neural networks by including the goal of interpretability in the optimization process, thereby foregoing the need for post-hoc explanations. In this context, we evaluated and developed model guidance techniques that allow us to ensure that the models are indeed right for the right reasons. Finally, we explored how to increase the inherent interpretability of conventional models: for this, we propose to fine-tune conventional DNNs to use more human-interpretable concepts by inserting semantic bottlenecks into the model architectures.

3D scene understanding aims to enable machines to perceive, interpret, and reason about the 3D world in a way that mimics human perception. As the world we live in is inherently three-dimensional, understanding the spatial structures and relationships between 3D objects is essential for a wide range of applications, such as robotics, autonomous driving, augmented and virtual reality, and smart city planning. In the reporting period we made various contributions to 3D scene understanding using point cloud data. In particular, we proposed a versatile and unified backbone architecture for 3D point cloud understanding, efficient methods to learning in point clouds, as well as object level understanding for 3D indoor scenes. Last, but not least, we have been researching novel and powerful approaches for 3D perception and prediction for autonomous driving, which won both the 2022 and the 2023 Waymo Open Dataset Motion Prediction Challenge.

Robustness is an essential property of deep neural networks. In particular in high-stakes applications robustness, proper uncertainty calibration and understanding misuse become essential for deployment. In this context we looked at various ways to both better understand but also improve robustness. For example, we proposed a novel approach to improve robustness of deep neural networks by enhancing weak subnets during training. Also, we investigated how adversarially robust generalization is related to flat minima during optimization.

**Vision for Autonomous Systems** *Investigator: Dengxin Dai*

Over the last two years, the Vision for Autonomous Systems Group has followed a coherent research line making visual perception algorithms more robust and scalable. As widely known, adverse weather and lighting conditions (e.g. fog, rain, snow, low-light, nighttime, glare and shadows) create visibility problems both for people and the sensors that power automated systems. Many real-world applications such as autonomous cars, agriculture robots, rescue robots, and security systems can hardly escape from ‘bad’ weather, challenging lighting conditions, dust, smoke, and so on. Our group has developed multiple novel approaches to increase the adaptability of visual recognition models to real-world adverse weather/lighting conditions such as Domain Flow (IJCV’21), Fog Simulation (ICCV’21) and Snowfall Simulation (CVPR’22) for LiDAR Data, and the adverse-condition dataset ACDC (ICCV’21). The approaches and dataset represent the state of the arts in this important research area.

Current robotic perception systems are typically trained in a rather fixed environment, allowing them to succeed in specific settings, but leading to failure in others. In the last two years, we have developed multiple influential work in domain adaptation, including learning domain adaptation with auxiliary tasks (ICCV’21), continual test-time domain adaptation (CVPR’22), new network architecture DAFormer (CVPR’22) and the multi-resolution domain adaptation framework HRDA (ECCV’22). They have all become fundamental approaches.

Machine learning has been advancing rapidly, and ever growing data is at the center of this evolution. While the recipe of learning with large-scale annotated datasets is still effective, this can hardly scale due to the high annotation cost. We have developed multiple innovative methods for the low data regime. These include LiDAR semantic segmentation with 3D scribbles (CVPR'22), universal pre-training with multiple self-supervised tasks (CVPR'22, supervision transfer between data modalities such as from video to audio (T-PAMI'22), and ZegFormer for zero-shot learning (CVPR'22). These approaches have advanced the research in this area significantly.

**Geometric Representation Learning** *Investigator: Jan Eric Lenssen*

The newly founded Geometric Representation Learning group is concerned with designing and exploring effective representations and algorithms to bridge the gap between 2D observations and 3D reconstruction. Most observable parts of the 3D world consists of visually structured objects and repeating patterns, following rules of composition. It is our current understanding that it is this inherent structure and repetitiveness that allows humans to map out their surroundings just from a few sparse 2D observations, by combining observations with structured a priori knowledge. The GRL group aims to replicate this ability in computer vision systems by exploring the most efficient ways of learning 3D data priors.

One part of such a system could include an explicit model of symmetries in natural or men-made objects. In our work SymNP (ICCV'23 submission), we show that we can represent an object category as a set of characteristic symmetries between local areas, which are learned from a large amount of images showing objects from this category. These learned symmetries allow to perform full 3D reconstruction just from a single image. We were able to show that this abstract data prior is more efficient and more capable than simply learning the full function space as existing deep learning methods do. The representation is able to better transfer details to symmetric parts and is better in fusing information from multiple views, if available. Of utmost importance for the presented method are novel neural point representations, which allow to encode symmetries explicitly as connections between local point embeddings. We hypothesize that research on efficient 3D representations is deeply entangled with the ability to learn data priors for vision, which is why it will be in the focus of future work.

**Data Priors for Pose Estimation** *Investigators: Gerard Pons-Moll and Jan Eric Lenssen*

Pose estimation from different data modalities using parameterizable models like SMPL or MANO has been of increasing importance for several applications in computer vision and graphics, such as creating and driving virtual avatars, action recognition, or 3D reconstruction. A big challenge in pose estimation from incomplete observations are ambiguities, i. e. multiple possible poses that would perfectly explain the given observation. Naturally, it is desired to solve these ambiguities by learning data priors from a vast amount of available data.

In the last year our group has developed multiple methods that aim to introduce data priors in pose estimation for human bodies and human hands. In our work Pose-NDF (ECCV'22), we show that manifold learning can be used to learn pose manifolds of plausible human poses and that it is a viable alternative to previously existing data priors based on VAEs and GANs. The presented method outlines a completely new way of modeling and using data priors, which was acknowledged by reviewers and committee through awarding a Best Paper Honorable Mention Award of ECCV'22. Further, our TOCH method (ECCV'22) demonstrates how estimation of hand-object interaction sequences in motion can also heavily benefit from learning on existing datasets. TOCH solves several technical challenges that arise from fitting a MANO hand model by introducing a novel, object-centric representation for hand-object interaction.

**Real Virtual Humans** *Investigator: Gerard Pons-Moll*

Modelling humans is vital to understand the world around us. It is also central to a wide range of applications in gaming, animation, robotics, augmented and mixed reality, etc. At Real Virtual Humans we aim to build models of digital humans that look, move and behave like the real ones, in order to better understand the humans and the world we live in. Broadly, speaking we have been focusing on (i) modelling the appearance of the digital humans, (ii) modelling the objects and scenes around them, and (iii) using our models of humans and surroundings to jointly model human-object interactions in 3D.

Our work (PoseNDF, Tiwari et al. ECCV'22) demonstrated that implicit functions can be used to model high dimensional data manifolds such as 3D poses and not just 3D shapes as is commonly understood. Our model can act as data prior, correct unrealistic poses and animate along the pose manifold. This work received "Best student paper award" at ECCV'22 as it opened new avenues for learning data distributions. Apart from this we improved upon classical tasks like 3D registration, proposed a generative model of people in clothing, and addressed real world tasks like character animation.

To better understand the world around us we've proposed novel methods for 3D scene segmentation with very sparse data annotations (Box2Mesh, Chibane et al. ECCV'22 oral), scene editing and novel view synthesis. Our focus in this direction is to learn scene representations that are generalisable and can be learnt with limited data, as obtaining annotated 3D data in real world setting is hard.

Studying 3D scenes and humans in isolation is limiting as it does not capture the continuous and rich interactions between them. Our research is at the forefront of this new and challenging research avenue. We proposed the first large scale dataset of human-object interactions in 3D, BEHAVE (Bhatnagar et al. CVPR'22). Apart from this we proposed the first learning based methods to track human-object interactions using multi-view depth, monocular RGB as well as body mounted sensor.

**Multimodal Deep Learning** *Investigator: Zeynep Akata*

Our Multimodal Deep Learning group aims to develop deep learning methods that observe and process multimodal input coming from the environment, make further connections through inference and communicate the system output to the user. The group's current research focus is exploring the interplay between vision and language for several machine learning tasks. The research of Multimodal Deep Learning group is broadly divided into three subfields: disentangled and compositional representation learning, learning with less supervision and explainable machine learning.

In disentangled and compositional representation learning, our premise is that if there exists some basic kind of compositional structure in the world (i.e., smaller parts that make up larger objects), then adaptive algorithms should exploit that structure. We aim to develop truly compositional models that combine the generalization and interpretability of symbolic structures with the learning capabilities of contemporary deep learning algorithms.

In learning with less supervision subfield, our aim is to learn a model about the environment given a set of observations that belongs to a certain set of classes. The main challenge of this task is that the set of classes at training and at test time are disjoint or the supervision signal is not always clean. As, the classic supervised learning algorithms that rely on the full set of class labels can not be employed for the extreme case of this task, i.e. zero-shot learning, we use language as auxiliary information to build a structure in the label space.

To enable trust of the user, one aspect of deep learning to strengthen that has been getting increasingly popular is to understand the internal decision process of a network. Another one is to justify the output of the decision maker given the input in human understandable terms. These explanations either come in the form of language or visual justifications such as machine attention. The group focuses on generating visual explanations and pointing to the evidence for a classification decision of a deep multimodal learning framework.

**Combinatorial Computer Vision** *Investigator: Paul Swoboda*

Combinatorial optimization is a fundamental tool in machine learning that offers the promise to improve performance of deep learning systems further by incorporating explicit prior knowledge in the optimization. The research group Combinatorial Computer Vision studies mathematical abstractions of computer vision tasks posed as combinatorial optimization problems. One focus is on developing efficient algorithms to solve the ensuing problems and to benchmark them w.r.t. metrics defined by the application. Another field of research is on integrating combinatorial optimization problems into deep networks to jointly train the network end-to-end. Application areas include tracking, clustering, segmentation and correspondence problems. Moreover, the group conducts basic research into algorithm design.

Achievements during the reporting period include: (i) A general purpose solver for combinatorial problems from structured prediction tasks and a highly parallel extension of it. (ii) Tracking systems that achieve state of the art results for tracking objects observed by a single or multiple cameras. (iii) Highly parallel as well as densely connected clustering with multicut. Our works have been published in top-tier conferences such as CVPR, ICML, NeurIPS, AAAI, ECCV and ICCV.

**Robust Visual Learning** *Investigator: Margret Keuper*

The susceptibility of neural network based models to common corruptions as well as targeted attacks is a crucial problem that hampers the real-world deployment of otherwise well performing models in safety critical environments. Facilitating the learning of robust models is thus a key challenge for my group. Therefore, we have studied robust and non-robust models with respect to various properties that relate to the models' training schemes as well as to their architecture, with the aim to identify key properties that allow to build more robust models.

Regarding architectures, we have investigated the impact of aliasing in the feature maps and have leveraged our findings to build better architectures for adversarially robust models (Springer Machine Learning, ECCV'22). We also contributed a dataset that relates neural architectures to their robustness (ICLR'23) and we proposed an approach that allows for efficient multi-objective architecture search (ECCV'22). This way, the architecture of a network can not only be optimized for the resulting validation accuracy but also for additional criteria such as energy consumption or latency on specific hardware.

Regarding training schemes, we have studied the impact of adversarial training on different model generalization properties and their calibration (CVPRw'23, NeurIPS'22). Further, we have investigated two specifically relevant aspects: the effect of realistic optical aberrations on a model's behavior (NeurIPS'22 workshop) and adverse weather conditions, and provide training data for both (WACV'23 and ICCV'23 submission).

Regarding generative models, we have contributed methods to shape the learned representation space of deterministic autoencoders using Gaussian mixture models to promote more robust behavior (NeurIPS'21, NeurIPS'22). We also investigated spectral artifacts and their impact on image generation and sustainable deepFake detection (AAAI'21, IJCAI'21).

In light of the rather low robustness and poor interpretability of most current deep neural models, we have also conducted research in a different direction: the combination of deep learning models and a graphical model based optimization. We have been working on improving the expressiveness of graph-based grouping formulations (TPAMI'22) as well as on the end-to-end optimization of neural networks in conjunction with graph-based grouping formulations (ECML-PKDD'22, GCPR'22).

## Publications and Awards

Chapter 28 contains a detailed report of the publications, cooperations and awards of the reporting period. From the journal publications 17 have been published or accepted at either IEEE PAMI, IJCV, ACM TOG, or TACL. From the conference publications, 64 have been published or accepted at one of the major computer vision conferences (CVPR, ICCV, ECCV). An additional 19 have been published at major conference in machine learning (15 at NeurIPS, ICML, and ICLR; 4 at AAAI).

Members of the group received a range of prizes including: (i) personal awards such as early career awards (the German Pattern Recognition Awards for both 2021 (Zeynep Akata) and 2022 (Dengxin Dai), as well as the ECVA Young Researcher Award 2022 (Zeynep Akata)), a dissertation award (Jan Eric Lenssen), as well as master level awards (Julian Chibane, Anurag Das, Anna Kukleva); (ii) best paper awards such as ECCV 2022 Best Paper Honorable Mention, Best Student Paper Awards 3DV 2022, and Best Paper Honorable mention 3DV 2022; (iii) fellowships and research awards such as a Meta PhD Fellowship Award, a Snap Research Award, and a Huawei Research Award; and (iv) winning competitions such as the WAYMO Motion Prediction Challenge 2022 and CVPR SHARP'21 Workshop Winner. Also, members of the group have been awarded 14 outstanding or top reviewer awards at conferences such as NeurIPS, CVPR, ICCV and ECCV.

# 4 D3: Internet Architecture

## History and Group Organization

The Research Group Internet Architecture (D3) was established in January 2018. The group's organizational structure and more details of the members are provided later in this report.

Due to the COVID-19 pandemic, the overall research group grew slower than may have been expected. During the reporting period, the group hosted three senior researchers/research group leaders (financed by MPI), namely, Oliver Gasser, Tobias Fiebig (from April 2022), and Savvas Zannettou (untill October 2021). Each of the group leaders supervise their own PhD-students (co-supervised by myself) and conduct research in their areas. Also the group hosted four postdocs (2 female) and 20 Ph.D. students (4 female). During the reporting period five new students joined and five Ph.D. students (2 female) handed in their theses at Saarland University (two still have to defend). Moreover, the "last" student from Berlin graduated.

Yiting, who joined MPI-INF as independent W2-tenure track faculty in Fall 2020, set up her research group. Anja Feldmann acts as her mentor and her research group is associated with D3 with regards to infrastructure and administrative support. Yiting's research follows a cross-layer approach and covers broad topics for optimizing the network stack, including switch hardware, network protocols, software systems, and cloud applications.

## Vision and Research Strategy

The Internet is an immensely successful human-made artifact that has fundamentally changed society. In becoming such an immensely successful infrastructure, the use of the Internet and, consequently, the Internet itself has changed and continues to change, as highlighted by this group's research efforts.

These changes are in part driven by the user or eye-ball interests as and how content, including user-generated data, is made available. The AS-level topology of the Internet has also experienced significant changes over time: It has evolved from a highly hierarchical topology to a flatter (non-hierarchical, simpler) topology. Service providers, e.g., IoT platforms, and content delivery networks are relying on sophisticated back-office infrastructures.

The future challenges in this context are (i) continual observation of the underlying infrastructure, (ii) locating and fixing current performance and functionality, bottlenecks, (iii) Internet security measurements, (iv) understanding and building dependable digital infrastructure as a socio-technical system, (v) online societies, and (vi) understanding the interactions of the infrastructure with the society.

We follow a data-driven systems research agenda to tackle these challenges: Collect data from operational networks, analyze them using big-data analytics to identify invariants, revisit assumptions, and detect and localize performance bottlenecks in the Internet. We also use simulation environments to validate our analysis and to support "what-if" studies. The insights obtained from the measurements form the foundation for shaping the future Internet via optimizations and alternative designs. Hereby, our output includes protocol enhancements, novel network management tools, concepts for software-defined networking, as well as studies of the impact of the technology on the society.

## Research Areas and Achievements

The main research areas of the department are Internet traffic analysis, innovative traffic control to future proof the Internet, Internet security measurement, dependable digital infrastructure, online social networks, as well as network neutrality and the sharing economy. The latter is addressed within the context of the Weizenbaum Institute.

**Internet Traffic Analysis** *Investigator: Anja Feldmann*

The Internet is a massively heterogeneous and also continuously evolving ecosystem. Naturally, not one vantage point can accurately capture the breadth of these changes. This limitation notwithstanding, there is a dire need to monitor and analyze the use of the Internet infrastructure as well as the characteristics of the infrastructure itself, especially given the constantly evolving nature of this ecosystem. Therefore, our research focuses on meticulously gathering measurements from diverse vantage points. We systematically analyze these measurements to characterize the performance and the operation of the Internet ecosystem.

**Research Highlight: Deep Dive into the IoT Backend Ecosystem** *Investigators: Said Jawad Saidi, Oliver Gasser, Anja Feldmann in cooperation with Srdjan Matic (IMDEA Software Institute) and Georgios Smaragdakis (TU Delft)*

Internet of Things (IoT) devices are becoming increasingly ubiquitous, e.g., at home, in enterprise environments, and in production lines. To support the advanced functionalities of IoT devices, IoT vendors as well as service and cloud companies operate IoT backends—the focus of this research project. In this project we follow up on our previous work of detecting IoT devices in the wild, and propose a methodology to identify and locate them IoT backends by (a) compiling a list of domains used exclusively by major IoT backend providers and (b) then identifying their server IP addresses. Our methodology relies on a fusion of information from public documentation, passive DNS, and active measurements.

**Innovative Traffic Control to Future-Proof the Internet** *Investigator: Anja Feldmann*

The continually evolving nature of the Internet ecosystem routinely introduces new, unforeseen challenges. With the increasing adoption of sensors and Internet of Things (IoT) devices, we are generating an unprecedented volume of data. As if this immense amount of data was not enough, such data now comes from divers endpoints, widely distributed throughout the network (at its edge). Still, this opens new opportunities, e.g., for scalable DDoS analysis.

The changes in the Internet ecosystem also provide new opportunities to revisit some long standing networking problems and design new practical, scalable solutions that exploit or leverage these recent changes, e.g., DDoS mitigation, revisiting congestion control, or exploiting new opportunities for video transmissions.

**Research Highlight: Collaborative DDoS Mitigation** *Investigators: Daniel Wagner, Anja Feldmann, and Christoph Dietzel, in cooperation with Daniel Kopp, Matthias Wichtlhuber (DE-CIX), Oliver Hohlfeld (Brandenburg University of Technology), and Georgios Smaragdakis (TU Delft)*

Despite ongoing research on DDoS detection and mitigation paired with improved understanding about adversary strategies, DDoS attacks are still on the rise and at an all time high. To date, attackers incorporate more sophisticated techniques and exploit by far more different mechanisms and protocols to form DDoS attacks at unparalleled threat levels.

In this project, we measure the ability of Internet Exchange Points (IXPs) to mitigate amplification DDoS attacks. Located in the heart of the Internet, they are closer to the source of attacks than conventional mitigation facilities that are typically located at the attack's destination. However, the IXPs' location usually lacks a holistic view on the attack traffic, as routes exist towards the target that bypass IXPs. The remaining fraction of the attack that crosses the IXP may not be large enough for local detection mechanisms to detect the traffic as malicious. To cope with this, we propose a collaboration between IXPs to get a more informed view on the attacks and improve the local attack detection. We unify the data of 11 IXPs across Europe and North America and identify 120k amplification DDoS attack events throughout a period of 6 months.

**Internet Security Measurements** *Investigator: Oliver Gasser*

As the Internet is becoming more and more ubiquitous in people's everyday lives—pushed even more so by the COVID-19 pandemic—aspects of security in Internet-connected systems are increasing in importance. In addition, the deployment of IPv6-enabled devices in the Internet is steadily increasing, with more than 40% of Google users using IPv6 as of March 2023. To analyze the security of Internet-connected devices, we devise measurement techniques to reach many different devices in the IPv4 and IPv6 Internet, and we conduct Internet measurements and characterize different aspects of security of these devices.

**Research Highlight: Illuminating Large-Scale IPv6 Scanning in the Internet**
*Investigators: Oliver Gasser in cooperation with Philipp Richter and Arthur Berger (Akamai)*

While scans of the IPv4 space are ubiquitous, today little is known about scanning activity in the IPv6 Internet. In this project, we present a longitudinal and detailed empirical study on large-scale IPv6 scanning behavior in the Internet, based on firewall logs captured at some 230,000 hosts of a major Content Distribution Network (CDN). We develop methods to identify IPv6 scans, assess current and past levels of IPv6 scanning activity, and study dominant characteristics of scans, including scanner origins, targeted services, and insights on how scanners find target IPv6 addresses. Where possible, we compare our findings to what can be assessed from publicly available traces. Our work identifies and highlights new challenges to detect scanning activity in the IPv6 Internet, and uncovers that today's scans of the IPv6 space show widely different characteristics when compared to the more well-known IPv4 scans.

**Dependable Infrastructure as a Socio-Technical System** *Investigator: Tobias Fiebig*

Over the past half century, the rise and accelerating evolution of digital technology shaped society as much as social dynamics shape this technology. Our research in this area focuses on the key questions in this interaction, ranging from how building digital infrastructure is shaped by social and societal interactions, to how technical developments ultimately shape how we live together as humans, from the smallest to the societal scale. Ultimately, the key-question of our work is: "How can we build IT systems that are reliable and secure, which support society in answering the practical problems of our time without introducing new threats?"

To tackle this question we combine the core-competencies of our group in Internet measurement with methods from social sciences, human factors, and the governance domain. This allows us to systematically leverage quantitative data for large scale perspectives, e.g., to assess societal shifts towards centralization, and combine this with qualitative methods to *explain* the root causes behind our quantitative findings, may it be interviews with operators and decision makers regarding root-causes for centralization or an in-depth protocol analysis to explain effects protocol's complexity had on the Internet as a whole.

**Research Highlight: Analyzing the Cloudification of Higher Education** *Investigators: Tobias Fiebig, Mannat Kaur, and Simran Munnot*

The digital transformation of academia is one of the major technical challenges of our time. However, as all challenges, this transformation is not without risks. Following common industry paradigms, universities now commonly look at infrastructure provided just by a few major cloud operators.

In this work, we investigate in how far universities depend on an small set of digital infrastructure providers. We were the first to conduct comprehensive measurements characterizing these developments from 2015 onwards, and published further analyses on the organizational implications of these developments. Our work illustrates how the progressing cloudification of academia impacts core-values like academic freedom, and ties differences in cloud adoption between several countries to stark differences in academic culture. Furthermore, we provide a clear long-term agenda to preserve academic freedom, as well as researchers and students privacy.

**Online Social Networks** *Investigator: Savvas Zannettou*

Over the past decade, Online Social Networks (OSN) have exploded in popularity, mainly because they help people in addressing their communication, information acquisition, and entertainment needs. At the same time, OSNs offer a fertile ground for the creation and amplification of important socio-technical issues like the spread of false information and hate speech. Therefore, it is imperative to analyze data from OSNs to understand, detect and mitigate these issues to minimize possible consequences both on the online and offline world (e.g., extensive dissemination of false information affecting people's voting decisions). In this research area, we aim to analyze activity on OSNs to understand emerging socio-technical issues, develop tools/techniques to detect potentially harmful information and develop and assess the effectiveness of various mitigation strategies.

**Research Highlight: Do Platform Migrations Compromise Content Moderation? Evidence from r/The_Donald and r/Incels** *Investigators: Savvas Zannettou in cooperation with Shagun Jhaver (Rutgers University, USA), Jeremy Blackburn (Binghamton University, USA), Emiliano De Cristofaro (University College United Kingdom), Gianluca Stringhini (Boston University), Robert West (EPFL, Switzerland), Krishna P. Gummadi (MPI-SWS, Germany)*

Analyzing content moderation online is important for several reasons. First, social media and other online platforms have become major sources of information and communication, shaping public discourse and opinions. The content shared on these platforms can have significant consequences for individuals, communities, and society as a whole. Thus, understanding the challenges and complexities of moderating online content is crucial for ensuring that these platforms are safe and inclusive spaces for all users. Second, content moderation involves a range of technical, social, and ethical issues that require interdisciplinary expertise. Studying content moderation online involves understanding the technical mechanisms used to identify and remove harmful content, as well as the social and cultural contexts in which these mechanisms operate. Moreover, content moderation online is a constantly evolving field, as new technologies and social dynamics emerge. In this line of work, our goal is to analyze and understand multiple aspects of content moderation, including how soft moderation interventions (i.e., warning labels are applied online) and effective they are, what happens after online platforms take moderation action on specific online communities (i.e., community bans), and how we can design systems to automatically identify accounts that are state-sponsored trolls and are involved in misinformative campaigns online.

## Projects and Cooperations

Since 2017, Anja Feldmann is PI at the Weizenbaum Institute, Berlin. The Weizenbaum Institute for Networked Society is the German Internet Institute, a place of excellent research on the transformation and design processes of digital change. In the spirit of Joseph Weizenbaum, we research the necessary framework conditions, means, and processes for individual and social self-determination in a networked society. The project is funded by the Federal Ministry of Education and Research.

In addition we are involved in the BMBF SupraCoNex project—a cooperation with Hochschule Nordhausen, Technische Universität Braunschweig, NewMedia-Net GmbH and Fraunhofer-Gesellschaft zur Förderung der angewandten Forschung e.V.; the 6G Research and Innovation Cluster (6G-RIC)—a research hub designed to lay the scientific and technical foundations for the next generation of mobile communications (6G) across all technology levels; the Quantum Internet Alliance—a collaboration between the TU Delft and TNO to develop a full-stack prototype network connecting distant cities.

## Publications, Awards, and Media Coverage

Despite all pandemic-related obstacles, members of the group were able to publish more than 60 papers, among them many at top-tier conferences and journals, e.g., 5(+7) papers at

Internet measurement conferences (IMC+PAM), 5 papers at security venues (CCS, USENIX, NDSS, S&P), and 9 papers at Web related conferences (WebSci, CSCW, WWW, ICWSM), win multiple awards for our research, including the IETF applied networking research prize for the CCS 2022 paper and the best paper award ad PAM 2023, receive excellent press coverage for some of our work, serve on steering committees of ACM ToN, ACM IMC, ACM CoNEXT, and PAM, conduct online/hybrid lectures and even an in-person seminar. Moreover, three former members of the group got their first assistant professorship positions.

Additionally, we (Anja Feldmann and Lars Prehn) contributed to the community by organizing a shadow TPC for the ACM Internet Measurement Conference 2022 as Shadow TPC Chairs. The feedback of the students was excellent. They all agreed that they are much better able to understand the process with all its pros and cons.

Anja Feldmann also contributes to the MPG environment by participating in a multitude of different MPG Commissions. She is, e.g., member of the presidential commission governance, the CPTS perspective commission, the advisory board digitalization, as well as the IT-Sicherheitskommission. Oliver Gasser is the elected Scientific Staff Member Representative of the Institute since mid 2021.

During the last two years, the result of multiple research works from our group was covered in the media including:

- Positionspapier zur Technologische Souveränität und die Rolle der Kommunikationssysteme in einer digitalisierten Gesellschaft Fachkreis Kommunikationstechnologien, Beraterkreis für das Referat „Vernetzung und Sicherheit digitaler Systeme“ im Bundesministerium und Bildung und Forschung, **April 2021**.
- A Year in Lockdown: How the Waves of COVID-19 Impact Internet Traffic by Anja Feldmann, Oliver Gasser, Franziska Lichtblau, Enric Pujol, Ingmar Poese, Christoph Dietzel, Daniel Wagner, Matthias Wichtlhuber, Juan Tapiador, Narseo Vallina-Rodriguez, Oliver Hohlfeld, Georgios Smaragdakis. Communications of the ACM, July 2021, Vol. 64 No. 7, Pages 101-108, **July 2021**.
- “Trotz Privacy Extensions: IPv6-Adressen für andauerndes Tracking nutzbar” by Stefan Krempl, heise online `https://www.heise.de/news/Trotz-Privacy-Extensions-IPv6-Adressen-fuer-andauerndes-Tracking-nutzbar-7186203.html`, **July 2022**.
- “Three-quarters of Dutch student data stored on American tech giants’ cloud” by Tobias Fiebig `https://nltimes.nl/2022/10/17/three-quarters-dutch-student-data-stored-american-tech-giants-cloud`, **October 2022**. This research led to an official request in the Dutch Parliament: `https://www.tweedekamer.nl/kamerstukken/kamervragen/detail?id=2022Z19631&did=2022D45952`

## Teaching

Our group is active in teaching and has offered the Data Networks Core Lecture in Summer Term 2021 (online), Summer Term 2022 (hybrid), and Summer Term 2023 (in person with hybrid option). In addition we offer the seminar Hot Topics in Data Networks Seminar in the Fall terms. In addition, group members have supervised 13 Master and Bachelor Thesis during the reporting period.

# 5 D4: Computer Graphics

## Group Overview

The computer graphics group (D4) was established in 1999 with the appointment of Hans-Peter Seidel. Karol Myszkowski joined in 2000. Christian Theobalt, who had been a tenured senior researcher with the group since 2009, was appointed director in 2021 and is now heading his own department on Visual Computing and Artificial Intelligence (D6).

Over the last two decades the group has attracted a highly talented pool of young researchers, and we have been home to an exceptional team of PhD students and postdocs. Meanwhile, more than 70 former group members hold faculty positions at other places.

At the time of writing, the group encompasses one senior researcher with tenure (K. Myszkowski), five group leaders and senior researchers without tenure (V. Babaei, G. Singh, R. Zayer, Paul Strohmeier and T. Leimkühler (co-appointed with D6)), three postdocs, and 12 PhD students. Three group members completed and handed in their PhD thesis during the reporting period, and several of our young researchers and former PhD students got offers for faculty appointments[1], or postdoc and senior postdoc positions[2]. Full details on current and recent group members are provided in the main part of the report.

## Vision and Research Strategy

During the last few decades computer graphics firmly established itself as a core discipline within computer science. New and emerging technologies such as digital media, social networks, digital television, digital photography and the rapid development of new sensing devices, telepresence, virtual and augmented reality, computational fabrication, further indicate its potential and pose new challenges in the years to come.

To address these challenges, and to seamlessly blend real and synthetic footage, we had early on adopted a new and more integrated scientific view of computer graphics as *3D Image Analysis and Synthesis* that takes into account the whole image processing pipeline from scene acquisition to scene reconstruction to scene editing to scene rendering. We also take into account human perception on all levels of the pipeline, and we exploit the abundance of digital visual data and novel concepts from machine learning to extract powerful priors that can assist us during the acquisition, reconstruction, editing, and image formation processes.

Our vision and long term goal are completely immersive, interactive, and visually rich environments with sophisticated scene representations and the highest visual quality, fused

[1] M. Chu (Beijing Univ., China, 2022), Y. Gryaditskaya (Univ. of Surrey, UK, 2022), P. Kellnhofer (TU Delft, NL, 2021), A. Serrano (Univ. Zaragoza, Spain, 2021), C. Tursun (Univ. Groningen, NL, 2021), R. Zayer (Univ. East Anglia, UK, 2023), Q. Zheng (Chinese Academy Sciences, Inst. Software, Beijing, China, 2021)

[2] J. Hladky (Huawei Research, Munich), C. Reed (King's College London, UK)

seamlessly with the real world. Standard 2D screens are being replaced with high dynamic range displays, stereo and automultiscopic screens, portable and wearable displays. Imaging algorithms with embedded perceptual models ensure that the perceived quality and viewing comfort is maximized. Interaction is intuitive and light weight. We are also reach out into new domains, such as computational fabrication, and sensorimotor interaction.

## Research Areas and Achievements

Our research is currently organized into the following six research areas (coordinators in brackets):

- Digital Geometry Processing (R. Zayer)
- Sampling, Image Synthesis, and Machine Learning (T. Leimkühler and G. Singh)
- Perception: HDR Imaging, VR, and Material Appearance (K. Myszkowski)
- Computational Design and Fabrication (V. Babaei)
- Sensorimotor Interaction (P. Strohmeier)

However, we are not organized into disjoint subgroups, and there is little hierarchy. While each of the areas has its specific focus, some of them also have significant overlaps. Likewise, the students and researchers working in each area are dynamically formed teams rather than spedifically dedicated staff. The senior researchers and group leaders together with Hans-Peter Seidel serve as an internal steering committee for the group. They also act as advisors or co-advisors of doctoral students.

During the reporting period we achieved fundamental results on correlated sampling and on bridging the gap between Monte Carlo rendering and neural radiance fields, and we substantially contributed towards improving the visual experience in VR and games and material appearance perception. We achieved groundbreaking results on neural inverse design in computational design, and we significantly advanced tactile rendering. We also continued our successful line of research on the mesh matrix formalism for geometric computing.

### Digital Geometry Processing

We continued our successful line of research on casting geometric computations and modifications in terms of concise linear algebra operations, based on the mesh matrix representation. During the reporting period we focused on the analysis of topological features on surfaces and developed a variational loop shrinking algorithm for handle and tunnel detection and Reeb Graph construction on surfaces. Our formulation tracks the evolution of a diffusion front randomly initiated on a single location on the surface. We develop a dynamic data structure for tracking the evolution on the surface encoded as a sparse matrix which serves for performing both diffusion numerics and loop detection and acts as the workhorse of our fully parallel implementation. As a byproduct, our approach can be used to construct Reeb graphs by diffusion (CGF'21, EG'23).

In a second line of research, together with researchers from D6, we investigated neural representations for the reconstruction and meaningful control of fluids. In particular, we proposed a novel data-driven conditional adversarial model that solves the challenging and theoretically ill-posed problem of deriving plausible velocity fields from a single frame of a density field (SIG'21). Besides density modifications, our generative model is the first to enable the control of the results using all of the following control modalities: obstacles, physical parameters, kinetic energy, and vorticity. Our method is based on a new conditional generative adversarial neural network that explicitly embeds physical quantities into the learned latent space, and a new cyclic adversarial network design for control disentanglement. In follow-up work we developed the first method to reconstruct dynamic fluid phenomena by leveraging the governing physics (ie, Navier -Stokes equations) in an end-to-end optimization from a mere set of sparse video frames (without taking lighting conditions, geometry information, or boundary conditions as input) (SIG'22).

## Sampling, Image Synthesis, and Machine Learning

Synthesizing realistic images typically involves computing high-dimensional light transport integrals. Traditional rendering algorithms aim towards designing sampling strategies that can ameliorate the error during the estimation of these integrals. Recent advances in neural rendering have opened an exciting line of research where the goal is to synthesize images with the help of deep learning. We develop state-of-the-art algorithms that operate at the intersection between traditional rendering and machine learning-based approaches. At the core, we focus on sampling problems for computer graphics, vision and machine learning, while integrating neural generative priors into the pipeline. We are also interested in interactive rendering and editing of traditional and neural scene representations.

View-dependent effects such as reflections pose a substantial challenge for image-based and neural rendering algorithms. Curved reflectors are particularly hard, as they lead to highly non-linear flows as the camera moves. We have introduced a new point-based representation to compute Neural Point Catacaustics allowing novel-view synthesis of scenes with curved reflectors, from a set of casually-captured input photos (SIGAsia'22).

Neural Radience Fields (NeRFs) have revolutionized novel-view synthesis for captured scenes, but their implicit representation makes them difficult to edit. We have introduced NeRFshop, a novel end-to-end method that allows users to interactively select and deform objects through cage-based transformations, thus enabling fine user control for the selection of regions or objects to edit (ACM I3D'23). Another limitation of NeRFs is that they are limited to synthesize images under the original fixed lighting conditions. To tackle relighting, several recent methods propose to disentangle reflectance and illumination from the radiance field. These methods can cope with solid objects with opaque surfaces, but participating media are neglected. We propose to learn neural representations for participating media with a complete simulation of global illumination (NeurIPS'21). We estimate direct illumination via ray tracing and compute indirect illumination with spherical harmonics. Our approach avoids computing the lengthy indirect bounces and does not suffer from energy loss. This project is a proof of concept work that demonstrates how to bridge the gap between traditional Monte Carlo rendering and neural radiance fields.

Current 2D Generative Adverarial Networks (GANs) produce photorealistic renderings of portrait images. Embedding real images into the latent space of such models enables high-level image editing. We have introduce FreeStyleGAN, a new approach that generates an image with StyleGAN defined by a precise 3D camera. Our solution provides the first truly free-viewpoint rendering of realistic faces at interactive rates, using only a small number of casual photos as input, while simultaneously allowing semantic editing capabilities, such as facial expression or lighting changes (SIGAsia'21).

We introduced QuadStream, a new streaming rendering content representation that reduces motion-to-photon latency by allowing clients to efficiently render novel views without artifacts caused by disocclusion events. The approach achieves superior quality compared both to video data streaming methods, and to geometry base streaming. (EG'21, SIGAsia'22)

Sampling is at the core of many applications in computer graphics. Our research aims at the theoretical understanding of sampling distributions, and we build tools to analyze the local and non-local correlations in sample patterns. We proposed a novel multi-class point optimization based on continuous Wasserstein barycenters. Our formulation is designed to handle hundreds to thousands of optimization objectives and comes with a practical optimization scheme (SIGAsia'22). For perceptual rendering, the samples are optimized irrespective of the scene. We therefore developed a novel perceptual error optimization scheme for Monte Carlo rendering that can be tailored to a given scene. The method distributes the error as visualy pleasing blue noise in image space (ACM TOG'22).

We reformulate Monte Carlo estimation as a regression-based problem where traditional Monte Carlo estimation can be seen as fitting a constant to the function evaluations. By using a polynomial function to fit the function values, the variance of the estimator is reduced, resulting in a provable improvement over the classical Monte Carlo estimator (SIG'22). We built a powerful pipeline for point-pattern synthesis using Gabor and random filters (requiring significantly less feature maps compared to existing VGG-19-based methods) (EGSR'22), and we are actively investigating the editing of point patterns by image manipulation.

### Perception: HDR Imaging, VR, and Material Appearance

Our research aims to advance our knowledge on image perception and to apply this knowledge to the development of imaging algorithms with embedded computational models of the human visual system (HVS). This approach offers significant improvements in both computational performance and perceived image quality. In particular, we aim for the exploitation of perceptual effects as a means of overcoming the physical limitations of display devices and enhancing the apparent image quality. During the reporting period we continued our successful and long standing line of research on HDR imaging, and we put special emphasis on Improving the Visual Experience in VR, and on Material Appearance.

While HDR content is becoming ubiquitous, the dynamic range of common display devices is still limited, therefore tone mapping (TM) remains a key challenge for image visualization. Neural networks have shown promise in TM, but their performance has been limited by training data. We proposed a novel self-supervised TM operator trained at test time for each HDR image using a carefully designed loss function based on contrast perception (by reformulating classic VGG feature maps into feature contrast maps) (CGF'22). We also addressed HDR video reconstruction (Comp.& Graph.'22).

Recently, neural scene representations have become a viable alternative for novel view synthesis, but the handling of transparent objects with complex refraction has been problematic.

In our work on eikonal fields for refractive novel-view synthesis we have demonstrated how to integrate the physical laws of eikonal light transport into the general NeRF setup and adapt an implicit representation that can model transparent and refractive objects with a spatially varying index of refraction, leading to high-quality novel view reconstruction of refractive objects (SIG'22).

The rendering performance is an important requirement in VR applications. Recently, there is a shift in real-time graphics from rendering with a fixed resolution and refresh rate to a more adaptive approach, in which we control spatiotemporal resolution in order to maximize the quality under a given rendering budget. We proposed a novel rendering method that takes advantage of the limitations of the visual system to reallocate the rendering budget to the most vital part of the spatiotemporal domain. The key component of the method is a novel metric, which considers how the judder, aliasing, and blur artifacts introduced by variable rate shading (VRS) at a given refresh rate are masked by hold-type blur, eye motion blur, and limited spatiotemporal sensitivity of the visual system (SIGAsia'21). Another important aspect of VR is image reproduction on head-mounted displays. Dimming a display can be beneficial for the VR experience as it reduces the visibility of flicker, saves power, prolongs battery life, and reduces the cost of the device. A downside of this approach is the reduced sensitivity to stereoscopic depth cues. This motivates our method for enhancing contrast at low luminance levels, intended to improve the reliability of stereoscopic depth cues. The proposed method improves the user experience for VR headsets that need to operate at low power and those that cannot achieve high refresh rates (SIG'22). Foveated rendering can greatly improve the computational performance in VR setups. The use of Generative Adversarial Networks (GANs) has recently been shown to be a promising solution for such a task, as they can successfully hallucinate missing image information. We addressed the problem of efficiently guiding the training of foveated reconstruction techniques such that they are able to recover perceptually important image features (ACM TAP'23). Faces are widely used in gaming, entertainment, and social applications, and humans are particularly sensitive to their appearance. In collaboration with Reality Labs (Meta), we generated a new dataset of high-quality and demographically-balanced face scans and conducted a large-scale perceptual study to subjectively evaluate them. Our dataset consists of over 84,000 quality comparisons, making it the largest-ever psychophysical dataset for geometric distortions. We demonstrated how our data can be used for applications like state-of-the-art metrics calibrations, compression, and level-of-detail rendering (SIGAsia'22).

Material appearance is determined not only by material reflectance, but also by surface geometry and illumination. We have analyzed these effects by collecting a large-scale dataset of perceptual ratings of appearance attributes with more than 215,680 responses for 42,120 distinct combinations of material, shape, and illumination, and by using this dataset to train a deep-learning architecture for predicting perceptual attributes that correlate with human judgments. We demonstrate the consistent and robust behavior of our predictor in various challenging scenarios, which, for the first time, enables perceived material attributes from general 2D images (SIG'21, TVC'21). In recent follow-up work, we investigated the problem of gloss appearance matching between the real world and display depiction as a function of material glossiness, surface geometry, scene illumination,, and display luminance (SIGAsia'22).

## Computational Design and Fabrication

The recent wide availability of advanced manufacturing hardware has triggered huge interest in academia and industry. We have a particular interest in the visual appearance of objects and strive for better algorithms for appearance 3D printing and for novel digital manufacturing platforms, such as laser marking. In addition, we develop algorithmic tools to evaluate, represent, and synthesize products with improved or completely novel functions, and we aim to integrate the fabrication hardware into the computational pipeline.

In computational design and fabrication, neural networks are becoming important surrogates for bulky forward simulations. A long-standing, intertwined question is that of inverse design: how to compute a design that satisfies a desired target performance? We demonstrated that the piecewise linear property, very common in everyday neural networks, allows for an inverse design formulation based on mixed-integer linear programming. Our mixed-integer inverse design uncovers globally optimal or near optimal solutions in a principled manner. Furthermore, our method significantly facilitates emerging, but challenging, combinatorial inverse design tasks, such as material selection (SIG'22).

The key property of a successful neural inverse method is the performance of its solutions when deployed in the real world, i.e., on the native forward process (and not only the learned surrogate). We developed Autoinverse, a highly automated approach for inverting neural network surrogates. Our main insight is to seek inverse solutions in the vicinity of reliable data which have been sampled form the forward process and used for training the surrogate model. Autoinverse finds such solutions by taking into account the predictive uncertainty of the surrogate and minimizing it during the inversion (NeurIPS'22 (spotlight)).

Dissolution processes are fascinating. Objects with the same mass but different shapes can dissolve via different dynamics, resulting in vastly different release profiles. We developed Shape from Release, a novel method for the inverse design and fabrication of controlled release structures. We start with a simple physically inspired differentiable forward model of dissolution and formulate our inverse design as a PDE-constrained topology optimization that has access to analytical derivatives obtained via sensitivity analysis (SIGAsia'22).

Despite the ongoing advances of computational methods in creating innovative content for manufacturing, the potential of these methods for optimizing or controlling the manufacturing processes in an intelligent way is largely untapped. Our research advocates the integration of the fabrication hardware into the computational pipeline and not treating the hardware merely as a "consumer" of the created content. We developed reinforcement learning based controllers for challenging additive manufacturing setups, such as direkt ink writing with low-viscosity materials (SIG'22), and multi-material fusion (ICRA'23).

## Sensorimotor Interaction

Our research aims to advance the understanding of sensorimotor interactions and apply this knowledge to the development of new technologies. We are currently exploring sensorimotor loops in three domains. The first is Mediation & Tacton Design, in which we explore ways in which technology can represent information. Next, we investigate vibrotactile rendering, particularly within the context of augmented tactile reality. Finally, we explore how we might enhance the experience of our body through human augmentation. Through the combination of these research areas, we try to develop a comprehensive understanding of how sensorimotor interaction can be used in the broader field of human-computer interaction (HCI).

We have developed TactJam, a software and hardware suite for designing vibrotactile icons on the body (TEI'22), and we have explored the use of such pre-reflective cues in the design of hermeneutic symbols, demonstrating that incorporating embodied mediation can enhance the effacy of symbol encoding and broaden the affective range of tactile symbols (CHI'23).

Many material experiences can be created using only vibration, including friction, compliance, torsion, bending, and elasticity. Our work on vibrotactile rendering demonstrates that these experiences all rely on the same perceptual mechanism, and suggests that the specifics of a material experience depend less on the rendering algorithm but largely on the action performed when interacting with the material (CHI'23). In HCI research, such haptic systems are seldom used, due to their complexity. We created Haptic Servos, accessible haptic rendering devices with low latency, which allows novices to set up a basic rendering system in minutes (CHI'23 (Honorable Mention)).

We have engaged in human augmentation by designing novel interfaces: In collaboration with Saarland University, we created a handheld printer for collaborative circuit design, where control is shared between user and computer (CHI'22). We have also worked on textile-based devices, such as SingingKnit, which measures muscle activity during singing to augment vocal performances and demonstrates how biofeedback can change body perception and artistic action (AH'22, 2xTEI'23)

### Software and Datasets

As part of our research we have developed several libraries, tools, and large corpora of reference data sets that are being made available to the research community at large. This includes our award-winning deeptech startup Oraclase for AI aided laser material processing, our software on profile detection for automatic door and window profile classification, the PFSTOOLS for processing high dynamic range images and video, the LocVis dataset of locally annotated images, the TactJam suite for creating and sharing low fidelity prototypes of on-body vibrotactile feedback, Haptic Servo for creating and augmenting material experiences, and our Interactive Shoe platform for augmented shoe design.

## Some Performance Indicators

We made significant progress in all of our research areas. In the reporting period from spring 2021 to spring 2023, we have published 58 peer-reviewd papers in high quality conferences and journals. This includes 18 papers at SIGGRAPH/SIGGRAPH Asia/ACM TOG, 5 papers at CHI, 14 papers at Eurographics/CGF/ACM TAP/TEI, and 10 papers at ICCV/ECCV/CVPR/NeuRIPS (see the main part of the report for details).

We actively participate in the program committees of major conferences (SIGGRAPH, SIGGRAPH Asia, CHI, Eurographics, Pacific Graphics, EGSR, TEI, AH, VMV, ICCV, ECCV, CVPR), and we hold editorial board seats with journals such as ACM TAP, CAGD, GMOD, J. Virtual Reality and Broadcasting, Int. J. Shape Modeling, and Visual Informatics. We have given numerous invited talks and tutorial presentations at major national and international events (see main part). Our software and datasets have been successfully used in a variety of projects (see main part), and many young researchers from the group have spread out to other institutions.

The group has been cooperating with a wide range of research groups worldwide. Cooperations that have led to joint publications during the reporting period include Stanford, MIT, Princeton, U. Waterloo, McGill, U. Maryland, UC David, Cambridge U., UCL London, CNRS, INRIA, TU Delft, IST Austria, TU Graz, USI Lugano, U. Zaragoza In addition, we also collaborated with some leading industrial research labs, including Google, Reality Labs (Meta), Ndivia, and Adobe. On a European level we cooperate with the CYENS Research Centre of Excellence in Cyprus (Member of the Board), and we participate in the Horizon 2020 Training Networks DISTRO and RealVision. There are also several collaborations within the institute and with other groups on campus.

## Awards and Selected Academic Activities

Hans-Peter Seidel became a member of the ACM SIGGRAPH Academy in 2022. He acted as a member of the DFG Senate and as a member of the Senate Evaluation Committee (SAE) of the Leibniz Association. Karol Myszkowski became Technical Papers Co-Chair for Eurographics'23. Vahid Babaei co-chairs Short Papers, and Gurprit Singh co-chairs the posters program for the event.

Vahid Babaei won the Hermann Neuhaus Prize 2022 of the Max Planck Society. His deeptech start-up effort (Oraclase) on laser marking won both a MAX!mize startup grant by the Max Planck Society and an EXIST Transfer Grant by the German Federal Ministry for Economic Affiars and Climate Action (BMWK). Our former postdoc Ana Serrano won the Eurographics Young Research Award 2023. Several group members won best paper awards at leading international venues.

# 6 D5: Databases and Information Systems

## History and Group Organization

D5 was established in 2003. With the upcoming retirement of its director (Gerhard Weikum: August 31, 2023), it has largely been phasing over the past years. At the time of the previous biennial report (May 2021), the group comprised 5 senior researchers, 19 doctoral students and 4 postdocs. As of now (May 2023), the group has shrunk to 1 senior researcher, 8 doctoral students and 1 postdoc. Except for some of the students whose graduation is expected in 2024, all researchers will leave the institute by the end of 2023.

## Scientific Vision and Research Areas

Our general objective is to develop methods for *knowledge discovery* in a broad sense: extracting, organizing, searching and exploring various kinds of knowledge from structured, semistructured, textual and multimodal information sources. Our approaches combine concepts, models and algorithms from several fields, including database systems, information retrieval, natural language processing, web science, data mining and machine learning.

Our long-term research has been driven by the overarching vision of automatically constructing, growing and curating large-scale and high-quality knowledge bases from Internet sources. We have spearheaded this research avenue, referring to it as *knowledge harvesting*. To boost search, data analysis and language understanding, machines need to be equipped with comprehensive knowledge about the world's entities, their semantic properties and their relationships among each other. In addition to such encyclopedic facts, machines should also have commonsense knowledge about properties of everyday objects and human activities, and should even capture socio-cultural contexts of propositions.

With the success of web-scale knowledge bases (KBs, aka KGs: knowledge graphs), from early blueprints (including our work) to wide industrial adoption, the research focus over the last years has shifted to advancing the scope and coverage of "machine knowledge" and supporting the KB life-cycle in terms of growth and quality assurance. Also, with the advent of new applications like question answering, conversational bots and other tasks that involve natural language understanding, major efforts have been going into leveraging KBs for these use cases. Methodologically, all this involves machine learning, often based on Transformers and large language models. The latter can be viewed as latent knowledge capturing the content of huge corpora into many billions of neural-network parameters. There is speculation that these trends could make explicit KBs obsolete. We do not share this opinion and rather believe that machine learning and machine knowledge are both needed and complement each other: the more a computer knows the better it can learn, and the better learner can acquire more and deeper knowledge.

Within this scope, our research has been organized into four technical areas, each coordinated by a senior researcher:

- Knowledge Base Construction and Curation (coordinated by Simon Razniewski)
- Information Retrieval and Content Analysis (coordinated by Andrew Yates)
- Question Answering (coordinated by Rishiraj Saha Roy)
- Responsible Data Science (coordinated by Gerhard Weikum)

## Contributions and Impact: Long-Term Results

Enhancing computers with "machine knowledge" that can power intelligent applications by an epistemic backbone, has been a long-standing goal of computer science. Major advances in *knowledge harvesting* from web contents, with our group as a trendsetter, have made this formerly elusive vision practically viable.

Our work on knowledge harvesting and automatic knowledge base construction was motivated by the objective of semantic search, starting in 2004. Later it became the Yago-Naga project, with the first release of the Yago knowledge base (https://yago-knowledge.org) in February 2007. The unique strength of Yago is its high-quality type system with hundred thousands of classes. When IBM Watson won the Jeopardy quiz show, it leveraged Yago's knowledge of fine-grained entity types for semantic type checking.

**Impact:** Knowledge harvesting has been adopted at big industrial stakeholders, and knowledge bases (or knowledge graphs, as industry calls them) have become a key asset in semantic search (for queries about entities), question answering, analytics (e. g., aggregating by entities or types), recommendations and data integration (i. e., to combine heterogeneous datasets). Examples are the knowledge graphs for search engines (e. g., Google, Bing, Baidu) as well as domain-specific knowledge bases (e. g., Amazon, Alibaba, Bloomberg, Mayo Clinic). In addition, knowledge bases have found wide use as a distant supervision source for a variety of tasks in natural language processing. Our Yago-Naga project has served as a blueprint for many of these follow-up endeavors. The original Yago paper at WWW 2007 has nearly 5000 citations, and the Yago2 paper from 2013 has about 1500 citations. The influence and value of Yago has been recognized by the research community through the AIJ Influential Paper Award 2017 (http://aij.ijcai.org/aij-awards-list-of-previous-winners) for the 2013 Yago2 paper in the Artificial Intelligence Journal, and the W3C Seoul Test of Time Award 2018 (https://www.iw3c2.org/ToT) for the original WWW 2007 paper on Yago. We published a 350-page survey article in the Foundations-and-Trends series 2021, informally called "the machine knowledge book" (https://www.nowpublishers.com/article/Details/DBS-064).

**High-Risk Research:** The Yago-Naga project has been a high-risk (and high-gain) endeavor. In the first few years, hardly any of our colleagues believed that large knowledge bases would become viable and make impact. Some dimensions of this research theme did not work out, though. We started with the goal of developing a search engine that understands semantic concepts for computing precise and concise answers to sophisticated queries. This branch of our research led to insights and novel methods, influential publications and advanced prototypes like Bingo!, TopX and Naga, but did not succeed in building a full-scale system that could be deployed for Internet search. We also had the ambition to build the envisioned search engine in a completely decentralized manner as a peer-to-peer system. On this theme as well, we were very successful in terms of insight and publications, but our advanced prototype system, Minerva, did not make the practical impact that we had aimed for.

**Further Highlights:** The Yago-Naga theme spun off various side projects which have been very influential. One of these is the RDF-3X database engine for efficient storage and querying of subject-predicate-object triples, primarily developed by Thomas Neumann who is now a professor at TU Munich and has received the VLDB Early Career Award 2014 for this work (he also won the DFG Leibniz Award in 2019). The four main papers on this research together have more than 2000 citations, and the RDF-3X system has been widely used in the Semantic Web community.

Another contribution with high impact is the AIDA method and software for Named Entity Recognition and Disambiguation. Here, the knowledge base is leveraged as a background asset for better language understanding and text analytics. The EMNLP 2011 paper has about 1200 citations; it is among the most-cited papers on named entity disambiguation. This work has spawned a startup called Ambiverse and the development of the ambiverseNLU software suite, available as open source code.

In terms of methodology, the Yago-Naga project involved bridging the worlds of structured data (DB methods) and unstructured text (IR methods). One of our foundational works, on language models for temporal expressions, won the ECIR Test of Time Award in 2020.

Finally, our recent work on fairness for classifiers and rankings (jointly with colleagues of the Max Planck Institute for Software Systems) has received great attention in the research community. Our publication in the SIGIR 2018 conference, with Asia Biega as lead author (graduated in 2019, now an independent group leader at the Max Planck Institute for Security and Privacy), is one of the most cited papers on fair ranking (ca. 350 citations).

## Contributions and Impact: Major Results 2021–2023

We publish our results in top-tier conferences in several communities: Web research (Web), data mining and machine learning (DM&ML), database systems (DB), information retrieval (IR) and natural language processing (NLP). In the two-year timeframe 2021–2023, the group had 11 papers in first-rate Web venues (WWW, WSDM, ISWC, ICWSM, IJCAI), 6 in DM&ML (KDD, ICDM, SDM, ICML, ICLR), 3 in DB (VLDB, SIGMOD), 14 in IR (SIGIR, CIKM, ECIR), and 6 in NLP (ACL, EMNLP). In addition, we are successful in building prototype systems and publishing demo papers at top venues like ACL, SIGIR, SIGMOD, WWW and more.

**Advancing Knowledge Base Coverage:** Despite their enormous size, even Web-scale knowledge bases (KBs, aka KGs: knowledge graphs) have major gaps in the scope and depth of their contents. In particular, they often lack knowledge about long-tail entities (e.g., African musicians), non-standard predicates for notable entities (e.g., cover versions of songs, or the key and instrumentation of songs), quantitative properties (e.g., energy efficiency of car models), and commonsense properties of everyday objects (e.g., saxophones having a shiny surface, or SUVs consuming more energy than compact cars). To overcome these limitations we have pursued a number of research directions, with good progress towards advancing KB coverage and quality. Our contributions have been published in premier venues (including EMNLP, TACL, CIKM, VLDB, WSDM and WWW), and we released open-source code and value-added data resources. Also, our senior researcher Simon Razniewski and his co-workers gave tutorials on these directions at flagship conferences like AAAI, VLDB and WSDM.

First, we devised new techniques for assessing the (in-)completeness of KBs and identifying the best web sources to specifically tap into, thus boosting the benefit/cost ratio of KB population and curation. Second, we spearheaded the direction of automatically identifying salient negative properties, to reduce the uncertainty between "unknown" and "false" under the open world assumption. Third, we advanced techniques for extracting statements from long non-standard texts, such as fiction books. Fourth, on the theme of counts and quantities, our contributions include methods to support search as well as KB population (see also below). Last but not least, we successfully explored new ways of representing commonsense KB statements (with faceted detail not captured in any prior work), extracting them at the scale of huge web crawls (like the C4 corpus that underlies the T5 language model), and inferring culture-specific statements (conditioned on the dimensions of geo regions, religions, age groups and professions).

**Entities with Quantities:** We have been among the first to address the issue of extracting and semantically organizing quantities that appear in ad-hoc tables and text sources on the Web. Quantities are numeric expressions that denote financial, physical and other measures with units and a reference frame, such as annual revenue of companies, battery-only range of hybrid cars, carbon footprint of data centers, or conductivity of thermoelectric materials. We developed methods for extracting quantity-centric facts from text corpora and web tables, and for answering queries with quantity-filter conditions such as "sprinters who ran 100 meters under 9.9 seconds". Internet search engines support looking up quantities for a given entity, but often fail on evaluating quantity filters (lacking the proper understanding of queries and web contents). Knowledge bases, such as Wikidata, are rich in entities but very sparse in capturing quantity properties. This work has been published in premier venues like ISWC, WSDM, SIGMOD and WWW. Demonstrator systems with large-scale data collections are publicly accessible on our website (`https://qsearch.mpi-inf.mpg.de/`).

**Question Answering over Knowledge Bases, Text and Tables:** Question answering (QA) is a major use case for knowledge bases (KBs) and pretrained language models (LMs) alike, and today's search engines are also geared for this task. Starting with our early work on question-to-query translation (using ILP for semantic parsing (EMNLP 2012) and learned templates (WWW 2018)), we have broadened our scope and advanced our methods by leveraging a suite of machine learning techniques.

We have been among the first to devise methods that can handle complex questions seamlessly over knowledge bases (with multi-hop inference), text corpora, and collections of tables (HTML, JSON, spreadsheets etc.). Our methods are able to compensate for the limited coverage of KBs by tapping into text and tables, and they are able to counter the inherent noise from text and tables by incorporating structured knowledge (e.g., for answer-type checking). Unlike much of the literature, we carefully devised our solutions for efficiency as well; our prototype systems provide interactive response times. We achieved this by combining information-retrieval techniques for obtaining candidate passages, KB triples and table rows, with neural Transformers for answer extraction and ranking. Our latest method is based on graph neural networks for uniform handling of KB, text and tables. An additional benefit over many baselines is that the answers of our systems come with explanatory evidence for transparency, interpretability and trustworthiness towards end-users.

A particular focus in the last two years has been on *conversational QA*, where the user's input for follow-up questions is often very informal and all but self-contained. For this setting, we have devised novel methods that capture the relevant conversational context towards computing good answers with user-comprehensible explanations. Our methods are based on graph neural networks (GNNs) as well as reinforcement learning (RL).

In addition to publications in top-tier conferences (SIGIR, WSDM, WWW, EMNLP), our senior researchers Rishiraj Saha Roy and Andrew Yates co-authored two state-of-the-art surveys in the Morgan&Claypool series. Prototype systems and benchmarking datasets for this line of research are publicly accessible on our website https://qa.mpi-inf.mpg.de.

**Transparency and Trustworthiness of AI Systems:** On the theme of explainability and scrutability of recommender systems, we have developed new ways of generating user-comprehensible explanations for recommended items, using only information about the user herself – that is, not disclosing any cues about other users, for privacy preservation. Additionally, we developed a method to leverage user feedback on both a newly recommended item and the item that is shown as explanatory support. A user may, for example, dislike the recommendation but express liking the explanation item. Our method learns latent aspects that the user pays attention to when comparing different items. It uses pair-wise feedback to enhance a matrix-factorization model for capturing item-item similarities as well as user profiles, with latent aspects. Experiments and user studies show the practicaly viability of our method. This work, published in SIGIR 2021 and WWW 2021, is part of Azin Ghazimatin's dissertation, which won the GI-DBIS dissertation award 2023.

Another project on transparency and robustness has focused on assessing and mitigating deployment risks of machine-learning models, where bias and errors can adversely affect people's lives. The risks can arise from different kinds of uncertainty: data variability, especially near classifiers' decision boundaries, limitations of the learning models (e.g., not being expressive enough in terms of non-linear elements, capturing long-range dependencies etc.), or data shifts between training and deployment as the characteristics of production data are evolving. We have developed information-theoretic methods that allow quantifying different risks, and thus being able to recommend specific mitigation steps towards trustworthy applications. The latter include judicious abstention in classifiers and obtaining additional training data for critical regions. This work, published in ICDM 2021 and the Machine Learning journal 2022, is part of Preethi Lahoti's dissertation.

## Young Researchers

Our group has a strong track record in promoting young researchers in their careers. The academic offspring of D5 includes internationally visible scientists.[1] A good fraction of our graduates joined leading research labs in industry.[2]

Among our alumni, 11 women have achieved faculty positions: Oana Balalau (INRIA, France), Asia Biega (MPI for Security and Privacy, Germany), Anna Guimaraes (Warwick University, UK), Katja Hose (at U Aalborg, Denmark), Georgiana Ifrim (at UC Dublin, Ireland), Mouna Kacimi (at U Bolzano, Italy), Ndapa Nakashole (at UC San Diego, USA), Koninika Pal (at IIT Palakkad, India), Nicoleta Preda (at U Versailles, France), Maya Ramanath (at IIT Delhi, India), Erisa Terolli (Stevens Institute of Technology, USA).

In the two-year timeframe 2021–2023, we had 12 doctoral students graduating, including 5 women (Azin Ghazimatin, Anna Guimaraes, Preethi Lahoti, Sreyasi Nag Chowdhury, Anna Tigunova).

## Awards

Our research on responsible data science has earned GI-DBIS dissertation awards (by the German Computer Society) for the doctoral theses of two of our female graduates: Asia Biega (2021) on fairness and Azin Ghazimatin (2023) on explainability. The long-term contributions and impact of our group have been honored by Gerhard Weikum receiving the Konrad Zuse Medal (2021).

## Teaching

A total of 12 doctoral disserations and 7 Bachelor's and Master's theses in the 2021-2023 timeframe were completed under the supervision of the group's senior researchers. Our course offers were rather limited, as the department is phasing out.

## Cooperations

**Collaborations within MPG:** Over two decades, the group has conducted a variety of joint research works with the other departments of MPI-INF as well as faculty of MPI-SWS. The group also participated in the MaxNet research network on Big Data Driven Materials Science, which involves 9 institutes of the MPG.

[1] Exemplary faculty are: Srikanta Bedathur (IIT Delhi, India), Mario Boley (Monash University, Australia), Gerard de Melo (Hasso Plattner Institute, Germany), Rainer Gemulla (U Mannheim, Germany), Sebastian Michel (TU Kaiserslautern, Germany), Ndapa Nakashole (UC San Diego, USA), Thomas Neumann (TU Munich, Germany), Maya Ramanath (IIT Delhi, India), Ralf Schenkel (U Trier, Germany), Fabian Suchanek (U Telecom ParisTech, France), Martin Theobald (U Luxemburg).

[2] Examples are: Ralitsa Angelova (Google), Maximilian Dylla (Google), Patrick Ernst (Amazon), Azin Ghazimatin (Spotify), Johannes Hoffart (SAP), Preethi Lahoti (Google), Subhabrata Mukherjee (Microsoft), Josiane Parreira (Siemens), Stephan Seufert (Amazon), Daria Stepanova (Bosch), Mohamed Yahya (Bloomberg), and others.

**External Partners and Competitive Grants:** The department has successfully collaborated with high-caliber partners, both industrial labs (e.g., Amazon, Bosch, Google, Microsoft, Siemens, Volkswagen) and academic organizations (e.g., CMU, IIT Delhi, U Glasgow, U Melbourne, U ParisTech, UW Seattle). In the last few years, the most important collaboration has been the ERC Synergy Grant 610150 on Privacy, Accountability, Compliance and Trust for Tomorrow's Internet (imPACT), with Michael Backes (Helmholtz Center CISPA), Peter Druschel (MPI for Software Systems), Rupak Majumdar (MPI for Software Systems) and Gerhard Weikum as principal investigators (with a total budget of 10 Million Euros for the timeframe 2015-2022). This joint project has been finalized in July 2022.

## Future Plans

With the official retirement of its director, Gerhard Weikum, the department D5 will be effectively closed by the end of 2023. A handful of doctoral students will likely need 2024 for completing their dissertations. They will be supported by the institute, and Gerhard Weikum will continue supervising them. As an emeritus, Gerhard Weikum personally plans to further advance the research on question answering, particularly for entities with quantities (see above) and including coding work for our prototype systems.

# 7 D6: Visual Computing and Artificial Intelligence

## Department Overview

The Visual Computing and Artificial Intelligence Department was established in March 2021 with the appointment of Christian Theobalt as its Scientific Director. The department investigates challenging research questions at the intersection of computer graphics, computer vision and artificial intelligence, notably modern machine learning methods. Before his Director appointment, Christian Theobalt was a tenured research group leader at the MPI for Informatics, where he headed the Graphics, Vision and Video research group.

Since its founding, with members of the former Graphics, Vision and Video research group forming the founding team, the department has further developed. Currently, five group leaders and senior researchers without tenure are part of the department (M. Elgharib, V. Golyanik, M. Habermann, V. Golyanik, A. Kortylewski, T. Leimkühler (co-appointed with D4)). In addition, four postdocs, 18 PhD students, one research engineer and one team assistant are part of the department. D6 also has one affiliated postdoc (A. Tewari, currently at MIT), and two affiliated PhD students (A. Ghosh and N. Cheema, both with primary affiliation at DFKI) who are closely collaborating with the department. Since the founding of the department, four PhD students have graduated or handed in their PhD theses for defense. Several department members already accepted offers for faculty positions, postdoc positions at leading institutions, as well as leading researcher positions in industry[1].

In the reporting period, the department made a high number of important contributions to computer graphics and computer vision, at the intersection to artificial intelligence and machine learning that are further explained in the following. The following are two additional developments, which we would like to particularly highlight. The ability to capture high-quality reference data (e.g. of shape, motion and appearance) of real-world scenes, is essential for the future research in the department. To enable this, the department is building the Real Virtual Lab (see also Sec. 32.16), a new lab space at the MPI for Informatics that hosts unique state-of-the-art capture and display setups. In its final stage, the Real Virtual Lab will feature 350 $m^2$ of lab space. Since 2021, the construction of the Real Virtual Lab has steadily progressed; several major experimental setups were newly built or further extended. The lab features multiple scanners and multi-camera setups, including a capture room with 120 4K video cameras. An additional lab hall extension to MPI for Informatics will be finished by the end of 2024. A further milestone achievement is the establishment of the Saarbrücken Research Center for Visual Computing, Interaction, and Artificial Intelligence (VIA), the result of a new strategic partnership between the MPI for Informatics and Google), see

[1] Lingjie Liu (Faculty at U. of Pennsylvania), Xingang Pan (Faculty at NTU), Ayush Tewari (Postdoc at MIT), Mengyu Chu (Peking University), K. Sarkar (Researcher at Google)

Sec. 32.17. Christian Theobalt is the director of this new center at MPI for Informatics which conducts openly published basic research in collaborative projects with Google.

## Vision and Research Strategy

In the future, we will see new generations of computing systems that have greatly advanced capabilities to support us in previously unseen ways in our daily lives. For example, future computing systems will interact and collaborate with us with near human-level performance, and also have the ability to interact with us through photo-real computer-generated human embodiments in merged reality environments. Immersive virtual spaces in which the real and computer-simulated static and dynamic world will blend, will offer entirely new ways in which we communicate with each other over long distances, or in which we innovate or even do business together. Future computing systems will also offer completely new means to creatively express ourselves, to create media, or to design objects and machines. They will offer unseen new possibilities to make complex facts and processes (visually) accessible to humans in an understandable way, and they will be able to autonomously make sense of complex real-world observations. Future computing systems will have entirely new form factors and permeate our environments. They will support us as new forms of personal assistants, or as intelligent autonomous systems that need to safely act and interact with us in the physical world.

Many profoundly hard and long-term research challenges are awaiting answers on the way to this future. Importantly, profound methodical advancements in the area of visual computing, in particular in the areas of computer graphics and computer vision will be essential. The Visual Computing and Artificial Intelligence Department takes on the many profound open challenges in this area. We investigate foundational research problems at the intersection of computer graphics, computer vision and artificial intelligence. It is our long-term vision to develop entirely new ways to capture, represent, synthesize and simulate models of the real world at the highest detail, robustness, explainability and efficiency using sensor observations. To achieve this long-term goal, we develop new concepts that rethink and unite established approaches from computer graphics and computer vision with concepts from artificial intelligence, in particular machine learning.

Our work lays the foundations for a new way of thinking about computer graphics, and a new way of uniting and enriching the real world with computer graphics technology. It also lays the foundations for advanced methods to better perceive, understand and interpret the complex real world in motion surrounding us from sensor observation, which is an essential capability of future intelligent computing systems that effectively, intuitively and safely act and interact with humans and the human world.

To follow our vision, we make advancements along multiple dimensions of this problem space. On the one hand, we contribute new theoretical and algorithmic concepts that rethink traditional approaches and concepts of visual computing, notably by exploring new methodical paradigms at the intersection of visual computing and machine learning. We also aim to build functioning prototype systems that enable us to validate the developed concepts in complex real-world settings and application scenarios. To reach our goals, we aim for an open and collaborative research environment in which far-reaching ideas can be openly developed and jointly investigated. The department offers a stimulating environment in which group leaders are defining their own research agendas together with groups of students.

The department has established itself as one of the world's leading groups in the area of visual computing and artificial intelligence. The following are very impactful example lines of work we received recognition for. We have become known for developing new concepts that unite the strengths of traditional explicit model-based approaches to visual computing with concepts from machine learning in new integrated models and methods for reconstruction, simulation and rendition of complex real-world scenes. Further, the department is recognized for its many years of seminal work on reconstructing, animating and visualizing virtual human models at highest fidelity. We are also proud of our seminal contributions to artificial intelligence-empowered graphics and real world modeling and reconstruction, notably in the area of neural rendering and neural scene representation, a field which we co-founded. Since its inception, the department has also further strengthened and opened up research in other areas. For instance, we made widely cited contributions to generative modeling, to inverse rendering, as well as to more foundational aspects of visual computing and machine learning, notably quantum visual computing. In pursuing our research vision, we also closely collaborate with the other research groups and departments at the institute. The closest collaborations exist with the Computer Graphics Department and the Computer Vision and Machine Learning Department.

## Research Areas and Achievements

The research in the department is structured into eleven research areas. Each of the research areas is lead by one coordinator or a team of coordinators. Research in these areas is highly collaborative and cuts across the boundaries of the research groups in the department. In the following, a selection of important research results from each of our main research areas is reviewed. The department also makes available a large range of reference datasets (e.g. on various aspects of human modeling), as well as reference implementations available through the department's asset internet portal. A more detailed discussion of our research results, the structure of the group, as well as facilities and datasets, is found in Chap. 32 of the report.

**Marker-less 3D Estimation of Full Human Body and Hand Poses** *Coordinators: M. Habermann, V. Golyanik and C. Theobalt*

Estimating 3D poses of humans is a fundamentally hard and often ill-posed machine perception problem of essential relevance to graphics and vision. In the reporting period, the department has advanced the state of the art in monocular 3D human pose estimation, 3D pose capture from egocentric body-worn camera setups, 3D motion capture with environmental constraints, and 3D hand reconstruction. The following are a few example results.

We proposed the first fully learning-based monocular marker-less motion capture approach that integrates explicit physics modeling and contact force modeling for starkly improved 3D pose capture (SIGGRAPH 2021). Further, we presented new egocentric motion capture methods using a head-mounted (fisheye) camera, allowing the user to move around a large space without restrictions. In ICCV 2021, we presented a new approach combining explicit optimization and neural reconstruction that, unlike earlier works, estimates poses globally in the 3D environment, not only in the local camera frame. In UnrealEGO (ECCV 2022), we presented a new large-scale synthetic dataset for learning-based egocentric pose estimation, alongside a new neural approach for state-of-the-art pose estimation from fisheye stereo.

We also developed new capture methods explicitly taking into account scene constraints, such as a new neural approach (ECCV 2022) to capture contacts between humans and scenes and a learned pose manifold to constrain captured poses. In addition, we presented state-of-the-art methods for capturing human pose and non-rigid environment deformations (3DV 2022), and for capturing 3D human-object interactions (CVPR 2022).

We further investigated marker-less hand motion capture. Particularly, we presented one of the first methods in the literature to estimate the full distribution of plausible poses of two interacting hands from monocular imagery (VMV 2022), and the first learning-based approach to capture very rapid 3D hand motion using a new event camera (ICCV 2021).

**Human Performance Capture** *Coordinators: C. Theobalt and M. Habermann*

While *motion capture* recovers the skeletal pose, *human performance capture* focuses on obtaining the dense 3D appearance and geometry of the entire human, including the deforming clothing, from (sparse) sensing devices. In the reporting period, we presented performance capture methods improving the state of the art by enabling faster runtimes, improved physical plausibility, and higher 3D surface detail.

In Habermann et al. (IEEE PAMI 2023), we presented a new graph neural network-based approach for monocular template-based human performance capture at state-of-the-art quality. A new approach for monocular real-time capture of the body pose, hand gestures, and facial expressions of a parametric human body model was also introduced (CVPR 2021). In Jiang et al. (BMVC 2022), we further improved the completeness of reconstruction on the basis of a personalized 3D template enabling more detailed cloth deformation capture. Both approaches exhibit state-of-the-art accuracy and performance.

We further developed new techniques to obtain parametric and controllable models of entire humans from multi-view video, which capture detailed deformations and appearance as a function of the skeletal pose and camera parameters. This enables full free-viewpoint rendering of humans at previously unseen quality and even in real-time (SIGGRAPH 2021). In a follow-up (arXiv 2022), we further improved the reconstructed detail of free-viewpoint appearance by introducing a hybrid neuro-explicit representation.

**Neural Rendering and Editing of Human Models** *Coordinators: C. Theobalt and L. Liu*

In the reporting period, we presented novel approaches for photo-realistic rendering of humans with control over scene properties, such as body pose, appearance and viewpoints. Approaching this with the classical computer graphics pipeline using explicit model design, motion capture, and physics-based rendering is very complex. We, therefore, developed new methods that combine the traditional explicit graphics pipeline with neural network-based image formation models or new neural dynamic scene representations in new ways.

In Kappel et al. (CVPR 2021), we presented a new multi-stage recurrent deep neural network architecture for high-fidelity monocular human motion transfer with natural loose garment motion. Our Neural Actor (NA) algorithm (SIGGRAPH Asia 2021) utilizes a coarse body model as proxy to unwarp the surrounding 3D space into a canonical space. A neural radiance field is then used to learn pose- and view-dependent deformations and appearance effects from multi-view video. The method yields previously unseen free-viewpoint rendering quality in real-time. In Yoon et al. (CVPR 2021), we presented a new neural method that allows synthesizing plausible human animations from a single human image at previously unseen quality. We also developed one of the first methods for neural free-viewpoint rendering of humans using a driving signal from an egocentric fisheye camera (ICCV 2021).

**3D Reconstruction, Neural Rendering and Editing of Human Faces** *Coordinators: M. Elgharib and C. Theobalt*

Reconstructing and rendering the human face in a controllable way is a long-standing problem in visual computing, with high importance for computer animation, AR/VR, but also scene perception and human-machine interaction, to name a few. In the reporting period, we developed new methods and scene representations for high-quality face reconstruction, and (neural) rendering and editing of human faces. On the one hand, we developed innovative new methods that combine explicit scene representations such as meshes (2xCVPR 2021, ICCV 2021) with machine learning elements. For instance, we presented a pioneering approach that learns comprehensive parametric 3D mesh models of face identity geometry, albedo, and expression entirely from large collections of in-the-wild 2D images and videos (CVPR 2021). By leveraging the vast amounts of community data and adopting a self-supervised learning-based strategy and using a new neuro-explicit method design, our approach can generate, in a self-supervised way from monocular imagery, face models that surpass the expressiveness of existing models. At ICCV 2021, we presented a new self-supervised algorithm that combines a CNN with a differentiable ray-tracer for reconstruction of highly detailed face geometry, face reflectance and complex scene illumination using semantically meaningful parameters.

In addition, we developed new face modeling and rendering approaches using novel neural implicit scene representations (CVPR 2021, SIGGRAPH Asia 2022, 3DV 2022, BMVC 2022). Here, milestone results were our i3DMM work (CVPR 2021), the first neural implicit 3D morphable model of the entire human head, including face interior and scalp hair, as well as the first implicit-based 3D morphable model of the human teeth and gum (SIGGRAPH Asia 2022). Further, we proposed a novel neural representation for face reflectance that can be estimated from a single monocular image (CVPR 2021). The method can re-render the face under any viewpoint and lighting condition. We also contributed a new dataset (FaceForensicsHQ) and a learning-based detection approach (ICME 2021) for disambiguation of real and (neurally) synthesized face video at high accuracy.

**Reconstructing and Modeling General Deformable Scenes** *Coordinators: V. Golyanik and C. Theobalt*

We investigate methods to capture and model general deformable objects and scenes, see our state-of-the-art report (EUROGRAPHICS 2023). Most of our work in this domain considers monocular 3D reconstruction from a single input video; the following are examples. Our $\phi$-SfT algorithm (CVPR 2022) is a new optimization-based method utilizing a template to reconstruct a temporally coherent sequence of 3D shapes from a monocular RGB video. It uses differentiable physics to account for physical fold formation, yielding a new level of physical fidelity and realism in reconstruction.

Non-Rigid NeRF (ICCV 2021) was one of the first works to capture dynamic neural radiance fields from monocular or multi-view video. It can photorealistically reconstruct and re-render 4D models of general scenes undergoing moderate deformation. MoNeRF (arXiv 2022) captures dynamic neural radiance fields by builds a volumetric 4D scene representation from multi-view data using *monocularization*: only one view per timestamp has to be recorded, stored, and provided to the method during training, which accelerates training and decreases memory needs. MoNeRF outperforms previous related approaches.

**Simulation, Image Synthesis and Inverse Rendering** *Coordinators: T. Leimkühler and C. Theobalt*

Computer graphics aims at synthesizing images from scene descriptions, while the goal of computer vision is to infer scene descriptions from images. Our research does justice to this by investigating computational light transport both from the synthesis and the analysis side, resulting in data-driven models for multi-modal simulation, image-synthesis and editing.

Simulation and reconstruction of complex physical processes pose a particularly challenging problem for both computer graphics and computer vision. Togehter with D4 (SIGGRAPH 2021), we developed a new approach that allows sophisticated control over fluid simulations by embedding physical quantities into the latent space of a neural generative model.

The color of each pixel in a real-world image is the result of complex light-material interactions. Recovering these constituents solely from images is referred to as inverse rendering. We have proposed the first approach for the decomposition of a monocular color video into direct and indirect illumination components in real-time (ACM TOG 2021), enabling new real-time illumination-aware video edits. We have also presented the first approach to recover a neural radiance transfer field (ECCV 2022) from a set of real images, enabling free-viewpoint relighting with realistic global illumination.

We also develop new differentiable rendering solutions. A differentiable renderer needs to compute gradients with respect to the image formation process, and allows the propagation of cues from 2D images towards scene parameters (geometry, material, etc.). This is essential for combining graphics models with data-driven approaches. We introduced a new method for differentiable visibility and soft shadow computation (ICCV 2021) that is significantly more efficient than related work. We also investigate new scene representations to effectively solve inverse problems with differentiable rendering, such as a novel neural point-based approach for the high-quality reconstruction of reflections (SIGGRAPH Asia 2022).

**Neural Scene Representations and Neural Rendering** *Coordinators: L. Liu and C. Theobalt*

*Neural scene representations* and *neural rendering* methods rethink the classical (inverse) computer graphics pipeline by showing new ways to incorporate neural learnable elements into the (explicit) modeling of scenes and respective rendering processes. Our department made seminal contributions to this emerging area (see EUROGRAPHICS 2022 STAR report, SIGGRAPH 2021 course). The following are example contributions.

Reconstructing and rendering single objects from a collection of multi-view images has been studied for decades. In our NeuS algorithm (NeurIPS 2021), we represent the scene as a neural radiance field and propose a new unbiased volume rendering formulation enabling high-quality surface reconstruction from multi-view imagery. The reconstructed results by far exceed the visual quality of classical multi-view stereo and competing neural methods. To further enhance the efficiency of neural surface reconstruction, we presented Voxurf (arXiv 2022), a voxel-based coarse-to-fine method that preserves the color-geometry dependency and can be trained 20x faster than NeuS while obtaining higher quality. In Liu et al. (CVPR 2022), we propose a new image-conditioned neural implicit modeling approach that explicitly models occlusion probabilities with respect to reference views, which drastically enhances novel-view rendering quality.

We also contributed new methods to neurally capture and render larger scenes. Examples are a new state-of-the-art method for neural indoor scene reconstruction using an SDF-based neural scene representation (ECCV 2022), and the first method for outdoor scene relighting based on neural radiance fields (ECCV 2022) that allows simultaneous direct editing of scene illumination and camera viewpoint. Finally, our BungeeNeRF algorithm (ECCV 2022) introduced a progressive architecture making it the first method able to render scenes from the scale of a building to the whole planet.

**Generative Models** *Coordinators: X. Pan and C. Theobalt*

We develop new learning-based generative models for high-quality visual synthesis. Specifically, we proposed novel generative models of humans, which can synthesize new 3D humans and human faces, videos of humans, humans and environments, and (text- and speech-conditioned) human gestures and motions with high quality.

StyleNeRF (ICLR 2022) uses a hybrid architecture with 3D-aware feature encoding and progressive up-sampling to synthesize state-of-the-art high-resolution images while maintaining multi-view consistency. We also studied the disentanglement of 3D shape and appearance. Disentangled3D (CVPR 2022) is first 3D generative model that allows us to modify the shape and appearance independently and that was learned in a self-supervised way. In HumanGAN (3DV 2021), we presented a new method for highly realistic synthesis of images of dressed humans that allow control over pose, local body part appearance, and garment style. The method outperforms baselines in terms of realism and image resolution.

While the aforementioned methods synthesize static imagery, StyleVideoGAN (BMVC 2021) was one of the first generative approaches to synthesize plausible videos of talking heads that were not actually recorded. In Playable Environments (CVPR 2022), we present a new approach that learns in an unsupervised way to synthesize 3D videos of both humans and the environments they act in (e.g. tennis games), while supporting user interaction through actions similar to playing a video game.

We also developed text- and speech-conditioned generative approaches to synthesize human gesture and motion, such as 3D pose sequences depicting multiple sequential or superimposed actions provided in long, compositional sentences (ICCV 2021). In MoFusion (arXiv 2022), we introduced a framework for performing multi-modal (e.g. text- or music-based) conditional motion synthesis on the basis of Denoising Diffusion Probabilistic Models (DDPM) at state-of-the-art quality. We also developed state-of-the-art learning-based approaches to synthesize body hand and face gestures conditioned on speech input (IVA 2021, SIGGRAPH 2022).

**Robust World Perception and Recognition** *Coordinator: A. Kortylewski*

Over the last decade, we have seen a tremendous increase in the performance of computer vision systems due to advances in deep learning. However, Deep Neural Networks (DNNs) are still far from reaching human-level performance at visual recognition tasks. The most important limitation of DNNs is that they fail to give reliable predictions in unseen or adverse viewing conditions, which would not fool a human observer. They are unreliable when objects are partially occluded, seen in an unusual pose or context, or in bad weather. This lack of robustness in DNNs is generally acknowledged but largely remains unsolved. In the reporting period, we made several contributions towards improving DNN performance in that respect. We presented new benchmark datasets to measure and new methods featuring better out-of-distribution robustness (e.g. ECCV 2022, CVPR 2023), and presented perception methods with improved robustness through neural analysis-by-synthesis (e.g. ECCV 2022, MOVI 2022, WACV 2023). We also proposed new approaches for part-based modeling and reconstruction of complex scenes that advanced the state of the art (e.g. ECCV 2022, AAAI 2022).

**Quantum Visual Computing** *Coordinator: V. Golyanik*

Our department develops new quantum computing methods for solving hard problems in visual computing and machine learning. Our focus lies on adiabatic quantum computers (AQC) since modern realizations of the latter (*i.e.,* D-Wave Advantage with >5000 qubits) allow experimental evaluation of the developed methods on real quantum hardware at relevant problem sizes. AQC require the target problem to be provided in the form of a quadratic unconstrained binary optimization (QUBO) problem. The following are example works.

We investigated the randomness properties of superconducting qubits used in the D-Wave quantum annealers proposing a new approach to generate truly random numbers (IEEE Access 2022) with application in randomized sampling, Monte Carlo rendering and neural network initialization. Our Q-Sync algorithm (CVPR 2021) is the first approach for permutation synchronization using AQC. Its objective is to consolidate input pairwise (and, perhaps, noisy) matches into globally-consistent correspondences for multiple sets of input features (extracted on images or shapes). We further also proposed the first quantum method for motion segmentation (ECCV 2022), and introduced Q-Match (ICCV 2021), i.e. a new hybrid approach for quantum-based alignment of two meshes using cyclic alpha expansion that outperforms the previous quantum state of the art. We also developed a quantum version of the Frank-Wolfe algorithm for constrained convex optimization (ECCV 2022).

**Foundational Methods for Visual Real-World Reconstruction and Artificial Intelligence**
*Coordinators: V. Golyanik and C. Theobalt*

In the reporting period, our department also contributed solutions to foundational, often lower-level, problems in visual computing and machine learning that are relevant across specific problem instances, and often provide building blocks for more complex frameworks. For instance, we developed a new method for that jointly performs matching and clustering of graphs (3DV 2021). We also proposed a new method for image correspondence pruning called Laplacian Motion Coherence Network (LMCNet) (CVPR 2021), which uses a novel formulation to capture global motion coherence from sparse putative correspondences. A new shape-regularized semi-supervised landmark detection approach was also presented in the reporting period (Neurocomputing 2022). Our department also investigated new depth map estimation methods. Examples are a new method for video-based depth estimation modeling geometric and temporal coherence among the frames (CVPR 2021), and a new single-image approach utilizing normal constraints to enforce geometric consistency (ICCV 2021).

## Selected Performance Indicators

We made significant progress in all of our research areas. In the reporting period from spring 2021 to spring 2023, we have published 75 peer-reviewed papers in high quality conferences and journals of computer graphics, computer vision, and machine learning. This includes 11 papers in flagship graphics conferences SIGGRAPH and SIGGRAPH Asia, 37 papers in the flagship vision conferences (CVPR, ICCV, ECCV), 13 papers in NeurIPS/ICLR/3DV/BMVC, and 11 papers in top-tier journals of vision and graphics. In the reporting period, the department entertained collaborations with leading academic and industrial research labs worldwide, that resulted in a high number of joint publications. The department is supported by major grants, such as an ERC Consolidator Grant (C. Theobalt), and an Emmy Noether Starting Grant (A. Kortylewski); a detailed list can be found in Sec. 32.21.

Department members widely served as program and senior program committee members of major conferences in graphics, vision and machine learning, and have served as editorial board member in leading journals. They also gave many invited talks and tutorial presentations at major international institutes and events (details in Chap. 32). Our department has attracted top international talents as researchers or visiting researchers on all levels of experience. Several former members of the department moved on to new positions in leading academic and industrial institutions worldwide. Further, some of our research results have served as basis for commercial products. Examples are our startup the Captury commercializing a leading marker-less motion capture technology, and Flawess.AI which develops a revolutionary new technology to visually dub movies that is partially based on our research (see Sec. 32.15).

Department members were honored with awards (see Sec. 32.20 for complet list) and special commissions of trust (see Sec. 32.18 for complete list). For example, in the reporting period, Christian Theobalt was elected Fellow of the EUROGRAPHICS organization. He also served as chair of the ERC Consolidator Grant Selection Panel in Computer Science, is director of a new ELLIS research program on *Learning for Graphics and Vision*, and director of the new VIA center in collaboration with Google. Adam Kortylewski was awarded a DFG Emmy Noether group. Ayush Tewari (2022) and Marc Habermann (2023) received the *Otto-Hahn-Medal* of the Max-Planck Society. Marc Habermann received the *EUROGRAPHICS Dissertation Award* and the *DAGM MVTec Dissertation Award*, both in 2022. Several publications from the department received best paper awards, honorable mentions, or were award finalists at major conferences (e.g. at 3DV, BMVC, CVPR, etc.).

# 8 RG1: Automation of Logic

## History

The Automation of Logic Group was established in September 2005 and is headed by Christoph Weidenbach. The group covers the complete pipeline from basic research on logics and their automation up to applications in research and industry.

There are currently four researchers, four PhD students, and four long-term guests in the group. Fajar Haifani has finished his PhD thesis during the reporting period. Martin Bromberger has recently been appointed senior researcher.

## Group Organization and Development

Our four PhD students are co-supervised by two senior members of the group, including our long-term guests. We have weekly meetings with the PhD Students. There are additional meetings for the joint software development, where we are currently focusing on implementing SCL (Clause Learning from Simple Models).

Thomas Sturm, working in computer algebra, Jasmin Blanchette, working in interactive theorem proving and its mechanization, Sophie Tourret working on first-order logic reasoning and its formalization, and Mathias Fleury, working on CDCL and its formalization are long-term guests of the group. They interact with the group through regular meetings, joint projects such as VeriDis and Matrioshka, and joint PhD student supervision. Martin Bromberger has started his own subgroup on the combination of theories and is supervising one PhD student. Uwe Waldmann is continuing his close collaboration with Alexander Bentkamp and Jasmin Blanchette on adding higher order reasoning techniques to superposition. Sibylle Moehle has recently joined the group and continues her work on propositional logic model counting.

## Vision and Research Strategy

The vision of the group is to increase the productivity of formal analysis/verification/problem solving technology through a higher degree of automation of the underlying logics. The following challenges motivate main parts of our work: (i) drive the development of first-order, and beyond first-order reasoning calculi, (ii) understand the mechanics of reasoning with respect to concrete models, (iii) build specific reasoning procedures for arithmetic theories, (iv) show applicability of our methods to reasoning challenges from other areas, (v) formally verify our own results, and (vi) scale the applicability of our methods to the size of real world industrial applications.

About half of our work is of a theoretical nature and the other half is experimental, in particular on the basis of developed tools. The considered problems are typically NP-hard and beyond. Therefore, implementation of our methods is important to check their automation potential in practice, to increase the impact of our theoretical results by providing our software to other people and eventually to detect further challenges in theory development. In 2012 we founded the company "Logic for Business" (L4B), providing consulting and software for the overall life-cycle management of complex products. It was sold in 2021.

We have structured our research along the following topics: (i) Foundations of Automated Reasoning, (ii) Arithmetic, (iii) Saturation Theorem Proving, (iv) Formalizing Logic, (v) Software and Applications. The structure does not impose a structure on the group. In fact, most of us contribute to several areas and almost all of us to the development of software.

A selection of scientific results for topics (i), (ii), (iii), and (iv), respectively, are:

**Applications of SCL to Real-World Scenarios (TACAS 2022, VMCAI 2021, FroCos 2021)** The SCL variant modulo linear arithmetic SCL(LA) and its underlying logic BS(LA), the Bernays-Schönfinkel fragment extended with linear arithmetic constraints, turns out to be a perfect base for the formalization and automatic verification of so called Supervisors. Supervisors are simple, high-level software controllers for technical systems. We studied a lane-change supervisor and a supervisor controlling a combustion engine. In both cases we could formalize the respective software and automatically prove correctness. This was established by a reduction to pure first-order Horn logic, called our Datalog Hammer. Eventually, a Datalog engine did the verification work.

**Verification and Development of SCL (CADE 2023, ArxiV 2023)** There is a long tradition in the group to formally verify our pen and paper results on automated reasoning calculi in Isabelle. We developed several variants of the SCL calculus from 2019-2021 including pen and paper proofs for correctness. Our goal was to unify all aspects of the calculi for the case of first-order logic without equality. In parallel the calculus was formalized and verified in Isabelle. Both the formalization and the unification of the variants were done in parallel and contributed to each other. Through the formalization we found one bug in a former SCL variant and minimized invariants on the calculus compared to the pen and paper proofs. The formalization profited a lot from our new proof technique on termination which no longer relies on an overall measure but on the fact that SCL learns only non-redundant clauses.

**Connection-Minimal Abduction (IJCAR 2022, DL 2022, SOQE 2021)** As automated reasoning becomes more and more a part of the overall life cycle of systems, being able to explain reasoning results, becomes more and more important. We developed a new notion on abduction that in particular rules out hypotheses that are not connected to the actual formulas. It is implemented through a tool chain and a first-order logic translation for the description logic $\mathcal{EL}$. This work won the best student paper award at DL 2022.

**Parametric Qualitative Analysis of Reaction Networks and Epidemic Models (CASC 2021)** Chemical reaction network theory provides constructive mathematical modeling approaches for a wide range of biological networks, e.g., reaction networks and epidemic models. Examples are the well-studied MAP kinese and SEIR models, respectively. Variants of SEIR are used, e.g., for various subtypes of influenza A and quite recently also for Covid-19.

The dynamics of the networks are described by ordinary differential equations (ODE) in time, assuming mass action kinetics. Traditional approaches have been based on numerical simulations of the ODE for fixed measured or estimated parameters. Our principal approach, in contrast, combines symbolic methods and automated reasoning with model reduction methods aiming at a *qualitative* analysis of biological networks.

In the previous report we have reported on a comprehensive study, which exhibited that a significant number of models from the BioModels.net repository has steady state regions that are geometrically toric. Technically, we had applied decision methods to characterizations of toricity as first-order sentences. We have now taken an important step from fixed measured or estimated parameters to formal parameters. Quantifer elimnation takes the place of decision procedures and provides necessary and sufficient constraints for toricity in terms of the parameters. We have studied both the real domain and the complex domain, providing rigorous formal foundations along with paractial computations on models from our previous study.

## Projects and Cooperations

Together with Stephan Merz's group (Inria Nancy) we constitute the Inria project VeriDis (http://www.inria.fr/en/teams/veridis), where we investigate automated reasoning support for the verification of distributed algorithms. In the interdisciplinary project SYMBIONT we apply reasoning in non-linear arithmetic to problems from biology. We constitute the core of the ERC starting grant Matryoshka on the mechanization of higher-order logic. In the transregional collaborative research center TRR 248, "Perspicuous Systems", established in January 2019 and finished in December 2022, we investigate automated reasoning in the context of perspicuous software systems in close cooperation with researchers from TU Dresden.

## Prizes and Awards

Fajar Haifani won the best student paper award at the Description Logic workshop 2022. It was a joint paper with Patrick Koopmann from Dresden, Sophie Tourret and Christoph Weidenbach.

## Teaching

We regularly teach and supervise bachelor and master students at Saarland university. We are responsible for the core lecture "Automated Reasoning" of the master curriculum. In addition, we provided the lecture "Competitive Programming" together with colleagues from the CS department in the reporting period.

# 9 RG2: Network and Cloud Systems

## History

Yiting Xia joined MPI-INF as a tenure-track faculty in September 2020. She is currently leading the Network and Cloud Systems (NCS) research group. Yiming Lei joined NCS as a Ph.D. student in October 2021. Jialong Li joined NCS as a postdoctoral researcher in November 2021. The NCS research group is actively growing.

## Group Organization and Development

At present, the research group consists of one tenure-track faculty, one postdoctoral researcher, one Ph.D. student, and a secretary. A new Ph.D. student will join the research group in October, 2023.

Yiting Xia is leader of the group. Her research interests include computer networks, cloud systems, and machine learning systems. Jialong Li is a postdoctoral researcher. His research interests include optical networks, data center networks, and optical communications. Yiming Lei is a second-year Ph.D. student. He is interested in programmable switches and network system implementation. Iris Wagner is now the secretary of the group.

Besides, two Master's students are doing their Master's theses in the group, and we will host three interns in the summer.

Federico De Marchi is a Master's student at Saarland University. He is doing his Master's thesis in the group. His topic is designing transport protocols for fast-switched optical data center networks. He will join the research group as a Ph.D. student after he obtains the Master's degree. Vadim Farutin is also a Master's thesis student from Saarland University. His topic is improving GPU sharing efficiency for distributed training jobs.

Xiaoxiang Shi, a Master's student from Shanghai Jiao Tong University, will join the group in July as an intern. He will work on the system implementation for improving GPU sharing efficiency. Haotian Gong, an undergraduate student from the University of British Columbia, will join the group in May as an intern. He will work on the topology design for optical data center networks. Parham Chavoshian, an undergraduate student from Sharif University of Technology, will join the group as an intern in July. He will work on congestion control for data center networks.

We had one Master's student and one intern in the past year. Zhengqing Liu, was from Ecole Polytechnique, Paris and now is a Ph.D. student in Imperial College London. He finished his Master's thesis—*System Implementation for Hop-On Hop-Off Routing via Programmable Switches*. Rui Pan, was from University of Wisconsin–Madison and now is a Ph.D. student at Princeton University. He finished the internship project on network abstraction for distributed deep learning training and published a HotNets paper as the first author.

## Vision and Research Strategy

Emerging services and hardware advancements are reshaping the landscape of cloud computing. One trend is edge computing, where small data centers known as cloudlets are built close to users as easy access points for Internet of Things (IoT). The other trend is the increasing modular placement of hardware resources. For example, computing racks with disaggregated functions have become fundamental building blocks of data centers; and a number of data centers are interconnected as a region in dense metro areas. These changes motivate networking research that blends the boundary of traditional data center networks and backbone networks as well as distributed systems research that involves new hardware, network design, and cloud systems. The scope of the Network and Cloud Systems group is to build network infrastructure and distributed systems for cloud computing. We follow a cross-layer approach and cover broad topics for optimizing the network stack, including switch hardware, network protocols, software systems, and cloud applications.

## Research Areas and Achievements

**Optical Data Center Networks.** Optical data center network (DCN) fabrics are renovating the infrastructure design in the cloud. However, there is a gap between the diverse optical hardware architectures and system integration work to realize the architectures as end-to-end workable systems. This research direction is to design and implement practical systems to enable different optical architectures in production DCNs. Towards that, we abstract fundamental building blocks for optical DCNs, including global time synchronization with nanosecond-scale accuracy, generic routing regardless of the optical hardware, and an application-agnostic host stack. Up till now, we have implemented a prototype system with P4 on Tofino2 switches and libvma on Mellanox NICs. Extensive micro-benchmark studies with production DCN traffic show that our system keeps synchronization errors under 15ns and ensures zero packet loss with 99.93% achievable network utilization. We demonstrate three optical architectures on the system with real DCN applications and observe similar flow completion times for mice flows compared to electrical DCNs. We are open-sourcing the system, as a tool for the networking community to test, improve, and deploy optical DCNs.

**Network-Accelerated Machine Learning Systems.** Recent years have witnessed the rapid development of deep learning. Various parallel strategies have been adopted by distributed deep learning training (DDLT) frameworks to accommodate the ever-growing model sizes. As a result, communication among distributed workers, especially over a shared, highly dynamic network with competing training jobs, has become a notable bottleneck of the training process. We aim to accelerate inter-node communications in machine learning systems. In one project, we propose the first network abstraction for DDLT and devise a generic method to model the drastically different computation patterns across training paradigms. We use the abstraction for flow scheduling in DDLT jobs and demonstrate its effective with case studies. In another project, we introduce network-aware GPU sharing to improve efficiency of job placement and GPU scheduling in machine learning clusters. Compared to previous scheduling mechanisms that assume fixed data transmission time, we for the first time model the network dynamically and provide tighter bounds on the data transmission time. Simulation results show our scheduling method achieves high GPU utilization with minimal slowdown of training time.

**Reliable Network Management.** The complexity of large networks makes their management a daunting task. State-of-the-art network management systems program workflows of operational steps with arbitrary scripts, which pose substantial challenges to reliability. We leverage the fact that most modern network management systems are backed with a source-of-truth database and customize database techniques to the context of network management. The network management framework exposes a programming model to network operators for conveying the key management logic. Then the operators are completely shielded from reliability concerns, such as distributed devices, operational conflicts, task atomicity, and failures, which are instead handled by the runtime system using database techniques. Our simulation evaluation and production case studies demonstrate the system's effectiveness in minimizing network vulnerable time and resolving task conflicts. We open-source our simulator and task traces for academic researchers to contribute to this industrial problem.

## Projects and Cooperations

We have collaborated with National University of Singapore and Vrije Universiteit Amsterdam on optical data center networks, and we have collaborated with ETH Zürich on network-accelerated machine learning systems. We have collaborated with Meta and Rice University on reliable network management. We have an ongoing collaboration with Microsoft on transport protocols for optical data centers.

## Prizes and Awards

Yiting Xia won the N2Women Rising Star Award in 2021.

## Teaching

Yiting Xia co-taught Distributed Systems with Krishna Gummadi and Jonathan Mace in summer 2021, and with Deepak Garg and Laurent Binschaedler in summer 2023. Yiting Xia and Jialong Li co-taught Data Networks with Anja Feldmann and Oliver Gasser in summer 2022. Yiting Xia and Jialong Li co-organized the Hot Topics in Data Networks seminar with Anja Feldmann and Oliver Gasser in winter 2021 and 2022.

# 10 Vahid Babaei: Computational Design and Fabrication

## Group Development

Our group was established in September 2018, originally funded by the Max Planck Center for Visual Computing and Communication (MPC-VCC). Currently, the team is led by Vahid Babaei and employs two PhD student (Navid Ansari and Emiliano Luci), two research engineers (Sebastian Cucerca and Muhammad Yaseen), and three master students (Parham Zolfaghari, Siddhartha, and Balaji Venkatesan). During the reporting we hosted a postdoctoral researcher (Thibault Tricard), a visiting PhD student (Kang Liao), two research interns (Haleh Mohammadian and Nima Ferdowsi). Also, a master's thesis, by Nikan Doosti, was completed in the group.

## Vision and Research Strategy

We are deeply interested in developing computational methods that address the digitalization of manufacturing. Inspired by computer graphics methods, in my research on *computational design*, we develop algorithmic tools to *evaluate*, *represent*, and *synthesize* products with improved or completely novel functions. Our research on *intelligent manufacturing* strives to radically change the manufacturing process through bringing the fabrication hardware inside the computation loop, and also empowering this hardware with on-the-fly sensing and computational decision making.

The applications of this research span from patient-specific ocular and facial prosthetics with a detailed appearance that matches that of human eyes and skin, to digital manufacturing solutions that enable novel products through deep integration of on-the-fly sensing and computational decision making, and data-driven simulation toolboxes that unify the discovery of novel designs in mechanics, photonics, acoustics, and beyond.

## Research Areas and Achievements

**Computational Design.** We have introduced pioneering physically-inspired design *evaluation* methods. For example, in appearance 3D printing the design evaluation concerns predicting different appearance features of a digital, multi-material composite. Notably, we developed an evaluation engine relying on a light transport simulator implemented via the Monte Carlo volume rendering. We have also been developing data-driven evaluation methods using *neural surrogate models* (NSMs) – neural networks that learn and replace physics-based simulations.

In numerous computational design problems, design representation is the make or break factor of a practical solution. We have recently introduced *design representation networks* (DRNs), inspired by the coordinate-based neural networks used in computer vision community for scene representation. So far, we have successfully integrated DRNs into two computational design problems: 3D printing of static light-field displays, and structural topology optimization. In both cases, we formulated problems with domain sizes which would be impossible to approach using a traditional voxel representation.

One of the most ambitious aspirations of computational design is to automatically *synthesize* fabricable designs given the functional goals. We have been in the forefront of research on design synthesis using neural surrogate models (NSMs). We recently introduced a general framework for solving combinatorial inverse design problems in a data-driven fashion using *piecewise linear* surrogates. Our main insight is that given a piecewise linear NSM, e.g., the omnipresent neural networks with ReLU activation functions, we can formulate the neural inverse design as a mixed integer linear programming and solve it using powerful mixed-integer solvers. In another recent work on neural inverse design, *Autoinverse*, we take into account the predictive uncertainty of the surrogate and minimize it during the inversion. Apart from high accuracy, our method enforces the feasibility of solutions automatically, comes with embedded regularization, and does not require careful initialization.

**Intelligent Manufacturing.** We have developed a set of novel manufacturing processes imporoved by computatioanl methods. For example, we introduced a Bayesian optimization framework for laser-material processing. At each iteration of the framework, we brought the laser marking hardware into our algorithmic loop by marking and measuring a set of trial laser parameters recommended by our algorithm. With important implications for the surface activation and functionalization industry, in project *Oraclase* funded by the German Federal Ministry of Economic Affairs and Climate Action (€777,000), we are commercializing this technology. In another effort to enable direct ink writing printers to accurately deposit difficult-to-print materials, we built a 3D printer with a reinforcement learning closed-loop controller. We searched for control policies that deposit the materials uniformly. These policies were learned through a numerical simulation that models a computational fluid dynamics environment. We successfully transferred the learned policies to our real fabrication system, a self-built 3D printer equipped with an integrated, precise vision system.

## Projects and Cooperations

During the current reporting period, we have collaborated with Wojciech Matusik (MIT), Szymon Rusinkiewicz (Princeton), Piotr Didyk (USI), Michal Piovarci and Bernd Bickel (IST), and Julian Panetta (UCD).

## Prizes and Awards

– *Hermann Neuhaus Prize* of the Max Planck Foundation (€25,000).

– *Max!mize Award* of the Max Planck Society (€50,000).

# 11 Martin Bromberger: Combination of Theories

## Group Development

The Combination of Theories group is a freshly created subgroup of Automation of Logic, established in April 2023. In addition to Martin Bromberger, the group includes two PhD students: Lorenz Leutgeb and Simon Schwarz, both co-supervised by Christoph Weidenbach.

## Vision and Research Strategy

The Combination of Theories group investigates various first-order theories and ways to combine them. Contrary to classical first-order logic, first-order theories can be seen as specialized logics that impose a fixed semantic interpretation on certain function, predicate, and sort symbols. Due to the historical focus on software verification in automated reasoning, these theories typically correspond to domains found in programming languages, e.g., the theories of integers, strings, and arrays. Naturally, it is necessary to combine several such theories in order to model most programs. Satisfiability modulo theories (SMT) solvers are one category of automated theorem provers (ATPs) that are especially well-suited to handling formulas over a combination of theories. SMT solvers are nowadays ubiquitous wherever safe software is developed. They are used for instance by Amazon, Google, Intel, and Microsoft.

The group's research focuses on three main topics: developing and improving techniques for first-order theories, developing new ways to combine theory solvers into one theorem prover, and investigating the complexity of first-order fragments based on combinations of theories. The research is of both theoretical and practical nature: the group not only develops new techniques and calculi, but it also implements them in/as state-of-the-art automated theorem provers, and uses them to solve real world problems.

## Research Areas and Achievements

**New Techniques for Arithmetic and SMT Solvers:** The theory of linear arithmetic is of particular interest in the context of software verification because almost every piece of non-trivial software performs at least some arithmetic operations. Martin Bromberger has developed an SMT solver called SPASS-SATT that for now focuses on quantifier-free linear arithmetic. SPASS-SATT contains not only efficient adaptions of techniques used in other SMT solvers, but also several new techniques developed by Martin Bromberger. These new techniques efficiently handle classes of formulas, e.g., unbounded formulas, that were particularly hard—most of the time even unsolvable—for other SMT solvers. As a result, SPASS-SATT became one of the leading SMT solvers for linear arithmetic and won several awards at the International Satisfiability Modulo Theories Competitions. In the meantime, most professional SMT solvers have adopted at least some of SPASS-SATT's techniques.

**Supervisor Verification:** A supervisor is a component in a technical system that controls the systems functionality. One example for a supervisor is the electronic control unit (ECU) of a combustion engine. Recently, the ECU gained questionable fame when it turned out that car companies intentionally programmed them to cheat during laboratory emissions testing. Detecting hacks like these with standardized laboratory tests is very hard and rather costly. As part of the CPEC project, Martin Bromberger worked on verifying the correctness of a supervisor's software directly. The research focused on two use cases: a prototype supervisor for a lane change assistant (LCA) and an ECU. It turned out that both can be modeled in a logical fragment called Horn Bernays-Schönfinkel over simple linear arithmetic (HBS(SLA)). Martin Bromberger managed to show that each HBS(SLA) clause set can be transformed into an HBS clause set, also called a Datalog program, for which there already exist highly efficient reasoning engines. The combination of the transformation, called Datalog Hammer, and the Datalog reasoner VLog was implemented in a new ATP called SPASS-SPL. SPASS-SPL managed to decide verification conditions for both examples. In the case of the LCA prototype, SPASS-SPL managed to detect several bugs and produced corresponding input examples. This helped the developers of the prototype to fix it. In the meantime, SPASS-SPL verified the latest version of the LCA as bug-free. For more complex supervisors that cannot be modeled in HBS(SLA), the group has developed a new reasoning approach: the SCL(T) calculus. A prototype is currently under development.

**Saturation Based Theorem Proving Modulo Theories:** For pure first-order logic, saturation based theorem provers are considered state-of-the-art. Implementations are only competitive if clauses that are redundant, and therefore do not inform progress, are deleted to free up capacity. In practice, this redundancy deletion focuses on subsumed formulas and for pure first-order logic there exist various heuristics for efficient detection of subsumed clauses. However, the heuristics for pure first-order logic are in general not applicable when first-order logic is combined with a background theory. The reason is that for pure first-order logic purely syntactic criteria are sufficient to check for subsumption, whereas most theories (e.g. the theory of linear arithmetic) require a semantic component to decide subsumption. Lorenz Leutgeb developed the sample point heuristic to fix this gap. It stores a solution per clause, which can later be used in attempts to quickly falsify subsumption. He showed in experiments that the heuristic is very specific, i.e. 94% out of all true negatives are predicted correctly, and effective (speed-up by a factor of 124).

## Projects and Cooperations

*CPEC, Center for Perspicuous Computing*, the collaborative research center 248 foundations of perspicuous software systems, funded by the DFG. This project involves, in addition to the Max-Planck Institute for Informatics, also Saarland University, the Max-Planck Institute for Software Systems, and the Technische Universität Dresden. In the context of the Combination of Theories group the collaboration was with Christof Fetzer's team (TU Dresden), Markus Kroetsch's team (TU Dresden), and Holger Hermann's team (UdS).

# 12 Dengxin Dai: Robust and Adaptive Visual Perception

## Group Development

Our group, called Vision for Autonomous Systems, was created in August 2021. We have one PostDoc researcher, three PhD students, and three Master's students. Furthermore, we have multiple co-supervised PostDoc researchers and PhD students. The group has been grown rapidly.

## Vision and Research Strategy

The scientific mission of my research group is to develop robust and scalable perception systems for real-world applications. We focus on deep learning-based perception for autonomous systems such as autonomous driving. We are especially fascinated about scaling existing visual perception models to novel domains, to more data modality, and to new tasks.

## Research Areas and Achievements

Specifically, our group mainly works on the following three areas: Vision for all Seasons, Domain Adaptation, and Learning with Time-efficient Labels.

**Vision for All Seasons**: As widely known, adverse weather and lighting conditions (e.g. fog, rain, snow, low-light, nighttime, glare and shadows) create visibility problems both for people and the sensors that power automated systems. Many real-world applications such as autonomous cars, agriculture robots, rescue robots, and security systems can hardly escape from 'bad' weather, challenging lighting conditions, dust, smoke, and so on. Therefore, vision algorithms that can robustly adapt to adverse weather and lighting conditions are strongly needed for real-world applications. Over the last two years, our group has developed multiple novel approaches to increase the adaptability of visual recognition models to real-world adverse weather/lighting conditions. For instance, we have developed a method called Domain Flow (IJCV'21) to synthesize data of intermediate domains given data of the source domain and data of the target domain. Bad weather also heavily affects other types of sensors such as LiDARs. Oour group has developed methods to train robust LiDAR-based perception algorithms under bad weather conditions. In particular, we took a new, exciting journey to simulate the effects of fog (ICCV'21) and snowfall (CVPR'22) into LiDAR imaging. Besides developing effective learning methods, our group has also made significant contribution in developing new benchmarks and organizing international workshops for visual perception under adverse conditions https://vision4allseason.net/. Our work ACDC (ICCV'21) features high-quality annotations created by using a specialized annotation protocol. The dataset has become the de facto benchmark to benchmark the performance of visual perception algorithms in adverse weather and lighting conditions.

**Domain Adaptation**: Current robotic perception systems are typically trained in a rather fixed environment, allowing them to succeed in specific settings, but leading to failure in others. One of the key challenge now is to build adaptive perception systems that can work well even when encountering domain changes. Our group has made significant contributions for general domain adaptation, in particular unsupervised domain adaptation (UDA).

Our group has made great contributions in designing novel network architecture and training strategies for domain adaptation. In our recent CVPR'22 work, the influence of recent network architectures has been systematically studied to reveal the potential of Transformers for UDA semantic segmentation, and we proposed a novel UDA method DAFormer. It represents a major advance in UDA. In our ECCV'22 work, we further proposed a novel method HRDA to incorporate more context information and at the same time preserve high-resolution details for UDA. HRDA is the first work on UDA semantic segmentation that systematically studies the influence of resolution and crop size and exploits HR inputs for adapting fine details.

**Learning with Time-efficient Labels**: Machine learning has been advancing rapidly, and ever growing data is at the center of this evolution. While the recipe of learning with large-scale annotated datasets is still effective, this can hardly scale due to the high annotation cost. Therefore, it has become very important now to develop methods for the low data regime. In our CVPR'22 work, we proposed using scribbles to annotate LiDAR point clouds and released ScribbleKITTI, the first scribble-annotated dataset for LiDAR semantic segmentation. In our another CVPR'22 work, we have developed a novel method to combine multiple self-supervised learning tasks so that the learned feature representations generalise well to all downstream tasks. Cross modal distillation for supervision transfer is another example of label efficient learning (T-PAMI). Here, the goal is to transfer supervision between different modalities, for instance from images – where larger amount of annotations are available – to audios – where smaller amount of annotations are available.

## Projects and Cooperations

- One Grant as PI from Toyota: 137,183 EURO, Multi-task Learning, 2022
- One Gift Grant from Meta Research: 50,000 USD, Robust Object Detection, 2022
- One EU Horizon Grant as co-PI: 7.3M in total, our part is 300,000 EURO, Robust Visual Perception for Autonomous Driving

## Prizes and Awards

Award Winner of the German Pattern Recognition Award 2022.

# 13 Mohamed Elgharib: Vision-based Graphics

## Group Development

The research group Vision-based Graphics is led by Dr. Mohamed Elgharib. The focus of the group is neural reconstruction and rendering with a special interest in face digitization and relighting. Over the reporting period the group included 2 associated PhD students, 3 Master's thesis students and 2 long-term interns. Co-supervision included 1 post-doc, 2 PhD students and 3 Master's thesis students.

## Vision and Research Strategy

3D reconstruction and rendering is the cornerstone of digitizing our world. This has applications in several domains, including extended reality, movie and media production, medicine, robotics and others. The group aims to build digital models that are controllable and can be rendered in a highly photorealistic manner. The human face is at the center of our visual communications, and hence its digitization is of a special interest to us. Our work addresses all stages of the digitization pipeline, including data capture, modelling, reconstruction and rendering. We also have efforts in detecting synthetic content. While the group primarily works with RGB cameras, it is also interested in other types of input data including depth and audio. We are also very interested in event cameras, which are biologically inspired and neuromorphic based sensors. While earlier works of the group used explicit means of representing scenes, the group's work in the reporting period shifted more towards learnable implicit scene representations. Work presented by the group won several awards. This includes the best paper award in IVA 2021 and the best Paper Award Honourable Mention at BMVC 2022.

## Research Areas and Achievements

While the group has a special interest in digitizing the human face, it also developed methods for digitizing other objects and regions including the human body, the human hands, buildings and others. In this section, we will focus primarily on our work on human face digitization.

**Relighting** The group has a strong presence in facial relighting using light stage. This is marked by several top-tier publications. To this end, the group presented the first approach for facial relighting using learnable implicit scene representations that can take as few as a single image as input. This work learns a latent space for the facial identity and reflectance. Reflectance is modelled using one-light-at-a-time (OLAT) images captured using a light stage. This work was accepted in BMVC as an oral presentation, won the Best Paper Award Honourable Mention and got invited to a special issue of the Internal Journal of

Computer Vision (IJCV). The group also presented the first approach for high-quality facial relighting using generative models and light stage data. The main novelty here was to perform reflectance learning in the strong latent space of StyleGAN, which is a generative model of human faces. We plan to keep pursuing our efforts in light stage based facial relighting, especially after the installation of the Department's 3.8 meter multi-spectral light stage.

**Implicit 3D Head Morphable Models** The group pioneered the first approach for learning a 3D morphable model of the human intraoral region using neural implicit representations. This model learns a component-wise representation for each tooth and the gum and is controllable via individual latent codes. The model is trained using 3D dental scans and allows novel applications such as interpolation between two 3D scans, one taken before treatment and another after treatment. This line of work has the potential of bringing several strong impacts to the field of medical imaging.

**Other works in human face digitization** We developed several other methods for human face digitization. This includes methods for learning 3D morphable models of the human face and methods for learning generative models of still and moving faces, The group also largely contributed in building sophisticated hardware devices for face capture. This includes for instance building a multiview rig for recording moving faces. The rig is equipped with 24 4K Sony RXO II cameras that are hardware synced. We are now collecting state of the art facial dataset to foster further research in human face digitization.

## Projects and Cooperations

The group has strong collaborations with several institutes including MIT, Harvard, UCL, University of Erlangen-Nuremberg, IST-Austria and the University of York.

# 14 Tobias Fiebig: Sustainable, Secure, and Dependable Digital Infrastructure

## Group Development

Tobias Fiebig joined the MPI-INF in April 2022 as a researcher after holding a permanent facutly position at TU Delft. Since then he continued the supervision of two doctoral researchers from TU Delft on their path to a PhD and concluded the supervision of two master theses at TU Delft. At MPI-INF, he is currently building his group to further converge network measurements and socio-technial perspectives, with Florian Steurer joining his group (co-supervised with Prof. Feldmann). In addition his former intern Dr. Dao now joined MPI-INF as a post-doc from the National Institute of Informatics Japan, and Mannat Kaur joined as a researcher from TU Delft. Furthermore, he is supervising a bachelor (Jonathan Binkle) and master thesis (Simran Munnot).

## Vision and Research Strategy

Over the past half century, the rise and accelerating evolution of digital technology shaped society as much as society shaped technology. We now find ourselves in a world where digital infrastructure brings as much harm as it brings remidies, ranging from security issues in critical infrastructure to public values being threatened by platformization. Yet, we are still far from a true end-to-end understanding of these complex interaction effects in digital infrastructures, ranging from delayering technical complexity to socio-technical questions towards understanding those who ultimately build and maintain digital infrastructure.

With my research, I fill this knowledge gap about the technological bedrock of society by attaining an end-to-end understanding of building, maintaining, and ultimately living with digital infrastructure. To accomplish this, I developped a unique approach leveraging an interdisciplinary convergence of network measurements and computer security with methods from governance research and human-factors perspectives from the social sciences. This mixed-methods approach lets me synthesizes insights stretching beyond each individual field.

Drawing from my system and network engineer background and research experience as a security researcher allows me to thoroughly analyze digital infrastructure, and identify threats and points for improvement alike. Using qualitative methods from the social sciences enables me to explain how human factors interact with computer security and reliability. Using large scale network measurements I design, enables me to assess the impact and prevalence of issue, and serves as the foundation of analyses of societal developments that remain hidden without this quantitative perspective. Rooted in these observations, I use an analytical approach to design practical solutions—which may be technical, organizational, or policy driven—and empirically evaluate their efficacy.

## Research Areas and Achievements

**Metrics and Overview:** Since joining the MPI in April 2022, I co-authored 12 papers (6×A*/A, 2×B CORE) that are now either peer-reviewed and accepted for publication, or already published. Additionally, I gave the keynotes at two major international conferences of the networking community, disseminating my research to tens of thousands of professionals.
**Newtwork Measurement:** Over the past year I developped measurement methodologies enabling novel empirical studies of cloudification, DNS, censorship, and email. My empirical studies using these techniques led to several top-tier publications. Additionally, I valorize these results towards the public good via self-assessment platforms. Furthermore, following earlier work on email, I inform standardization processes, and work on a new instrument for the scientific community to increase the methodological, ethical, and research-data governance feasibility of active network measurements. Furthermore, my work on cloudification of higher education shaped the ongoing academic discourse, was solicited as expert input for the coalition negotiations of the current German government, and lead to an official inquiry in the Dutch parliament. Finally, I support internal digital infrastructure efforts of the MPG.
**Social and Human Factors:** Since joining MPI, my work on human factors in system administration and computer security with Mannat Kaur enabled us to demonstrate the connection between computer security, safety science, and feminism. This leads to a reframing from security as a goal to security as an emergent effect. Furthermore, my work on mental models of (non-)experts highlighted a connection between bad security/privacy outcomes and diverging mental models. In line with this, I currently lead an explanatory study on how academic cloudification connects to personal, organizational, and societal factors.

## Projects and Cooperations

I am currently involved in more than ten differnet research activities, ranging from measurement work on cloudification and the historic centralization of the Internet and supporting qualitative studies (collaborating with TU Delft, TU Wien), the development of increasing complexity in network protocols, e.g., DNS, Email (collaborting with Virginia Tec, SBA Research, and TU Wien), inquiries in censorship infrastructure (collaborating with TU Delft and TU Wien), IPv6 and DNS measurements (collaborating with TU Delft and TU Munich), human factors in security (collaborating with University of Paderborn, TU Delft, CISPA, Utrecht University), gender and social factors in system administration (collaborating with TU Delft and University of Paderborn), and finally methodological work on measurement techniques (tightly interacting with several operations communities).

## Prizes and Awards

*"How Ready Is DNS for an IPv6-Only World?"*, which I co-authored, received the best paper award at PAM2023. Furthermore, I received an outstanding reviewer award at CCS2022. Before joining the MPI, I was awarded 'Teacher of the Year' at TU Delft, and received a commendation by the European Data Protection Supervisor for my support of privacy preserving and self-sovereign digital infrastructure in the wake of the COVID-19 pandemic.

# 15 Oliver Gasser: Internet Security Measurements

## Group Development

Oliver Gasser joined MPI-INF as a postdoctoral researcher in February 2020. In July 2021 he was appointed Senior Researcher and is since then leading the "Internet Security Measurement" research group. He is currently supervising two PhD students (Ali Rasaii and Fahad Hilal) and co-advising three PhD students (Aniss Maghsoudlou, Zubair Sediqi, and Fariba Osali). He has supervised three student theses in the past two years and is currently also supervising one Bachelor's and one Master's thesis. Since December 2021, Oliver is also serving as the institute's scientific staff representative within the Max Planck Society.

## Vision and Research Strategy

The deployment of IPv6 in the Internet is slowly but steadily increasing. This brings many new possibilities when it comes to addressing the needs of today's Internet: billions of connected devices, scalable services and protocols, and resilient networks. The vast IPv6 address space, however, does not only have benefits, but also comes with novel challenges and risks. These range from privacy risks when devices use identifiable addresses instead of randomized ones, to new kinds of possible denial-of-service attacks when using billions of different addresses. One main research pillar of the group is to investigate possible risks stemming from the increased use of IPv6 and propose possible mitigations against these.

Moreover, a large part of today's Internet measurement research is done as a one-off study. This leads to outdated data with decreased value for other researchers and a lot of duplicated effort when redoing similar studies five years down the road. To remedy that, the group has strived to conduct research with a continuously running measurement service already in mind from the start. This has lead to multiple new measurement services over the past two years, covering areas such as SNMPv3, Multipath TCP, and hyper-specific BGP prefixes. Moreover, new tools such as BannerClick or Yarrpbox have been published for the research community to use. The group will aim to further pursue the goal of reproducible research, measurement services, and publicly available tools in the coming years.

## Research Areas and Achievements

**IPv6:** The continuous increase in IPv6 deployment in the Internet poses multiple research and deployment challenges. As part of a collaboration with researchers from TU Munich, we analyze the development of the IPv6 hitlist—which was published by Oliver Gasser and his

colleagues in 2018. We identify and ameliorate multiple artifacts in the hitlist data, analyze the effect of aliased prefixes on different types of scans, and add 5.6 M new responsive addresses to the hitlist. Moreover, a research collaboration with the Contend Distribution Network Akamai lead to the investigation of scanning in the IPv6 Internet. We find that, unlike IPv4 scans, large-scale IPv6 scans are still comparably rare events (60 ASes) and target a large number of ports (sometimes exceeding 100). This more closely resembles general and unspecific penetration testing behavior, as opposed to scanning patterns of botnets trying to spread laterally by exploiting individual vulnerabilities. Finally, together with colleagues from TU Munich and TU Delft we analyze how DNS would fare in an IPv6-only world. We find DNS to be heavily concentrated, with a single operator enabling IPv6 DNS resolution for around 15.5% of all non-resolving zones in our dataset until January 2017. Even today, 10% of DNS operators are responsible for more than 97.5% of all zones that do not resolve using IPv6. Research in this area has resulted in three papers: two published at ACM IMC 2022 and one published at PAM 2023, which was awarded with the best paper award.
**Security & Privacy:** With the steadily increasing deployment of Internet-connected devices, security and privacy are becoming more and more important nowadays. In a collaboration with researchers from TU Delft, we unveil that around 19% of end-users' privacy can be at risk due to a single device at home (mostly IoT devices) that encodes its MAC address into the IPv6 address—even if other devices are using privacy extensions. Furthermore, a research project between our group and KU Leuven investigated the prevalence of tracking cookies in the Web from a multitude of perspectives. For this we develop a tool to automatically interact with cookie banners and find that websites send, on average, 5.5 times more third-party cookies after clicking "accept". Finally, we developed a simple yet effective technique to fingerprint routers with SNMPv3 probe packets and we show that our technique can fingerprint more than 4.6 million devices of which around 350k are network routers. Research in this area has resulted in three papers published at ACM IMC 2021, in the ACM SIGCOMM CCR journal in 2021, and at PAM 2023.
**Protocols & Devices:** To better understand the use of different network protocols and devices, we analyzed the deployment of SRv6, BGP hyper-specific prefixes, the IoT backend ecosystem, the VPN ecosystem, middlebox deployments, and developed a system to correlate Netflow and DNS streams at scale. Research in this area has resulted in six papers published at WTMC 2022, ACM SIGCOMM CCR journal in 2022, ACM IMC 2022, ACM CoNEXT 2022, PAM 2023, and CoNEXT 2023.

## Projects and Cooperations

Our group is involved in multiple collaborative projects with universities, research institutes, and industry. This includes collaborations with researchers from TU Delft, KU Leuven, TU Munich, the University of Edinburgh, the CNRS, and the Global Cyber Alliance (GCA).

## Prizes and Awards

Oliver Gasser's co-authored paper "How Ready Is DNS for an IPv6-Only World?" was awarded the Best Paper Award at the 2023 Passive and Active Measurement Conference.

# 16 Vladislav Golyanik: 4D and Quantum Computer Vision

## Group Development

Vladislav Golyanik was appointed as a research group leader in May 2021; he established the *4D and Quantum Computer Vision (4DQV)* research group at the new D6 Department. As of April 2023, 4DQV includes two affiliated PhD students (Viktor Rudnev, since May 2021; Navami Kairanda, since February 2022). Vladislav Golyanik also co-supervises several other doctoral students at D6. During the reporting period, 4DQV accommodated four remote or on-site interns (Willi Menapace, a doctoral student at the University of Trento; Macrel Seelbach Benkner, a doctoral student at the University of Siegen; Rishabh Dabral, a doctoral student at IIT Bombay at the time; and Harshil Bhatia, an undergraduate student at IIT Jodhpur). Several bachelor's and master's theses were accomplished under the (co-)supervision of Vladislav Golyanik (Jalees Nehvi, Maximilian Krahn, Navami Kairanda, Hiroyasu Akada, Artur Jesslen, Zhouyingcheng Liao and Eric C. M. Johnson).

## Vision and Research Strategy

The primary research areas of 4DQV include 1) 3D reconstruction of general non-rigid scenes with different types of prior knowledge, and 2) quantum visual computing (QVC). We are also looking at closely related domains such as novel view synthesis and generative vision. QVC is an emerging research area that investigates how visual computing can leverage quantum hardware. Regarding QVC, we predominantly focus on quantum annealers, since they are able to sample combinatorial objectives of moderate sizes arising in visual computing and allow experimental evaluation of the developed methods on real quantum hardware.

One of 4DQV's long-term goals is the development of techniques for dense 4D reconstruction of the new generation that offer increased robustness and operate under less restricted assumptions than previous approaches. Next, we aim at revealing opportunities of the quantum computational paradigm, which, we believe, has high potential to reshape not only visual computing but also other domains of computer science. 4DQV actively uploads and maintains datasets and source codes for the accomplished works (*i.e.,* source codes of 15 methods and datasets for nine projects were released during the last two years).

## Research Areas and Achievements

4DQV advanced state of the art and published 25 papers during the reporting period (including 6xCVPR, 10xECCV/ICCV, 3x3DV, 1xTPAMI, 1xSIGGRAPH and 1xEG STAR).

**3D Reconstruction and Modeling of Deformable Scenes.** We focused on the 3D reconstruction of general deformable scenes, 3D human motion capture, 3D human pose estimation and 3D-aware (generative) video modeling of humans and complex scenes. Our interest in formulations involving physics-based constraints in inverse problems (*e.g.,* monocular 3D human motion capture and 3D reconstruction of general deformable objects) resulted in several pivotal contributions in the field (*e.g.,* SIGGRAPH'21, ICCV'21, ECCV'22, 2xCVPR'22). We also proposed innovative methods for 3D hand tracking from depth and event cameras (*e.g., HandVoxNet++* and *EventHands*) and our works on 3D-aware video modeling of humans and composite scenes received broad visibility (*e.g.,* CVPR'21, 3DV'21 and CVPR'22). Moreover, one of our neural rendering approaches (NR-NeRF, ICCV'21) was the first to enable photo-realistic novel view synthesis of general deformable scenes from monocular RGB videos. Last but not least, we published a comprehensive survey on advances in neural rendering at Eurographics (EG) 2022 that was coordinated by Vladislav Golyanik and co-authored by the leading researchers in the field.
**Quantum Visual Computing (QVC).** During the reporting period, we published four papers on QVC at primary computer vision conferences and one article in a journal studying low-level qubit properties. Several of our works address correspondence problems (*e.g.,* synchronizing permutations, CVPR'21, or iteratively matching 3D shapes, ICCV'21) and one method segments points in different frames into independent motions (ECCV'22). We also developed a general constrained solver for quadratic binary optimization using quantum annealers, called Quantum Frank-Wolfe (Q-FW) algorithm (ECCV'22). All these methods push the quantum state of the art in QVC and illuminate new ways to solve challenging problems in the field.

## Projects and Cooperations

We work closely with Christian Theobalt and the members of D6 at MPI-INF and collaborate with other researchers from different countries; we next present examples of recent collaborations initiated at 4DQV. In Germany, we had joint projects with Zorah Lähner, Michael Moeller (Uni Siegen), Florian Bernard (Uni Bonn), Matthias Nießner (TU Munich) and Justus Thies (MPI-IS). Internationally, we published with Tolga Birdal (ICL); Yifan Wang, Gordon Wetzstein, Leonidas Guibas (Stanford University); Pascal Fua (EPFL); Federica Arrigoni (Polytechnic University of Milan); Elisa Ricci (University of Trento); Masaki Takahashi (Keio University); Arjun Jain (IISc Bangalore); Sergey Tulyakov, Aliaksandr Siarohin (Snap Research); Franziska Mueller, Ben Mildenhall, Pratul Srinivasan, Ricardo Martin-Brualla, Jon Barron (Google Research) and Stéphane Lathuilière (Telecom Paris).

## Prizes and Awards

Vladislav Golyanik is a member of the ELLIS Society (since January 2023). He received the yearly Dissertation Award of the German Association for Pattern Recognition (DAGM) in 2020. One of his works (*MoCapDeform*) published during the reporting period received *Best Student Paper Award* at 3DV 2022, and another one (*Physical Inertial Poser*) was nominated for the best paper award at CVPR 2022 (<2% of the accepted papers).

# 17 Marc Habermann: Graphics and Vision for Digital Humans

## Group Development

Marc Habermann has obtained his PhD at MPI-INF in October 2021. He was appointed research group leader of the *Graphics and Vision for Digital Humans* group at MPI-INF and the head of the *Real Virtual Lab* in November 2021. The group has grown from 0 to 2 PhD students: Heming Zhu (joined in February 2022) and Guoxing Sun (joined in November 2022), and from 0 to 3 Master students. 1 Master thesis and 1 internship were completed during the reporting period. Moreover, Marc Habermann currently co-supervises 6 PhD students and 5 Master students.

## Vision and Research Strategy

Humans interacting with each other and their environment is a key element in everyone's life. This holds true from the beginning of humankind and has never changed since then. However, what has changed is the digitization of society over the last decades. This imposes major possibilities but also significant challenges in the way of how digital systems can sense humans, but also how humans can communicate with each other in the future.

Our group develops new foundational algorithms on the basis of Computer Graphics, Computer Vision, and Machine Learning to let computers perceive humans in front of digital sensing devices to learn a higher-level, robust, and detailed understanding of humans. Ideally, such algorithms should be able to sense every aspect of the human, e.g. gestures, facial expression, body pose, and the entire surface geometry. However, this starkly conflicts with the fact that sensing devices are typically sparse in our daily life. To address this challenge, our research focuses on 1) combining model-based and data-driven priors of humans, 2) exploring neuro-explicit representations for modeling humans, and 3) integrating the fundamental laws of physics in the capture process.

However, sensing and analyzing humans is only one aspect. Digitization also allows us to explore new and immersive types of human communication, which may partially happen in a virtual counterpart of the real world. Such technology could bring people closer together even though they are thousands of miles away from each other. One major challenge is that communication and interaction has to feel real in every aspect requiring photorealistic rendering methods of the human and drivable virtual avatars reflecting all the nuances that define you as an individual. We address those challenges by 1) exploring deeply entangled implicit and explicit representations, 2) reasoning about geometry, material and light transport, and 3) developing efficient algorithms scaling to high resolution data.

## Research Areas and Achievements

During the reporting period, our group has achieved major milestones concerning this long-term vision. In the following, we highlight the scientific achievements in each sub-category.

**Motion Capture.** We demonstrated the first real-time method for joint capture of the body pose, hand gestures, and facial expressions (CVPR'21). Further, we proposed the first physics-aware motion capture method (CVPR'22) using only six body-worn inertia measurement units, which we further extended for joint sensing of the scene geometry (SIGGRAPH'23). Moreover, we presented a novel method for joint multi-person pose and scene estimation (Eurographics'23), which disambiguates monocular ambiguities by combined reasoning of the scene and the humans.

**Performance Capture.** In addition to human motion capture, we also proposed the first weakly-supervised method for space-time coherent capture of the dense geometry of the human from monocular video (TPAMI'21). We further pioneered a new method integrating physical cloth simulation into deep learning frameworks for accurate cloth capture (3DV'21). Recently, we presented a novel method for joint capturing of body pose, clothing geometry, facial expressions, and hand gestures from single RGB videos (BMVC'22).

**Human Rendering.** Concerning high-quality human rendering, we proposed the first real-time approach for controllable and photorealistic avatars learned from multi-view video (SIGGRAPH'21). We further demonstrated how such avatars can be learned without requiring a template while even improving the rendering quality (SIGGRAPH Asia'21).

**Inverse Graphics.** We also conducted research on analyzing and rendering general scenes. We proposed an efficient and differentiable representations for shadows (ICCV'21) enabling better recovery of geometry, texture, and lighting. Moreover, we demonstrated the first neural method for global illumination-aware object relighting (ECCV'22). Recently, we demonstrated a method for non-rigid geometry recover from a single RGB video of a general object (CVPRW'23) and we contributed to a survery on non-rigid reconstruction from monocular video (Eurographics'23).

## Projects and Cooperations

In the reporting period, there have been joint publications with Feng Xu (Tsinghua University). Moreover, there are ongoing collaborations with Pascal Fua (EPFL), Henry Fuchs (University of North Carolina at Chapel Hill), Michael Zollhoefer (Meta), and Philipp Slusallek (DFKI).

## Prizes and Awards

Marc Habermann has received numerous awards for his PhD thesis. This includes the Eurographics PhD Award 2022, which is the highest PhD award in the area of Computer Graphics in Europe. Moreover, he received the DAGM MVTec Dissertation Award 2022, which is the highest PhD award in the area of Computer Vision and Pattern Recognition in Germany. The Max Planck Society awarded Marc Habermann with the Otto-Hahn Medal 2022 for outstanding scientific achievements. The work *Physical Interial Poser* has been nominated for best paper candidate at CVPR 2022 (1.6% of accepted papers).

# 18 Andreas Karrenbauer: Optimization

## Group Development

The group was established in January 2013 with the support of the Max Planck Center for Visual Computing and Communication. In this report period, Gao Yuan and Paolo Luigi Rinaldi joined as PhD students, both in October 2021.

## Vision and Research Strategy

Optimization problems are ubiquitous. Whenever there is a choice between several possibilities, there is an inherent optimization problem. Our core competence in this area is to recognize such optimization problems, establish mathematical models, tackle them by state-of-the-art methods, and invent new techniques to obtain satisfactory results. To this end, we combine methods from discrete and continuous optimization. Our research is application-driven, which also includes inter-disciplinary research with collaborators from areas such as engineering, physics, chemistry, biology, and economics. Furthermore, connections and synergies with the field of machine learning are being explored.

## Research Areas and Achievements

For this report period, we highlight application-driven and inter-disciplinary research projects.

### Optimization in Car Manufacturing

Cars are manufactured on assembly lines, that is, they are attached to conveyors that move them through the plant at a certain speed so that they leave the factory one after the other at a fixed rate (between half a minute and a minute). Since customers of new cars can configure a wide variety of features (color, engine, tow-bar, GPS, etc.) individually, there are rarely two identical *orders* on the same production day.

The build sequence for a production day is typically generated a few days in advance to allow the suppliers to deliver the corresponding parts just in time. However, a sequence that allows for large batches of the same color in the paint shop, might cause a line stop in the final assembly when the workload for a certain feature of back to back cars exceeds the capacity of the corresponding station, for instance, it might only be possible to mount a tow-bar to every other car in the allotted time.

Therefore, a plant is partitioned into three main stages: welding of the body, painting of the body, assembly of the car (engine, seats, wheels, and so on). These stages are separated by buffers so that a line stop in one area does not immediately affect the production in another area. These buffers can consist of multiple parallel lanes so that they can not only be used for storage, but also allow for a limited reorganization of the build sequence. This is particularly helpful when the initial sequence is distorted due to extra work on some cars. In particular, paint is not perfect and hence a certain fraction of the cars needs additional work, so the original sequence is disturbed after paint and needs to be restored as much as possible.

In the first stage, bodies have no paint and no particular features. Therefore, at this point in the process, a body is compatible with many possible *orders*. The *order* is the set of all the features that a car is expected to have at the end of the process. Each *order* is identified with a *Vehicle Identification Number* (VIN). Since a body can be paired with several possible VINs, we aim to exploit this flexibility to produce a sequence that results in minimal costs for the plant. As these bodies proceed with the next steps, more features are applied to them, and the degrees of freedom decrease. Perfect sequencing could result in an efficiency gain of several percent.

In a project together with the Ford Motor Company, we have developed algorithms for restoring order in a buffer with multiple parallel lanes while minimizing the costs for color changes in the paint shop and adhering to given production constraints. Moreover, we have designed a software architecture to facilitate the integration into the production process, which was introduced in a plant of our industrial partner in Q1 of 2023.

The development of our effective algorithms for practice would not have been possible without a profound understanding of the limited sorting capabilities of such lane buffers. To this end, we addressed this sorting problem in an ideal and formal framework.

### Computational Interactivity

We are continuing our work in in the area of *Computational Interactivity*, that is, the use of computational methods in the field of Human-Computer-Interaction. Clearly, there are multiple alternative designs for a user-interface for a given task so that we are faced with an optimization problem. However, in contrast to classical optimization where constraints and objectives are mathematically well-defined, the big challenge here is that the notion of a good (better/worse) design is highly subjective. Typically, many stakeholders are involved having competing opinions, which may even evolve over time.

One of our research highlights in this area is the successful engineering of global optimization techniques to optimize non-linear optimization problems that appear in the context of the placement of electro-physiological sensors on the human body, which was published in *Nature Communications* in November 2021.

## Projects and Cooperations

The project MoDigPro is concerned with the modernization and digitalization of the production scheduling in the Ford plant in Saarlouis. In this project, we collaborate with Prof. Verena Wolf and her group at Saarland University.

Our work on computational interactivity was partly done in collaboration with Prof. Antti Oulasvirta and his group at Aalto University, Finland. The results on electro-physiological sensors came from a collaboration with Prof. Tobias Kraus from the Leibniz Institute for New Materials and Prof. Jürgen Steimle and his group at Saarland University.

# 19 Thomas Leimkühler: Image Synthesis and Machine Learning

## Group Development

Thomas Leimkühler obtained his PhD at MPI for Informatics in 2019, spent two years as a Postdoc at Inria Sophia-Antipolis, France, and was appointed group leader of the *Image Synthesis and Machine Learning* group in September 2021, where he is jointly affiliated with D4 and D6. Currently, Thomas is advising two PhD students – one as the main supervisor (Ntumba Elie Nsampi), and one in a co-advising role (Linjie Lyu) – as well as a Masters student (Adarsh Djeacoumar). Further, the group has hosted a research intern (Alireza Javanmardi) during the reporting period.

## Vision and Research Strategy

Our group conducts foundational research at the intersection between computer graphics and machine learning with a focus on rendering. We seek to develop algorithms and data structures that enable fast and high-quality image generation for interactive virtual environments. To this end, we consider a broad range of approaches for synthesizing images and for reasoning about light transport.

Traditional rendering frameworks usually employ a mesh-based scene representation and use ray-tracing or rasterization techniques to produce a discrete pixel grid of color values. We explore alternative and complementary designs of the image synthesis pipeline, especially (but not exclusively) in view of recent advances in machine learning. Meshes as scene representations are complemented by implicit fields, point clouds, mere images, or expressive latent spaces emerging from training data. Ray-tracing and rasterization are augmented with neural networks, while pixel grids transition into continuous multi-scale image representations. To maintain the efficiency required for interactive virtual environments, we develop natively parallel solutions.

## Research Areas and Achievements

### Diverse Rendering and Light Transport Analysis

We consider all types of image synthesis algorithms and explore the entire continuum of techniques. This includes physically-based, image-based, and data-driven/neural rendering – each of which offer different advantages we seek to reconcile. In this context, we have developed a method to render reflections of curved surfaces based on multi-view observations, using a novel point-based neural rendering framework (SIGGRAPH Asia/TOG 2022). We have further refined the approach (SIGGRAPH/TOG 2023) to yield state-of-the-art novel-view synthesis quality at real-time rates. Additionally, we have looked into global illumination-aware inverse rendering for relighting, combining traditional graphics approaches with modern neural representations for light transport (ECCV 2022 Oral).

### Generative Models

Generative image models using deep-learning techniques are approaching photo-realism and have received a lot of attention also in the general public. We investigate how these expressive models can be included into the image synthesis pipeline in a controllable manner. For example, we have developed a method that augments a pre-trained 2D Generative Adversarial Network with precise 3D camera control (SIGGRAPH Asia/TOG 2021). In ongoing work, we are investigating the application of generative models into inverse-rendering and relighting pipelines, image generation with an orders-of-magnitude variety of scales, as well as learning high-dynamic-range images from low-dynamic-range data.

### Efficient and Expressive Scene and Image Representations

The representation of a scene or an image is tightly linked to the algorithm used for generating synthetic imagery. We therefore explore different representations, including meshes, implicit fields, point clouds, etc. in combination with their corresponding properties in terms of image quality, efficiency, sparsity, and controllability. Many rendering applications, such as virtual and mixed reality, require visual feedback in the order of milliseconds. To meet these extreme constraints, we conduct algorithm development with in-built parallelism, such that we can benefit from specialized hardware such as GPUs. In recent work, we have developed a method to perform complex geometric edits in a Neural Radiance Field (i3D 2023). Thanks to a GPU-friendly implementation, the system enables real-time editing of large-scale scenes. In ongoing work, we are developing methods to perform signal processing in implicit neural representations.

## Projects and Cooperations

In addition to a diverse set of projects across D4 and D6, our group is involved in collaborations with Inria Sophia-Antipolis, University College London, Massachusetts Institute of Technology, Universidad de Zaragoza, Meta, and Saarland University.

# 20 Jan Eric Lenssen: Geometric Representation Learning

## Group Development

Jan Eric Lenssen obtained his PhD from TU Dortmund University in February 2022 and joined MPI-INF as a postdoctoral researcher in November 2021. After one year of co-supervising PhD students in multiple projects, he was appointed group leader of the *Geometric Representation Learning* group within D2 in March 2023, and is in the process of becoming a Senior Researcher. He is main superviser of one PhD students (Christopher Wewer), one RIL student (Ayçe İdil Aytekin), three Master students, and co-supervises six PhD students (Keyang Zhou, Verica Lazova, Julian Chibane, Philipp Schröppel, Yannan He, Raza Yunus).

## Vision and Research Strategy

Most observable parts of the 3D world consists of visually structured objects and repeating patterns, following rules of composition. It is our current understanding that it is this inherent structure and repetitiveness that allows humans to map out their surroundings just from a few sparse 2D observations, by combining observations with a structured priori knowledge. Replicating this ability has been a longstanding goal in computer vision with increasing success in recent years through advances in deep learning.

Traditionally, computer vision algorithms were designed down to the last detail, carefully defining how data is processed. Deep learning, in contrast, allows to define how a priori knowledge is learned, represented and processed on a higher level of abstraction, by designing architectural inductive biases. This paradigm shift allows us to automatically discover patterns from large amounts of data, which are often too complex and of too high dimension for the human mind to be grasped and expressed in algorithms. Models learned in such a way are often referred to as *data prior*. Our current methods are far from perfect in combining data priors with new observations. While we believe that we soon have the computational capacities to implement a near human-level vision system, we lack the methods to successfully make the step from 2D observations to 3D representations. Our research aims to bridge this gap by finding optimal representations for 3D data priors that come with a favorable and controllable prior vs. observation trade-off.

## Research Areas and Achievements

**Geometric Representations** Current research on 3D representations is mostly centered around neural fields. They range from methods working purely as compression, such as NeRF, with a weak inductive bias from continuous Multi-Layer Perceptrons (MLPs), to strong category-level data priors, e.g. Scene Representation Networks that model the full space of radiance functions describing instances of one category. However, there has yet to appear a representation that allows to learn data priors about structural correspondences and patterns from arbitrary in-the-wild images on a large scale, which would allow, for example, to reuse reoccurring patterns in full scenes or to automatically learn symmetries in objects.

We recently did a first step in that direction by creating a neural point representation modeling category-level symmetries (ICCV'23 submission) that allow for high detail single-view reconstructions. Given that neural point representations are a very young research area, there is still much to learn about capabilities and applications. Future research in this direction can build upon past contributions and experience in the areas of graph neural networks, geometric deep learning and modelling symmetries, which are necessary to process irregular structured data such as neural points with deep neural networks.

**Data Prior vs. Observation** Especially in the area of 3D reconstruction we face the intricacies of the data prior vs. observation trade-off when performing inference with a learned data prior. If the data prior overpowers the observation, the method mostly performs retrieval from the prior. The opposite leads to noisy, incomplete and inconsistent results. Commonly, feed-forward deep neural networks learn a function, which is evaluated using an observation as input, e.g. in our work for denoising hand-object interaction sequences (ECCV'22). However, we recently explored alternatives, which expose much more control over data prior influence. We revisited test-time optimization, sometimes referred to as autodecoding, in a recent work about neural point representations (ICCV'23 submission). This formulation allows explicit control of how the data prior acts by designing inductive biases of the data prior function. Further, we proposed to use an unsigned distance function (ECCV'22, Best Paper Honorable Mention), representing a high-dimensional manifold as data prior, indicating plausibility as distance. This formulation allows explicit control over prior strength by weighting and might be applicable to other domains as well. Exploring the trade-off and the respective methods is of importance for the formulated research goal. It is part of our research agenda to further extend knowledge in this area.

## Projects and Cooperations

Our group is currently involved in collaborative projects with Saarland University (Prof. Dr. Eddy Ilg), University of Tübingen (Prof. Dr. Gerard Pons-Moll), University of Freiburg (Prof. Dr. Thomas Brox), Imperial College London (Prof. Dr. Tolga Birdal), and Kumo.ai (Prof. Jure Leskovec from Stanford University), a startup Jan Eric Lenssen is involved in as a Founding Engineer and which is partially built upon his previous work on GNNs.

## Prizes and Awards

Jan Eric Lenssen co-authored the paper *Pose-NDF: Modeling Human Pose Manifolds with Neural Distance Fields*, which received a Best Paper Honorable Mention Award at ECCV 2022. Further, Jan Eric Lenssen received the Dissertation Award of TU Dortmund University for the best computer science dissertation in 2022.

# 21 Karol Myszkowski: Perception in HDR Imaging, VR, and Material Appearance

## Group Development

The research group was founded in June 2000 as a part of the Computer Graphics Department and initially, it was focused on the human perception aspects in global illumination and rendering. The scope of its research gradually expanded towards High Dynamic Range Imaging (HDRI), whose goal is the precise representation of real-world light intensity and the color gamut at all stages of image and video processing, from acquisition to display. Recently, we are interested in more general aspects of computational photography and virtual reality (VR). We return also to global illumination simulation but this time in the context of computational fabrication, material appearance perception, and perception-driven Monte Carlo sampling.

Currently, the group consists of two postdocs (Bin Chen and Lingyan Ruan) and five Ph.D. students (Krzysztof Wolski, Mojtaba Bemana, Chao Wang, Uğur Çoğalan, and Martin Balint). The Ph.D. thesis of Mojtaba Bemana is currently under evaluation, and Krzysztof Wolski will submit his Ph.D. thesis in the coming weeks.

## Vision and Research Strategy

The common goal of all group research efforts is the advancement of knowledge on image perception and the development of imaging algorithms with embedded computational models of the human visual system (HVS). This approach offers significant improvements in both the computational performance and the perceived image quality. Often, we refer to *perceptual effects* rather than *physical effects*, which puts more emphasis on the *experience* of the observers than *physical measurements*. In particular, we aim for the exploitation of perceptual effects as a means of overcoming the physical limitations of display devices and enhancing the apparent image quality.

Seamless matching of visual content from the virtual and real world is a long-standing goal in computer graphics. However, strict reproduction of luminance and contrast levels from the real world typically is not feasible, even for modern displays. Although VR setups reproduce important visual cues such as binocular vision and motion parallax, some constraints can be imposed on depth and motion ranges. Having a model of the HVS which accounts for all these limitations enables dynamic reallocation of rendering resources to achieve the best visual quality given computing hardware and display characteristics.

We consider 3D printing as another aspect of the seamless transition between virtual and physical objects, we are aiming at developing a domain-specific perception model for faithful appearance reproduction in fabrication. When objects are 3D printed their physical appearance is expected to match their virtual models as closely as possible. The perceptual aspects of such appearance matching have not been investigated thoroughly so far, given that each printing technology might impose some unique constraints, such as strong light diffusion within the material that washes out the texture details of the object.

One of the key goals in our research vision is spreading awareness of the importance of HVS modeling for computer graphics and image processing. We achieve this by co-authoring textbooks, such as the one on High Dynamic Range Imaging (over 2,500 Google Scholar citations). Another important mission of our group is training of Ph.D. students and postdocs in applied perception, who proliferate such expertise in new research areas in their future scientific activities. Many former Ph.D. students, for example, Rafał Mantiuk (now Professor at Cambridge University), Piotr Didyk (now Associate Professor at USI Lugano), Yulia Gryadistskaya (now Lecturer at Surrey University), Petr Kellnhofer (now Assistant Professor at Delft University), Tunç Aydın (now Research Scientist at Disney Research), Dawid Pająk (now Distinguished Engineer at Nvidia), who started their independent research careers, published their work with strong perceptual components at top venues ranging from Nature to ACM SIGGRAPH/TOG (over 50 papers). Also, several former post-docs who accepted faculty/research positions, for example, Vlastimil Havran (now Full Professor at CTU Prague), Martin Cadik (now Associate Professor at Brno University), Belen Masia (now Associate Professor at the University of Zaragoza), Cara Tursun (now Assistant Professor at the University of Groningen), Peter Vangorp (now Assistant Professor at Utrecht University), Ana Serrano (now Assistant Professor at the University of Zaragoza), Gurprit Singh (now Group Leader at MPI for Informatics), and Hyeonseung Yu (now Researcher at Samsung), often include perception-aware components to their research.

## Research Areas and Achievements

**High Dynamic Range Imaging** High Dynamic Range (HDR) Imaging is a well-established research area, where core problems such as tone mapping and inverse tone mapping have been recently revisited using machine learning tools. The existing approaches typically rely on fully supervised solutions that in the case of tone mapping require pairs of high-quality tone-mapped and HDR images. This constitutes a problem as given the specific HDR image content a suitable tone mapping must be selected and its parameters intensively tuned for the best visual result. We advocate for self-supervised tone mapping solutions that are trained at test time and this way become optimized for each specific HDR image. To this end, we experiment with findings on contrast perception that we transfer into the feature space. For example, in our recent work, we achieve this goal by reformulating classic VGG feature maps into feature contrast maps that normalize local feature differences by their average magnitude in a local neighborhood. This allows us to formulate the loss function that accounts for contrast masking effects. We are also active in inverse tone mapping research, where we attempt to develop a generative framework to produce directly HDR images. Interestingly, our framework is fully trained exclusively on low dynamic range images that are abundantly present in numerous repositories. This means that we overcome completely the need for any HDR image in our training.

Another track of research is revisiting specialized sensors for HDR video capturing such as dual-exposure sensors. We demonstrate that such sensors in combination with proper CNN-based solutions greatly facilitate such tasks as denoising, deblurring, and view interpolation that otherwise might be difficult for temporally interleaved multi-exposure video using traditional single-exposure sensors. On the other hand, CNN-based methods are great in compensating for notorious problems of dual-exposure sensors with a reduced resolution that is trivial to solve by modern super-resolution techniques. An exciting area of future work is a more systematic investigation of how sensor weaknesses can be compensated by CNN solutions and vice versa, which should lead to novel sensor designs and new challenges for CNN techniques to process such data.

**Improving Visual Experience in VR** The rendering performance is an important requirement in VR applications. Recently, there is a shift in real-time graphics from rendering with a fixed resolution and refresh rate to a more adaptive approach, in which we control spatiotemporal resolution in order to maximize the quality under a given rendering budget. Variable Rate Shading (VRS) and adaptive refresh-rate monitors provide an opportunity for such fine control, which becomes particularly important as the increasing resolution and refresh rate of displays place a high demand on the GPU, especially on mobile devices. We propose a novel rendering method that takes the advantage of the limitations of the visual system to reallocate the rendering budget to the most vital part of the spatiotemporal domain. The key component of the method is our new metric, which considers how the judder, aliasing, and blur artifacts introduced by the VRS at a given refresh rate are masked by hold-type blur, eye motion blur, and limited spatiotemporal sensitivity of the visual system. The metric is shown to explain multiple datasets with only a few fitted parameters.

Another important aspect of VR is image reproduction on head-mounted displays. Dimming a display can be beneficial for VR experience as it reduces the visibility of flicker, saves power, prolongs battery life, and reduces the cost of the device. The major downside of this approach is the reduced sensitivity to stereoscopic depth cues. We find increased difficulty and lower precision in assessing 3D shapes based on binocular cues. This motivates our method for enhancing contrast at low luminance levels, intended to improve stereoscopic depth cues' reliability. The proposed method can improve the user experience for VR headsets that need to operate at low power or those that cannot achieve high refresh rates.

**Neural rendering** Traditional photorealistic rendering of real-world scenes proves tedious and challenging due to the need to reconstruct all physical parameters describing the rendered scenes. Recently, implicit scene representations have become a viable alternative to this task, where the entire scene is encoded into the parameters of a neural network. Despite recent advances in such scene representations, existing approaches cannot properly reconstruct novel views of transparent objects with complex refraction and require special treatment. We propose a solution to this problem by lifting the assumption that light rays are traversing in straight lines and adapting a physically correct approach to bend the light rays when they intersect with a refractive object in the scene. Specifically, we integrate the physical laws of the eikonal light transport with a state-of-the-art novel view synthesis method (NeRF) and adapt an implicit representation that can model refractive objects with a spatially varying index of refraction, leading to high-quality novel view reconstruction of refractive objects without requiring a dedicated capturing setup or scene configuration.

We have also developed a novel low-pass pyramidal kernel prediction scheme to denoise path-traced images. To this end, we learn partitioning the input radiance between pyramidal layers and apply learnable denoising and upsampling kernels.

**Appearance Reproduction in Computational Fabrication** Material appearance is determined not only by material reflectance, but also by surface geometry and illumination. The same material may have a different appearance for different combinations of lightings and shapes, or conversely, different materials may appear similar for specific illumination and shape choices. We analyze these effects by collecting a large-scale dataset of perceptual ratings of appearance attributes with more than 215,680 responses for 42,120 distinct combinations of material, shape, and illumination. We use the collected dataset to train a deep-learning architecture for predicting perceptual attributes that correlate with human judgments. We demonstrate the consistent and robust behavior of our predictor in various challenging scenarios, which, for the first time, enables estimating perceived material attributes from general 2D images.

In another research track, we investigate the problem of gloss appearance matching between the real world and display depiction as a function of material glossiness, surface geometry, scene illumination, and display luminance. To this end, we fabricate a wide range of differently painted glossy objects and create a dataset of the corresponding HDR photographs and rendered images. We use the dataset in a large-scale perceptual experiment, where we systematically investigate the gloss-matching task as a function of essential factors influencing gloss perception. The collected data enables deriving the quantitive measurements of the gloss difference between the real and virtual worlds. We find a strong dependency of such corrections on the scene illumination, display luminance, and a weaker influence of the object's geometry. Finally, we propose a model which predicts the correction to minimize the gap between real samples and displayed counterparts. We show that our gloss correction can significantly reduce the appearance gap in the digital product design, fabrication, and digitalization of physical artifacts.

**Publications** In the reporting period, the group members published their work at many top venues such as ACM TOG/SIGGRAPH/SIGGRAPH Asia (8), EUROGRAPHICS (2), ACM TAP (1), TVC (1), and Comp&Graph (1).

## Projects and Cooperations

Our close collaboration with the groups of Piotr Didyk (USI Lugano), Tobias Ritschel (UCL), Rafał Mantiuk (Cambridge University), Gurprit Singh (MPI-INF), and Vahid Babaei (MPI-INF) resulted in multiple joint papers. We have been also working with external partners on multiple short- and long-term projects: with Ana Serrano (Zaragoza University) on *on neural tone mapping, and perceived gloss reproduction in real (fabricated) and virtual (rendered) objects*, and with Iliyan Georgiev (Adobe) *perception-based Monte Carlo sampling*. Our research has been supported by Horizon 2020 Training Networks *RealVision: Hyperrealistic Imaging Experience*, which funded one Ph.D. student for three years (completed in 2022).

# 22 Rishiraj Saha Roy: Question Answering

## Group Development

The research group on *Question Answering* is led by Rishiraj Saha Roy, and currently involves Gerhard Weikum, Magdalena Kaiser (Ph.D. student), and Philipp Christmann (Ph.D. student). One Ph.D. student (Azin Ghazimatin) graduated during the reporting period. An intern (Jingjing Xu, University of Luxembourg) was hosted by the group in the summer of 2022. An invited speaker (Prof. Arijit Khan, Aalborg University Denmark) visited the group in November 2023.

## Vision and Research Strategy

Research on question answering (QA) aims to provide direct answers to natural language (NL) utterances, either over structured knowledge bases (KB), Web tables, infoboxes, unstructured text documents or combinations of these. In our group, we have tried to push the state-of-the-art in QA along multiple dimensions, the current driving criteria being handling complex and conversational questions. The general guiding principles behind the developed solutions are that they should be robust, explainable, and efficient. Robustness is defined as generalizability beyond a handful of domains or benchmarks, explainability is geared towards personas like end-users or system developers, and efficiency considers both memory and runtimes. To provide an introduction to factual QA, and to bring fragmented perspectives of QA in the the IR, NLP and AI communities closer together, Rishiraj (jointly with Prof. Avishek Anand from TU Delft, The Netherlands) co-authored a book on QA published by Springer. All team members delivered several invited talks on the research topics listed below.

In a parallel research thread, we have also explored non-standard and emerging problems about transparency in online forums like community question answering (CQA) sites and e-commerce platforms. The vision here has been to enable users with interpretable and actionable explanations to items recommended by such online service providers.

## Research Areas and Achievements

**Learning from question reformulations.** Information needs are rarely one-off: users often pose follow-up questions on a topic with the context left implicit in later turns, simulating dialogue with another human. We proposed the first mechanism for learning from user feedback in such conversational question answering (ConvQA) systems, using reinforcement learning with question reformulations. We also released a benchmark with question reformulations, where real users interacted with a live ConvQA system at the backend. We published our work in SIGIR 2021.

**Answering over heterogeneous sources.** Existing ConvQA systems tap into a single information source: a KB, or a text corpus, or a collection of Web tables: these methods miss out on the benefits of answer coverage and answer redundancy, that can be obtained by tapping into these sources together. We propose an explainable pipeline using intermediate structured representations and iteratively applied graph neural networks, as the first systems for ConvQA over heterogeneous sources. We also released the first benchmarks for heterogeneous QA, containing questions over five to ten turns, and ranging over five domains. We published our work in: SIGIR 2022 and SIGIR 2023.

**Tackling complexity in information needs.** Another key challenge arises from complexities in the information needs. We proposed a method that can answer complex questions with multiple entities and relations directly from heterogeneous Web sources, by computing compact subgraphs over dynamically constructed noisy quasi KBs. We also looked at the special case of questions with temporal intent, where a combination of dense-subgraph algorithms and relational graph convolutional networks was found to be effective. We published our work in CIKM 2021.

**Search space reduction for question answering.** A key challenge in most QA models is to reduce the volume of the search space inside a KB, to look for candidate answers. We provided the first dedicated method for such search space reduction, outperforming the most common baseline of using a named entity disambiguation system. As an auxiliary contribution, we created an API that enables fast and easy access to large knowledge bases that are Terabytes in size. The API already has more than 20 million external calls. We published our work in WSDM 2022 and BTW 2023.

**Providing explanations for online recommendations.** We showed that explainability in recommender models, a topic of high contemporary interest in the ML and AI communities, can lead to both user satisfaction and performance improvement. Through the course of a PhD thesis, we showed, over a variety of online forums, how to generate user-side and provider-side explanations, followed by their application to model enhancement. Further, we provided the first method for generating counterfactual explanations for neural recommenders. We published our work in WWW 2021 and SIGIR 2021.

## Awards

**GI DBIS Award.** Azin Ghazimatin was awarded the German Computer Science Society's best Dissertation Award (2023) in the field of Databases and Information Systems for her PhD thesis titled "Enhancing Explainability and Scrutability of Recommender Systems".

## Projects and Cooperations

Current external collaborators include Soumajit Pramanik (IIT Bhilai, India), Jesujoba Alabi (Saarland University, Germany), Zhen Jia (Southwest Jiaotong University, China), Khanh Hiep Tran (SAP, Germany), Azin Ghazimatin (Spotify, Germany), and Avishek Anand (TU Delft, The Netherlands).

# 23 Gurprit Singh: Sampling & Rendering

## Group Development

Gurprit Singh joined MPI-INF as a post-doctoral researcher in September, 2017. He is currently leading the Sampling and Rendering Group in the computer graphics department. Gurprit is supervising two PhD students (Corentin Salaün and Xingchang Huang), co-advising a PhD student from DFKI (Misa Korac) and a research intern (Sascha Holl) who will start his PhD in April. In the past two years, he has also supervised two research interns (Pacome Luton and Sylvain Hofs) and two Master thesis students (Shishir Reddy and Varshini Muthukumar). Another prospective PhD candidate is planning to join his group in the Fall 2023.

## Vision and Research Strategy

Over the past two years, the research direction of the group has expanded largely from the frontiers of Monte Carlo sampling for rendering to machine learning, texture synthesis, perception-oriented rendering and point cloud reconstruction. In the area of machine learning, almost all architectures rely on random samples. The impact of correlated sampling is not studied since the theory of Monte Carlo and Quasi-Monte Carlo (MC/QMC) sampling cannot be directly leveraged in machine learning setting. The group is investigating different machine learning models (diffusion, GANs) to bridge this gap and tap the vast knowledge of MC/QMC sampling into these models. Another closely related area is differentiable rendering. Recently, the group is exploring the impact of point correlations on haptic sensing in collaboration with Paul Strohmeier (MPI). The goal is to develop tools to reproduce haptic experiences of certain materials using simply the point correlations. The next two year goal for the group is to leverage benefits of correlated sampling in machine learning and differentiable rendering.

## Research Areas and Achievements

Our research areas are broadly categorized under *sampling* and *rendering*. Over the past two years we have published in both categories.

**Sampling** Randomness can be avoided by introducing meaningful correlations among samples. These samples could be the points used for stippling art, object placement, better data visualization or Monte Carlo estimation. We developed solutions for all the applications and the work has been published in the top-tier conferences. At EG 2021, we introduced the concept of *blue noise* to visualizing data samples. Blue noise samples maintain a minimum distance from each other avoiding clusters. We developed a multi-class blue noise sampling strategy to generate blue noise plots. At EGSR 2022, we first propose a data-less optimization strategy that allows point pattern expansion. The idea is to expand a small point pattern used as an exemplar on a large canvas. This has direct applications in object placement and element texture sythesis. At SIGGRAPH Asia 2022, we presented a framework based on optimal transport that unifies sampling strategies for stippling art, object placement and Monte Carlo rendeirng of global illumination.

**Rendering** We have been actively working towards bridging the gap between machine learning and traditional rendering strategies. At NeurIPS 2021, we proposed a differentiable pipeline that can render participating media volumetric effects from images. The idea is to use ray tracing for direct illumination and spherical harmonics for indirect illumination. The resulting pipeline allows editing and rendering volumetric effects like smoke and clouds. At SIGGRAPH 2022, we presented two more papers. The first paper propose to regress point samples to obtain a new Monte Carlo estimator with provable improvements over traditional estimators. The second paper takes a closer look at error distribution on image space in Monte Carlo rendering. We propose a perception-oriented optimization framework that gives visually pleasing error distribution for rendered images.

## Projects and Cooperations

Over the past two years, the group is actively collaborating with industry and academia. This includes University College London (Tobias Ritschel), University of Waterloo (Toshiya Hachisuka), McGill University (Adrien Gruson), VinAI Research (Binh-Son Hua), AutoDesk Arnold renderer and Adobe (Iliyan Georgiev). Internally, we are collaborating with other MPI-INF members (Paul Strohmeier, Karol Myszkowski, Hans-Peter Seidel).

# 24 Paul Strohmeier: Sensorimotor Interaction

## Group Development

The Sensorimotor Interaction Group was established in July 2021 and is led by Paul Strohmeier. The group consists of PhD student Nihar Sabnis, PostDoc Courtney Reed, Visiting Researcher Dennis Wittchen, returning interns Gaby Vega and Valentin Martinez Missir, and HiWi-turned-IMPRS student Sina Mavali. The team has expanded with the addition of Yuran Ding, a shared PhD student with the Maryland Max Planck Ph.D. Program in Computer Science, in May 2023.

## Vision and Research Strategy

The Group aims to advance the understanding of sensorimotor interactions and apply this knowledge to the development of new technologies. The group's current focus is on creating platforms to build upon in later research. This work has already paid off: The open-source platform for vibrotactile rendering we developed was not only deployed as a research tool in two further publications but is also being used by other research groups in Denmark, Canada, and the USA and has won an award at CHI 2023.

We are currently exploring sensorimotor loops in three domains. The first is **mediation**, in which we explore ways in which technology can represent information. Next, we investigate **vibrotactile rendering**, particularly within the context of Augmented Tactile Reality. Finally, we explore how we might enhance the experience of our body and agency in the world through **human augmentation**. Through the combination of these research areas, we are developing a comprehensive understanding of how sensorimotor interaction can be used in the broader field of human-computer interaction (HCI).

## Research Areas and Achievements

**Mediation & Tacton Design** Our interactions with computers are mostly explicit, symbolic, and declarative. We explore complementary methods involving implicit, embodied, pre-reflective, and non-declarative experiences, focusing on the whole body rather than just visual communication. We have explored the use of such pre-reflective cues in the design of hermeneutic symbols, demonstrating that incorporating embodied mediation can enhance the efficacy of symbol encoding and broaden the affective range of tactile symbols.

**Vibrotactile Rendering** Many material experiences can be created using only vibration, including friction, compliance, torsion, bending, and elasticity. We demonstrate that these experiences all rely on the same perceptual mechanism. Our work suggests that the specifics of a material experience depend less on the rendering algorithm but largely on the action performed when interacting with the material.

In HCI research, such haptic systems are seldom used, due to their complexity. We created Haptic Servos, accessible haptic rendering devices with low latency, which allows novices to set up a basic system rendering system in minutes. Haptic Servos were deployed in our research, demonstrating their effectiveness in creating material experiences and influencing action perception. Haptic Servos received a Best Paper – Honorable Mention award at CHI 2023.

**Human Augmentation** The Sensorimotor Interaction Group has participated in efforts to better understand the relationship between technology and the human body, identifying four analytical lenses: material, morphological, sensorimotor, and experiential views. These views help understand the emerging fields of Human Computer Integration and Embodied Interaction

We have also engaged in Human Augmentation by designing novel interfaces: In collaboration with Saarland University, we created a handheld printer for collaborative circuit design, where control is shared between user and computer. We have also worked on textile-based devices, such as SingingKnit, which measures muscle activity during singing to augment vocal performances and demonstrates how biofeedback can change body perception and artistic action. The group has contributed to the experiential aspect of integration, exploring new methods of studying introspective pre-reflective experiences using micro-phenomenology interviews.

## Publications

Group members published their works at top international HCI venues including CHI (5), TEI (4), Augmented Humans (2), the Routledge Handbook of Bodily Awareness (1) and Foundations and Trends in Human-Computer Interaction (1).

## Projects and Cooperations

We are engaged in ongoing collaborations within MPI-INF, where we are exploring synergies with Gurprit Singh and Vahid Babaei, and within the Saarland Informatics Campus, where we work closely with Jürgen Steimle. The are also many international collaborations, including with Andrew McPherson (Imperial College London), Katta Spiel (TU Wien), Kasper Hornbaek (University of Copenhagen), Jarrod Knibbe (Melbourne University), Oliver Schneider (University of Waterloo), and others.

Through the Maryland Max Planck Ph.D. Program in Computer Science, we are also establishing strong ties with the University of Maryland, exploring shared interests through co-supervision of PhD students with Jun Nishida.

# 25 Paul Swoboda: Combinatorial Computer Vision

## Group Development

The research group Combinatorial Computer Vision was established in March 2018. Its current members are Paul Swoboda and the doctoral student Ahmed Abbas. The former member Andrea Hornakova was successful in her research and completed her doctorate in 2022 and is now working as a data scientist. Ahmed Abbas is on track to finish his PhD at the end of this year. The doctoral students have published at NeurIPS, CVPR, ICCV, ECCV and AAAI.

## Vision and Research Strategy

Combinatorial optimization is a fundamental tool in machine learning that offers the promise to improve performance of deep learning systems further by incorporating explicit prior knowledge in the optimization and allows for theoratically principled solutions to basic machine learning tasks such as clustering, correspondence and others. The research group Combinatorial Computer Vision studies mathematical abstractions of computer vision tasks posed as combinatorial optimization problems. One focus is on developing efficient algorithms to solve the ensuing problems and to benchmark them w.r.t. metrics defined by the application. Another field of research is on integrating combinatorial optimization problems and deep networks. Application areas include tracking, clustering, segmentation and correspondence problems. Moreover, the group conducts basic research into algorithm design, how to make algorithms massively parallel and differentiable and how to augment them by machine learning.

## Research Areas and Achievements

### Optimization Algorithm Development

We research efficient general purpose algorithms for structured prediction problems that are formulated as combinatorial optimization problems. To this end we have proposed sequential CPU-based as well as massively parallel GPU-capable solvers. This has lead to the first generally applicable GPU-based integer linear program solver that can solve a wide range of structured prediction tasks without custom solver development. On some problems we even outperform specialized solvers for narrow problem classes. This work has been published at ICML 2021 and CVPR 2022.

### Clustering

The multicut problem, also known as correlation clustering, is a fundamental optimization problem for clustering with a large number of applications in machine learning and computer vision. We have studied image segmentation by end-to-end differentiable clustering that improves upon non-end-to-end learned baselines. In order to solve very large scale multicut problems fast we have proposed a massively parallel GPU-based algorithm for multicut that is an order of magnitude faster than previous CPU solvers. Another work concerns multicut with complete interactions between any two data points, whose complexity would grow quadratically when approached naively. We have investigated factorization formulations and efficient heuristics for this case and improved upon ordinary, sparse multicut w.r.t. solution quality and time. These works have been published at NeurIPS 2021 and CVPR 2022.

### Tracking

The widely studied problem of tracking multiple objects across a series of timeframes is well served by a combination of combinatorial optimization for solving a global association problem for connecting individual detections into tracks and deep learning for computing association costs. For data association in multiple object tracking we have previously proposed a new expressive optimization problem called the lifted disjoint paths problem. We have developed a scalable solver and deep learning pipeline based on algorithmic techniques also researched in our group that can routinely solve long and crowded sequences fast. We have developed this work further for multi-camera tracking. This research has been published at ICCV 2021 and CVPR 2022.

### Correspondence: Shape & Graph Matching

Correspondence problems arise prominently in computer vision for shape and keypoint matching. We have proposed a fast and scalable algorithm for a combinatorial shape matching formulation. For keypoint matching a popular combinatorial formulation is the quadratic assignment problem. It is notorious for being one of the empirically hardest optimization problems. In order to bring clarity into which one of the many developed algorithms to use and which algorithmic techniques are really working we have conducted a large-scale benchmark study involving the most promising algorithms tested on a wide range of problems covering many problem sizes and difficulties. Our research has been published at CVPR 2022 and ECCV 2022.

## Projects and Cooperations

We collaborate inside the MPI with the Theobalt group from the graphical department for correspondence problems. Outside the MPI for Informatics we have worked with groups from University of Hannover, Bonn University, University of Heidelberg and with the Max-Planck Institute for Intelligent Systems (MPI-IS).

# 26 Rhaleb Zayer: High Performance Digital Geometry Processing

## Group Development

The HPDGP group was formed in April 2019 and is led by Rhaleb Zayer. During the reporting period, postdoc Mengyu Chu accepted an Assistant Professor position at Beijing University, China. Rhaleb Zayer accepted an Associate Professor position at the University of East Anglia, UK. The team regularly hosts interns and master students locally, e.g., Divesh Kumar (U. Saarland), Duarte David (U. Saarland), Lisa Heidmann (U. Saarland), or remotely, e.g. Ömer Köse (Middle East Technical University, Turkey).

## Vision and Research Strategy

As the computing landscape is being dramatically reshaped by ubiquitous parallelism, big data, and modern machine learning methods, extracting both meaning and performance from existing data representations gives rise to formidable challenges which get further amplified by differentiability requirements, the unstructured nature of data (e.g., meshes and graphs) and by the serial nature of prevalent algorithmic solutions. As a result, the high performance promise of modern hardware seems elusive.

Going against the common trend of viewing performance as a mere post-processing step left to "Ninja coders" or automatic code optimization software, we seek to make high performance accessible at a high level of abstraction without requiring practitioners to delve into the low level intricacies of parallelism. To facilitate deployment, our solutions need to be independent of the underlying silicon or granularity, i.e. the same algorithms can be operated on clusters as well as on single graphics cards (GPUs) independently of the vendor. To this end, we target a seamless transition between data structures and numerics, more precisely, we explores the potential of channeling performance through sparse matrix algebra. This methodology is demonstrated on selected problems in computer graphics, computational geometry, and scientific computing, all of which are believed to be hard to parallelize. While performance is key towards achieving interactivity, intuitive control of the modeling workflow is equally important. Typical scenarios have been explored in the contexts of classical shape modeling and additive manufacturing. In view of the rapidly growing field of machine learning, particular attention is paid to data-driven modeling.

## Research Areas and Achievements

### Topological feature discovery

A key ingredient in our research has been the use of novel representations and abstractions to cast new light on the problems at hand. For instance, the coupling of sparse matrix algebra with our diffuse interface methods allowed us, most recently, to provide an elegant solution to some notably challenging problems such as Reeb graph extraction and topological handle and tunnel detection. This was achieved by going back to first principles, that is, using the Poincaré shrinking loop itself instead of subsequent elaborate algebraic topology constructs (Eurographics 2023).

### Machine learning based modeling and interaction

**Learning meaningful controls for fluids.** Modern fluid simulation methods achieve high-quality results, however controlling and interpreting motion from visual cues familiar and meaningful to humans, such as the advected marker density is still a big challenge. We proposed a novel data-driven conditional adversarial model that solves the challenging, and theoretically ill-posed problem of deriving plausible velocity fields from a single frame of a density field. Besides density modifications, our generative model is the first to enable the control of the results using all of the following control modalities: obstacles, physical parameters, kinetic energy, and vorticity (Siggraph 2021).

**Reconstruction of dynamic fluids.** High-fidelity reconstruction of dynamic fluids from sparse multiview RGB videos remains a formidable task, due to the complexity of the underlying physics as well as the severe occlusion and complex lighting in the captured data. We have explored reconstructing smoke from sparse multiview RGB videos of real-world scenes with unknown lighting conditions and arbitrary obstacles. By augmenting time-varying neural radiance fields with physics-informed deep learning, our method benefits from the supervision of images and physical priors. It exhibits high-quality results with relaxed constraints and strong flexibility on synthetic and real flow captures (Siggraph 2022).

## Projects and Cooperations

Within the department, we collaborate with Hans-Peter Seidel and the group Vahid Babaei (MPC-VCC) as well as with Christian Theobalt (Visual Computing and AI Department). Externally, we collaborate with Nils Thuerey (TU Munich), Markus Steinberger (TU Graz), and Quan Zheng (Chinese Academy of Sciences).

# Part III

# Research Units in Detail

# 27 D1: Algorithms and Complexity

## 27.1 Personnel

**Head of Group**

Prof. Danupon Nanongkai, Ph.D. (since August 2022)

**Researchers**

Prof. Dr. Kurt Mehlhorn (since December 1990, Director Emeritus since September 2019)
Dr. Andreas Karrenbauer (since January 2013)
Dr. Roohani Sharma (since October 2020)
Dr. Philip Wellnitz (since October 2021)
Dr. Tomasz Kociumaka (since April 2022)
Dr. Adam Polak (since December 2022)

**Ph.D. Students (as of March 1st, 2023)**

Nick Fischer (since April 2019)
Hannaneh Akrami (since October 2019)
Dr. Corinna Coupette (since January 2020)
Golnoosh Shahkarami (since April 2021)
Alejandro Cassis (since June 2021)
Yonggang Jiang (since October 2021)
Yuan Gao (since October 2021)
Paolo Luigi Rinaldi (since October 2021)
Martin Herold (since October 2022)
Zihang Wu (since November 2022)

**Long-term Guests**

Prof. Dr. Karl Bringmann (since January 2016; Guest since November 2019)
Prof. (emeritus) Dr. Hans Ulrich Simon (since October 2020)
Dr. Karol Węgrzycki (since October 2020)
Dr. Evangelos Kipouridis (since September 2022)

**Former (Recent) Staff**

- Dr. Amir Zandieh (September 2020 – April 2021), Applied Scientist, Zalando.
- Dr. Christoph Lenzen (July 2014 – June 2021), Faculty Member, CISPA.
- Dr. Sándor Kisfaludi-Bak (September 2019 – August 2021), Assistant Professor, Aalto University, Finland.
- Dr. Bhaskar Ray Chaudhury (February 2017 – August 2021), Future Faculty Fellow, University of Illinois at Urbana-Champaign, US.

- Dr. Ben Wiederhake (March 2017 – August 2021), Postdoc, CISPA.
- Dr. Themistoklis (Themis) Gouleakis (April 2019 – September 2021), Senior Research Fellow, National University of Singapore.
- Leonie (Leo) Krull (August 2020 – September 2021), Ph.D. Student, Goethe-Universität Frankfurt.
- Dr. André Nusser (June 2017 – November 2021), Postdoc, University of Copenhagen.
- Dr. habil. Joachim Spoerhase (March 2022 – February 2023), Lecturer, University of Sheffield, UK.

**Secretaries**

Christina Fries (until July 2022)
Stefanie Katja Harres (since August 2022)

## 27.2 Visitors

From March 2021 to February 2023, the following researchers visited our group:

| | | |
|---|---|---|
| Kamkari, Hamidreza | 27.01.2022–06.03.2022 | Sharif University |
| van den Brand, Jan | 01.06.2022–31.07.2022 | Georgia Institute of Technology |
| Gadekar, Ameet | 15.06.2022–15.08.2022 | Aalto University |
| Ghafari, Mohadese | 01.07.2022–30.09.2022 | Sharif University |
| Shoghi, Mansooreh | 02.07.2022–31.07.2022 | Wroclaw University |
| Rahimi, Siavash | 05.07.2022–22.09.2022 | Sharif University |
| Danaei, Alireza | 07.07.2022–22.09.2022 | Sharif University |
| Mahdavifar, Ali | 10.07.2022–22.09.2022 | Sharif University |
| Zarsav, Minoo | 25.07.2022–15.08.2022 | Aalto University |
| Khodamoradi, Kamyar | 25.07.2022–15.08.2022 | Aalto University |
| Haghi, Alireza | 26.07.2022–23.09.2022 | Sharif University |
| Abbasi, Fateme | 01.08.2022–05.08.2022 | Wroclaw University |
| Banerjee, Sandip | 01.08.2022–05.08.2022 | Wroclaw University |
| Byrka, Jaroslaw | 01.08.2022–05.08.2022 | Wroclaw University |
| Blikstad, Joakim | 12.09.2022–19.12.2022 | KTH Royal Institute of Technology |
| Chalermsook, Parinya | 16.10.2022–25.10.2022 | Aalto University |
| Cheriyan, Joseph | 19.10.2022–20.10.2022 | University of Waterloo |
| Tu, Ta-Wei | 10.12.2022–31.08.2023 | ETH Zürich |
| Livanos, Vasilis | 21.11.2022–28.11.2022 | University of Illinois |
| Lassota, Alexandra | 21.11.2022–30.11.2022 | EPFL |
| Nusser, André | 25.11.2022–01.12.2022 | BARC Denmark |
| Gadekar, Ameet | 01.12.2022–31.01.2023 | Aalto University |
| Seddighin, Masoud | 03.12.2022–01.02.2023 | Sharif University |
| Kumar, Nikhil | 04.12.2022–08.12.2022 | Hasso-Plattner Institut |
| Christiansen, Alexander | 12.12.2022–14.12.2022 | DTU Denmark |
| Rothenberg, Eva | 12.12.2022–14.12.2022 | DTU Denmark |
| Van der Hoog, Ivor | 12.12.2022–14.12.2022 | DTU Denmark |

| | | |
|---|---|---|
| Yuval, Efron | 12.01.2023–29.01.2023 | Columbia University |
| Hé, Zhongtian | 13.01.2023–21.01.2023 | Princeton University |
| Chaudhury, Baskar Ray | 28.02.2023–06.03.2023 | University of Illinois |

## 27.3 Group Organization

The whole group meets twice a week for a scientific talk and discussions. In addition, the subgroups have their meetings. The whole group meets every other week to discuss organizational matters. The selection of new postdocs is made by all group members holding a Ph.D.

### Long-term Guests

**Karl Bringmann.** Karl Bringmann is professor at Saarland University and associated with the institute. Karl leads a group consisting of Karol Węgrzycki, Evangelos Kipouridis, Alejandro Cassis, and Nick Fischer, who are also associated with the institute. Karl's group is working broadly on algorithm design and complexity theory.

On the algorithm side, his group develops combinatorial insights that yield efficient algorithms for problems from various domains such as graph algorithms, string algorithms, discrete optimization, and computational geometry. His group's results range from sublinear-time approximation algorithms for string edit distance, via nearest-neighbor data structures on curves, to pseudopolynomial algorithms for Subset Sum.

On the complexity side, Karl Bringmann's group has made significant contributions to fine-grained complexity theory—the young subfield of complexity theory that proves quantitative running time lower bounds based on plausible conjectures on certain core problems such as Satisfiability. In this area, Karl Bringmann was the first to prove a quadratic-time lower bound for a dynamic programming problem. Recently he has been developing novel techniques to prove fine-grained hardness of approximation for polynomial-time problems. The general goal of his group is to develop best-possible algorithms, that is, to design efficient algorithms and prove matching fine-grained lower bounds showing that the developed algorithms are likely to be optimal.

Karl Bringmann has received an ERC Starting Grant in 2019 and a Busy Beaver teaching award by the students council at Saarland University in 2023.

## 27.4 Algorithmic Game Theory

We worked on two subjects: sampling from the Gibbs distribution (see Section 27.4.3 and Fair Allocation of Goods and Chores. The bulk of our work is on the latter subject and we concentrate on this work in the introduction to this section.

A set of items has to be allocated to a set of agents. The items are either divisible or indivisible and either goods or bads. In the case of indivisible items, each item is fully allocated to some agents, and in the case of divisble goods, an item may be split into parts and then the parts are allocated to the agents. The goal is to find a *fair* and *efficient* allocation. We discuss notions of fairness and efficiency below.

Goods have positive utility for the agents and agents are interested in high utility. Chores have negative utility (= positive disutility) and the agents are interested in small disutility. The allocation of chores seems to be much harder than the allocation of goods. Why? Goods can always be allocated; at worst, the good has no value for the agent. However, chores cannot always be allocated. A chore with infinite disutility for an agent cannot be allocated to the agent.

For divisible goods (and additive valuations), the problem is well-understood. It reduces to a Fisher market with equal incomes and can be solved in polynomial time. For divisible chores, the corresponding market problem is much harder. We developed an FPTAS, see Section 27.4.1.

We turn to indivisible goods. Each agent is given a bundle of the goods and each agent has a valuation for bundles. In the most general case, the valuation is only required to be monotone (getting more does not hurt). In the simplest case, valuations are additive, that is, agents value individual goods and the value of a bundle is the sum of the values of its goods. We use $v_i$ to denote the valuation function of the $i$-th agent and $A_i$ for the bundle assigned to them. Then, $v_i(A_i)$ is the utility for the $i$-th agent.

There are three main notions of fairness:

*Envy-Freeness:* An allocation is *envy-free*, if no agent envies another agent, that is, $v_i(A_j) \leq v_i(A_i)$ for all $i$ and $j$. Envy-freeness is too much to ask for; consider two agents and a single good that is liked by both. The good has to be assigned to one of the agents and the other will envy. The weakest relaxation is *envy-freeness up to any good* (EFX), that is, we must have $v_i(A_j \setminus g) \leq v_i(A_i)$ for all $i$ and $j$ and all goods $g \in A_j$. We showed in the previous reporting period that EFX-allocations exist for three agents and additive valuations and for any number of agents and general valuations, if a linear number of goods can be set aside (= given to charity). In this period, we simplified the proof of the first result and extended it to a larger class of valuations, and we reduced the number of goods that have to be given to charity.

We also studied a relaxation EF$k$X, where $k$ goods can be removed to remove envy.

We give more information on these results in Section 27.4.1.

A criticism against envy-based notions of fairness it that envy-freeness does not come with any guarantee for efficiency. For example, if we have two agents and two pairs of shoes and give the first agent two left shoes and the second agent two right shoes, the both bundles have value zero for both agents (assuming that only a proper pair has value) and any allocation of the two bundles will be envy-free.

*Share-Based:* The MaxMin-Share (MMS) of an agent with valuation function $v$ is defined as

$$\text{MMS}(v) = \max_{(A_1,\ldots,A_n)} \min_i v(A_i),$$

where the maximization is over all partitions of the items into $n$ bundles. The MMS is a natural generalization of the cut-and-choose protocol for two agents. An agent divides the goods into $n$ bundles and then the other $n-1$ agents choose (in some order). The agent is guaranteed the minimum value of any bundle in the partition. Note that the MMS of an agent depends only on the valuation of the agent.

Instances with additive valuations are known for which no allocation $(A_1, \ldots, A_n)$ exists, where $v_i(A_i) \geq \text{MMS}(v_i)$ for all $i$. So approximation algorithms are called for. An improved approximation algorithm is the subject of current work. Allocations that approximate the MMS-value are efficient in the sense that each agent is guaranteed a certain share.

*Nash Social Welfare:* Nash Social Welfare (NSW) is the oldest notion of fairness (Nash, 1950). It is defined as the geometric mean of the individual utilities, that is,

$$\text{NSW}(A_1, \ldots, A_n) = \left( \prod_i v_i(A_i) \right)^{1/n} .$$

An allocation maximizing NSW is Pareto-optimal, that is, there is no other allocation in which no agent is worse-off and some agent is strictly better-off.

Unfortunately, finding an allocation maximizing NSW is NP-complete even for additive valuations. We delineated the boundary of tractability for two-valued additive valuation, that is, $v_i(g)$ $1, s$ for all agents $i$ and goods $g$, where $s$ is a fixed rational number greater than 1. We showed that an allocation maximizing NSW can be computed in polynomial time if $s$ is a multiple of $1/2$ and maximizing NSW is NP-complete otherwise. We give more information in Section 27.4.2.

### 27.4.1 Envy-Based Notions of Fairness

#### Finding EFX-Allocations

*Investigators: Hannaneh Akrami, Bhaskar Ray Chaudhury (now UIUC), Kurt Mehlhorn, Pranabendu Misra (now Chennai Mathematical Institute), in collaboration with Noga Alon (Princeton), Jugal Garg, and Ruta Mehta (UIUC),*

Envy-freeness up to any good (EFX) is arguably the most compelling fairness notion in the context of dividing indivisible goods. However, the existence of EFX allocations has not been settled and is one of the most important problems in fair division. Toward resolving this problem, relaxations and special cases were investigated, for instance, the existence of EFX for three agents [3], the existence of 0.618-EFX allocations [2], and the existence of EFX with at most $n-1$ unallocated goods (charity) [5].

Reducing the number of unallocated goods for arbitrary number of agents is a systematic way to settle the big question. To this end, we show that for every $\varepsilon$ $(0, 1/2)$, there exists a $(1-\varepsilon)$-EFX allocation with a sublinear number of unallocated goods. Our proof is algorithmic (and therefore constructive) and it establishes an intriguing connection to a problem in *zero sum extremal combinatorics* [4, 1]. In the latter paper, we also give a simplified proof for the existence of EFX for three agents. The proof requires only one of the valuations to be additive; the other two valuations may be general.

#### References


[1] H. Akrami, N. Alon, B. Ray Chaudhury, J. Garg, K. Mehlhorn, and R. Mehta. *EFX Allocations: Simplifications and Improvements*, 2022. arXiv: 2205.07638.

[2] S. Barman, S. K. K. Murthy, and R. Vaish. Finding fair and efficient allocations. *CoRR*, abs/1707.04731, 2017. E-Text: 1707.04731. to appear in EC 2018.

[3] B. R. Chaudhury, J. Garg, and K. Mehlhorn. EFX exists for three agents. In *EC '20*, 2020, pp. 1–19. ACM.

• [4] B. R. Chaudhury, J. Garg, K. Mehlhorn, R. Mehta, and P. Misra. Improving EFX guarantees through rainbow cycle number. In P. Biró, S. Chawla, F. Echenique, and E. Sodomka, eds., *EC '21, 22nd ACM Conference on Economics and Computation*, Budapest, Hungary (Virtual), 2021, pp. 310–311. ACM.

[5] B. R. Chaudhury, T. Kavitha, K. Mehlhorn, and A. S. itsa. A little charity guarantees almost envy-freeness. *SIAM J. Comput.*, 50(4):1336–1358, 2021.

### An EF2X Allocation Protocol for Restricted Additive Valuations

*Investigators: Hannaneh Akrami, in cooperation with Masoud Seddighin (Tehran Institute for Advanced Studies) and Rojin Rezvan (University of Texas at Austin),*

We study the problem of fairly allocating a set $M$ of $m$ indivisible goods to a set of $n$ agents. Each agent $i$ is equipped with a valuation function $v_i : 2^M \rightarrow \mathbb{R}_{\geq 0}$. The envy-freeness up to any good (EFX) criterion—which requires that no agent prefers the bundle of another agent after removal of any single good—is known to be a remarkable analogous of envy-freeness when the resource is a set of indivisible goods. In [1], we investigate the EFX notion for restricted additive valuations, that is, every good $g$ has some non-negative value $v(g)$, and each agent $i$ either is not interested in $g$ at all or derives value $v(g)$ from it. Formally, $v_i(\{g\}) \in \{0, v(g)\}$.

We introduce a natural relaxation of EFX called EF$k$X which requires that no agent envies another agent after removal of any $k$ goods. Our main contribution is an algorithm that finds a complete (that is, no good is discarded) EF2X allocation for the setting with restricted additive valuations. We also give an algorithm which finds an EFX allocation for restricted additive valuations that discards at most $\lfloor n/2 \rfloor - 1$ goods.

The high level idea of our algorithms is to start from an empty allocation and update the allocation using some updating rules as long as one is applicable. Designing these updating rules carefully, we prove that the final allocation is EFX and the number of unallocated goods is at most $\lfloor n/2 \rfloor - 1$. For the EF2X result, we use different updating rules and prove that by the time no updating rule is applicable, the allocation is EFX. Then we run an extra step of allocating the remaining goods while making sure that the final allocation is EF2X.

## References

• [1] H. Akrami, R. Rezvan, and M. Seddighin. An EF2X allocation protocol for restricted additive valuations. In L. de Raedt, ed., *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence (IJCAI 2022)*, Vienna, Austria, 2022, pp. 17–23. IJCAI.

### Fair Division of Divisble Bads

*Investigators: Bhaskar Ray Chaudhury (now UIUC), in collaboration with Ruta Mehta, and Shant Boodaghians (UIUC).*

*Fair Division of Divisible Bads:* Competitive equilibrium with equal income (CEEI) is considered one of the best mechanisms to allocate a set of items among agents fairly and efficiently. We study the computation of CEEI when items are chores that are disliked (negatively valued) by agents, under 1-homogeneous and concave utility functions which include linear utility functions. It is well-known that, even with linear utilities, the set of CEEI may be non-convex and disconnected, and the problem is PPAD-hard in the more general Arrow-Debreu economic model. In contrast to these negative results, we design the first FPTAS: A polynomial-time algorithm to compute $\varepsilon$-approximate CEEI where the running-time depends polynomially on $\varepsilon$ [1].

Our algorithm relies on the recent EG-like characterization of CEEI which represents the set of CEEI as exactly the KKT points of a non-convex minimization problem that have all coordinates non-zero. Due to this non-zero constraint, naive gradient-based methods fail to find the desired local minima as they are attracted toward zero. We develop an exterior-point method that alternates between guessing non-zero KKT points and maximizing the objective along supporting hyperplanes at these points. We show that this procedure must converge quickly to an approximate KKT point which then can be mapped to an approximate CEEI; this exterior point method may be of independent interest. When utility functions are linear, we give explicit procedures for finding the exact iterates, and as a result show that a stronger form of approximate CEEI can be found in polynomial time. Finally, we note that our algorithm extends to the setting of unequal incomes (CE), and to mixed manna with linear utilities where each agent may like (positively value) some items and dislike (negatively value) others.

### References


- [1] S. Boodaghians, B. Ray Chaudhury, and R. Mehta. Polynomial time algorithms to find an approximate competitive equilibrium for chores. In S. Naor and N. Buchbinder, eds., *Proceedings of the Annual ACM-SIAM Symposium on Discrete Algorithms (SODA 2022)*, Virtual, 2022, pp. 2285–2302. SIAM.


## 27.4.2 Maximizing Nash Social Welfare: Delineating Tractability

*Investigators: Hannaneh Akrami, Bhaskar Ray Chaudhury (now UIUC), Kurt Mehlhorn, and Golnoosh Shahkarami, in cooperation with Quentin Vermande, Ernest van Wijland (École Normale Supérieure, Paris), Martin Hoefer, Marco Schmalhofer, and Giovanna Varricchio (Goethe University Frankfurt, Institute for Computer Science),*

We study the problem of allocating a set of indivisible goods among a set of agents with *2-value additive valuations.* In this setting, each good is valued either 1 or $p/q$, for some fixed co-prime numbers $p, q$ $\mathbb{N}$ such that $1 \leq q < p$, and the value of a bundle is the sum of the values of the contained goods. Our goal is to find an allocation which maximizes the *Nash social welfare* (NSW), that is, the geometric mean of the valuations of the agents. In this work, we give a complete characterization of polynomial-time tractability of NSW maximization that solely depends on the values of $q$.

In [1], we design an algorithm to compute an optimal allocation in polynomial time if $q$ divides $p$, that is, when $q = 1$ and $p$ $\mathbb{N}$ after appropriate scaling. We show that such an improvement cannot be further extended to the case $q = 3$; indeed, we prove that it is NP-hard to compute an allocation with maximum NSW whenever $q \geq 3$. In terms of approximation, we present positive and negative results for general $p$ and $q$. We show that our algorithm obtains an approximation ratio of at most 1.0345. Moreover, we prove that the problem is APX-hard, with a lower bound of 1.000015 achieved at $p/q = 5/4$.

The case $q = 2$ and odd $p$ was left open in the reported paper. In [2], we show that this case is tractable. We define a collection of improvement rules based on alternating paths and alternating walks and show that any allocation can be transformed into an optimal allocation using these rules. For the polynomial time algorithm, we exploit a connection to perfect matchings with parity constraints.

## References


- [1] H. Akrami, B. Ray Chaudhury, M. Hoefer, K. Mehlhorn, M. Schmalhofer, G. Shahkarami, G. Varricchio, Q. Vermande, and E. van Wijland. Maximizing Nash social welfare in 2-value instances. In *Proceedings of the 36th AAAI Conference on Artificial Intelligence*, Virtual Conference, 2022, pp. 4760–4767. AAAI.
- [2] H. Akrami, B. Ray Chaudhury, M. Hoefer, K. Mehlhorn, M. Schmalhofer, G. Shahkarami, G. Varricchio, Q. Vermande, and E. van Wijland. *Maximizing Nash Social Welfare in 2-Value Instances: The Half-Integer Case*, 2022. arXiv: 2207.10949.


### 27.4.3 Sampling from the Gibbs Distribution in Congestion Games

*Investigator: Pieter Kleer (now Tilburg University)*

Logit dynamics is a form of randomized game dynamics where players have a bias towards strategic deviations that give a higher improvement in cost. Logit dynamics is used extensively in practice. In congestion (or potential) games, the dynamics converges to the so-called Gibbs distribution over the set of all strategy profiles, when interpreted as a Markov chain. In general, logit dynamics might converge slowly to the Gibbs distribution, but beyond that, not much is known about their algorithmic aspects, nor that of the Gibbs distribution. In [2] we are interested in the following two questions for congestion games:

i) Is there an efficient algorithm for sampling from the Gibbs distribution?

ii) If yes, do there also exist natural randomized dynamics that converges quickly to the Gibbs distribution?

We first study these questions in extension parallel congestion games, a well-studied special case of symmetric network congestion games. As our main result, we show that there is a simple variation on the logit dynamics (in which we in addition are also allowed to randomly interchange the strategies of two players) that converges quickly to the Gibbs distribution in such games. This answers both questions above affirmatively. We also address the first question for the class of so-called capacitated $k$-uniform congestion games.

To prove our results, we rely on the recent breakthrough work of Anari, Liu, Oveis-Gharan and Vinzant [1] concerning the approximate sampling of the base of a matroid according to strongly log-concave probability distribution.

### References


[1] N. Anari, K. Liu, S. O. Gharan, and C. Vinzant. Log-concave polynomials II: High-dimensional walks and an FPRAS for counting bases of a matroid. In *STOC 2019*, 2019, pp. 1–12. ACM.

• [2] P. Kleer. Sampling from the Gibbs distribution in congestion games. In P. Biró, S. Chawla, F. Echenique, and E. Sodomka, eds., *EC '21, 22nd ACM Conference on Economics and Computation*, Budapest, Hungary (Virtual), 2021, pp. 679–680. ACM.


## 27.5 Algorithms and Complexity on Graphs

Graph algorithms provide a powerful framework for solving a wide range of problems in diverse domains such as social networks, transportation networks, biology, and many more. Complexity theory, on the other hand, provides a theoretical foundation for analyzing the efficiency and effectiveness of algorithms for solving computational problems. The study of graph algorithms and complexity aims to design efficient algorithms for solving graph-related problems and understand the limits of algorithmic solvability for such problems.

Our results cover a wide range of computational problems related to graphs and networks. During the last reporting period, one of the primary areas of focus has been developing efficient algorithms for solving foundational problems in graph theory such as shortest path and minimum cut problems. We also explored several other aspects of efficient graph algorithms, such as computations in a dynamic or distributed context, algorithms on particular graph classes, and describing similarity of graphs in various applications. We briefly introduce the main results below.

**Foundational Problems: Paths, Matchings, and Matroids.** One of the most fundamental problems in graph theory is finding the shortest path between vertices. This problem has widespread applications in both theoretical and practical contexts, ranging from routing in computer networks to finding the most efficient way to travel between two locations on a map. Additionally, the problem of finding large matchings in graphs is a core topic in graph theory, with a variety of real-world applications such as scheduling, resource allocation, and social network analysis. Matroid intersection is a powerful generalization of matchings in bipartite graphs that has important theoretical implications. Our recent work achieved significant progress in these areas.

For more details on these results, consult Section 27.5.1.

**Fast Algorithms for Graph Cuts and Connectivity.** Given a weighted graph $G$, a cut is a set of edges whose removal disconnects the graph. The *minimum cut* or *min-cut* is the cut with minimum total edge weight.

When the graph is unweighted, this problem is sometimes called Edge Connectivity.

The Vertex Connectivity (or just Connectivity) problem is when we want to remove (unweighted) vertices instead of edges.

These problems are classic graph optimization problems with countless applications. In the last reporting period, we developed many fast algorithms for these problems in many models of computation. For more details on these results, consult Section 27.5.2.

**Distributed Computing.** Distributed computing is a field of computer science that deals with the design and analysis of algorithms for systems consisting of multiple processors or computers that work together to solve a problem. The main theoretical goals of research in distributed computing are to develop efficient algorithms for solving complex computational problems and to understand the fundamental principles that govern the behavior of distributed systems. To achieve these goals, researchers in distributed computing focus on designing algorithms that are scalable, fault-tolerant, and able to operate in a highly distributed and dynamic environment. Distributed computing presents various challenges and new objectives that are not present in centralized sequential computing. For example, the need to minimize the number of rounds of communication or resolving collisions are central goals in distributed computing.

In particular, in the last reporting period, we studied variants of Contention Resolution and of Minimum Spanning Tree. For more details on these results, consult Section 27.5.3.

**Algorithms on Treelike Structures.** Restricted graph classes are a subject of great interest in graph theory, as they provide a way to study the properties and structures of graphs with specific characteristics. One such class is "treelike" graphs, which have a treelike hierarchical structure. There are many ways this concept can be formalized. For example, one can consider a graph treelike if it can be made acyclic with the deletion of a few vertices (that is, has small feedback vertex set number). The notion of treewidth defines a more general class of treelike graphs, whose combinatorial and algorithmic properties were studied extensively in the literature. Understanding bounded-treewidth graphs is usually considered as the first step of understanding graphs with excluded minors, opening up a rich area in structural graph theory. Chordal graphs offer a very different way of generalizing some of the advantageous algorithmic properties of trees, leading to a graph class that includes also dense graphs.

In the last reporting period, among other results, we obtained faster algorithms for Perfect Code and other (generalized) domination problems on tree-like graphs and graphs with bounded leafage. For more details on these results, consult Section 27.5.4.

**Analyzing Graphs and Networks.** Graph analysis is an area of research in computer science that deals with the study of the properties of graphs arising in different applications. This area encompasses a broad range of topics including community detection, centrality measures, graph clustering, similarity measures, graph embeddings any many more. Many real-world datasets exhibit higher-order interactions that cannot be modeled using simple graphs. In such cases, hypergraphs, which are a generalization of graphs, can be used to model more complex relationships.

In the last reporting period, we introduced new tools and techniques for the analysis of graphs and hypergraphs; in particular, with legal networks as a prime application in mind. For more details on these results, consult Section 27.5.5.

### 27.5.1 Foundational Problems: Paths, Matchings, and Matroids

#### Negative-Weight Single-Source Shortest Paths in Near-linear Time

*Investigators: Karl Bringmann, Alejandro Cassis, Nick Fischer, and Danupon Nanongkai, in cooperation with Aaron Bernstein (Rutgers University) and Christian Wulff-Nilsen (University of Copenhagen)*

In [2], we present a randomized algorithm that solves Single-Source Shortest Path (sssp) in $\mathcal{O}(m\log^8(n)\log W)$ time when edge weights are integral and can be negative, where $n$ and $m$ denote the number of vertices and edges, respectively, and $W \geq 2$ is such that every edge weight is at least $-W$. $\widetilde{\mathcal{O}}$ hides polylogarithmic factors. This essentially resolves the classic Negative-Weight sssp problem.

The previous bounds are $\widetilde{\mathcal{O}}((m+n^{1.5})\log W)$ [3] and $m^{4/3+o(1)}\log W$ [1]. Near-linear time algorithms were known previously only for the special case of planar directed graphs [5].

In contrast to all recent developments that rely on sophisticated continuous optimization methods and dynamic algorithms, our algorithm is simple: it requires only a simple graph decomposition (which we call "Low-diameter Decomposition") and elementary combinatorial tools. In fact, ours is the first combinatorial algorithm for negative-weight sssp to break through the classic $\widetilde{\mathcal{O}}(m\sqrt{n}\log W)$ bound from over three decades ago [6].

In a follow-up work [4], we improve the number of logarithmic factors in the running time of [2], resulting in a running time of $\mathcal{O}(m\log^2(n)\log(nW)\log\log n)$. Our improvements also yield an algorithm to compute the minimum cycle mean in the same running time, as well as an alternative construction algorithm for obtaining a Low-diameter Decomposition.

#### References

[1] K. Axiotis, A. Madry, and A. Vladu. Circulation control for faster minimum cost flow in unit-capacity graphs. In *FOCS*, 2020, pp. 93–104. IEEE.

• [2] A. Bernstein, D. Nanongkai, and C. Wulff-Nilsen. Negative-weight single-source shortest paths in near-linear time. In *FOCS 2022, IEEE 63rd Annual Symposium on Foundations of Computer Science*, Denver, CO, USA, 2022, pp. 600–611. IEEE.

[3] J. v. d. Brand, Y. T. Lee, D. Nanongkai, R. Peng, T. Saranurak, A. Sidford, Z. Song, and D. Wang. Bipartite matching in nearly-linear time on moderately dense graphs. In *FOCS*, 2020, pp. 919–930. IEEE.

• [4] K. Bringmann, A. Cassis, and N. Fischer. *Negative-Weight Single-Source Shortest Paths in Near-Linear Time: Now Faster!*, 2023. arXiv: 2304.05279.

[5] J. Fakcharoenphol and S. Rao. Planar graphs, negative weight edges, shortest paths, and near linear time. *J. Comput. Syst. Sci.*, 72(5):868–889, 2006. Announced at FOCS'01.

[6] H. N. Gabow and Y. Xu. Efficient algorithms for independent assignments on graphic and linear matroids. In *FOCS*, 1989, pp. 106–111. IEEE Computer Society.

### Fine-Grained Lower Bounds for Distance Oracles

*Investigators: Karl Bringmann and Nick Fischer in cooperation with Or Zamir (Institute for Advanced Study, Princeton), Seri Khoury (UC Berkeley), and Amir Abboud (Weizmann Institute of Science)*

A *distance oracle* is a data structure that efficiently preprocesses an undirected graph and can then quickly return the distance between any given pair of nodes—up to a small error. Specifically, given two vertices $u, v$, it computes an approximation $\widetilde{d}(u, v)$ of the distance $d(u, v)$ satisfying $d(u, v) \leq \widetilde{d}(u, v) \leq \alpha \cdot d(u, v)$, where $\alpha$ is the *stretch* of the distance oracle. In their seminal paper from 2001, Thorup and Zwick [3] introduced the Distance Oracle problem suggesting that it is perhaps the most natural formulation of the classical All-Pairs Shortest Paths problem.

In sparse graphs (that is, graphs with $m \approx n$ edges), the Thorup-Zwick distance oracle achieves stretch $2k - 1$, preprocessing time $\mathcal{O}(kn^{1+1/k})$ and query time $\mathcal{O}(k)$. Despite various later improvements, this remains the state of the art for constant $k$. Even the first question one might ask remains poorly understood: *What is the best stretch $f(k)$ we can achieve if we insist on preprocessing time $\mathcal{O}(n^{1+1/k})$ and almost-constant query time $n^{o(1)}$?*

In two papers, we have made significant progress toward this question. In the first work [2], we prove that $f(k) \geq k/6.3772 - \mathcal{O}(1)$ conditioned on the 3Sum or apsp Hypotheses from fine-grained complexity theory. We thereby (conditionally) prove that $f(k)$ grows linearly with $k$. To obtain this result, we introduce the *short cycle removal* technique. Roughly speaking, using this technique we can prove conditional hardness of certain graph problems even if the input graph contains few short cycles. This requires insights on a structure-versus-randomness dichotomy on graphs: Either the given is graph is random-like and therefore has few short cycles, or the graph is structured in which case we can find structured pieces and solve the problem faster on these pieces.

In the second work [1], we strengthen the lower bound and prove that $f(k) \geq k$, conditioned on the 3Sum Hypothesis. hWhile the high-level idea remains similar, our tool set changes substantially: We apply the same structure-versus-randomness ideas on the 3Sum instances directly. This new proof involves heavy machinery from additive combinatorics.

Determining the optimal $f(k)$ $[k, 2k - 1]$ remains an intricate open problem.

### References


• [1] A. Abboud, K. Bringmann, and N. Fischer. Stronger 3-SUM lower bounds for approximate distance oracles via additive combinatorics. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing (STOC 2023)*, Orlando, FL, USA, 2023. ACM. Accepted.

• [2] A. Abboud, K. Bringmann, S. Khoury, and O. Zamir. Hardness of approximation in P via short cycle removal: Cycle detection, distance oracles, and beyond. In S. Leonardi and A. Gupta, eds., *STOC '22, 54th Annual ACM Symposium on Theory of Computing*, Rome, Italy, 2022, pp. 1487–1500. ACM.

[3] M. Thorup and U. Zwick. Approximate distance oracles. In J. S. Vitter, P. G. Spirakis, and M. Yannakakis, eds., *Proceedings on 33rd Annual ACM Symposium on Theory of Computing, July 6-8, 2001, Heraklion, Crete, Greece*, 2001, pp. 183–192. ACM.

### Dynamic Matching

*Investigators: Peter Kiss in cooperation with David Wajc (Google Reseaerch), Joakim Blikstad (KTH), Thathaphol Saranurak (University of Michigan, Ann Arbor), and Sayan Bhattacharya (University of Warwick),*

In the Fully-Dynamic Matching problem, we are given a graph undergoing edge insertions and deletions. Our task is to maintain a matching of the graph with small approximation ratio while minimizing the time spent handling the updates. A long line of research has lead to two natural barriers for the problem which have stood for more than a decade: no algorithm has been found which can maintain a better than 2-approximate matching in polylogarithmic update time or a $(1+\varepsilon)$ approximate matching in polynomially sub-$n$ update time. The prior problem of beating 2-approximation has been pointed out as one of the most important problems within the dynamic literature by multiple authors, for instance in [10, 2, 11, 3]. Observe that we can obtain a $(1+\varepsilon)$-approximate algorithm for the problem with $\mathcal{O}(n)$ update time through straightforward periodic rebuilds. Through relaxing the problem to the maintenance of a matching size estimate instead of an approximate matching we have bypassed both of these longstanding barriers.

In [7] (independently and concurrently with [1]), we obtain the first fully dynamic algorithm for maintaining a better than 2-approximation of the maximum matching size ($\sim 1.7$ and $\sim 1.97$ for bipartite and general graphs respectively). Furthermore, in [4] we show the first algorithm with polynomially sub-$n$ update time which maintains a $(1+\varepsilon)$-approximate matching size estimate.

As the ultimate goal of dynamic matching research of a $(1+\varepsilon)$-approximate Fully Dynamic Matching algorithm with polylogarithmic update time seems far out of reach, a number of papers focused on more restricted settings. In the incremental setting where the graph may only undergo edge insertions prior algorithms maintaining a $(1+\varepsilon)$-approximate matching either has update times depending on $n$ [9, 5] or exponentially on $\varepsilon^{-1}$ [8]. In [6], we present the first algorithm for the problem with update time polynomial in $\varepsilon^{-1}$. Furthermore, we show that the same algorithm can handle vertex deletions in addition to edge insertions contrast to previous algorithms for the problem.

### References

[1] S. Behnezhad. Dynamic algorithms for maximum matching size. *CoRR*, abs/2207.07607, 2022. E-Text: 2207.07607.

[2] S. Behnezhad, J. Lacki, and V. S. Mirrokni. Fully dynamic matching: Beating 2-approximation in $\delta^{\epsilon}$ update time. In S. Chawla, ed., *Proceedings of the 2020 ACM-SIAM Symposium on Discrete Algorithms, SODA 2020, Salt Lake City, UT, USA, January 5-8, 2020*, 2020, pp. 2492–2508. SIAM.

[3] A. Bernstein and C. Stein. Faster fully dynamic matchings with small approximation ratios. In R. Krauthgamer, ed., *Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2016, Arlington, VA, USA, January 10-12, 2016*, 2016, pp. 692–711. SIAM.

• [4] S. Bhattacharya, P. Kiss, and T. Saranurak. *Dynamic* $(1+\epsilon)$*-Approximate Matching Size in Truly Sublinear Update Time*, 2023. arXiv: 2302.05030.

• [5] S. Bhattacharya, P. Kiss, and T. Saranurak. Dynamic algorithms for packing-covering LPs via multiplicative weight. In N. Bansal and V. Nagarajan, eds., *Proceedings of the 2023 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA 2023)*, Florence, Italy, 2023, pp. 1–47. SIAM.

• [6] S. Bhattacharya, P. Kiss, and T. Saranurak. Sublinear algorithms for $(1.5 + \epsilon)$-approximate matching. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing (STOC 2023)*, Orlando, FL, USA, 2023. ACM. Accepted.

• [7] S. Bhattacharya, P. Kiss, T. Saranurak, and D. Wajc. Dynamic matching with better-than-2 approximation in polylogarithmic update time. In N. Bansal and V. Nagarajan, eds., *Proceedings of the 2023 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA 2023)*, Florence, Italy, 2023, pp. 100–128. SIAM.

[8] F. Grandoni, S. Leonardi, P. Sankowski, C. Schwiegelshohn, and S. Solomon. $(1 + \epsilon)$-approximate incremental matching in constant deterministic amortized time. In T. M. Chan, ed., *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2019, San Diego, California, USA, January 6-9, 2019*, 2019, pp. 1886–1898. SIAM.

[9] M. Gupta. Maintaining approximate maximum matching in an incremental bipartite graph in polylogarithmic update time. In V. Raman and S. P. Suresh, eds., *34th International Conference on Foundation of Software Technology and Theoretical Computer Science, FSTTCS 2014, December 15-17, 2014, New Delhi, India*, 2014, LIPIcs 29, pp. 227–239. Schloss Dagstuhl - Leibniz-Zentrum für Informatik.

[10] K. Onak and R. Rubinfeld. Maintaining a large matching and a small vertex cover. In L. J. Schulman, ed., *Proceedings of the 42nd ACM Symposium on Theory of Computing, STOC 2010, Cambridge, Massachusetts, USA, 5-8 June 2010*, 2010, pp. 457–464. ACM.

[11] D. Wajc. Rounding dynamic matchings against an adaptive adversary. In K. Makarychev, Y. Makarychev, M. Tulsiani, G. Kamath, and J. Chuzhoy, eds., *Proccedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing, STOC 2020, Chicago, IL, USA, June 22-26, 2020*, 2020, pp. 194–207. ACM.

### Fast Algorithms via Dynamic-Oracle Matroids

*Investigators: Danupon Nanongkai and Ta-Wei Tu in cooperation with Joakim Blikstad (KTH) and Sagnik Mukhopadhyay (University of Sheffield)*

Matroid problems are powerful generalizations that can model many fundamental problems such as Disjoint Spanning Trees and Bipartite Matching. However, a fast algorithm for a matroid problem does not necessarily imply a fast algorithm for the problems it generalizes.

In [1], we bridge this gap by initiating the study of matroid problems in a new oracle model called *dynamic oracle*. Our algorithms in this model lead to new bounds for some classic problems, and a "unified" algorithm whose performance matches previous results developed in various papers for various problems. We also show a lower bound that answers some open problems from a few decades ago. Concretely, our results are as follows.

**Improved algorithms for matroid union and disjoint spanning trees.** We show an algorithm with $\widetilde{\mathcal{O}}_k(n + r \quad \overline{r})$ dynamic-rank-query and time complexities for the matroid union problem over $k$ matroids, where $n$ is the input size, $r$ is the output size, and $\widetilde{\mathcal{O}}_k$ hides poly$(k, \log(n))$ terms. This implies all of the following.

- An improvement over the $\widetilde{\mathcal{O}}_k(n\sqrt{r})$ bound implied by [2] for matroid union in the traditional rank-query model.
- An $\widetilde{\mathcal{O}}_k(E + V\sqrt{V})$-time algorithm for the $k$-Disjoint Spanning Tree problem. This is nearly linear for moderately dense input graphs and improves the $\widetilde{\mathcal{O}}_k(V\sqrt{E})$ bounds of Gabow-Westermann [3] and Gabow [4].

Consequently, this gives improved bounds for Shannon Switching Game and Graph Irreducibility, and more.

**Matroid intersection.** We show a matroid intersection algorithm with $\widetilde{\mathcal{O}}(n\sqrt{r})$ dynamic-rank-query and time complexities. This implies new bounds for some problems (for instance for Maximum Forest with Deadlines) and bounds that match the classic ones obtained in various papers for various problems, for instance for Colorful Spanning Tree [5], for Graphic Matroid Intersection [6], for Simple Scheduling Matroid Intersection [9], and for Hopcroft-Karp Combinatorial Bipartite Matching.

More importantly, this is done via a "unified" algorithm in the sense that an improvement over our dynamic-rank-query algorithm would imply improved bounds for *all* the above problems simultaneously.

**Lower bounds.** We show simple super-linear ($\Omega(n \log n)$) query lower bounds for Matroid Intersection and union problems in our dynamic-rank-oracle and the traditional independence-query models; the latter improves the previous $\log_2(3)n - o(n)$ bound by Harvey [7] and answers an open problem raised by, for instance, Welsh [8] and Chakrabarty et al. [2].

## References


[1] J. Blikstad, T.-W. Tu, D. Nanongkai, and S. Mukhopadhyay. Fast algorithms via dynamic-oracle matroids. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing (STOC 2023)*, Orlando, FL, USA, 2023. ACM. Accepted.

[2] D. Chakrabarty, Y. T. Lee, A. Sidford, S. Singla, and S. C. Wong. Faster matroid intersection. In *FOCS*, 2019, pp. 1146–1168. IEEE Computer Society.

[3] H. Gabow and H. Westermann. Forests, frames, and games: Algorithms for matroid sums and applications. In *STOC*, 1988, pp. 407–421.

[4] H. N. Gabow. A matroid approach to finding edge connectivity and packing arborescences. In *STOC*, 1991, pp. 112–122. ACM.

[5] H. N. Gabow and M. F. M. Stallmann. Efficient algorithms for graphic matroid intersection and parity (extended abstract). In *ICALP*, 1985, LNCS 194, pp. 210–220. Springer.

[6] H. N. Gabow and Y. Xu. Efficient algorithms for independent assignments on graphic and linear matroids. In *FOCS*, 1989, pp. 106–111. IEEE Computer Society.

[7] N. J. A. Harvey. Algebraic algorithms for matching and matroid problems. *SIAM J. Comput.*, 39(2):679–702, 2009.

[8] D. J. Welsh. *Matroid Theory.* Academic Press, London, New York, 1976.

[9] Y. Xu and H. N. Gabow. Fast algorithms for transversal matroid intersection problems. In *ISAAC*, 1994, LNCS 834, pp. 625–633. Springer.


### 27.5.2 Fast Algorithms for Graph Cuts and Connectivity

#### Finding a Small Vertex Cut on Distributed Networks

*Investigators: Yonggang Jiang in cooperation with Sagnik Mukhopadhyay (University of Sheffield)*

In [2], we present an algorithm for distributed networks to efficiently find a small vertex cut in the CONGEST model. Given a positive integer $\kappa$, our algorithm can, with high probability, either find $\kappa$ vertices whose removal disconnects the network or return that such $\kappa$ vertices do not exist. Our algorithm takes $\kappa^3 \cdot \widetilde{\mathcal{O}}(D + \sqrt{n})$ rounds, where $n$ is the number of vertices in the network and $D$ denotes the network's diameter. This implies $\widetilde{\mathcal{O}}(D + \sqrt{n})$ round complexity whenever $\kappa = polylog(n)$.

Prior to our result, a bound of $\widetilde{\mathcal{O}}(D)$ was known only when $\kappa = 1, 2$ [5]. For $\kappa \geq 3$, this bound can be obtained only by an $\mathcal{O}(\log n)$-approximation algorithm [1], and the only known exact algorithm takes $\mathcal{O}\left((\kappa \Delta D)^{\mathcal{O}(\kappa)}\right)$ rounds, where $\Delta$ is the maximum degree [4]. Our result answers an open problem by Nanongkai, Saranurak, and Yingchareonthawornchai [3].

## References


[1] K. Censor-Hillel, M. Ghaffari, and F. Kuhn. Distributed connectivity decomposition. In M. M. Halldórsson and S. Dolev, eds., *ACM Symposium on Principles of Distributed Computing, PODC '14, Paris, France, July 15-18, 2014*, 2014, pp. 156–165. ACM.

• [2] Y. Jiang and S. Mukhopadhyay. Finding a small vertex cut on distributed networks. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing (STOC 2023)*, Orlando, FL, USA, 2023. ACM. Accepted.

[3] D. Nanongkai, T. Saranurak, and S. Yingchareonthawornchai. Breaking quadratic time for small vertex connectivity and an approximation scheme. In M. Charikar and E. Cohen, eds., *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing, STOC 2019, Phoenix, AZ, USA, June 23-26, 2019*, 2019, pp. 241–252. ACM.

[4] M. Parter. Small cuts and connectivity certificates: A fault tolerant approach. In J. Suomela, ed., *33rd International Symposium on Distributed Computing, DISC 2019, October 14-18, 2019, Budapest, Hungary*, 2019, LIPIcs 146, pp. 30:1–30:16. Schloss Dagstuhl - Leibniz-Zentrum für Informatik.

[5] M. Parter and A. Petruschka. Near-optimal distributed computation of small vertex cuts. In C. Scheideler, ed., *36th International Symposium on Distributed Computing, DISC 2022, October 25-27, 2022, Augusta, Georgia, USA*, 2022, LIPIcs 246, pp. 31:1–31:21. Schloss Dagstuhl - Leibniz-Zentrum für Informatik.

### Near-Linear Time Approximations for Cut Problems via Fair Cuts

*Investigators: Danupon Nanongkai in cooperation with Debmalya Panigrahi (Duke University), Jason Li (UC Berkeley), and Thatchaphol Saranurak (University of Michigan, Ann Arbor)*

In [3], we introduce the notion of *fair cuts* as an approach to leverage approximate $(s,t)$-MINCUT (equivalently $(s,t)$-MAXFLOW) algorithms in undirected graphs to obtain near-linear time approximation algorithms for several cut problems. Informally, for any $\alpha \geq 1$, an $\alpha$-fair $(s,t)$-cut is an $(s,t)$-cut such that there is an $(s,t)$-flow that uses a $1/\alpha$ fraction of the capacity of *every* edge in the cut. (So, any $\alpha$-fair cut is also an $\alpha$-approximate mincut, but not vice-versa.) We give an algorithm for $(1+\varepsilon)$-FAIR $(s,t)$-CUT in $\widetilde{\mathcal{O}}(m)$-time, thereby matching the best runtime for $(1+\varepsilon)$-APPROXIMATE $(s,t)$-MINCUT [5]. We then demonstrate the power of this approach by showing that this result almost immediately leads to several applications:

- the first nearly-linear time $(1+\varepsilon)$-approximation algorithm that computes all-pairs maxflow values (by constructing an approximate Gomory-Hu tree). Prior to our work, such a result was not known even for the special case of STEINER MINCUT [2, 1];
- the first almost-linear-work subpolynomial-depth parallel algorithms for computing $(1+\varepsilon)$-approximations for all-pairs maxflow values (again via an approximate Gomory-Hu tree) in unweighted graphs;
- the first near-linear time expander decomposition algorithm that works even when the expansion parameter is polynomially small; this subsumes previous incomparable algorithms [4, 7, 6].

### References

[1] R. Cole and R. Hariharan. A fast algorithm for computing steiner edge connectivity. In L. L. Larmore and M. X. Goemans, eds., *Proceedings of the 35th Annual ACM Symposium on Theory of Computing, June 9-11, 2003, San Diego, CA, USA*, 2003, pp. 167–176. ACM.

[2] Y. Dinitz and A. Vainshtein. The connectivity carcass of a vertex subset in a graph and its incremental maintenance. In F. T. Leighton and M. T. Goodrich, eds., *Proceedings of the Twenty-Sixth Annual ACM Symposium on Theory of Computing, 23-25 May 1994, Montréal, Québec, Canada*, 1994, pp. 716–725. ACM.

• [3] J. Li, D. Nanongkai, D. Panigrahi, and T. Saranurak. Near-linear time approximations for cut problems via fair cuts. In N. Bansal and V. Nagarajan, eds., *Proceedings of the 2023 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA 2023)*, Florence, Italy, 2023, pp. 240–257. SIAM.

[4] D. Nanongkai and T. Saranurak. Dynamic spanning forest with worst-case update time: Adaptive, las vegas, and o($n^{1/2 - \epsilon}$)-time. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2017, Montreal, QC, Canada, June 19-23, 2017*, 2017, pp. 1122–1129. ACM.

[5] R. Peng. Approximate undirected maximum flows in o (m polylog (n)) time. In *Proceedings of the twenty-seventh annual ACM-SIAM symposium on Discrete algorithms*, 2016, pp. 1862–1867. SIAM.

[6] T. Saranurak and D. Wang. Expander decomposition and pruning: Faster, stronger, and simpler. In T. M. Chan, ed., *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2019, San Diego, California, USA, January 6-9, 2019*, 2019, pp. 2616–2635. SIAM.

[7] C. Wulff-Nilsen. Fully-dynamic minimum spanning forest with improved worst-case update time. In *STOC*, 2017, pp. 1130–1143. ACM.

### Fully Dynamic Exact Edge Connectivity in Sublinear Time

*Investigators: Danupon Nanongkai in cooperation with Gramoz Goranci (ETH Zurich), Mikkel Thorup, Christian Wulff-Nilsen (University of Copenhagen), Thatchaphol Saranurak (University of Michigan, Ann Arbor), and Monika Henzinger (University of Vienna),*

In [1], we present a dynamic algorithm for computing Edge Connectivity.

Given a simple $n$-vertex, $m$-edge graph $G$ undergoing edge insertions and deletions, we give two new fully-dynamic algorithms for exactly maintaining the edge connectivity of $G$ in $\widetilde{\mathcal{O}}(n)$ worst-case update time and $\widetilde{\mathcal{O}}(m^{1-1/16})$ amortized update time, respectively. Prior to our work, all dynamic Edge Connectivity algorithms assumed bounded edge connectivity, guaranteed approximate solutions, or were restricted to edge insertions only. Our results answer in the affirmative an open question posed by Thorup [2].

## References

• [1] G. Goranci, M. Henzinger, D. Nanongkai, T. Saranurak, M. Thorup, and C. Wulff. Fully dynamic exact edge connectivity in sublinear time. In N. Bansal and V. Nagarajan, eds., *Proceedings of the 2023 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA 2023)*, Florence, Italy, 2023, pp. 70–86. SIAM.

[2] M. Thorup. Fully-dynamic min-cut. *Comb.*, 27(1):91–127, 2007.

### Cut Query Algorithms with Star Contraction

*Investigators: Danupon Nanongkai in cooperation with Simon Apers (Université de Paris, CNRS, IRIF), Yuval Efron (University of Columbia), Sagnik Mukhopadhyay (University of Sheffield), and Paweł Gawrychowski (University of Wrocław)*

In [1], we study the complexity of determining the edge connectivity of a simple graph with cut queries. We show that **(i)** there is a bounded-error randomized algorithm that computes Edge Connectivity with $\mathcal{O}(n)$ cut queries, and **(ii)** there is a bounded-error quantum algorithm that computes Edge Connectivity with $\widetilde{\mathcal{O}}(\sqrt{n})$ cut queries. To prove these results, we introduce a new technique, called *star contraction*, to randomly contract edges of a graph while preserving non-trivial minimum cuts. In star contraction, vertices randomly contract an edge incident on a small set of randomly chosen "center" vertices. In contrast to the related 2-out contraction technique of Ghaffari, Nowicki, and Thorup [3], star contraction only contracts vertex-disjoint star subgraphs, which allows it to be efficiently implemented via cut queries.

The $\mathcal{O}(n)$ bound from item (i) was not known even for the simpler problem of Connectivity, and it improves the $\mathcal{O}(n \log^3 n)$ upper bound by Rubinstein, Schramm, and Weinberg [5].

The bound is tight under the reasonable conjecture that the randomized communication complexity of Connectivity is $\Omega(n \log n)$, an open question since the seminal work of Babai, Frankl, and Simon [2]. The bound also excludes using Edge Connectivity on simple graphs to prove a superlinear randomized query lower bound for minimizing a symmetric submodular function. The quantum algorithm from item (ii) gives a nearly-quadratic separation with the randomized complexity, and addresses an open question of Lee, Santha, and Zhang [4]. The algorithm can alternatively be viewed as computing the Edge Connectivity of a simple graph with $\widetilde{\mathcal{O}}(\sqrt{n})$ matrix-vector multiplication queries to its adjacency matrix.

Finally, we demonstrate the use of star contraction outside of the cut query setting by designing a one-pass semi-streaming algorithm for computing Edge Connectivity in the complete vertex arrival setting. This contrasts with the edge arrival setting where two passes are required.

## References


[1] S. Apers, Y. Efron, P. Gawrychowski, T. Lee, S. Mukhopadhyay, and D. Nanongkai. Cut query algorithms with star contraction. In *FOCS 2022, IEEE 63rd Annual Symposium on Foundations of Computer Science*, Denver, CO, USA, 2022, pp. 507–518. IEEE.

[2] L. Babai, P. Frankl, and J. Simon. Complexity classes in communication complexity theory. In *Proceedings of the 27th IEEE Annual Symposium on Foundations of Computer Science (FOCS '86)*, 1986, pp. 337–347. IEEE.

[3] M. Ghaffari, K. Nowicki, and M. Thorup. Faster algorithms for edge connectivity via random 2-out contractions. In *Proceedings of the 31st ACM-SIAM Symposium on Discrete Algorithms (SODA '20)*, 2020, pp. 1260–1279. SIAM.

[4] T. Lee, M. Santha, and S. Zhang. Quantum algorithms for graph problems with cut queries. In *Proceedings of the 32nd ACM-SIAM Symposium on Discrete Algorithms (SODA '21)*, 2021, pp. 939–958. SIAM.

[5] A. Rubinstein, T. Schramm, and S. M. Weinberg. Computing exact minimum cuts without knowing the graph. In *Proceedings of the 9th Innovations in Theoretical Computer Science Conference (ITCS '18)*, 2018, LIPIcs 94, pp. 39:1–39:16. Schloss Dagstuhl - Leibniz-Zentrum für Informatik.


### Parameterized Complexity of Multicut Problems

*Investigators: Roohani Sharma, in cooperation with Dániel Marx, Philipp Schepper (CISPA), Prafullkumar Tale (Indian Institute of Science Education and Research Pune), Paloma T. Lima (IT University of Copenhagen), Eun Jung Kim (Paris-Dauphine University), Meike Hatzel (National Institute of Informatics, Tokyo), Magnus Wahlström (Royal Holloway, University of London), Esther Galby (TU Hamburg), Manuel Sorge (TU Wien), Lars Jaffke (University of Bergen), Tomáš Masařík, and Marcin Pilipczuk (University of Warsaw),*

The Edge Multicut problem is a classical cut problem where given an undirected graph $G$, a set of pairs of vertices $\mathcal{P}$, and a budget $k$, the goal is to determine if there is a set $S$ of at most $k$ edges such that for each $(s, t) \in \mathcal{P}$, $G - S$ has no path from $s$ to $t$. The parameterized complexity of this problem in undirected graphs, after being a long-standing open problem for a while, has been resolved around 2010 independently by two groups of researchers Marx and Razgon [10], and Bousquet, Daligault and Thomassé [2].

In the weighted version of the problem, called WEIGHTED EDGE MULTICUT one is additionally given a weight function $\mathrm{WT} : E(G) \quad \mathbb{N}$ and a weight bound $w$, and the goal is to determine if there is a solution of size at most $k$ and weight at most $w$. Both the FPT algorithms for EDGE MULTICUT by Marx et al. and Bousquet et al. fail to generalize to the weighted setting. In fact, the weighted problem is non-trivial even on trees and determining whether WEIGHTED EDGE MULTICUT on trees is FPT was posed as an open problem by Bousquet et al. [3]. In [5], we answer this question positively by designing an algorithm which uses a very recent result by Kim et al. [7] about directed flow augmentation as a subroutine.

We also study a variant of this problem where there is no bound on the size of the solution, but the parameter is a structural property of the input, for example, the number of leaves of the tree. We strengthen our results by stating them for the more general vertex deletion version.

In a follow-up work [8], along with other problems, we resolve the parameterized complexity question of WEIGHTED MULTICUT on *general* graphs in positive.

Finally, in [6], we study the EDGE MULTICUT problem in directed graphs. In this setting, the problem is called DIRECTED MULTICUT. In directed graphs, the problem in full generality was quickly observed to be W[1]-hard [10]. However, some restrictions turned out to be tractable: the case of directed acyclic graphs [9], DIRECTED MULTIWAY CUT, where we are given just a set of terminals and we ask to cut all paths between every pair of distinct terminals [4], or DIRECTED MULTICUT with two terminal pairs [4]. Observe that the one-terminal-pair case is just the classic MINIMUM CUT problem, which is polynomail-time solvable. In 2015, Pilipczuk and Wahlström [11] provided a hardness reduction for the four-terminal-pairs case, leaving the three-terminal-pairs case open until now.

In [6], we show fixed-parameter tractability of the DIRECTED MULTICUT problem with three terminal pairs (with a randomized algorithm). On the technical side, we use two recent developments in parameterized algorithms. Using the technique of *directed flow-augmentation* [7] we cast the problem as a CSP problem with few variables and constraints over a large ordered domain. We observe that this problem can be in turn encoded as an first-order model-checking task over a structure consisting of a few 0-1 matrices. We look at this problem through the lenses of twin-width, a recently introduced structural parameter [1]: By a recent characterization [1] the said first-order model-checking task can be done in FPT time if the said matrices have bounded grid rank. To complete the proof, we show an irrelevant vertex rule: If any of the matrices in the said encoding has a large grid minor, a vertex corresponding to the "middle" box in the grid minor can be proclaimed irrelevant—not contained in the sought solution—and thus reduced.

## References

[1] É. Bonnet, E. J. Kim, S. Thomassé, and R. Watrigant. Twin-width I: tractable FO model checking. *Journal of the ACM*, 69(1):3:1–3:46, 2022.

[2] N. Bousquet, J. Daligault, and S. Thomassé. Multicut is FPT. *SIAM Journal on Computing*, 47(1):166–207, 2018.

[3] N. Bousquet, J. Daligault, S. Thomassé, and A. Yeo. A polynomial kernel for multicut in trees. In S. Albers and J. Marion, eds., *26th International Symposium on Theoretical Aspects of Computer Science, STACS 2009, February 26-28, 2009, Freiburg, Germany, Proceedings*, 2009, LIPIcs 3, pp. 183–194. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, Germany.

[4] R. H. Chitnis, M. Hajiaghayi, and D. Marx. Fixed-parameter tractability of directed multiway cut parameterized by the size of the cutset. *SIAM J. Comput.*, 42(4):1674–1696, 2013.

• [5] E. Galby, D. Marx, P. Schepper, R. Sharma, and P. Tale. Parameterized complexity of weighted multicut in trees. In M. A. Bekos and M. Kaufmann, eds., *Graph-Theoretic Concepts in Computer Science (WG 2022)*, Tübingen, Germany, 2022, LNCS 13453, pp. 257–270. Springer.

• [6] M. Hatzel, L. Jaffke, P. T. Lima, T. Masařík, M. Pilipczuk, R. Sharma, and M. Sorge. Fixed-parameter tractability of DIRECTED MULTICUT with three terminal pairs parameterized by the size of the cutset: Twin-width meets flow-augmentation. In N. Bansal and V. Nagarajan, eds., *Proceedings of the 2023 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA 2023)*, Florence, Italy, 2023, pp. 3229–3244. SIAM.

[7] E. J. Kim, S. Kratsch, M. Pilipczuk, and M. Wahlström. Directed flow-augmentation. In S. Leonardi and A. Gupta, eds., *STOC '22: 54th Annual ACM SIGACT Symposium on Theory of Computing, Rome, Italy, June 20 - 24, 2022*, 2022, pp. 938–947. ACM.

• [8] E. J. Kim, T. Masařík, M. Pilipczuk, R. Sharma, and M. Wahlström. *On Weighted Graph Separation Problems and Flow-Augmentation*, 2022. arXiv: 2208.14841.

[9] S. Kratsch, M. Pilipczuk, M. Pilipczuk, and M. Wahlström. Fixed-parameter tractability of multicut in directed acyclic graphs. *SIAM Journal on Discrete Mathematics*, 29(1):122–144, 2015.

[10] D. Marx and I. Razgon. Fixed-parameter tractability of multicut parameterized by the size of the cutset. *SIAM Journal on Computing*, 43(2):355–388, 2014.

[11] M. Pilipczuk and M. Wahlström. Directed multicut is W[1]-hard, even for four terminal pairs. *ACM Transactions on Computation Theory*, 10(3):13:1–13:18, 2018.

### 27.5.3 Distributed Computing

#### Approximate Minimum Directed Spanning Trees under Congestion

*Investigators: Christoph Lenzen, Hossein Vahidi*

Until recently, the MINIMUM DIRECTED SPANNING TREE (MDST) problem has not been studied in distributed computing models. This fundamental task generalizes the well-studied MINIMUM SPANNING TREE problem, by asking for a minimum weight spanning tree rooted at some specified node of a directed network. In [1], Fischer and Oshman reduce the MDST problem to the SINGLE-SOURCE SHORTEST PATH (SSSP) problem, with a polylogarithmic increase in running time. This holds both in the CONGEST and CONGESTED CLIQUE models. Fischer and Oshman further suggest the possibility that an approximate SSSP algorithm could be leveraged in computing an approximate MDST.

We extend their analysis to show that this is indeed the case [2]: For $\varepsilon > 0$, using a $(1+\varepsilon)$-approximation to SSSP running in $R$ rounds we can compute a $(1+\varepsilon)$-approximate MDST in $\widetilde{\mathcal{O}}(R)$ rounds (where $\widetilde{\mathcal{O}}$ hides $\log^{\mathcal{O}(1)} n$ factors). In particular, this implies the following improvements in the state of the art for $(1+\varepsilon)$-approximation of MDST.

- An $\widetilde{\mathcal{O}}(n^{1-2/\omega+o(1)}) \subset \widetilde{\mathcal{O}}(n^{0.158})$-round Congested Clique algorithm, where $\omega < 2.373$ is the fast matrix multiplication exponent.
- An $\widetilde{\mathcal{O}}(\lambda^2)$-round Congested Clique algorithm in graphs where each edge has an at most factor $\lambda \geq 1$ heavier reverse edge.
- An $\widetilde{\mathcal{O}}(\lambda^2( \ \overline{n} + D))$-round Congest algorithm in the same family of graphs. For $\lambda \quad \log^{\mathcal{O}(1)} n$, the resulting running time of $\widetilde{\mathcal{O}}( \ \overline{n} + D)$ is unconditionally tight up to a polylogarithmic factor.

## References


[1] O. Fischer and R. Oshman. A distributed algorithm for directed minimum-weight spanning tree. In J. Suomela, ed., *33rd International Symposium on Distributed Computing, DISC 2019, October 14-18, 2019, Budapest, Hungary*, 2019, LIPIcs 146, pp. 16:1–16:16. Schloss Dagstuhl - Leibniz-Zentrum für Informatik.

• [2] C. Lenzen and H. Vahidi. Approximate minimum directed spanning trees under congestion. In T. Jurdziński and S. Schmid, eds., *Structural Information and Communication Complexity (SIROCCO 2021)*, Wrocław, Poland (Online), 2021, LNCS 12810, pp. 352–369. Springer.


### Robust and Optimal Contention Resolution without Collision Detection

*Investigators: Yonggang Jiang, in cooperation with Chaodong Zheng (Nanjing University)*

Contention Resolution on a multiple-access communication channel is a classical problem in distributed and parallel computing. In this problem, a set of nodes arrives over time, where each node comes with a message it intends to send. Time proceeds in synchronous slots, and in each slot, each node can broadcast its message or remain idle. If in a slot one node broadcasts alone, it succeeds; otherwise, if multiple nodes broadcast simultaneously, messages collide and none succeeds. Nodes can differentiate collision and silence (that is, no node broadcasts) only if a collision detection mechanism is available. Ideally, a contention resolution algorithm should satisfy at least three criteria: (a) low time complexity (that is, high throughput), meaning it does not take too long for all nodes to succeed; (b) low energy complexity, meaning each node does not make too many broadcast attempts before it succeeds; and (c) strong robustness, meaning the algorithm can maintain good performance even if interference is present. Such interference is often modeled by jamming—a jammed slot always generates collision.

Previous work [1] has shown, with collision detection, there are "perfect" algorithms for Contention Resolution that satisfy all three criteria. However, without collision detection, it was not until 2020 that an algorithm was discovered which can achieve optimal time complexity and low energy cost, assuming there is no jamming [2]. More recently, the trade-off between throughput and robustness was studied [3]. However, an intriguing and important question remains unknown: without collision detection, are there "perfect" Contention Resolution algorithms? In other words, when collision detection is absent and jamming is present, can we achieve both low total time complexity and low per-node energy cost?

In [4], we answer the above question affirmatively. Specifically, we develop a new randomized algorithm for Robust Contention Resolution, assuming collision detection is not available. With $n$ arrivals and $d$ jammed slots, our algorithm has $\mathcal{O}(n \log n + d)$ time complexity and $\mathcal{O}(\log^2 n + \log^2 d)$ energy complexity with high probability. Lower bound results demonstrate it achieves both optimal time complexity and optimal energy complexity. If all nodes start execution simultaneously—which is often referred to as the "static case" in literature—another algorithm is developed that achieves $\mathcal{O}(n + d)$ time complexity and $\mathcal{O}(\log^2 n + \log^2 d)$ energy complexity with high probability. The separation on time complexity suggests, for Robust Contention Resolution without collision detection, "batch" instances (that is, nodes start simultaneously) are inherently easier than "scattered" ones (that is, nodes arrive over time).

## References


[1] M. A. Bender, J. T. Fineman, S. Gilbert, and M. Young. Scaling exponential backoff: Constant throughput, polylogarithmic channel-access attempts, and robustness. *J. ACM*, 66(1):6:1–6:33, 2019.

[2] M. A. Bender, T. Kopelowitz, W. Kuszmaul, and S. Pettie. Contention resolution without collision detection. In K. Makarychev, Y. Makarychev, M. Tulsiani, G. Kamath, and J. Chuzhoy, eds., *Proccedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing, STOC 2020, Chicago, IL, USA, June 22-26, 2020*, 2020, pp. 105–118. ACM.

[3] J. Chen, W. Czerwinski, Y. Disser, A. E. Feldmann, D. Hermelin, W. Nadara, M. Pilipczuk, M. Pilipczuk, M. Sorge, B. Wróblewski, and A. Zych-Pawlewicz. Efficient fully dynamic elimination forests with applications to detecting long paths and cycles. In D. Marx, ed., *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms, SODA 2021*, 2021, pp. 796–809. SIAM.

• [4] Y. Jiang and C. Zheng. Robust and optimal contention resolution without collision detection. In K. Agrawal and I.-T. A. Lee, eds., *SPAA '22, 34th ACM Symposium on Parallelism in Algorithms and Architectures*, Philadelphia, PA, USA, 2022, pp. 107–118. ACM.


### 27.5.4 Algorithms on Treelike Structures

#### Counting Generalized Dominating Sets in Bounded-Treewidth Graphs

*Investigators: Philip Wellnitz in cooperation with Jacob Focke, Dániel Marx, Fionn Mc Inerney, Philipp Schepper (CISPA), Govind S. Sankar (Duke University), and Daniel Neuen (Simon Fraser University)*

Treewidth (and its underlying concept of tree decompositions) measures how tree-like a graph is and helps to generalize algorithms for trees to algorithms for general graphs with bounded treewidth. Usually the goal is to determine (under standard complexity assumptions) the best possible value $c > 0$ such that there is a $c^k \cdot n^{\mathcal{O}(1)}$-time algorithm (where $n$ is the input size) but no $(c - \varepsilon)^k \cdot n^{\mathcal{O}(1)}$-time algorithm for any $\varepsilon > 0$, always assuming a tree decomposition of width $k$ is given as input.

Based on such results for Perfect Matching and Generalized Matching [1, 3, 4], we study the related vertex-selection problem [2]. For sets $\sigma, \rho$ of non-negative integers, a $(\sigma, \rho)$-set of a graph $G$ is a set $S$ of vertices such that $N(u) \setminus S$ $\sigma$ for every $u$ $S$, and $N(v) \setminus S$ $\rho$ for every $v$ $S$. The problem of finding a $(\sigma, \rho)$-set (of a certain size) unifies standard problems such as Independent Set, Dominating Set, Independent Dominating Set, and many others even if we restrict ourselves to finite or cofinite $(\sigma, \rho)$.

Let $\sigma_{\text{top}}$ denote the largest element of $\sigma$ if $\sigma$ is finite, or the largest missing integer $+1$ if $\sigma$ is cofinite; $\rho_{\text{top}}$ is defined analogously for $\rho$. The currently fastest algorithm counts $(\sigma, \rho)$-sets in time $(\sigma_{\text{top}} + \rho_{\text{top}} + 2)^k \cdot n^{\mathcal{O}(1)}$ [5]. Surprisingly, we improve this time significantly for many $(\sigma, \rho)$. We say that $(\sigma, \rho)$ is $m$-structured if there is a pair $(\alpha, \beta)$ such that every integer in $\sigma$ equals $\alpha$ mod $m$, and every integer in $\rho$ equals $\beta$ mod $m$. Then, setting

- $c_{\sigma,\rho} = \sigma_{\text{top}} + \rho_{\text{top}} + 2$ if $(\sigma, \rho)$ is not $m$-structured for any $m \geq 2$,
- $c_{\sigma,\rho} = \max\ \sigma_{\text{top}}, \rho_{\text{top}}\ + 2$ if $(\sigma, \rho)$ is 2-structured, but not $m$-structured for any $m \geq 3$, and $\sigma_{\text{top}} = \rho_{\text{top}}$ is even, and
- $c_{\sigma,\rho} = \max\ \sigma_{\text{top}}, \rho_{\text{top}}\ + 1$, otherwise,

we provide algorithms counting $(\sigma, \rho)$-sets in time $c_{\sigma,\rho}^k \cdot n^{\mathcal{O}(1)}$. For example, for the Perfect Code problem (also known as Exact Independent Dominating Set) corresponding to $\sigma =\ 0$ and $\rho =\ 1$, this improves the $3^k \cdot n^{\mathcal{O}(1)}$ algorithm of van Rooij to $2^k \cdot n^{\mathcal{O}(1)}$.

Despite the unusually delicate definition of $c_{\sigma,\rho}$, we show that our algorithms are most likely optimal, that is, for any pair $(\sigma, \rho)$ of finite or cofinite sets where the problem is non-trivial, and any $\varepsilon > 0$, a $(c_{\sigma,\rho} - \varepsilon)^k \cdot n^{\mathcal{O}(1)}$-algorithm counting the number of $(\sigma, \rho)$-sets would violate the Counting Strong Exponential-Time Hypothesis(#SETH). For finite sets $\sigma$ and $\rho$, our lower bounds also extend to the decision version, showing that our algorithms are optimal in this setting as well. In contrast, for many cofinite sets, we show that further significant improvements for the decision and optimization versions are possible using the technique of representative sets.

## References


[1] R. Curticapean and D. Marx. Tight conditional lower bounds for counting perfect matchings on graphs of bounded treewidth, cliquewidth, and genus. In R. Krauthgamer, ed., *Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2016, Arlington, VA, USA, January 10-12, 2016*, 2016, pp. 1650–1669. SIAM.

• [2] J. Focke, D. Marx, F. M. Inerney, D. Neuen, G. S. Sankar, P. Schepper, and P. Wellnitz. Tight complexity bounds for counting generalized dominating sets in bounded-treewidth graphs. In N. Bansal and V. Nagarajan, eds., *Proceedings of the 2023 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA 2023)*, Florence, Italy, 2023, pp. 3664–3683. SIAM.

[3] D. Marx, G. S. Sankar, and P. Schepper. Degrees and gaps: Tight complexity results of general factor problems parameterized by treewidth and cutwidth. In N. Bansal, E. Merelli, and J. Worrell, eds., *48th International Colloquium on Automata, Languages, and Programming, ICALP 2021, July 12-16, 2021, Glasgow, Scotland (Virtual Conference)*, 2021, LIPIcs 198, pp. 95:1–95:20. Schloss Dagstuhl - Leibniz-Zentrum für Informatik.

[4] D. Marx, G. S. Sankar, and P. Schepper. Anti-factor is FPT parameterized by treewidth and list size (but counting is hard). In H. Dell and J. Nederlof, eds., *17th International Symposium on Parameterized and Exact Computation, IPEC 2022, September 7-9, 2022, Potsdam, Germany*, 2022, LIPIcs 249, pp. 22:1–22:23. Schloss Dagstuhl - Leibniz-Zentrum für Informatik.

[5] J. M. M. van Rooij. Fast Algorithms for Join Operations on Tree Decompositions. In F. V. Fomin, S. Kratsch, and E. J. van Leeuwen, eds., *Treewidth, Kernels, and Algorithms - Essays Dedicated to Hans L. Bodlaender on the Occasion of His 60th Birthday*, 2020, LNCS 12160, pp. 262–297. Springer.

### Metric Dimension Parameterized by Feedback Vertex Set and Other Structural Parameters

*Investigators: Roohani Sharma, in cooperation with Fionn Mc Inerney (CISPA), Prafullkumar Tale (Indian Institute of Science Education and Research Pune), Esther Galby (TU Hamburg), and Liana Khazaliya (TU Wien)*

For a graph $G$, a subset $S \subseteq V(G)$ is called a *resolving set* if for any two vertices $u, v$ $V(G)$, there is a vertex $w$ $S$ such that $d(w,u) = d(w,v)$. The Metric Dimension problem takes as input a graph $G$ and a positive integer $k$, and asks whether there is a resolving set of size at most $k$. This problem was introduced in the 1970s and is known to be NP-hard [3]. In the realm of parameterized complexity, Hartung and Nichterlein [4] proved that the problem is W[2]-hard when parameterized by the natural parameter $k$. They also observed that it is fixed-parameter tractable (FPT) when parameterized by the vertex cover number and asked about its complexity under *smaller* parameters, in particular the feedback vertex set number. We answer this question in [2] by proving that Metric Dimension is W[1]-hard when parameterized by the feedback vertex set number. This also improves the result of Bonnet and Purohit [1] which states that the problem is W[1]-hard parameterized by the treewidth.

Regarding the parameterization by the vertex cover number, in [2] we prove that Metric Dimension does not admit a polynomial kernel under this parameterization unless NP $\subseteq$ co-NP/poly. We observe that a similar result holds when the parameter is the distance to clique. On the positive side, we show that Metric Dimension is FPT when parameterized by either the distance to cluster or the distance to co-cluster, both of which are smaller parameters than the vertex cover number.

### References

[1] E. Bonnet and N. Purohit. Metric dimension parameterized by treewidth. *Algorithmica*, 83:2606–2633, 2021.

[2] E. Galby, L. Khazaliya, F. Mc Inerney, R. Sharma, and P. Tale. Metric dimension parameterized by feedback vertex set and other structural parameters. In S. Szeider, R. Ganian, and A. Silva, eds., *47th International Symposium on Mathematical Foundations of Computer Science (MFCS 2022)*, Vienna, Austria, 2022, Leibniz International Proceedings in Informatics 241, Article 51. Schloss Dagstuhl.

[3] M. R. Garey and D. S. Johnson. *Computers and Intractability - a Guide to Np-Completeness.* W.H. Freeman and Company, 1979.

[4] S. Hartung and A. Nichterlein. On the parameterized and approximation hardness of metric dimension. In *Proceedings of the 28th Conference on Computational Complexity, CCC*, 2013, pp. 266–276. IEEE Computer Society.

### Treedepth vs circumference

*Investigators: Roohani Sharma, in cooperation with Marcin Briański, Piotr Micek, Michał T. Seweryn (Jagiellonian University), Gwenaël Joret (Université libre de Bruxelles) and Konrad Majewski (University of Warsaw)*

The circumference of a graph $G$ is the length of a longest cycle in $G$, or $+$  if $G$ has no cycle. Birmelè [1] showed that the treewidth of a graph $G$ is at most its circumference minus 1. We strengthen this result in [2] for 2-connected graphs as follows: If $G$ is 2-connected, then its treedepth is at most its circumference. The bound is best possible and improves on an earlier quadratic upper bound due to Marshall and Wood [4].

An application of our results solves an open problem in [3]. More precisely, given an integer $d \geq 1$, Chen et al. [3] designed a data structure for maintaining an optimal elimination forest of a dynamic graph $G$ with worst case $2^{\mathcal{O}(d^2)}$ update time, under the promise that the treedepth of $G$ never exceeds $d$. Here, the graph $G$ is *dynamic* in the sense that edges can be added or removed, one at a time. An update time of $2^{\mathcal{O}(d^2)}$ is a natural barrier in this context, because the best known FPT algorithm for deciding whether an $n$-vertex graph $G$ has treedepth at most $d$ runs in time $2^{\mathcal{O}(d^2)} \cdot n$, see [5]. Any improvement on the $2^{\mathcal{O}(d^2)}$ update time would lead to a corresponding improvement of the latter result, by adding all edges of $G$ one at a time. One application of the above result that is developed in [3] is as follows: Given an integer $k \geq 1$, there is a data structure for answering queries of the following type on a dynamic graph $G$: *Does $G$ contain a cycle of length at least $k$?* Their data structure answers these queries in constant time and has an amortized update time of $2^{\mathcal{O}(k^4)} + \mathcal{O}(k \log n)$, assuming access to a dictionary on the edges of $G$. (Note that here $G$ is not required to have treedepth at most $k$, it is an arbitrary graph.) As can be checked in their proof, it turns out that the $2^{\mathcal{O}(k^4)}$ term in the latter result comes from their $2^{\mathcal{O}(d^2)}$ bound mentioned previously combined with the fact that the treedepth $d$ of a 2-connected graph with circumference $k$ is $\mathcal{O}(k^2)$. Using our result instead in their proof reduces the amortized update time down to $2^{\mathcal{O}(k^2)} + \mathcal{O}(k \log n)$. This solves an open problem from [3].

### References

[1] E. Birmelé. Tree-width and circumference of graphs. *J. Graph Theory*, 43(1):24–25, 2003.

• [2] M. Briański, G. Joret, K. Majewski, P. Micek, M. T. Seweryn, and R. Sharma. Treedepth vs circumference. *Combinatorica*. Accepted 2023.

[3] J. Chen, W. Czerwinski, Y. Disser, A. E. Feldmann, D. Hermelin, W. Nadara, M. Pilipczuk, M. Pilipczuk, M. Sorge, B. Wróblewski, and A. Zych-Pawlewicz. Efficient fully dynamic elimination forests with applications to detecting long paths and cycles. In D. Marx, ed., *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms, SODA 2021*, 2021, pp. 796–809. SIAM.

[4] E. A. Marshall and D. R. Wood. Circumference and pathwidth of highly connected graphs. *J. Graph Theory*, 79(3):222–232, 2015.

[5] F. Reidl, P. Rossmanith, F. S. Villaamil, and S. Sikdar. A faster parameterized algorithm for treedepth. In J. Esparza, P. Fraigniaud, T. Husfeldt, and E. Koutsoupias, eds., *Automata, Languages, and Programming - 41st International Colloquium, ICALP 2014, Proceedings, Part I*, 2014, LNCS 8572, pp. 931–942. Springer.

### Domination and Cut Problems on Chordal Graphs with Bounded Leafage

*Investigators: Roohani Sharma in cooperation with Dániel Marx, Philipp Schepper (CISPA), Prafullkumar Tale (Indian Institute of Science Education and Research, Pune), and Esther Galby (TU Hamburg)*

A graph is said to be an interval graph if we can assign each vertex an interval (of natural numbers) such that two vertices are adjacent if and only if their intervals intersects. Instead of considering intervals one can also think of a large path and then we consider intersecting subpaths (that is, subpaths sharing at least one vertex) of the larger path. Many classical problems such as Dominating Set, Connected Dominating Set, Steiner Tree, Multicut with Undeletable Terminals, or Multiway Cut with Undeletable Terminals are polynomial-time solvable on interval graphs.

Interestingly, when considering the slightly more general family of graphs defined as the intersection graphs of subtrees of a tree, these problems become NP-hard. Observe that this precisely defines the family of all chordal graphs, that is, every cycle of length at least four has a chord. The hardness for this graph class especially holds for the special case of split graphs, which can be defined as the intersection graphs of substars of a star. To interpolate between these two extremes of chordal graphs which are (the easy to handle) interval graphs and (the hard to handle) split graphs, the notion of leafage was introduced by Lin, McKee, and West [3].

The *leafage of a chordal graph G* is the minimum integer $\ell$ such that $G$ can be described by an intersection graph of subtrees of a tree with $\ell$ leaves. We consider the classical problems mentioned above and show fast (that is, FPT-time) algorithms when parameterizing by leafage [2]. For Dominating Set on chordal graphs with leafage $\ell$, we present an algorithm with running time $2^{\mathcal{O}(\ell)} \cdot n^{\mathcal{O}(1)}$ which is conceptually much simpler than the previously fastest algorithm with running time $2^{\mathcal{O}(\ell^2)} \cdot n^{\mathcal{O}(1)}$ [1]. Moreover, as the algorithm is actually designed to solve Red-Blue Dominating Set, we extend it to solve Connected Dominating Set and Steiner Tree while preserving the running time.

Additionally, we show that the classical cut problem MultiCut with Undeletable Terminals is W[1]-hard when parameterized by the leafage. We complement this result by presenting a simple $n^{\mathcal{O}(\ell)}$-time algorithm for the problem. For the related Multiway Cut with Undeletable Terminals, we show that the problem can be solved in time $n^{\mathcal{O}(1)}$ on general chordal graphs even without parameterizing by leafage.

## References


[1] F. V. Fomin, P. A. Golovach, and J. Raymond. On the tractability of optimization problems on h-graphs. *Algorithmica*, 82(9):2432–2473, 2020.

• [2] E. Galby, D. Marx, P. Schepper, R. Sharma, and P. Tale. Domination and cut problems on chordal graphs with bounded leafage. In H. Dell and J. Nederlof, eds., *17th International Symposium on Parameterized and Exact Computation (IPEC 2022)*, Potsdam, Germany, 2022, Leibniz International Proceedings in Informatics 249, Article 14. Schloss Dagstuhl.

[3] I. Lin, T. A. McKee, and D. B. West. The leafage of a chordal graph. *Discuss. Math. Graph Theory*, 18(1):23–48, 1998.

### Derandomization of Isolation Schemes on Decomposable Graphs

*Investigators: Karol Węgrzycki, in cooperation with Céline M. F. Swennenhuis (Eindhoven University of Technology), Michał Pilipczuk (University of Warsaw), and Jesper Nederlof (Utrecht University)*

The Isolation Lemma of Mulmuley, Vazirani and Vazirani [2] provides a self-reduction scheme that allows one to assume that a given instance of a problem has a unique solution, provided a solution exists at all. Since its introduction, much effort has been dedicated toward derandomization of the Isolation Lemma for specific classes of problems. So far, the focus was mainly on problems solvable in polynomial time.

In [3] we study a setting that is more typical for NP-complete problems, and obtain partial derandomizations in the form of significantly decreasing the number of required random bits. In particular, motivated by the advances in parameterized algorithms, we focus on problems on decomposable graphs. For example, for the problem of detecting a Hamiltonian cycle, we build upon the rank-based approach from [1] and design isolation schemes that use

- $\mathcal{O}(t \log n + \log^2 n)$ random bits on graphs of treewidth at most $t$;
- $\mathcal{O}(\sqrt{n})$ random bits on planar or $H$-minor free graphs; and
- $\mathcal{O}(n)$-random bits on general graphs.

In all these schemes, the weights are bounded exponentially in the number of random bits used. As a corollary, for every fixed $H$ we obtain an algorithm for detecting a Hamiltonian cycle in an $H$-minor-free graph that runs in deterministic time $2^{\mathcal{O}(\sqrt{n})}$ and uses polynomial space; this is the first algorithm to achieve such complexity guarantees. For problems of more local nature, such as finding an independent set of maximum size, we obtain isolation schemes on graphs of treedepth at most $d$ that use $\mathcal{O}(d)$ random bits and assign polynomially-bounded weights.

We also complement our findings with several unconditional and conditional lower bounds, which show that many of the results cannot be significantly improved.

## References

[1] H. L. Bodlaender, M. Cygan, S. Kratsch, and J. Nederlof. Deterministic single exponential time algorithms for connectivity problems parameterized by treewidth. *Inf. Comput.*, 243:86–111, 2015.

[2] K. Mulmuley, U. V. Vazirani, and V. V. Vazirani. Matching is as easy as matrix inversion. *Comb.*, 7(1):105–113, 1987.

• [3] J. Nederlof, M. Pilipczuk, C. M. F. Swennenhuis, and K. Węgrzycki. Isolation schemes for problems on decomposable graphs. In P. Berenbrink and B. Monmege, eds., *39th International Symposium on Theoretical Aspects of Computer Science (STACS 2022)*, Marseille, France (Virtual Conference), 2022, Leibniz International Proceedings in Informatics 219, Article 50. Schloss Dagstuhl.

## 27.5.5 Analyzing Graphs and Networks

### Graph Similarity Description

*Investigators: Corinna Coupette, in cooperation with Sebastian Dalleiger and Jilles Vreeken (CISPA Helmholtz Center for Information Security)*

Generalizing graph isomorphism tests, Graph Similarity Assessment, the task of comparing two or more graphs, is commonly treated as a *measurement* problem. However, when the input graphs represent snapshots of real-world systems, numerical answers provide limited insight. Therefore, we argue that if the goal is to gain understanding, we should treat Graph Similarity Assessment as a *description* problem instead. To enable descriptive comparisons between two or more graphs [2], we formalize the Graph Similarity Description problem as a model selection task using the Minimum Description Length principle, capturing the similarity of the input graphs in a common model and the differences between them in transformations to individual models. To discover good models, we propose Momo, which breaks the problem into two parts and introduces efficient algorithms for each, and we demonstrate that Momo works well in practice.

When we are given not just two or a handful but hundreds or thousands of graphs, and each graph is associated with a discrete feature, we are faced with a related description task, which we call graph group analysis [1]: Given a set of graphs and a partition of these graphs into groups, discover what graphs in one group have in common, how they systematically differ from graphs in other groups, and how multiple groups of graphs are related. To perform Graph Group Analysis, we introduce Gragra, which uses maximum entropy modeling to identify a non-redundant set of subgraphs with statistically significant associations to one or more graph groups, and we confirm experimentally that Gragra works well in practice.

### References


- [1] C. Coupette, S. Dalleiger, and J. Vreeken. Differentially describing groups of graphs. In *Proceedings of the 36th AAAI Conference on Artificial Intelligence*, Virtual Conference, 2022, pp. 3959–3967. AAAI.
- [2] C. Coupette and J. Vreeken. Graph similarity description: How are these graphs similar? In F. Zhu, C. Ooi, Beng, C. Miao, G. Cong, J. Tang, and T. Derr, eds., *KDD '21, 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, Virtual Event, Singapore, 2021, pp. 185–195. ACM.


### Hypergraph Analysis

*Investigators: Corinna Coupette, in cooperation with Sebastian Dalleiger, Jilles Vreeken (CISPA Helmholtz Institute for Information Security), and Bastian Rieck (Institute of AI for Health, Helmholtz Munich)*

Hypergraphs generalize graphs by allowing any number of nodes to participate in an edge. They enable us to faithfully represent complex relations, such as co-authorship of scientific papers, multilateral interactions between chemicals, or group conversations, which cannot be adequately captured by graphs. While hypergraphs are more expressive than graphs and other relational objects like simplicial complexes, how exactly they should be used to model higher-order relational data is hardly understood. Further, hypergraphs are more difficult to analyze than graphs, both theoretically and empirically, and many concepts that have proven useful for understanding graphs have yet to be transferred to the hypergraph setting.

To explore the intricacies of *modeling* relational data as hypergraphs, we introduce Hyperbard, a dataset of diverse relational data representations derived from Shakespeare's plays [2]. Our representations range from simple graphs capturing character co-occurrence in single scenes to hypergraphs encoding complex communication settings and character contributions as hyperedges with edge-specific node weights. Leveraging Hyperbard, we show that many solutions to popular graph mining problems are highly dependent on the representation choice, thus calling current graph curation practices into question.

To facilitate hypergraph analysis, we generalize the concept of *curvature* to hypergraphs. Bridging geometry and topology, curvature has established itself as a powerful characteristic of Riemannian manifolds. For graphs, *graph curvature* measures to what extent the neighborhood of an edge deviates from certain idealized model spaces, such as cliques, grids, or trees. One prominent notion of graph curvature is the *Ollivier-Ricci curvature* (orc). The orc compares random walks based at specific nodes, revealing differences in the information diffusion behavior in the graph. As the sizes of edges and edge intersections can vary in hypergraphs, there are many ways to generalize orc to hypergraphs. While some notions of hypergraph orc have been previously studied in isolation, a unified framework for their definition and computation has been lacking. Against this background, we develop Orchid, a flexible framework generalizing Ollivier-Ricci curvature to hypergraphs [1]. We prove that the resulting curvatures have favorable theoretical properties, and demonstrate that Orchid curvatures are both scalable and useful to perform a variety of hypergraph tasks in practice.

## References


- [1] C. Coupette, S. Dalleiger, and B. Rieck. Ollivier-Ricci curvature for hypergraphs: A unified framework. In *Eleventh International Conference on Learning Representations (ICLR 2023)*, Kigali, Rwanda, 2023. OpenReview.net. Accepted.
- [2] C. Coupette, J. Vreeken, and B. Rieck. *All the World's a (Hyper)Graph: A Data Drama*, 2022. arXiv: 2206.08225.


### Legal Networks

*Investigators: Corinna Coupette, in cooperation with Janis Beckedorf, Dirk Hartung (Bucerius Law School), Maximilian Böther (ETH Zurich), Holger Spamann (Harvard Law School), Michael Bommarito, Daniel Martin Katz (Illiois Tech), and Jyotsna Singh (Saarland University)*

In recent years, network analysis has proven critical in advancing our understanding of complex adaptive systems. As such, it has also been used to gain a quantitative understanding of *law as a complex adaptive system*. However, most research has focused on legal documents of a single type, and a unified framework for quantitative legal document analysis using network-analytical tools has been lacking. Against this background, we introduce a comprehensive framework for analyzing legal documents as multi-dimensional, dynamic document networks [1, 2]. We demonstrate the utility of this framework by applying it to an original dataset of statutes and regulations from two different countries, the United States and Germany, spanning more than twenty years (1998–2019). Our framework provides tools for assessing the size and connectivity of the legal system as viewed through the lens of specific document collections as well as for tracking the evolution of individual legal documents over time.

Integrating graph-analytic with text-analytic approaches, and building on the concept of *code smells* from the refactoring literature, we initiate the study of *law smells*, that is, patterns in legal texts that pose threats to the comprehensibility and maintainability of the law [4, 5]. With five intuitive law smells as running examples, we develop a comprehensive law smell taxonomy. This taxonomy classifies law smells by when they can be detected, which aspects of law they relate to, and how they can be discovered. We introduce graph-based and text-based methods to identify instances of law smells, confirming their utility in practice using the United States Code as a test case. Our work demonstrates how ideas from software engineering can be leveraged to assess and improve the quality of legal code.

Based on our experience with conducting and reviewing computational legal studies especially in the realm of legal data science, we also develop recommendations for ensuring substantial and reproducible progress in the field [3].

### References


- [1] C. Coupette, J. Beckedorf, D. Hartung, M. Bommarito, and D. M. Katz. Measuring law over time: A network analytical framework with an application to statutes and regulations in the United States and Germany. *Frontiers in Physics*, 9, Article 658463, 2021.
- [2] C. Coupette and D. Hartung. Rechtsstrukturvergleichung. *Rabels Zeitschrift für ausländisches und internationales Privatrecht*, 86(4):935–975, 2022.
- [3] C. Coupette and D. Hartung. *Sharing and Caring: Creating a Culture of Constructive Criticism in Computational Legal Studies*, 2022. arXiv: 2205.01071.
- [4] C. Coupette, D. Hartung, J. Beckedorf, M. Bother, and D. M. Katz. Law smells – defining and detecting problematic patterns in legal drafting. *Artificial Intelligence and Law*, 2022.
- [5] C. Coupette, J. Singh, and H. Spamann. Simplify your law: Using information theory to deduplicate legal documents. In B. Xue, M. Pechenizkiy, and Y. S. Koh, eds., *21st IEEE International Conference on Data Mining Workshops (ICDMW 2021)*, Virtual Conference, 2021, pp. 631–638. IEEE.


## 27.5.6 Miscellaneous Topics on Graphs

### Faster Exponential-Time Approximation Algorithms Using Approximate Monotone Local Search

*Investigators: Roohani Sharma in cooperation with Barış Can Esmer, Ariel Kulik, Dániel Marx (CISPA), and Daniel Neuen (Simon Fraser University)*

In [2], we generalize the monotone local search approach of Fomin, Gaspers, Lokshtanov and Saurabh [3], by establishing a connection between parameterized approximation and exponential-time approximation algorithms for Monotone Subset Minimization problems. In a Monotone Subset Minimization problem, the input implicitly describes a non-empty set family over a universe of size $n$ which is closed under taking supersets. The task is to find a minimum cardinality set in this family. Broadly speaking, we use *approximate monotone local search* to show that a parameterized $\alpha$-approximation algorithm that runs in $c^k \cdot n^{\mathcal{O}(1)}$ time, where $k$ is the solution size, can be used to derive an $\alpha$-approximation randomized algorithm that runs in $d^n \cdot n^{\mathcal{O}(1)}$ time, where $d$ is the unique value in $d \quad 1, 1 + (c-1)/\alpha$ such that $\mathcal{D}\left(1/\alpha \;\; (d-1)/(c-1)\right) = \ln(c)/\alpha$ and $\mathcal{D}\left(a \;\; b\right)$ is the Kullback-Leibler divergence. This running time matches that of Fomin et al. for $\alpha = 1$, and is strictly better when $\alpha > 1$, for any $c > 1$. Furthermore, we also show that this result can be derandomized at the expense of a sub-exponential multiplicative factor in the running time.

We use an approximate variant of the exhaustive search as a benchmark for our algorithm. We show that the classic $2^n \cdot n^{\mathcal{O}(1)}$ exhaustive search can be adapted to an $\alpha$-approximate exhaustive search that runs in time $\left(1 + \exp\left(-\alpha \cdot \mathcal{H}\left(1/\alpha\right)\right)\right)^n \cdot n^{\mathcal{O}(1)}$, where $\mathcal{H}$ is the binary entropy function. Furthermore, we provide a lower bound stating that the running time of this $\alpha$-approximate exhaustive search is the best achievable running time in an oracle model. When compared to approximate exhaustive search, and to other techniques, the running times obtained by approximate monotone local search are strictly better for any $\alpha \geq 1,\ c > 1$.

We demonstrate the potential of approximate monotone local search by deriving new and faster exponential approximation algorithms for Vertex Cover, 3-Hitting Set, Directed Feedback Vertex Set, Directed Subset Feedback Vertex Set, Directed Odd Cycle Transversal and Undirected Multicut. For instance, we get a 1.1-approximation algorithm for Vertex Cover with running time $1.114^n \cdot n^{\mathcal{O}(1)}$, improving upon the previously best known 1.1-approximation running in time $1.127^n \cdot n^{\mathcal{O}(1)}$ by Bourgeois et al. [1].

In an ongoing follow-up work we extend the setting and show how an exponential time $\beta$-approximation algorithm can be obtained from a parameterized $\alpha$-approximation algorithm that runs in $c^k \cdot n^{\mathcal{O}(1)}$. Surprisingly, for any $\beta > 1$ and $\alpha, c \geq 1$, the running time of the resulting algorithm is better than that of the approximate exhaustive search. Furthermore, we show the running time of the resulting algorithm is optimal in an oracle model.

## References

[1] N. Bourgeois, B. Escoffier, and V. T. Paschos. Approximation of max independent set, min vertex cover and related problems by moderately exponential algorithms. *Discret. Appl. Math.*, 159(17):1954–1970, 2011.

[2] B. C. Esmer, A. Kulik, D. Marx, D. Neuen, and R. Sharma. Faster exponential-time approximation algorithms using approximate monotone local search. In S. Chechik, G. Navarro, E. Rotenberg, and G. Herman, eds., *30th Annual European Symposium on Algorithms (ESA 2022)*, Berlin/Potsdam, Germany, 2022, Leibniz International Proceedings in Informatics 244, Article 50. Schloss Dagstuhl.

[3] F. V. Fomin, S. Gaspers, D. Lokshtanov, and S. Saurabh. Exact algorithms via monotone local search. *J. ACM*, 66(2):8:1–8:23, 2019.

### Approximate Sampling and Counting Graphs with Degree Constraints

*Investigators: Pieter Kleer, in cooperation with Leen Stougie (Centrum Wiskunde & Informatica), Catherine Greenhill (UNSW Sydney), James Ross (The University of Chicago Booth School of Business), Georgios Amanatidis (University of Essex) and Martin Dyer (Unviversity of Leeds)*

We study the problem of approximately uniformly at random sampling graphs with given degree constraints. This is an important task, for instance, to test (statistical) hypotheses on collected data in the fields of social network analysis and ecology.

We consider various types of degree constraints:

- In [2], we consider the problem of sampling graphs with given degree and joint degree constraints. Here the task is to approximately sample a graph from the set of all graphs for which we are given a degree for every node, and a joint degree constraint matrix, that specifies how many edges there are between nodes of the same degree. Joint degree constraints can be used to model, for instance, the property that in social networks nodes with high degrees tend to connect more often to nodes with low degree. We give a polynomial time algorithm for sampling graphs with two degree classes and a specified joint degree constraint between them. This is the first result beyond the case of one degree class, which corresponds to sampling a regular graph with a given common degree for all the nodes.
- In [3], we consider the problem of sampling hypergraphs with a given degree sequence. This is a generalization of the well-known problem of sampling a graph with a given degree sequence. We provide a polynomial time rejection sampling-based approach for sampling hypergraphs with various given degree sequences in [3], which is the first algorithm for sampling such graphs. The rejection-sampling approach reduces the problem to that of sampling a bipartite graph with given degrees.
- In [1], we consider another generalization of the problem of sampling graphs with a given degree sequence, namely the problem of sampling graphs with given degree intervals. This is a more realistic assumption in case an observed network (on which one wants to test a hypothesis) is only partially observed or has missing data points. We provide a polynomial time algorithm for sampling graphs with given degree intervals in case all intervals are close to being regular, thereby generalizing the problem of sampling regular graphs with a given common degree.

## References


- [1] G. Amanatidis and P. Kleer. *Approximate Sampling and Counting of Graphs with Near-Regular Degree Intervals*, 2021. arXiv: 2110.09068.
- [2] G. Amanatidis and P. Kleer. Rapid mixing of the switch Markov chain for 2-class joint degree matrices. *SIAM Journal on Discrete Mathematics*, 36(1):118–146, 2022.
- [3] M. Dyer, C. Greenhill, P. Kleer, J. Ross, and L. Stougie. Sampling hypergraphs with given degrees. *Discrete Mathematics*, 344(11), Article 112566, 2021.

### Computing Graph Hyperbolicity Using Dominating Sets and Enumeration of Far-Apart Pairs by Decreasing Distance

*Investigators: André Nusser in cooperation with Laurent Viennot (Paris University, Inria, CNRS, Irif) and David Coudert (Université Côte d'Azur, Inria, CNRS, I3S)*

Hyperbolicity is a graph parameter which indicates how much the shortest-path distance metric of a graph deviates from a tree metric. It is used in various fields such as networking, security, and bioinformatics for the classification of complex networks, the design of routing schemes, and the analysis of graph algorithms. Despite recent progress, computing the hyperbolicity of a graph remains challenging. Indeed, the best known algorithm has time complexity $\mathcal{O}(n^{3.69})$, which is prohibitive for large graphs, and the most efficient algorithms in practice have space complexity $\mathcal{O}(n^2)$. Thus, time as well as space are bottlenecks for computing the hyperbolicity.

In [2], we design a tool for enumerating all far-apart pairs of a graph by decreasing distances. A node pair $(u, v)$ of a graph is far-apart if both $v$ is a leaf of all shortest-path trees rooted at $u$ and $u$ is a leaf of all shortest-path trees rooted at $v$. This notion was previously used to drastically reduce the computation time for hyperbolicity in practice. However, it required the computation of the distance matrix to sort all pairs of nodes by decreasing distance, which requires an infeasible amount of memory already for medium-sized graphs. We present a new data structure that avoids this memory bottleneck in practice and for the first time enables computing the hyperbolicity of several large graphs that were far out-of-reach using previous algorithms. For some instances, we reduce the memory consumption by at least two orders of magnitude. Furthermore, we show that for many graphs, only a very small fraction of far-apart pairs has to be considered for the hyperbolicity computation, explaining this drastic reduction of memory. As iterating over far-apart pairs in decreasing order without storing them explicitly is a very general tool, we believe that our approach might also be relevant to other problems.

In a follow-up work [1], we propose and evaluate an approach that uses a hierarchy of distance-$k$ dominating sets to reduce the search space. This technique, compared to [2], enables us to compute the hyperbolicity of graphs with unprecedented size (up to a million nodes).

### References


- [1] D. Coudert, A. Nusser, and L. Viennot. Computing graph hyperbolicity using dominating sets. In C. A. Phillips and B. Speckman, eds., *Proceedings of the Symposium on Algorithm Engineering and Experiments (ALENEX 2022)*, Alexandria, VA, USA, 2022, pp. 78–90. SIAM.

- [2] D. Coudert, A. Nusser, and L. Viennot. Enumeration of far-apart pairs by decreasing distance for faster hyperbolicity computation. *ACM Journal of Experimental Algorithmics*, 27, Article 1.15, 2022.

### The Complexity of Contracting Bipartite Graphs into Small Cycles

*Investigators: Roohani Sharma, in cooperation with Prafullkumar Tale (Indian Institute of Science Education and Research Pune) and R. Krithika (Indian Institute of Technology Palakkad)*

Due to the ubiquitous presence of graphs in modeling real-world networks, many problems of practical importance can be posed as editing problems on graphs. In this work, we focus on modifying a graph by only performing edge contractions. Contracting an edge in a graph results in the addition of a new vertex adjacent to the neighbors of its endpoints followed by the deletion of the endpoints. As graphs typically represent binary relationships among a collection of objects, edge contractions naturally correspond to merging two objects into a single entity or to treating two objects as indistinguishable.

The problem of determining whether a graph $G$ is contractible to a graph $H$ may be seen as the task of determining if the "underlying structure" of $G$ is $H$. One of the related graph parameters in this context is cyclicity. The cyclicity of a graph is the largest integer $\ell$ for which the graph is contractible to the induced cycle on $\ell$ vertices. This parameter was introduced in the study of another important graph invariant called circularity [1]. Ever since, there have been efforts towards understanding the complexity of computing cyclicity.

For a positive integer $\ell \geq 3$, the $C_\ell$-Contractibility problem takes as input an undirected simple graph $G$ and determines whether $G$ can be transformed into a graph isomorphic to $C_\ell$ (the induced cycle on $\ell$ vertices) using only edge contractions. Brouwer and Veldman [2] showed that $C_4$-Contractibility is NP-complete in general graphs. It is easy to verify that that $C_3$-Contractibility is polynomial-time solvable. Dabrowski and Paulusma [3] showed that $C_\ell$-Contractibility is NP-complete on bipartite graphs for $\ell = 6$ and posed as open problems the status of the problem when $\ell$ is 4 or 5. In [4], we show that both $C_5$-Contractibility and $C_4$-Contractibility are NP-complete on bipartite graphs.

## References

[1] D. J. Blum. *Circularity of Graphs.* PhD thesis, Virginia Polytechnic Institute and State University, 1982.

[2] A. E. Brouwer and H. J. Veldman. Contractibility and NP-completeness. *J. Graph Theory*, 11(1):71–79, 1987.

[3] K. K. Dabrowski and D. Paulusma. Contracting bipartite graphs to paths and cycles. *Inf. Process. Lett.*, 127:37–42, 2017.

• [4] R. Krithika, R. Sharma, and P. Tale. The complexity of contracting bipartite graphs into small cycles. In M. A. Bekos and M. Kaufmann, eds., *Graph-Theoretic Concepts in Computer Science (WG 2022)*, Tübingen, Germany, 2022, LNCS 13453, pp. 356–369. Springer.

### Balanced Substructures in Bicolored Graphs

*Investigators: Roohani Sharma, in cooperation with P. S. Ardra, R. Krithika (Indian Institute of Technology Palakkad), and Saket Saurabh (The Institute of Mathematical Sciences, University of Bergen)*

Ramsey Theory is a branch of combinatorics that deals with patterns in large arbitrary structures. In the context of edge-colored graphs where each edge is colored with one color from a finite set of colors, a fundamental problem in the area is concerned with the existence of *monochromatic* subgraphs of a specific type. Here, monochromatic means that all edges of the subgraph have the same color. For simplicity, we discuss only undirected graphs where each edge is colored either red or blue. In [1], we study questions related to the existence of and finding *balanced* subgraphs instead of monochromatic subgraphs, where by a balanced subgraph we mean one which has an equal number of edges of each color. These problems come under a subarea of Ramsey Theory known as Zero-Sum Ramsey Theory. Here, given a graph whose vertices/edges are assigned weights from a set of integers, one looks for conditions that guarantee the existence of a certain subgraph having total weight zero. For example, one may ask when is a graph whose all the edges are given weight $-1$ or 1 guaranteed to have a spanning tree with total weight of its edges 0. This is equivalent to asking when a red-blue graph is guaranteed to have a balanced spanning tree. Necessary and sufficient conditions have been established for complete graphs, triangle-free graphs and maximal planar graphs [2]. However, we do not think that simple characterizations will exist for all graphs. This brings us to study the algorithmic questions concerning balanced connected subgraphs.

Given a graph $G$ whose edges are colored using two colors and a positive integer $k$, the objective in the Edge Balanced Connected Subgraph problem is to determine if $G$ has a balanced connected subgraph containing at least $k$ edges. We first show that this problem is NP-complete and remains so even if the solution is required to be a tree or a path. Then, we focus on the parameterized complexity of Edge Balanced Connected Subgraph and its variants (where the balanced subgraph is required to be a path/tree) with respect to $k$ as the parameter. Toward this end, we show that if a graph has a balanced connected subgraph/tree/path of size at least $k$, then it has one of size at least $k$ and at most $f(k)$ where $f$ is a linear function. We use this result combined with dynamic programming algorithms based on *color coding* and *representative sets* to show that Edge Balanced Connected Subgraph and its variants are FPT. Further, using polynomial-time reductions to the Multilinear Monomial Detection problem, we give faster randomized FPT algorithms for the problems. In order to describe these reductions, we define a combinatorial object called *relaxed-subgraph*. We define this object in such a way that balanced connected subgraphs, trees and paths are relaxed-subgraphs with certain properties. This object is defined in the spirit of branching walks known for the Steiner Tree problem and may be of independent interest.

## References


- [1] P. S. Ardra, R. Krithika, S. Saurabh, and R. Sharma. Balanced substructures in bicolored graphs. In L. Gąsieniec, ed., *SOFSEM 2023: Theory and Practice of Computer Science*, Nový Smokovec, Slovakia, 2023, LNCS 13878, pp. 177–191. Springer.

[2] Y. Caro, A. Hansberg, J. Lauri, and C. Zarb. On zero-sum spanning trees and zero-sum connectivity. *Electron. J. Comb.*, 29(1), 2022.

### Coloring Mixed and Directional Interval Graphs

*Investigators: Joachim Spoerhase in cooperation with Grzegorz Gutowski (Jagiellonian University), Ignaz Rutter (University of Passau), Florian Mittelstädt, Alexander Wolff, Johannes Zink (University of Würzburg)*

In [1], we prove directional variants of classic results concerning efficient recognition and colorability of interval graphs. More formally, a *mixed graph* has a set of vertices, a set of undirected edges, and a set of directed arcs. A *proper coloring* of a mixed graph $G$ is a function $c$ that assigns to each vertex in $G$ a positive integer such that, for each edge $u, v$ in $G$, $c(u) = c(v)$ and, for each arc $(u, v)$ in $G$, $c(u) < c(v)$. For a mixed graph $G$, the *chromatic number* $\chi(G)$ is the smallest number of colors in any proper coloring of $G$. A *directional interval graph* is a mixed graph whose vertices correspond to intervals on the real line. Such a graph has an edge between every two intervals that are contained in each other and an arc between every two overlapping intervals, directed towards the interval that starts and ends to the right.

Coloring such graphs has applications in routing edges in layered orthogonal graph drawing according to the Sugiyama framework, which is extensively used in drawing layered graphs; the colors correspond to the tracks for routing the edges. We show how to recognize directional interval graphs, and how to compute their chromatic number efficiently. On the other hand, for *mixed interval graphs*, that is, graphs where two intersecting intervals can be connected by an edge or by an arc in either direction arbitrarily, we prove that computing the chromatic number is NP-hard.

Our results show that classic results regarding recognition and colorability of interval graphs carry over to the directional setting. However, the underlying geometry differs notably. For example, the known result for vanilla interval graphs rely heavily on a simple characterization of cliques as subsets of intervals intersecting a single point. Cliques in the directional setting may have an significantly more intricate structure.

## References

• [1] G. Gutowski, F. Mittelstädt, I. Rutter, J. Spoerhase, A. Wolff, and J. Zink. Coloring mixed and directional interval graphs. In P. Angelini and R. von Hanxleden, eds., *Graph Drawing and Network Visualization (GD 2022)*, Tokyo, Japan, 2022, LNCS 13764, pp. 418–431. Springer.

### Hypergraph Representation of Axis-Aligned Point-Subspace Cover

*Investigators: Joachim Spoerhase in cooperation with Oksana Firman (University of Würzburg)*

Our work [1] concerns understanding the combinatorial structure of incidences between points and axis-parallel lines and more generally between points and axis-aligned affine sub-spaces in higher dimensions. The question arises in the study of geometric set cover (for instance, axis-parallel point-line cover) and geometric intersection graphs. We give a structural characterization of hypergraphs that represent these incidences. This allows a polynomial-time recognition algorithm.

More formally, consider the following representation via $k$-partite $k$-uniform hypergraphs ($k$-hypergraphs for short). Let $P$ be a finite set of points in $\mathbb{R}^d$, and let $\ell \leq d-1$ be a parameter. Each point in $P$ is covered by $k = \binom{d}{\ell}$ many axis-aligned affine $\ell$-dimensional subspaces of $\mathbb{R}^d$, which we call $\ell$-subspaces for brevity. We interpret each point in $P$ as a hyperedge that contains each of the covering $\ell$-subspaces as a vertex. The class of $(d, \ell)$-*hypergraphs* is the class of $k$-hypergraphs (where $k = \binom{d}{\ell}$) that can be represented in this way. The resulting classes of hypergraphs are fairly rich: Every $k$-hypergraph is a $(k, k-1)$-hypergraph. On the other hand, $(d, \ell)$-hypergraphs form a proper subclass of the class of all $\binom{d}{\ell}$-hypergraphs for $\ell < d - 1$.

In this paper we give a natural structural characterization of $(d, \ell)$-hypergraphs based on vertex cuts. This characterization leads to a polynomial-time recognition algorithm that decides for a given $\binom{d}{\ell}$-hypergraph whether or not it is a $(d, \ell)$-hypergraph and that computes a representation if existing. We assume that the dimension $d$ is constant and that the partitioning of the vertex set is prescribed.

### References


• [1] O. Firman and J. Spoerhase. Hypergraph representation via axis-aligned point-subspace cover. In P. Mutzel, M. S. Rahman, and Slamin, eds., *WALCOM: Algorithms and Computation*, Jember, Indonesia, 2022, LNCS 13174, pp. 328–339. Springer.


## 27.6 Algorithms and Complexity on Numbers and Geometric Objects

In this research area we investigate the structures and principles that are fundamental for the understanding and design of efficient algorithms on mathematical objects such as numbers and polynomials and on geometric objects such as points, lines and curves. We develop techniques for designing efficient exact as well as approximation algorithms for polynomial-time and NP-hard problems. We complement algorithm design with fine-grained complexity theory, a modern approach from complexity theory that yields quantitative running time lower bounds, and thus allows us to provide evidence that our designed algorithms are best-possible.

**Numbers and Polynomials.** Mathematical objects such as numbers and polynomials are ubiquitous in the design of discrete algorithms. For instance, many algorithmic improvements have been obtained from encoding data in the coefficients of polynomials and exploiting fast algorithms for polynomial multiplication.

In the last reporting period, we developed algorithms for multiplying sparsely represented polynomials with nonnegative coefficients, and we developed new algorithms for the Subset Sum problem and variants of the Knapsack problem. For more details on these results, consult Section 27.6.1.

**Foundations of Fine-Grained Complexity Theory.** The classic theory of NP-hardness provides a coarse-grained classification of problems into efficient (that is, polynomial-time solvable) and intractable (that is, NP-hard). In contrast, modern fine-grained complexity theory yields quantitative running time lower bounds, and thus provides a more fine-grained classification of problems according to their time complexity. Naturally, such stronger lower bounds rely on stronger hypotheses than P = NP, for instance, the Strong Exponential Time Hypothesis (SETH) for SATISFIABILITY.

In the last reporting period, we contributed to the foundations of fine-grained complexity theory, for instance, by developing proof techniques for hardness of approximation in P and by studying fine-grained complexity classes. For more details on these results, consult Section 27.6.2.

**Fine-Grained Complexity in Computational Geometry** In addition to the foundations of fine-grained complexity, we also developed fine-grained complexity in several application areas. In the reporting period, the main focus was computational geometry, where we proved several lower bounds based on the 3Sum Hypothesis or the Strong Exponential Time Hypothesis for the time complexity of various geometric problems including POLYGON CONTAINMENT, the diameter of intersection graphs, and similarity measures between point sets.

For more details on these results, consult Section 27.6.3.

**Curve Similarity Measures in Computational Geometry** Many tasks in computational geometry involve curves, for instance, given by GPStrajectories, sensor measurements, or stock price history. The two most popular distance measures for judging the similarity of two curves are the Fréchet distance and DYNAMIC TIME WARPING.

In the past reporting period, we tackled a wide range of algorithmic questions on these similarity measures, including NEAREST NEIGHBOR data structures and algorithms for translation-invariant variations of these measures. For more details on these results, consult Section 27.6.4.

**NP-Hard Problems in Computational Geometry** NP-hard problems are ubiquitous, also in computational geometry. In the last reporting period, we developed methods to cope with NP-hardness, including approximation algorithms and exact algorithms with mildly exponential running time.

In particular, we designed an improved approximation algorithm for the classic TRAVELING SALESPERSON PROBLEM on points in the plane, and we explored NP-hard problems on geometric intersection graphs. For more details on these results, consult Section 27.6.5.

### 27.6.1 Numbers and Polynomials

#### Algorithms for Sparse Convolution

*Investigators: Karl Bringmann, Nick Fischer, and Vasileios Nakos*

The *convolution* of two integer vectors $A, B$ is the vector $A \star B$ which is defined coordinate-wise by $(A \star B)[k] = \sum_{i+j=k} A[i] \cdot B[j]$. Computing convolutions of integer vectors is a fundamental computational primitive which arises in several disciplines of science and engineering and has been a vital component in fields like computer algebra (computing convolutions is equivalent to multiplying polynomials), signal processing, computer vision and deep learning (convolutional neural networks). By the Fast Fourier Transform (FFT), we can compute $A \star B$ in time $\mathcal{O}(n \log n)$, where $n$ is the length of the given vectors.

In this line of research, we focus on *sparse* convolutions, where the input vectors $A, B$ and the output vector $A \star B$ are promised to contain at most $t$ nonzero entries. This variant has also found several applications in algorithm design, specifically for the Subset Sum and $k$-Sum problems and various string problems. About 20 years ago, Cole and Hariharan [5] designed the first randomized algorithm computing $A \star B$ in near-linear time $\mathcal{O}(t \log^2 n)$, assuming that all vectors have nonnegative entries. This assumption was later removed by Nakos [6].

We further improve the state of the art for Sparse Nonnegative Convolution in several directions. First, in [1] we improve the logarithmic factors and obtain a randomized algorithm with running time $\mathcal{O}(t \log t)$. This running time is asymptotically optimal, assuming the widely believed conjecture that FFT requires time $\Omega(n \log n)$. Moreover, even if there is an improved FFT algorithm, we show that Sparse Nonnegative Convolution can be solved in time $\mathcal{O}(T_{\text{FFT}}(t) + t(\log\log t)^2)$. This result can be seen as a *fine-grained reduction* from Sparse Nonnegative Convolution to Dense Nonnegative Convolution.

In addition, in [2] we design a randomized algorithm that matches Cole and Hariharan's original running time [5], but is considerably simpler and thus possibly more practical.

In another direction we study *deterministic* algorithms for Sparse Nonnegative Convolution. Based on ideas by Chan and Lewenstein [4], in [3] we design a deterministic algorithm for sparse nonnegative convolution with running time $t \cdot 2^{\mathcal{O}(\sqrt{\log t \log\log n})}$. In [2] we further improve this algorithm to near-linear running time $t \operatorname{polylog}(n)$, using insights from linear algebra. This result allows us to derandomize several algorithms for the Subset Sum and $k$-Sum problems and various string problems.

## References


- [1] K. Bringmann, N. Fischer, and V. Nakos. Sparse nonnegative convolution is equivalent to dense nonnegative convolution. In S. Khuller and V. Vassilevska Williams, eds., *STOC '21, 53rd Annual ACM SIGACT Symposium on Theory of Computing*, Virtual, Italy, 2021, pp. 1711–1724. ACM.
- [2] K. Bringmann, N. Fischer, and V. Nakos. Deterministic and Las Vegas algorithms for sparse nonnegative convolution. In S. Naor and N. Buchbinder, eds., *Proceedings of the Annual ACM-SIAM Symposium on Discrete Algorithms (SODA 2022)*, Virtual, 2022, pp. 3069–3090. SIAM.
- [3] K. Bringmann and V. Nakos. Fast $n$-fold Boolean convolution via additive combinatorics. In N. Bansal, E. Merelli, and J. Worrell, eds., *48th International Colloquium on Automata, Languages, and Programming (ICALP 2021)*, Glasgow, UK (Virtual Conference), 2021, Leibniz International Proceedings in Informatics 198, Article 41. Schloss Dagstuhl.

[4] T. M. Chan and M. Lewenstein. Clustered integer 3sum via additive combinatorics. In R. A. Servedio and R. Rubinfeld, eds., *Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing, STOC 2015, Portland, OR, USA, June 14-17, 2015*, 2015, pp. 31–40. ACM.

[5] R. Cole and R. Hariharan. Verifying candidate matches in sparse and wildcard matching. In J. H. Reif, ed., *Proceedings on 34th Annual ACM Symposium on Theory of Computing, May 19-21, 2002, Montréal, Québec, Canada*, 2002, pp. 592–601. ACM.

[6] V. Nakos. Nearly optimal sparse polynomial multiplication. *IEEE Trans. Inf. Theory*, 66(11):7231–7236, 2020.

### Computing Generalized Convolutions

*Investigators: Karol Węgrzycki in cooperation with Barış Can Esmer, Ariel Kulik, Dániel Marx, and Philipp Schepper (CISPA)*

Convolutions occur naturally in many algorithmic applications, especially in exact and parameterized algorithms. The most prominent example is a subset convolution procedure, for which an efficient $\mathcal{O}^*(2^n)$-time algorithm exists. Researchers considered a plethora of other variants of convolutions, such as: Cover Product, Xor Product, Packing Product, Generalized Subset Convolution, or Discriminantal Subset Convolution. These subroutines are crucial ingredients in the design of efficient algorithms for many exact and parameterized algorithms such as Hamiltonian Cycle, Feedback Vertex Set, Steiner Tree, Connected Vertex Cover, Chromatic Number, Max $k$-Cut, or Bin Packing.

Usually, in the process of algorithm design, the researcher needs to design a different type of convolution from scratch to solve each of these problems. Often this is a highly technical and laborious task. Ideally, we would like to have a single tool that can be used as a black-box in all of these cases.

In [1], our goal is to unify convolution procedures under one general umbrella. Toward this goal, we consider the problem of computing the $f$-*Generalized Convolution* introduced by van Rooij [2]. Let $\mathcal{D}$ be a finite domain, let $f : \mathcal{D} \times \mathcal{D} \quad \mathcal{D}$ be an arbitrary function and let $\oplus_f$ be a coordinate-wise application of the function $f$. For two functions $g, h : \mathcal{D}^n \quad \mathbb{Z}$ the $f$-Convolution denoted by $(g \otimes h) : \mathcal{D}^n \quad \mathbb{Z}$, is defined for all $\mathbf{v} \quad \mathcal{D}^n$ as

$$(g \otimes h)(\mathbf{v}) := \sum_{\substack{\mathbf{v}_g, \mathbf{v}_h \in \mathcal{D}^n \\ \text{such that } \mathbf{v} = \mathbf{v}_g \oplus_f \mathbf{v}_h}} g(\mathbf{v}_g) \cdot h(\mathbf{v}_h).$$

For an arbitrary function $f$ and domain $\mathcal{D}$ one can compute $f$-Convolution via brute-force enumeration in time $\widetilde{\mathcal{O}}( \mathcal{D}^{2n})$. Our main result is an improvement over this naive algorithm. We show that $f$-Convolution can be computed exactly in time $\widetilde{\mathcal{O}}((0.75 \cdot \mathcal{D}^{2})^n)$ when $\mathcal{D}$ has even cardinality. Our main observation is that a *cyclic partition* of a function $f : \mathcal{D} \times \mathcal{D} \quad \mathcal{D}$ can be used to speed up the computation of $f$-Convolution, and we show that an appropriate cyclic partition exists for every $f$.

### References

• [1] B. C. Esmer, A. Kulik, D. Marx, P. Schepper, and K. Węgrzycki. Computing generalized convolutions faster than brute force. In H. Dell and J. Nederlof, eds., *17th International Symposium on Parameterized and Exact Computation (IPEC 2022)*, Potsdam, Germany, 2022, Leibniz International Proceedings in Informatics 249, Article 12. Schloss Dagstuhl.

[2] J. M. M. van Rooij. A Generic Convolution Algorithm for Join Operations on Tree Decompositions. In R. Santhanam and D. Musatov, eds., *Computer Science - Theory and Applications - 16th International Computer Science Symposium in Russia, CSR 2021, Sochi, Russia, June 28 - July 2, 2021, Proceedings*, 2021, LNCS 12730, pp. 435–459. Springer.

### Faster Algorithms for Unbounded Knapsack and 0/1 Knapsack

*Investigators: Karl Bringmann, Alejandro Cassis*

In [3], we develop improved exact and approximation algorithms for two variants of the KNAPSACK problem: 0/1-KNAPSACK and UNBOUNDED KNAPSACK. In both of these, we are given $n$ items where each item has a profit $p_i$ and a weight $w_i$, along with a knapsack capacity $W$. In the 0/1 variant, the goal is to select a subset of items with total weight at most $W$ that maximizes the total profit. In the unbounded variant the goal is the same, but any item can be included multiple times in a solution.

Our results are the following:

- *Exact Algorithm for 0-1-Knapsack:* 0/1-KNAPSACK has known algorithms running in time $\widetilde{\mathcal{O}}(n + \min\; n \cdot \mathrm{OPT}, n \cdot W, \mathrm{OPT}^2, W^2\; )$ [2], where $n$ is the number of items, $W$ is the weight budget, and OPT is the optimal profit. We present an algorithm running in time $\widetilde{\mathcal{O}}(n + (W + \mathrm{OPT})^{1.5})$. This improves the running time in case $n, W, \mathrm{OPT}$ are roughly equal.

- *Exact Algorithm for Unbounded Knapsack:* UNBOUNDED KNAPSACK has known algorithms running in time $\widetilde{\mathcal{O}}(n + \min\; n \cdot p_{\max}, n \cdot w_{\max}, p_{\max}^2, w_{\max}^2\; )$ [1, 7, 4], where $n$ is the number of items, $w_{\max}$ is the largest weight of any item, and $p_{\max}$ is the largest profit of any item. We present an algorithm running in time $\widetilde{\mathcal{O}}(n + (p_{\max} + w_{\max})^{1.5})$, giving a similar improvement as for 0/1-KNAPSACK.

- *Approximating Unbounded Knapsack with Resource Augmentation:* UNBOUNDED KNAPSACK has a known FPTAS with running time $\widetilde{\mathcal{O}}(\min\; n/\varepsilon, n + 1/\varepsilon^2\; )$ [6]. We study *weak* approximation algorithms, which approximate the optimal profit but are allowed to overshoot the weight constraint (that is, resource augmentation). We present the first approximation scheme or UNBOUNDED KNAPSACK in this setting, achieving running time $\widetilde{\mathcal{O}}(n + 1/\varepsilon^{1.5})$. Along the way, we also give a simpler FPTAS with lower order improvement in the standard setting.

For all of these problem settings, the previously known results had matching conditional lower bounds. We avoid these lower bounds in the approximation setting by allowing resource augmentation, and in the exact setting by analyzing the time complexity in terms of weight and profit parameters (instead of only weight or only profit parameters).

Our algorithms can be seen as reductions to Min-Plus-Convolution on monotone sequences with bounded entries. These structured instances of Min-Plus-Convolution can be solved in time $\mathcal{O}(n^{1.5})$ [5] (in contrast to the conjectured $n^{2-o(1)}$ lower bound for the general case). We complement our results by showing reductions in the opposite direction, that is, we show that achieving our results with the constant 1.5 replaced by any constant less than 2 implies subquadratic algorithms for MIN-PLUS-CONVOLUTION on monotone sequences with bounded entries.

### References

[1] K. Axiotis and C. Tzamos. Capacitated dynamic programming: Faster knapsack and graph algorithms. In C. Baier, I. Chatzigiannakis, P. Flocchini, and S. Leonardi, eds., *46th International*

*Colloquium on Automata, Languages, and Programming, ICALP 2019, July 9-12, 2019, Patras, Greece*, 2019, LIPIcs 132, pp. 19:1–19:13. Schloss Dagstuhl - Leibniz-Zentrum für Informatik.

[2] R. Bellman. *Dynamic Programming.* Princeton University Press, Princeton, NJ, USA, 1957.

• [3] K. Bringmann and A. Cassis. Faster Knapsack algorithms via bounded monotone Min-Plus-Convolution. In M. Bojańczyk, E. Merelli, and D. P. Woodruff, eds., *49th EATCS International Conference on Automata, Languages, and Programming (ICALP 2022)*, Paris, France, 2022, Leibniz International Proceedings in Informatics 229, Article 31. Schloss Dagstuhl.

[4] T. M. Chan and Q. He. More on change-making and related problems. *J. Comput. Syst. Sci.*, 124:159–169, 2022.

[5] S. Chi, R. Duan, T. Xie, and T. Zhang. Faster min-plus product for monotone instances. In S. Leonardi and A. Gupta, eds., *STOC '22: 54th Annual ACM SIGACT Symposium on Theory of Computing, Rome, Italy, June 20 - 24, 2022*, 2022, pp. 1529–1542. ACM.

[6] K. Jansen and S. E. J. Kraft. A faster FPTAS for the unbounded knapsack problem. *Eur. J. Comb.*, 68:148–174, 2018.

[7] K. Jansen and L. Rohwedder. On integer programming and convolution. In A. Blum, ed., *10th Innovations in Theoretical Computer Science Conference, ITCS 2019, January 10-12, 2019, San Diego, California, USA*, 2019, LIPIcs 124, pp. 43:1–43:17. Schloss Dagstuhl - Leibniz-Zentrum für Informatik.

### Online Fractional Knapsack

*Investigators: Andreas Karrenbauer, in cooperation with Jeff Giliberti (ETH Zürich, Switzerland)*

The Fractional Knapsack problem is one of the classical problems in combinatorial optimization, which is well understood in the offline setting. However, the corresponding online setting has been handled only briefly in the theoretical computer science literature so far, although it appears in several applications. Even the previously best known guarantee for the competitive ratio was worse than the best known for the integral problem in the popular random order model.

In [1], we show that there is an algorithm for the Online Fractional Knapsack problem that admits a competitive ratio of 4.39. Our result significantly improves over the previously best known competitive ratio of 9.37 and surpasses the current best 6.65-competitive algorithm for the integral case. Moreover, our algorithm is deterministic in contrast to the randomized algorithms achieving the results mentioned above.

### References

• [1] J. Giliberti and A. Karrenbauer. Improved online algorithm for fractional knapsack in the random order model. In J. Koenemann and B. Preis, eds., *Approximation and Online Algorithms*, Lisbon, Portugal, 2021, LNCS 12982, pp. 188–205. Springer.

### Knapsack and Subset Sum Parameterized by the Maximum Item

*Investigators: Adam Polak, Karol Węgrzycki, in cooperation with Lars Rohwedder (Maastricht University)*

Knapsack and Subset Sum are fundamental NP-hard problems in combinatorial optimization. Recently there has been a growing interest in understanding the best possible pseudopolynomial running times for these problems with respect to various parameters. In [7], we focused on parametrization by the maximum item size $s$ and the maximum item value $v$.

For Subset Sum, already the work reported in previous report [2] implied an algorithm running in time $\widetilde{\mathcal{O}}(n + s^{3/2})$, but only int the special case where all the input items are different. Now, in [7], we gave an algorithm running in time $\widetilde{\mathcal{O}}(n + s^{5/3})$ for the general case. Our algorithm works for the more general problem variant with multiplicities, where each input item comes with a (binary encoded) multiplicity, which succinctly describes how many times the item appears in the instance. In this variant, $n$ denotes the (possibly much smaller) number of distinct items. Our Knapsack algorithm also works for the variant with multiplicities.

For Knapsack, we gave algorithms running in time $\mathcal{O}(n + s^3)$ and $\mathcal{O}(n + v^3)$. This shows that the previous $\widetilde{\mathcal{O}}(n+st)$ time algorithms [6, 1], which cannot be improved in general under the Min-Plus Convolution Hypotheses [3], are not optimal in certain parameter regimes.

We obtained our results by combining and optimizing several diverse lines of research: integer linear programming proximity arguments [4], fast convex $(\min, +)$-Convolution [6], and additive combinatorics methods for Subset Sum [5, 2].

### References


[1] M. Bateni, M. Hajiaghayi, S. Seddighin, and C. Stein. Fast algorithms for knapsack via convolution and prediction. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2018, Los Angeles, CA, USA, June 25-29, 2018*, 2018, pp. 1269–1282. ACM.

[2] K. Bringmann and P. Wellnitz. On near-linear-time algorithms for dense Subset Sum. In D. Marx, ed., *Proceedings of the Thirty-Second ACM-SIAM Symposium on Discrete Algorithms (SODA 2021)*, Alexandria, VA, USA (Virtual Conference), 2021, pp. 1777–1796. SIAM.

[3] M. Cygan, M. Mucha, K. Węgrzycki, and M. Wlodarczyk. On problems equivalent to (min, +)-convolution. *ACM Trans. Algorithms*, 15(1):14:1–14:25, 2019.

[4] F. Eisenbrand and R. Weismantel. Proximity results and faster algorithms for integer programming using the steinitz lemma. *ACM Trans. Algorithms*, 16(1):5:1–5:14, 2020.

[5] Z. Galil and O. Margalit. An almost linear-time algorithm for the dense subset-sum problem. In *Automata, Languages and Programming, 18th International Colloquium, ICALP91, Madrid, Spain, July 8-12, 1991, Proceedings*, 1991, LNCS 510, pp. 719–727. Springer.

[6] H. Kellerer and U. Pferschy. Improved dynamic programming in connection with an FPTAS for the knapsack problem. *J. Comb. Optim.*, 8(1):5–11, 2004.

• [7] A. Polak, L. Rohwedder, and K. Węgrzycki. Knapsack and subset sum with small items. In N. Bansal, E. Merelli, and J. Worrell, eds., *48th International Colloquium on Automata, Languages, and Programming (ICALP 2021)*, Glasgow, UK (Virtual Conference), 2021, Leibniz International Proceedings in Informatics 198, Article 106. Schloss Dagstuhl.

### Improving Schroeppel and Shamir's Algorithm for Subset Sum

*Investigators: Karol Węgrzycki in cooperation with Jesper Nederlof (Utrecht University)*

In the Subset Sum problem we are given a set of $n$ integers and a target integer $t$, and the task is to find a subset of these numbers that sum to precisely $t$. In [3] we present an $\mathcal{O}^*(2^{0.5n})$ time and $\mathcal{O}^*(2^{0.249999n})$ space randomized algorithm for solving worst-case Subset Sum instances. This is the first improvement over the long-standing $\mathcal{O}^*(2^{n/2})$-time and $\mathcal{O}^*(2^{n/4})$-space algorithm due to Schroeppel and Shamir [4].

We breach this gap in two steps: (1) We present a space efficient reduction to the Orthogonal Vectors Problem (ov), one of the most central problem in fine-grained complexity theory. The reduction is established via an intricate combination of the method of Schroeppel and Shamir, and the representation technique introduced by Howgrave-Graham and Joux [2] for designing Subset Sum algorithms for average case instances. (2) We provide an algorithm for ov that detects an orthogonal pair among $N$ given vectors in $\{0,1\}^d$ with support size $d/4$ in time $\widetilde{\mathcal{O}}(N \cdot 2^d / \binom{d}{d/4})$. Our algorithm for ov is based on and refines the representative families framework developed by Fomin, Lokshtanov, Panolan and Saurabh [1].

Our reduction uncovers a curious tight relation between Subset Sum and ov, because any improvement of our algorithm for ov would imply an improvement over the running time of Schroeppel and Shamir, which is also a long standing open problem.

### References


[1] F. V. Fomin, D. Lokshtanov, F. Panolan, and S. Saurabh. Efficient Computation of Representative Families with Applications in Parameterized and Exact Algorithms. *J. ACM*, 63(4):29:1–29:60, 2016.

[2] N. Howgrave-Graham and A. Joux. New Generic Algorithms for Hard Knapsacks. In *Advances in Cryptology - EUROCRYPT 2010, 29th Annual International Conference on the Theory and Applications of Cryptographic Techniques. Proceedings*, 2010, pp. 235–256.

• [3] J. Nederlof and K. Węgrzycki. Improving Schroeppel and Shamir's algorithm for subset sum via orthogonal vectors. In S. Khuller and V. Vassilevska Williams, eds., *STOC '21, 53rd Annual ACM SIGACT Symposium on Theory of Computing*, Virtual, Italy, 2021, pp. 1670–1683. ACM.

[4] R. Schroeppel and A. Shamir. A T=$O(2^{n/2})$, S=$O(2^{n/4})$ Algorithm for Certain NP-Complete Problems. *SIAM J. Comput.*, 10(3):456–464, 1981.


### Polynomial Witnesses for Coverability in 2-VASS

*Investigators: Karol Węgrzycki in cooperation with Filip Mazowiecki (University of Warsaw) and Henry Sinclair-Banks (University of Warwick)*

Coverability in Petri nets finds applications in verification of safety properties of reactive systems. We study coverability in an equivalent model known as Vector Addition Systems with States (vass). A $k$-vass can be seen as $k$ counters and a finite automaton whose transitions are labeled with $k$ integers. Counter values are updated by adding the respective transition labels. A configuration in this system consists of a state and $k$ counter values. Importantly, the counters are never allowed to take negative values. The coverability problem

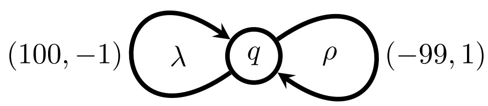



Figure 27.1: Example 2-VASS with one unary counter $V$. Consider the instance of coverability consisting of $V$, the initial configuration $q(0,1)$, and the target configuration $q(0,10)$. Consider the path $\pi = \lambda\rho\ \lambda\rho \cdots \lambda\rho\ \rho \cdots \rho$ which induces a run in $V$ from the initial configuration $q(0,1)$. There are 990 repetitions of the pair of cycles $\lambda\rho$ to witness the configuration $q(990,1)$. The cycles alternate so both counters remain non-negative throughout the run. This is followed by 10 iterations of the cycle $\rho$ so the configuration $q(0,11)$ is witnessed, achieving coverability of the target configuration $q(0,10)$.

asks whether one can traverse the $k$-VASS from the initial configuration to a configuration with counter values bounded from below by the target.

In a well-established line of work on $k$-VASS, it is known that coverability in 2-VASS is PSPACE-hard when the integer updates are encoded in binary. This lower bound limits the practicality of applications, so it is natural to focus on restrictions. In [1] we initiate the study of 2-VASS with one unary counter. Here, one counter receives binary encoded updates and the other receives unary encoded updates. We prove that that coverability in 2-VASS with one unary counter is in NP. This improves upon the inherited state-of-the-art upper bound of being contained in PSPACE. Our main technical contribution is that one only needs to consider runs in a certain compressed linear form.

### References


- [1] F. Mazowiecki, H. Sinclair-Banks, and K. Węgrzycki. Coverability in 2-VASS with one unary counter is in NP. In O. Kupferman and P. Sobocinski, eds., *Foundations of Software Science and Computation Structures (FoSSaCS 2023)*, Paris, France, 2023, LNCS 13992, pp. 196–217. Springer.


## 27.6.2 Foundations of Fine-Grained Complexity Theory

### Hardness of Detecting Subgraphs of Bounded Treewidth

*Investigators: Karl Bringmann in cooperation with Jasper Slusallek (Saarland University)*

In Graph Pattern Detection, we wish to decide whether a fixed pattern graph $H$ appears as a subgraph of a large host graph $G$ with $n$ vertices. We examine this problem when the pattern graph $H$ has bounded treewidth $\mathrm{tw}(H)$, as occurs in a variety of applications. This problem has a well-known algorithm via color-coding that runs in time $\mathcal{O}(n^{\mathrm{tw}(H)+1})$ [1]. Improved algorithms are known for some patterns, but do they exist for all pattern graphs?

The only known lower bound rules out time $n^{o(\mathrm{tw}(H)/\log(\mathrm{tw}(H)))}$ for any class of patterns of unbounded treewidth assuming the Exponential-Time Hypothesis [3].

In [2], we demonstrate the existence of maximally hard pattern graphs $H$ that require time $\mathcal{O}(n^{\mathrm{tw}(H)+1})$. Specifically, under the Strong Exponential-Time Hypothesis, a standard conjecture from fine-grained complexity theory, we prove the following: for any $\varepsilon > 0$ there exists $t \geq 3$ and a pattern graph $H$ of treewidth $t$ such that detecting pattern graph $H$ has no algorithm running in time $\mathcal{O}(n^{t+1-\varepsilon})$. Under the more recent 3-Uniform Hyperclique Hypothesis, we even obtain tight lower bounds for each specific treewidth $t \geq 3$.

In addition to these main results, we explore (1) colored and uncolored problem variants (and why they are equivalent for most cases), (2) treewidth 1 and 2, (3) graph pattern detection parameterized by pathwidth instead of treewidth, and (4) weighted problem variants. For many of these settings we obtain similarly tight upper and lower bounds.

## References


[1] N. Alon, R. Yuster, and U. Zwick. Color-coding. *J. ACM*, 42(4):844–856, 1995.

• [2] K. Bringmann and J. Slusallek. Current algorithms for detecting subgraphs of bounded treewidth are probably optimal. In N. Bansal, E. Merelli, and J. Worrell, eds., *48th International Colloquium on Automata, Languages, and Programming (ICALP 2021)*, Glasgow, UK (Virtual Conference), 2021, Leibniz International Proceedings in Informatics 198, Article 40. Schloss Dagstuhl.

[3] D. Marx. Can you beat treewidth? *Theory Comput.*, 6(1):85–112, 2010.


### Tight Fine-Grained Bounds for Direct Access on Join Queries

*Investigators: Karl Bringmann in cooperation with Stefan Mengel (Université d'Artois, CNRS) and Nofar Carmeli (Université Montpellier, Inria, CNRS)*

Building on the success of fine-grained complexity theory in determining the optimal time complexity of many graph problems, a next step is to study problems on hypergraphs. Such problems arise naturally in database theory in the form of conjunctive queries on relational databases. Therefore, in cooperation with experts on database theory, we have explored fine-grained lower bounds for database problems.

Specifically, in [1] we considered the task of lexicographic direct access to query answers. That is, we want to simulate an array containing the answers of a join query on a relational database sorted in a lexicographic order chosen by the user. A recent dichotomy showed for which queries and orders this task can be done in polylogarithmic access time after quasilinear preprocessing [2], but this dichotomy does not tell us how much time is required in the cases classified as hard. We determine the preprocessing time needed to achieve polylogarithmic access time for all join queries and all lexicographical orders. To this end, we propose a decomposition-based general algorithm for direct access on join queries. We then explore its optimality by proving lower bounds for the preprocessing time based on the hardness of a certain online Set-Disjointness problem, which shows that our algorithm's bounds are tight for all lexicographic orders on join queries. Then, we prove the hardness of Set-Disjointness based on the Zero-Clique Hypothesis which is an established conjecture from fine-grained complexity theory. This way, for every lexicographic direct access problem we obtain tight bounds on the preprocessing time needed to achieve polylogarithmic access time.

## References


- [1] K. Bringmann, N. Carmeli, and S. Mengel. Tight fine-grained bounds for direct access on join queries. In L. Libkin, P. Barceló, and A. Grez, eds., *PODS '22, 41st ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems*, Philadelphia, PA, USA, 2022, pp. 427–436. ACM.
- [2] N. Carmeli, N. Tziavelis, W. Gatterbauer, B. Kimelfeld, and M. Riedewald. Tractable orders for direct access to ranked answers of conjunctive queries. In *PODS*, 2021, pp. 325–341. ACM.


### Fine-Grained Classification of Optimization Problems in P

*Investigators: Karl Bringmann, Alejandro Cassis, Nick Fischer, and Marvin Künnemann*

Inspired by the first completeness results for decision problems in P [3], we systematically studied completeness and classification theorems for optimization problems in the polynomial-time regime. In particular, we define polynomial-time analogues of the classic classes MaxSNP and MinSNP, which we call MaxSP and MinSP. These new classes contain a number of natural optimization problems in P, including Maximum and Minimum Inner Product, general forms of Nearest and Furthest Neighbor Search, and optimization variants of the $k$-Xor problem. Specifically, we define MaxSP as the class of problems definable as

$$\max_{x_1,\ldots,x_k} \#\ (y_1,\ldots,y_\ell) : \phi(x_1,\ldots,x_k,y_1,\ldots,y_\ell)\ ,$$

where $\phi$ is a quantifier-free first-order property over a given relational structure. That is, one can think of the input as a relational database and $\phi$ as a first-order database query. MinSP is defined similarly.

In [1], we establish the following completeness result: if the Maximum Inner Product problem can be solved faster than the brute-force time $\mathcal{O}(m^{k+\ell-1})$ (where $m$ is the size of the given database), then all problems in MaxSP can be solved faster than this running time. Thus, Maximum Inner Product is *complete* for the class MaxSP. Similar statements hold for approximation algorithms for Maximum Inner Product, and for the minimization variants.

In [2], we set out to understand the *non-complete* problems. Here, we restrict ourselves to graph problems, that is, problems for which the database consists of a single binary table and can therefore be seen as a graph with $m$ edges. We study the approximability of these problems in great detail and obtain a classification into four categories:

- the problem is optimizable to exactness in better-than-brute-force time, or
- there is (only) a $(1+\varepsilon)$-approximation in better-than-brute-force time, or
- there is (only) a fixed constant-factor approximation better-than-brute-force time, or
- there is (only) an $m^\varepsilon$-factor approximation in better-than-brute-force time.

In particular, none of these problems has an optimal approximation factor of, for instance, $\log m$.

## References


- [1] K. Bringmann, A. Cassis, N. Fischer, and M. Künnemann. Fine-grained completeness for optimization in P. In M. Wootters and L. Sanità, eds., *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2021)*, Seattle, WA, USA, 2021, Leibniz International Proceedings in Informatics 207, Article 9. Schloss Dagstuhl.
- [2] K. Bringmann, A. Cassis, N. Fischer, and M. Künnemann. A structural investigation of the approximability of polynomial-time problems. In M. Bojańczyk, E. Merelli, and D. P. Woodruff, eds., *49th EATCS International Conference on Automata, Languages, and Programming (ICALP 2022)*, Paris, France, 2022, Leibniz International Proceedings in Informatics 229, Article 30. Schloss Dagstuhl.

[3] J. Gao, R. Impagliazzo, A. Kolokolova, and R. R. Williams. Completeness for first-order properties on sparse structures with algorithmic applications. In P. N. Klein, ed., *Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2017, Barcelona, Spain, Hotel Porta Fira, January 16-19*, 2017, pp. 2162–2181. SIAM.


### Fine-Grained Complexity of Multidimensional Ordering Properties

*Investigators: Marvin Künnemann in cooperation with Mohit Gurumukhani (Cornell University), Russell Impagliazzo and Maria Paula Parga Nina (UC San Diego), Haozhe An (University of Maryland), and Michael Jaber (University of Texas at Austin)*

Continuing our line of research toward building a *structural* fine-grained complexity theory, we define a class of polynomial-time problems based on multidimensional ordering properties. This class captures a large variety of tasks where the input consists of an $n$-sized set of $d$-dimensional vectors and the problem is first-order definable using comparisons between coordinates. Examples include interesting generalizations of orthogonal range searching, model-checking first-order properties on certain geometric intersection graphs, and elementary questions on multidimensional data such as verifying Pareto optimality of a choice of data points.

In [1, 2], we give a baseline algorithm that solves any $(k+1)$-quantifier $d$-dimensional problem in this class in time $\mathcal{O}(n^k \log^{d-1} n)$. We also prove that this algorithm is optimal up to subpolynomial factors by showing that for every $k \geq 2$ there is a $(k+1)$-quantifier $3k$-dimensional problem in this class requiring time $n^{k-o(1)}$ assuming the 3-uniform Hyperclique Hypothesis.

Toward identifying a single representative problem for this class, we study the existence of complete problems, in the sense that polynomial-factor improvements for such a complete problem would improve the baseline algorithm for all problems in this class. Specifically, we define a problem Vector Concatenated Non-Domination (vcnd) and determine it as the essentially unique candidate to be complete for this class, under certain fine-grained assumptions.

## References


- [1] H. An, M. Gurumukhani, R. Impagliazzo, M. Jaber, M. Künnemann, and M. P. Parga Nina. The fine-grained complexity of multi-dimensional ordering properties. In P. A. Golovach and M. Zehavi, eds., *16th International Symposium on Parameterized and Exact Computation (IPEC*

*2021)*, Lisbon, Portugal, 2021, Leibniz International Proceedings in Informatics 214, Article 3. Schloss Dagstuhl.

- [2] H. An, M. Gurumukhani, R. Impagliazzo, M. Jaber, M. Künnemann, and M. P. Parga Nina. The fine-grained complexity of multi-dimensional ordering properties. *Algorithmica*, 84:3156–3191, 2022.

### 27.6.3 Fine-Grained Complexity in Computational Geometry

#### Polygon Placement Revisited

*Investigators: André Nusser and Marvin Künnemann*

In [2], using the lens of fine-grained complexity theory we revisit a classic geometric problem: Given simple polygons $P$ and $Q$, determine the largest copy of $P$ that can be placed into $Q$. Algorithmic results for various versions of this problem have been obtained since the early 1980s, generally leading to rather high polynomial running times. In contrast, the only previous conditional lower bound due to Barequet and Har-Peled [1] gives a 3Sum-based lower bound of $n^{2-o(1)}$ for (convex) polygons $P$ and $Q$ consisting of $\Theta(n)$ vertices each. This leaves open whether we can establish (1) hardness beyond quadratic time and (2) any superlinear bound for constant-sized $P$ or $Q$.

Among other results, we affirmatively answer these questions under the $k$-Sum Hypothesis, by proving natural hardness results that increase with each degree of freedom (scaling, $x$-translation, $y$-translation, rotation): When allowing scaling and $x$-translations, we prove a 3Sum-based lower bound of $n^{2-o(1)}$ already for orthogonal polygons $P$ and $Q$ with $O(1)$ and $n$ vertices, respectively. If additionally $y$-translations are permitted, we prove the same lower bound under the 4Sum hypothesis. Interestingly, exploiting an offline dynamic algorithm due to Overmars and Yap, we give an improved algorithm for this setting that matches our lower bound up to an $n^{1/2+o(1)}$-factor. Finally, when scaling, translations *and* rotations are permitted, we prove a 5Sum-based lower bound of $n^{3-o(1)}$, giving the first hardness result beyond quadratic time for the polygon placement problem.

To the best of our knowledge, this gives the first natural (degrees of freedom + 1)-Sum hardness for a geometric optimization problem.

#### References

[1] G. Barequet and S. Har-Peled. Polygon containment and translational min-Hausdorff-distance between segment sets are 3SUM-hard. *Int. J. Comput. Geom. Appl.*, 11(4):465–474, 2001.

- [2] M. Künnemann and A. Nusser. Polygon placement revisited: (degree of freedom + 1)-SUM hardness and an improvement via offline dynamic rectangle union. In S. Naor and N. Buchbinder, eds., *Proceedings of the Annual ACM-SIAM Symposium on Discrete Algorithms (SODA 2022)*, Virtual, 2022, pp. 3181–3201. SIAM.

### On Diameter Computation in Geometric Intersection Graphs

*Investigators: Karl Bringmann, Sándor Kisfaludi-Bak, Marvin Künnemann, André Nusser, and Zahra Parsaeian*

The diameter of a graph is the largest distance between any two vertices. In the context of communication networks, the diameter is the maximum number of links that information must traverse in order to reach its destination. Computing the diameter of a graph allows to estimate the time required for information to spread throughout the network, which can be important in various applications such as network design and analysis, network security, etc.

In a seminal result, Cabello [2] showed that the diameter of planar graphs can be computed in subquadratic time. In [1], we ask whether this result can be generalized to geometric intersection graphs. For several classes of intersection graphs the state of the art are near-quadratic-time algorithms [3].

Our results are mostly negative, showing that subquadratic algorithms are unlikely to exist for several classes of ihntersection graphs. Specifically, under standard hypotheses from fine-grained complexity, we prove that computing the diameter of intersection graphs of line segments, unit line segments, congruent equilateral triangles, unit ball graphs in $\mathbb{R}^3$, axis-parallel unit segments, and hypercube graphs in $\mathbb{R}^{12}$ all require at least quadratic time. Many of these results hold even restricted to graphs with a small constant diameter. In several cases our lower bounds match or almost match the near-quadratic state of the art [3].

As a positive result we design a near-linear-time algorithm for deciding whether the diameter of a unit square graph (that is, intersection graph of axis-parallel unit squares) has diameter at most 2.

### References


- [1] K. Bringmann, S. Kisfaludi-Bak, M. Künnemann, A. Nusser, and Z. Parsaeian. Towards subquadratic diameter computation in geometric intersection graphs. In X. Goaoc and M. Kerber, eds., *38th International Symposium on Computational Geometry (SoCG 2022)*, Berlin, Germany, 2022, Leibniz International Proceedings in Informatics 224, Article 21. Schloss Dagstuhl.

[2] S. Cabello. Subquadratic algorithms for the diameter and the sum of pairwise distances in planar graphs. *ACM Trans. Algorithms*, 15(2):21:1–21:38, 2019.

[3] T. M. Chan and D. Skrepetos. All-pairs shortest paths in geometric intersection graphs. *J. Comput. Geom.*, 10:27–41, 2019.


### Hardness of Hausdorff Distance Under Translation

*Investigators: Karl Bringmann and André Nusser*

Computing the similarity of two point sets is a ubiquitous task in medical imaging, geometric shape comparison, trajectory analysis, and many more settings. Arguably the most basic distance measure for this task is the Hausdorff distance, which assigns to each point from one set the closest point in the other set and then evaluates the maximum distance of any assigned pair. A drawback is that this distance measure is not translational invariant, that is, comparing two objects just according to their shape while disregarding their position in space is impossible.

Fortunately, there is a canonical translational invariant version, the Hausdorff distance under translation, which minimizes the Hausdorff distance over all translations of one of the point sets. For point sets of size $n$ and $m$, the Hausdorff distance under translation can be computed in time $\widetilde{\mathcal{O}}(nm)$ for the $L_1$ and $L_\infty$ norm [3] and $\widetilde{\mathcal{O}}(nm(n+m))$ for the $L_2$ norm [4].

As these bounds have not been improved for over 25 years, in [1, 2] we approach the Hausdorff distance under translation from the perspective of fine-grained complexity theory. We show (1) a matching lower bound of $(nm)^{1-o(1)}$ for $L_1$ and $L_\infty$ (and all other $L_p$ norms) assuming the Orthogonal Vectors Hypothesis and (2) a matching lower bound of $n^{2-o(1)}$ for $L_2$ in the imbalanced case of $m = \mathcal{O}(1)$ assuming the 3Sum Hypothesis.

## References


- [1] K. Bringmann and A. Nusser. Translating Hausdorff is hard: Fine-grained lower bounds for Hausdorff distance under translation. In K. Buchin and É. Colin de Verdière, eds., *37th International Symposium on Computational Geometry (SoCG 2021)*, Buffalo, NY, USA (Virtual Conference), 2021, Leibniz International Proceedings in Informatics 189, Article 18. Schloss Dagstuhl.
- [2] K. Bringmann and A. Nusser. Translating Hausdorff is hard: Fine-grained lower bounds for Hausdorff distance under translation. *Journal of Computational Geometry*, 13(2):30–50, 2021.

[3] L. P. Chew and K. Kedem. Improvements on geometric pattern matching problems. In O. Nurmi and E. Ukkonen, eds., *Algorithm Theory - SWAT '92, Third Scandinavian Workshop on Algorithm Theory, Helsinki, Finland, July 8-10, 1992, Proceedings*, 1992, LNCS 621, pp. 318–325. Springer.

[4] D. P. Huttenlocher, K. Kedem, and M. Sharir. The upper envelope of voronoi surfaces and its applications. *Discret. Comput. Geom.*, 9:267–291, 1993.


## 27.6.4 Curve Similarity Measures in Computational Geometry

### Approximate Near Neighbor Searching for Time Series

*Investigators: Karl Bringmann and André Nusser in cooperation with Anne Driemel and Ioannis Psarros (University of Bonn)*

In [1] we study the $c$-APPROXIMATE NEAR NEIGHBOR PROBLEM under the continuous Fréchet distance. Given a set of $n$ polygonal curves with $m$ vertices, a radius $\delta > 0$, and a parameter $k \leq m$, we want to preprocess the curves into a data structure that, given a query curve $Q$ with $k$ vertices, either returns an input curve with Fréchet distance at most $c \cdot \delta$ to $Q$, or returns that there exists no input curve with Fréchet distance at most $\delta$ to $Q$. We focus on the case where the input and the queries are one-dimensional polygonal curves—also called time series—and we give a comprehensive analysis for this case. We obtain new upper bounds that provide different trade-offs between approximation factor, preprocessing time, and query time.

Our data structures improve upon the state of the art in several ways. We show that for any $0 < \varepsilon \leq 1$ an approximation factor of $1 + \varepsilon$ can be achieved within the same asymptotic time bounds as the previously best result for $2+\varepsilon$. Moreover, we show that an approximation factor of $2 + \varepsilon$ can be obtained by using preprocessing time and space $\mathcal{O}(nm)$, which is linear in the input size, and query time in $\mathcal{O}(1/\varepsilon)^{k+2}$, where the previously best result used preprocessing time $n \cdot \mathcal{O}(m/\varepsilon k)^k$ and query time $\mathcal{O}(1)^k$.

We complement our new data structures with matching conditional lower bounds based on the Orthogonal Vectors Hypothesis. Interestingly, some of our lower bounds already hold for any super-constant value of $k$. This is achieved by proving hardness of a one-sided sparse version of the ORTHOGONAL VECTORS problem as an intermediate problem, which we believe to be of independent interest.

## References


- [1] K. Bringmann, A. Driemel, A. Nusser, and I. Psarros. Tight bounds for approximate near neighbor searching for time series under the Fréchet distance. In S. Naor and N. Buchbinder, eds., *Proceedings of the Annual ACM-SIAM Symposium on Discrete Algorithms (SODA 2022)*, Virtual, 2022, pp. 517–550. SIAM.


### Approximating Dynamic Time Warping Under Translation

*Investigators: Karl Bringmann, Sándor Kisfaludi-Bak, Marvin Künnemann, and André Nusser in cooperation with Dániel Marx (CISPA)*

The Dynamic Time Warping (DTW) distance is a popular measure of similarity for a variety of sequence data. For comparing polygonal curves $\pi, \sigma$ in $\mathbb{R}^d$, it provides a robust, outlier-insensitive alternative to the Fréchet distance. However, like the Fréchet distance, the DTW distance is not invariant under translations. Can we efficiently optimize the DTW distance of $\pi$ and $\sigma$ under arbitrary translations, to compare the *shape* of the curves irrespective of their absolute location?

There are surprisingly few works in this direction, which may be due to its computational intricacy: For the Euclidean norm, this problem contains as a special case the geometric median problem, which provably admits no exact algebraic algorithm (that is, no algorithm using only addition, multiplication, and $k$-th roots).

In [1], we thus investigate exact algorithms for non-Euclidean norms as well as approximation algorithms for the Euclidean norm. For the $L_1$ norm in $\mathbb{R}^d$, we present an $\mathcal{O}(n^{2(d+1)})$-time algorithm for curves with $n$ vertices, that is, we provide an exact polynomial-time algorithm for constant dimension $d$. For the Euclidean norm in $\mathbb{R}^d$ with fixed dimension $d$, we show that a simple problem-specific insight leads to a $(1+\varepsilon)$-approximation algorithm running in time $\mathcal{O}(n^3/\varepsilon^d)$. We then show how to obtain a subcubic $\widetilde{\mathcal{O}}(n^{2.5}/\varepsilon^d)$-time algorithm using novel ideas. This time bound comes close to the well-known quadratic-time barrier for computing DTW for a fixed translation. Technically, the algorithm is obtained by speeding up repeated DTW distance estimations using a dynamic data structure for maintaining shortest paths in weighted planar digraphs. Crucially, we show how to traverse a candidate set of translations using space-filling curves in a way that incurs only few updates to the data structure.

## References


- [1] K. Bringmann, S. Kisfaludi-Bak, M. Künnemann, D. Marx, and A. Nusser. Dynamic time warping under translation: Approximation guided by space-filling curves. In X. Goaoc and M. Kerber, eds., *38th International Symposium on Computational Geometry (SoCG 2022)*, Berlin, Germany, 2022, Leibniz International Proceedings in Informatics 224, Article 20. Schloss Dagstuhl.

#### Continuous Dynamic Time Warping of Time Series

*Investigators: André Nusser in cooperation with Kevin Buchin (Technical University of Dortmund) and Sampson Wong (University of Sydney)*

Dynamic Time Warping is arguably the most popular similarity measure for time series (that is, one-dimensional polygonal curves). The drawback of Dynamic Time Warping is that it is sensitive to the sampling rate of the time series. The Fréchet distance is an alternative that has gained popularity, however, its drawback is that it is sensitive to outliers.

Continuous Dynamic Time Warping (OV) is a recently proposed alternative that does not exhibit the aforementioned drawbacks. OV combines the continuous nature of the Fréchet distance with the summation of Dynamic Time Warping, resulting in a similarity measure that is robust to sampling rate and to outliers. In experimental work of Brankovic et al. [1], it was demonstrated that clustering under OV avoids the unwanted artifacts that appear when clustering under Dynamic Time Warping or Fréchet distance. Despite its advantages, the major shortcoming of OV is that there is no exact algorithm for computing OV, in polynomial time or otherwise.

In [2], we present the first exact algorithm for computing OV of time series. Our algorithm runs in time $\mathcal{O}(n^5)$ on a pair of time series consisting of $n$ vertices. In our algorithm, we propagate continuous functions along the dynamic program for OV. The main difficulty lies in bounding the complexity of the functions. We believe that our result is an important first step towards OV becoming a practical similarity measure of curves.

### References

[1] M. Brankovic, K. Buchin, K. Klaren, A. Nusser, A. Popov, and S. Wong. (k, l)-medians clustering of trajectories using continuous dynamic time warpin. In C.-T. Lu, F. Wag, G. Trajcevski, Y. Huang, S. Newsam, and L. Xiong, eds., *28th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (ACM SIGSPATIAL GIS 2020)*, Seattle, WA, USA (Online), 2020, pp. 99–110. ACM.

• [2] K. Buchin, A. Nusser, and S. Wong. Computing continuous dynamic time warping of time series in polynomial time. In X. Goaoc and M. Kerber, eds., *38th International Symposium on Computational Geometry (SoCG 2022)*, Berlin, Germany, 2022, Leibniz International Proceedings in Informatics 224, Article 22. Schloss Dagstuhl.

#### On the VC Dimension of Fréchet and Hausdorff Distance

*Investigators: André Nusser in cooperation with Ioannis Psarros (National and Kapodistrian University of Athens), Anne Driemel (University of Bonn), and Jeff M. Phillips (University of Utah)*

The Vapnik-Chervonenkis (VC) dimension provides a notion of complexity for set systems. If the VC dimension is small, then this drastically simplifies fundamental computational tasks such as classification, range counting, and density estimation through the use of sampling bounds. In [1], we analyze set systems where the ground set is a set of polygonal curves and the ranges are metric balls defined by curve similarity metrics, such as the Fréchet distance and the Hausdorff distance, as well as their discrete counterparts. We prove upper and lower bounds on the VC dimension that imply useful sampling bounds in the setting that the number of curves is large, but the complexity of the individual curves is small. Our upper and lower bounds are either near-quadratic or near-linear in the complexity of the curves that define the ranges and they are logarithmic in the complexity of the curves that define the ground set.

## References


• [1] A. Driemel, A. Nusser, J. M. Phillips, and I. Psarros. The VC dimension of metric balls under Fréchet and Hausdorff distances. *Discrete & Computational Geometry*, 2021.


## 27.6.5 Exponential-Time Algorithms in Computational Geometry

### Approximation Schemes for Euclidean TSP

*Investigators: Sándor Kisfaludi-Bak and Karol Węgrzycki in cooperation with Jesper Nederlof (Utrecht University)*

In [2] we revisit the classic task of finding the shortest traveling salesperson tour of $n$ points in $d$-dimensional Euclidean space, for any fixed constant $d \geq 2$. We determine the optimal dependence on $\varepsilon$ in the running time of an algorithm that computes a $(1+\varepsilon)$-approximate tour, under a plausible assumption. Specifically, we give an algorithm that runs in time $2^{\mathcal{O}(1/\varepsilon^{d-1})} n \log n$. This improves the previously smallest dependence on $\varepsilon$ in the running time $(1/\varepsilon)^{\mathcal{O}(1/\varepsilon^{d-1})} n \log n$ of the algorithm by Rao and Smith [3]. We also show that a $2^{o(1/\varepsilon^{d-1})}\mathrm{poly}(n)$-time algorithm would violate the Gap-Exponential-Time Hypothesis (Gap-ETH).

Our new algorithm builds upon the celebrated quadtree-based methods initially proposed by Arora [1]. We add a new idea that we call *sparsity-sensitive patching*, which allows us to set the granularity with which we simplify the tour depending on its local sparsity. We demonstrate that our technique extends to other problems, by designing algorithms with the same running time for Steiner Tree and Rectilinear Steiner Tree, and we prove a matching Gap-ETH-based lower bound for Rectilinear Steiner Tree.

## References


[1] S. Arora. Polynomial Time Approximation Schemes for Euclidean Traveling Salesman and other Geometric Problems. *J. ACM*, 45(5):753–782, 1998.

• [2] S. Kisfaludi-Bak, J. Nederlof, and K. Węgrzycki. A Gap-ETH-tight approximation scheme for Euclidean TSP. In *FOCS 2021, IEEE 62nd Annual Symposium on Foundations of Computer Science*, Denver, CO, USA (Virtual Conference), 2022, pp. 351–362. IEEE.

[3] S. Rao and W. D. Smith. Approximating Geometrical Graphs via “Spanners” and “Banyans”. In *Proceedings of the Thirtieth Annual ACM Symposium on the Theory of Computing (STOC 1998)*, 1998, pp. 540–550. ACM.

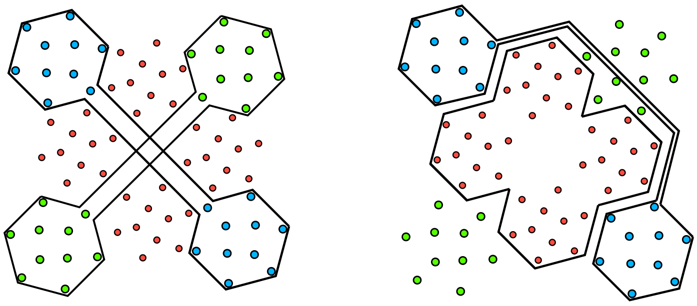


Figure 27.2: Illustration of the RED-BLUE-GREEN SEPARATION problem. The solution on the left is not feasible because the two polygons are crossing. The solution on the right is feasible because polygons are noncrossing and points in different color classes are separated.

### Noncrossing TSP and Red-Blue-Green Separation

*Investigators: Karol Węgrzycki in cooperation with François Dross (Université Bordeaux, CNRS), Krzysztof Fleszar, and Anna Zych-Pawlewicz (University of Warsaw)*

In [2] we study the following problems of connecting or separating classes of points via noncrossing structures.

In the BICOLORED NONCROSSING TRAVELING SALESPERSON problem we are given red and blue points in the Euclidean plane. The goal is to find a tour connecting all red points and a tour connecting all blue points with the additional constraint that these tours cannot cross.

In the RED-BLUE-GREEN SEPARATION problem we are given a set of point in the Euclidean plane. Each point is assigned one of three colors. The goal is to find two noncrossing polygons of minimum total length that separate points of different colors.

On the algorithmic side, we give a Gap-ETH-tight EPTAS for the BICOLORED NONCROSSING TRAVELLING SALESPERSON problem as well as for the RED-BLUE-GREEN SEPARATION problem. This improves the work of Arora and Chang [1] who gave a slower PTAS for the simpler RED-BLUE SEPARATION PROBLEM. These results are based on a new patching procedure that might be of independent interest.

On the negative side, we show that the problem of connecting terminal pairs with noncrossing paths is NP-hard in the Euclidean plane, and that the problem of finding two noncrossing spanning trees is NP-hard on plane graphs.

## References

[1] S. Arora and K. L. Chang. Approximation schemes for degree-restricted MST and red-blue separation problems. *Algorithmica*, 40(3):189–210, 2004.

• [2] F. Dross, K. Fleszar, K. Węgrzycki, and A. Zych-Pawlewicz. Gap-ETH-tight approximation schemes for red-green-blue separation and bicolored noncrossing Euclidean travelling salesman tours. In N. Bansal and V. Nagarajan, eds., *Proceedings of the 2023 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA 2023)*, Florence, Italy, 2023, pp. 1433–1463. SIAM.

### Clique-Based Separators for Geometric Intersection Graphs

*Investigators: Sándor Kisfaludi-Bak in cooperation with Mark de Berg, Morteza Monemizadeh, and Leonidas Theocharous (TU Eindhoven)*

Separator theorems have been instrumental in the design of algorithms for several graph classes, in particular they have been used to design efficient divide-and-conquer algorithms for various NP-hard graph problems. In case of intersection graphs, classical separators do not exist due to the possibility of large cliques. To design subexponential algorithms for problems such as Independent Set on intersection graphs, however, one can also work with a separator consisting of a small number of cliques instead of a small number of nodes. Such *clique-based separators* were introduced recently by de Berg et al. [1]. Formally, a clique-based separator of a graph $G$ is a collection $\mathcal{S}$ of vertex-disjoint cliques whose union is a balanced separator of $G$. The *weight* of $\mathcal{S}$ is defined as $\text{weight}(\mathcal{S}) := \sum_{C\in\mathcal{S}} \log(|C|+1)$. De Berg et al. [1] proved that the intersection graph of a set of $n$ convex fat objects in the plane admits a clique-based separator of weight $\mathcal{O}(\sqrt{n})$, and they used this to obtain subexponential algorithms for many classic NP-hard problems on such graphs, which are optimal assuming the Exponential-Time Hypothesis (ETH).

The goal of this work [2] was to investigate whether similar results are possible for non-fat objects in the plane. Note that not all intersection graphs admit clique-based separators of small weight, for instance, those that have large complete bipartite graphs as induced subgraphs.

We show the existence of sublinear-weight clique-based separators for several graph classes: map graphs (that is, the intersection graphs of interior-disjoint regions in the plane), intersection graphs of pseudo-disks, geodesic disk graphs of a polygonal region, and so-called visibility-restricted unit-disk graphs. As applications, we provide subexponential algorithms for Independent Set, Feedback Vertex Set and $q$-Coloring for constant $q$ on all of these graph classes, always matching or even slightly improving the state-of-the-art algorithms.

## References

[1] M. de Berg, H. L. Bodlaender, S. Kisfaludi-Bak, D. Marx, and T. C. van der Zanden. A framework for exponential-time-hypothesis-tight algorithms and lower bounds in geometric intersection graphs. *SIAM Journal on Computing*, 49(6):1291–1331, 2020.

• [2] M. de Berg, S. Kisfaludi-Bak, M. Monemizadeh, and L. Theocharous. Clique-based separators for geometric intersection graphs. In H.-K. Ahn and K. Sadakane, eds., *32nd International Symposium on Algorithms and Computation (ISAAC 2021)*, Fukuoka, Japan, 2021, Leibniz International Proceedings in Informatics 212, Article 22. Schloss Dagstuhl.

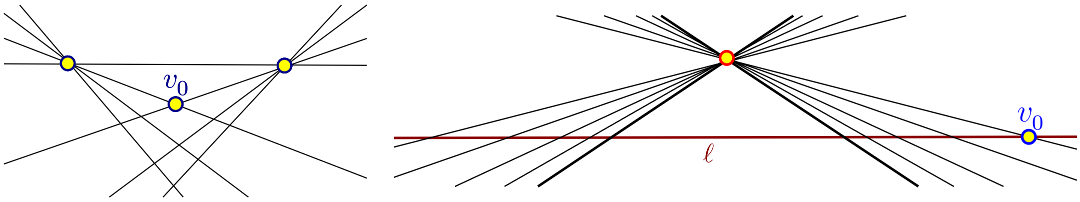



Figure 27.3: (left) A set $L$ of lines for which the vertex of $\mathcal{A}(\mathcal{L})$ of maximum level, which is $v_0$, does not lie on the upper envelope. (right) A more "substantial" construction, in which the horizontal line $\ell$ contains all but one of the vertices of $\mathcal{A}(\mathcal{L})$, all at upper level $\Theta(n)$. The maximum-level vertex is $v_0$ (as well as its symmetric counterpart on the other side of $\ell$).

### 27.6.6 Miscellaneous Topics in Computational Geometry

#### The Maximum-Level Vertex in an Arrangement of Lines

*Investigators: Kurt Mehlhorn in cooperation with Eunjin Oh (POSTECH, Korea) Dan Halperin, Micha Sharir (Tel Aviv University), and Sariel Har-Peled (University of Illinois)*

Let $L$ denote a set of $n$ lines in the plane, not necessarily in general position. The goal is to compute all the vertices of the arrangement $\mathcal{A}(\mathcal{L})$ of maximum level, where the level of a vertex $v$ is the number of lines of $L$ that pass strictly below $v$. The problem is posed as Exercise 8.13 in the computational geometry book by de Berg et al. [1] and the reader is supposed to find an $\mathcal{O}(n \log n)$ algorithm. If the lines are in general position, the problem is easily solved in time $\mathcal{O}(n \log n)$ as the desired vertices are exactly the vertices on the upper envelope. However, if the lines are not in general position, these vertices might not be on the upper envelope of the lines as shown in Figure 27.3, and the problem seems to be much harder.

In [2] we present an efficient algorithm for finding all the vertices of the arrangement $\mathcal{A}(\mathcal{L})$ of maximum level. We first assume that all lines of $L$ are distinct, and distinguish between two cases, depending on whether or not the upper envelope of $L$ contains a bounded edge. In the former case, we show that the number of lines of $L$ that pass *above* any maximum level vertex $v_0$ is only $\mathcal{O}(\log n)$. In the latter case, we establish a similar property that holds after we remove some of the lines that are incident to the single vertex of the upper envelope. We present algorithms that run, in both cases, in optimal time $\mathcal{O}(n \log n)$.

We then consider the case where the lines of $L$ are not necessarily distinct. This setup is more challenging, and the best we have is an algorithm that computes all maximum-level vertices in time $\mathcal{O}(n^{4/3} \log^3 n)$.

### References

[1] M. .de Berg, M. van Kreveld, M. Overmars, and O. Schwarzkopf. *Computational Geometry*. Springer, 1991.

• [2] D. Halperin, S. Har-Peled, K. Mehlhorn, E. Oh, and M. Sharir. The maximum-level vertex in an arrangement of lines. *Discrete & Computational Geometry*, 67:439–461, 2022.

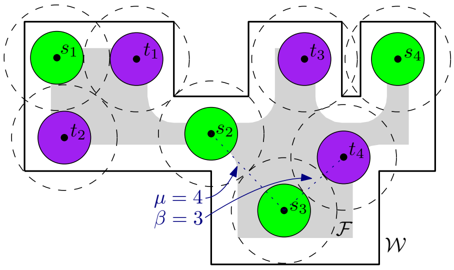



Figure 27.4: An example instance of the robot motion planning problem. The workspace $\mathcal{W}$ is the rectilinear polygon, the free space $\mathcal{F}$ is the inner gray area. The unit-disc robots satisfy the separation $\mu = 4$ for start-start pairs and for target-target pairs and the separation $\beta = 3$ for start-target pairs.

### Unlabeled Multi-Robot Motion Planning

*Investigators: Karl Bringmann in cooperation with Stijn Slot (Adyen), Bahareh Banyassady (FU Berlin), Dan Halperin (Tel Aviv University), Kevin Buchin (TU Dortmund), Mark de Berg and Irina Kostitsyna (TU Eindhoven), Yoshio Okamoto (UEC Japan), and Henning Fernau (University of Trier),*

We consider the Unlabeled Motion-Planning Problem of $m$ unit-disc robots moving in a simple polygonal workspace of $n$ edges. The goal is to find a motion plan that moves the robots to a given set of $m$ target positions. For the unlabeled variant, it does not matter which robot reaches which target position as long as all target positions are occupied in the end. If the workspace has narrow passages such that the robots cannot fit through them, then the free configuration space, representing all possible unobstructed positions of a single robot, consists of multiple connected components. Even if in each component of the free space the number of targets matches the number of start positions, the motion-planning problem does not always have a solution when the robots and their targets are positioned very densely. It is an open problem whether this problem is polynomial-time solvable or NP-hard.

In [2], we prove tight bounds on how much separation between start and target positions is necessary to always guarantee a solution, improving the bounds by Adler et al. [1]. Specifically, we prove that the following separation is sufficient: any two start positions are at least distance 4 apart, any two target positions are at least distance 4 apart, and any pair of a start and a target positions is at least distance 3 apart. Moreover, we describe an algorithm that always finds a solution in time $\mathcal{O}(n \log n + mn + m^2)$ if the separation bounds are met.

### References

[1] A. Adler, M. de Berg, D. Halperin, and K. Solovey. Efficient multi-robot motion planning for unlabeled discs in simple polygons. *IEEE Trans Autom. Sci. Eng.*, 12(4):1309–1317, 2015.

• [2] B. Banyassady, M. de Berg, K. Bringmann, K. Buchin, H. Fernau, D. Halperin, I. Kostitsyna, Y. Okamoto, and S. Slot. Unlabeled multi-robot motion planning with tighter separation bounds. In X. Goaoc and M. Kerber, eds., *38th International Symposium on Computational Geometry (SoCG 2022)*, Berlin, Germany, 2022, Leibniz International Proceedings in Informatics 224, Article 12. Schloss Dagstuhl.

## 27.7 Algorithms on Strings

Texts (or strings) are one of the fundamental objects studied in Theoretical Computer Science. Algorithms on strings have a direct and indirect impact on countless areas of our daily lives: from simple tools such as `grep` or `awk`, to spell checkers, to bioinformatics—every day, billions of people use string algorithms in some form or another.

In the reporting period, our group has developed into one of the leading groups working on fundamental problems on strings that appeal to a general theory audience. We obtained breakthrough results and substantially improved algorithms for approximate pattern matching problems and for computing string similarity measures; we discuss these areas in more detail below. Further, we obtained results on the compressibility of repetitive sequences, as well as faster algorithms for constructing compressed suffix arrays.

**Approximate Pattern Matching.** Perhaps the most natural task on strings is to find one pattern string in another text string as a substring—the classical string (or *pattern*) matching problem. For many real-world applications, we need to allow for some errors in the pattern or text: think of spelling mistakes, noisy data, or transmission errors, to name just a few examples. Hence, the classical *Approximate* Pattern Matching problem emerged, where the task is to find all substrings that are *similar* to the pattern. We studied different notions of similarity, namely the Hamming distance (the number of mismatches) and the edit or *Levenshtein* distance (the number of insertions, deletions, and substitutions to transform one string into the other).

In the reporting period, we greatly enhanced the understanding of Approximate Pattern Matching and closely related problems. When similarity to the pattern is taken to be the edit distance to the pattern, we obtained the first significantly improved algorithm since the seminal result by Cole and Hariharan from about 20 years ago. Further, we investigated Approximate *Circular* Pattern matching (for both Hamming distance and edit distance), where we also considered any cyclic rotation of the pattern to be a match. Again, we significantly improved upon the existing algorithms.

For more details on these results, consult Section 27.7.1.

**Edit Distance and Other String Similarity Measures.** In the last reporting period, we also deeply investigated algorithms for computing the edit distance and related string similarity measures between two strings. Using classical textbook algorithms, we can compute the edit distance in quadratic time—fine-grained complexity shattered any hopes for a significantly faster solution. As quadratic-time algorithms, often even linear-time algorithms, are too slow for practical applications, the focus shifted to approximating edit distance. In the last reporting period, we significantly moved the field forward by giving almost-optimal *sublinear-time* algorithms for approximating the edit distance between two strings.

We also studied extensions of the edit distance, including weighted edit distance (where the cost of an edit depends on the characters involved), tree edit distance (where we compare rooted ordered trees rather than strings), and Dyck edit distance (where the goal is to balance a sequence of parentheses of multiple types). For weighted edit distance and tree edit distance, we developed the first algorithms with a running time of the form $\widetilde{\mathcal{O}}(n + \text{poly}(k))$, where $n$ is the input size and $k$ is the computed edit distance. For Dyck edit distance, we contributed constant-factor approximation algorithms.

Further, we considered a spectrum of similarity measures that interpolate between the Hamming distance and the edit distance. Our results include an exact algorithm whose runtime is conditionally optimal (as a function of the input size, the distance, and the interpolation parameter) and a sublinear-time approximation algorithm.

Finally, we also considered dynamic data structures for the edit distance (among other problems) and we also gave linear-time approximation algorithms for the longest common subsequence of two strings, a measure the can be considered to be the dual of the edit distance.

For more details on these results, consult Section 27.7.2.

## 27.7.1 Approximate Pattern Matching

*Investigators: Tomasz Kociumaka and Philip Wellnitz, in cooperation with Panagiotis Charalampopoulos (Birkbeck, University of London), Solon P. Pissis, Wiktor Zuba (CWI, Amsterdam), Jakub Radoszewski, Wojciech Rytter, and Toamsz Waleń (University of Warsaw)*

Approximate pattern matching is a natural and well-studied problem on strings: Given a text $T$ of length $n$, a pattern $P$ of length $m$, and a threshold $k$, find (the starting positions of) all fragments of $T$ that are at distance at most $k$ from $P$. Under the Hamming distance, we search for fragments of $T$ that have at most $k$ mismatches with $P$, whereas under the edit distance, we search for fragments of $T$ that can be transformed to $P$ with at most $k$ edits. In the previous reporting periods [1, 4], our group obtained much insight into the structure of the $k$-mismatch and $k$-edit occurrences of $P$ in $T$, resulting in efficient *universal* algorithms applicable not only in the standard setting (where the strings are given explicitly) but also the fully-compressed setting (where strings are given as straight-line programs) and the dynamic setting (where we maintain a collection of strings under creation, splitting, and concatenation).

Unfortunately, these results did not bring any improvements for the standard setting: since the late 1990s, the state-of-the-art running time for pattern matching with $k$ edits remained at $\mathcal{O}(n + k^4 \cdot n/m)$ [6] (for small values of $k$) and $\mathcal{O}(nk)$ [7] (for large values of $k$). In [5], we presented an $\widetilde{\mathcal{O}}(n + k^{3.5} \cdot n/m)$-time algorithm, thus breaking through this long-standing barrier. For $m^{0.251} \leq k \leq m^{0.399}$, our solution is polynomially faster than all previous algorithms. Moreover, it is universal and, in particular, applicable to the fully-compressed and dynamic setting. To obtain our result, we observed that the bottleneck of our previous $\widetilde{\mathcal{O}}(n + k^4 \cdot n/m)$-time solution [4] is when the text and the pattern are (almost) periodic. Our new algorithm reduces this case to a new *dynamic puzzle matching* problem, which we solve by building on tools developed by Tiskin [8] for the so-called seaweed monoid of permutation matrices, which corresponds to the $(\min, +)$-product of unit-Monge matrices.

In [3], we investigated the complexity of approximate *circular* pattern matching, where we are looking for approximate occurrences (with up to $k$ mismatches or edits) of any *cyclic* *rotation* of the pattern $P$. All previous results for approximate circular pattern matching were either average-case upper bounds or heuristics, with the sole exception of [2], where an $\mathcal{O}(n + k^4 \cdot n/m)$-time algorithm for circular pattern matching with $k$ mismatches was provided. We improve upon that result and develop an $\widetilde{\mathcal{O}}(n + k^3 \cdot n/m)$-time procedure reporting (the starting positions of) all the circular occurrences (with up to $k$ mismatches) and an $\widetilde{\mathcal{O}}(n+k^2 \cdot n/m)$-time procedure deciding if at least one such occurrence exists. Notably, the running time of our decision algorithm is $\mathcal{O}(n)$ if $k = \mathcal{O}(\sqrt{m/\log m})$; in comparison, the fastest algorithm for (non-circular) pattern matching with $k$ mismatches runs in $\mathcal{O}(n)$ time only for $k = \mathcal{O}(\sqrt{m})$. Our algorithms for circular pattern matching with mismatches rely on a reduction to a geometric problem building on ideas from [4, 2].

As for circular pattern matching with $k$ edits, we give an $\mathcal{O}(nk^2)$-time algorithm for the reporting version and an $\widetilde{\mathcal{O}}(nk)$-time algorithm for the decision version. The latter complexity nearly matches the $\mathcal{O}(nk)$ runtime of the Landau–Vishkin algorithm [7], which remains the fastest known for (non-circular) pattern matching with $k = \Omega(m^{0.4})$ edits.

## References

[1] K. Bringmann, M. Künnemann, and P. Wellnitz. Few matches or almost periodicity: Faster pattern matching with mismatches in compressed texts. In T. M. Chan, ed., *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA 2019)*, San Diego, CA, USA, 2019, pp. 1126–1145. SIAM.

[2] P. Charalampopoulos, T. Kociumaka, S. P. Pissis, J. Radoszewski, W. Rytter, J. Straszyński, T. Waleń, and W. Zuba. Circular pattern matching with $k$ mismatches. *J. Comput. Syst. Sci.*, 115:73–85, 2021.

• [3] P. Charalampopoulos, T. Kociumaka, S. P. Pissis, J. Radoszewski, W. Rytter, T. Waleń, and W. Zuba. Approximate circular pattern matching. In S. Chechik, G. Navarro, E. Rotenberg, and G. Herman, eds., *30th Annual European Symposium on Algorithms (ESA 2022)*, Berlin/Potsdam, Germany, 2022, Leibniz International Proceedings in Informatics 244, Article 35. Schloss Dagstuhl.

[4] P. Charalampopoulos, T. Kociumaka, and P. Wellnitz. Faster approximate pattern matching: A unified approach. In *FOCS 2020, 61st Annual IEEE Symposium on Foundations of Computer Science*, Durham, NC, USA (Virtual Conference), 2020, pp. 978–989. IEEE.

• [5] P. Charalampopoulos, T. Kociumaka, and P. Wellnitz. Faster pattern matching under edit distance: A reduction to dynamic puzzle matching and the seaweed monoid of permutation matrices. In *FOCS 2022, IEEE 63rd Annual Symposium on Foundations of Computer Science*, Denver, CO, USA, 2022, pp. 698–707. IEEE.

[6] R. Cole and R. Hariharan. Approximate string matching: A simpler faster algorithm. *SIAM J. Comput.*, 31(6):1761–1782, 2002.

[7] G. M. Landau and U. Vishkin. Fast parallel and serial approximate string matching. *J. Algorithms*, 10(2):157–169, 1989.

[8] A. Tiskin. Fast distance multiplication of unit-monge matrices. *Algorithmica*, 71(4):859–888, 2015.

### 27.7.2 Edit Distance and Other String Similarity Measures

#### Almost-Optimal Sublinear-Time Algorithms for Approximating Edit Distance

*Investigators: Karl Bringmann, Alejandro Cassis, Nick Fischer, and Tomasz Kociumaka, in cooperation with Elazar Goldenberg (The Academic College of Tel Aviv-Yaffo), Vasileios Nakos (RelationalAI), Barna Saha (University of California, San Diego), and Robert Krauthgamer (Weizmann Institute of Science),*

The problem studied in the regime of sublinear-time edit distance approximation is the $(k, K)$-Gap Edit Distance problem, where we are given oracle access to two strings of length at most $n$ and the task is to distinguish whether their edit distance is at most $k$ or at least $K$.

Many efficient sublinear-time algorithms are *non-adaptive*, which means that they issue all queries to the access oracle in a single batch. In [5], we resolve the non-adaptive query complexity of Gap Edit Distance for $K = k^{1+\Omega(1)}$. Specifically, we design a non-adaptive algorithm with query complexity $\widetilde{\mathcal{O}}(nk^{0.5}/K)$, and we further prove that this bound is optimal up to polylogarithmic factors. Our algorithm also achieves optimal time complexity $\widetilde{\mathcal{O}}(nk^{0.5}/K)$ whenever $K = \Omega(k^{1.5})$. For $K = \mathcal{O}(k^{1.5})$, the running time of our algorithm is $\widetilde{\mathcal{O}}(nk^2/K^2)$. In the restricted case of $K = \Omega(n)$, our time complexity matches a known result of Batu et al. [2]; in all other (nontrivial) cases, it is strictly better compared to all previous algorithms, including the adaptive ones.

The most natural question stemming from [5] is whether adaptivity can be used to circumvent the $\tilde{\Omega}(nk^{0.5}/K)$ lower bound. In [3], we answer this question positively and develop an almost-optimal algorithm for the low-distance regime. Specifically, we show that in time $\mathcal{O}(n/k + k^{4+o(1)})$ we can even solve the $(k, k^{1+o(1)})$-Gap Edit Distance problem. This is the first sublinear-time algorithm to achieve a subpolynomial gap, and it is almost optimal in the following sense: In the low distance regime ($k \leq n^{0.19}$), our running time becomes $\mathcal{O}(n/k)$, which essentially matches a known $n/k^{1+o(1)}$ lower bound for the $(k, k^{1+o(1)})$-Gap Edit Distance problem up to lower order factors.

Our result also reveals a surprising similarity between Hamming distance and edit distance in the low distance regime: For both, the $(k, k^{1+o(1)})$-Gap Edit Distance problem has time complexity $n/k^{1\pm o(1)}$ for small $k$.

Our algorithm builds upon the algorithm of Andoni, Krauthgamer, and Onak [1], which approximates the edit distance in almost-linear time $\mathcal{O}(n^{1+\varepsilon})$ within a polylogarithmic factor. We first simplify their approach and then show how to effectively prune their computation tree in order to obtain a sublinear-time algorithm in the given time bound. To do this, we use a variety of structural insights on the (local and global) patterns that can emerge during this process and design appropriate property testers to effectively detect these patterns.

In follow-up work [4], we study a variant of this problem in a model that includes preprocessing. In this model, we can preprocess one or both of the strings (separately) in nearly linear time before solving the $(k, K)$-Gap Edit Distance problem. This extra information can then be used to solve the gap problem, with the hope of doing so faster. This model was introduced and studied by Goldenberg, Rubinstein, and Saha [6], motivated by applications where we need to estimate the edit distance of many long strings against each other.

We achieve the following results:

- After preprocessing *one* string in time $n^{1+o(1)}$, we can solve $(k, k \cdot n^{o(1)})$-Gap Edit Distance in time $(n/k + k) \cdot n^{o(1)}$.
- After preprocessing *both* strings separately in time $n^{1+o(1)}$, we can solve $(k, k \cdot n^{o(1)})$-Gap Edit Distance in time $k \cdot n^{o(1)}$.

Both results improve upon the state of the art with respect to either the gap, the query time, or the preprocessing time.

Our algorithms are based on our sublinear-time algorithm mentioned earlier [3]. By taking advantage of the preprocessing, we replace many complex components in that algorithm with faster and simpler solutions.

## References

[1] A. Andoni, R. Krauthgamer, and K. Onak. Polylogarithmic approximation for edit distance and the asymmetric query complexity. In *51th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2010, October 23-26, 2010, Las Vegas, Nevada, USA*, 2010, pp. 377–386. IEEE Computer Society.

[2] T. Batu, F. Ergün, J. Kilian, A. Magen, S. Raskhodnikova, R. Rubinfeld, and R. Sami. A sublinear algorithm for weakly approximating edit distance. In L. L. Larmore and M. X. Goemans, eds., *Proceedings of the 35th Annual ACM Symposium on Theory of Computing, June 9-11, 2003, San Diego, CA, USA*, 2003, pp. 316–324. ACM.

• [3] K. Bringmann, A. Cassis, N. Fischer, and V. Nakos. Almost-optimal sublinear-time edit distance in the low distance regime. In S. Leonardi and A. Gupta, eds., *STOC '22, 54th Annual ACM Symposium on Theory of Computing*, Rome, Italy, 2022, pp. 1102–1115. ACM.

• [4] K. Bringmann, A. Cassis, N. Fischer, and V. Nakos. Improved sublinear-time edit distance for preprocessed strings. In M. Bojańczyk, E. Merelli, and D. P. Woodruff, eds., *49th EATCS International Conference on Automata, Languages, and Programming (ICALP 2022)*, Paris, France, 2022, Leibniz International Proceedings in Informatics 229, Article 32. Schloss Dagstuhl.

• [5] E. Goldenberg, T. Kociumaka, R. Krauthgamer, and B. Saha. Gap edit distance via non-adaptive queries: Simple and optimal. In *FOCS 2022, IEEE 63rd Annual Symposium on Foundations of Computer Science*, Denver, CO, USA, 2022, pp. 674–685. IEEE.

[6] E. Goldenberg, A. Rubinstein, and B. Saha. Does preprocessing help in fast sequence comparisons? In K. Makarychev, Y. Makarychev, M. Tulsiani, G. Kamath, and J. Chuzhoy, eds., *Proccedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing, STOC 2020, Chicago, IL, USA, June 22-26, 2020*, 2020, pp. 657–670. ACM.

### Dynamic Data Structures for Bounded Edit Distance and Other Parameterized String Problems

*Investigators: Karol Węgrzycki, in cooperation with Jędrzej Olkowski, Michał Pilipczuk, Mateusz Rychlicki and Anna Zych-Pawlewicz (University of Warsaw)*

In [3], we study a dynamic version of the edit distance (among other problems). In particular, we study the bounded version, where we want to decide whether two strings are at distance at most $k$ or not. We give randomized data structures with the following operations:

- *Initialize* the data structure on two strings of length $\mathcal{O}(n)$ and a fixed $k$.
- *Update* any one of the strings by changing a single character.
- *Query* whether the current strings are at edit distance at most $k$.

We observe that from a result of Frandsen et al. [2], one can easily infer a meta-theorem that provides dynamic data structures for parameterized string problems with worst-case update time of the form $\mathcal{O}(f(k) \cdot \log\log n)$, where $k$ is the parameter in question (that is, for instance a bound on the edit distance) and $n$ is the length of the strings. Showcasing the utility of this meta-theorem, we give a first version of the aforementioned data structure for Bounded Edit Distance among others.

We then also give another data structure for Bounded Edit Distance with explicit running time bounds (in terms of $k$): we achieve an initialization time of $\mathcal{O}(kn)$, an update time of $\mathcal{O}(k^2 \log\log n)$, and a query time of $\mathcal{O}(1)$.

We also discuss how a lower bound methodology introduced by Amarilli et al. [1] can be used to show that obtaining update time $\mathcal{O}(f(k))$ (that is, without any dependency on $n$) for Bounded Edit Distance is unlikely already for a constant value of the parameter $k$.

We also consider the Closest String problem. Here, we are given an integer $d$ and strings $s_1, \ldots, s_n$ from some alphabet $\Sigma$ that have the same length $L$, and the task is to decide whether there is a word in $\Sigma^L$ that is at Hamming distance at most $d$ to each string $s_i$. Closest String is a classical problem that has received special focus in the area of parameterized complexity theory, as Closest String has many natural parameterizations: we may parameterize by $n$, $d$, $L$, or $\Sigma$.

We design randomized data structures (with the above operations) for Closest String with initialization time $2^{\mathcal{O}(d)} \cdot nL \, \Sigma^{1+o(1)}$, with amortized update time $2^{\mathcal{O}(d)}$, and with worst-case query time min $d^{\mathcal{O}(d)}, \Sigma^{\mathcal{O}(d)}$. We obtain said data structures by combining known static approaches to Closest String with color-coding.

## References

[1] A. Amarilli, L. Jachiet, and C. Paperman. Dynamic membership for regular languages. In *Proceedings of the 48th International Colloquium on Automata, Languages, and Programming, ICALP 2021*, 2021, LIPIcs 198, pp. 116:1–116:17. Schloss Dagstuhl - Leibniz-Zentrum für Informatik.

[2] G. S. Frandsen, P. B. Miltersen, and S. Skyum. Dynamic word problems. *J. ACM*, 44(2):257–271, 1997.

[3] J. Olkowski, M. Pilipczuk, M. Rychlicki, K. Węgrzycki, and A. Zych-Pawlewicz. Dynamic data structures for parameterized string problems. In P. Berenbrink, P. Bouyer, A. Dawar, and M. M. Kanté, eds., *40th International Symposium on Theoretical Aspects of Computer Science (STACS 2023)*, Hamburg, Germany, 2023, Leibniz International Proceedings in Informatics 254, Article 50. Schloss Dagstuhl.

### Linear-Time Algorithms for Weighted Edit Distance and Tree Edit Distance in the Low-Distance Regime

*Investigators: Tomasz Kociumaka, in cooperation with Debarati Das (Pennsylvania State University), Barna Saha (University of California, San Diego), Jacob Gilbert, MohammadTaghi Hajiaghayi, and Hamed Saleh (University of Maryland)*

Given two strings of length $n$ and an upper bound $k$ on their edit distance, the algorithms of Myers [7] and Landau–Vishkin [5] from the 1980s compute the edit distance in $\mathcal{O}(n+k^2)$ time. By now, this result is known to be optimal (under the Orthogonal Vectors Hypothesis), and it has been generalized in many ways, including an $\mathcal{O}(n+k^{4.55})$-time algorithm [4] computing the Dyck edit distance (the distance to the nearest balanced sequence of parentheses).

In contrast, no $\widetilde{\mathcal{O}}(n+\text{poly}(k))$-time algorithm was known for tree edit distance, which is defined for two ordered rooted (and possibly labeled) trees as the minimum number of node insertions, deletions, and relabels needed to transform one tree into the other. In this context, a node deletion removes a node and links its children directly to its parent, preserving their order; a node insertion allows us to select a consecutive set of siblings and bring them under a new node which appears at the previous position of the relocated nodes; a node relabel modifies the label of an existing node. The state-of-the-art running time for tree edit distance was $\widetilde{\mathcal{O}}(nk^2)$ for small $k$ [1] and $\mathcal{O}(n^{2.96})$ for large $k$ [6]. In [3], we develop a randomized $\widetilde{\mathcal{O}}(n+k^{15})$-time algorithm, improving upon the result of [1] in the low-distance regime.

While the classic Levenshtein distance is theoretically fundamental, most real-world applications utilize weighted edit distance, where different weights are assigned to different edit operations (insertions, deletions, and substitutions), and the weights may vary with the characters being edited. Formally, a weight function $w$ assigns a non-negative real weight to any two elements of the set $\Sigma \quad \lambda$ (consisting of the alphabet plus the 'empty' placeholder) so that $w(a,a)=0$ and $w(a,b)\geq 1$ for all $a,b \quad \Sigma \quad \lambda$ with $a = b$. The weighted edit distance is defined as a minimum-weight sequence of edits transforming one string into the other.

Except for the vanilla $\mathcal{O}(n^2)$-time dynamic-programming algorithm and its almost trivial $\mathcal{O}(nk)$-time variant [7], none of the aforementioned developments on the unweighted edit distance applies to the weighted variant. In [2], we propose the first $\widetilde{\mathcal{O}}(n+\text{poly}(k))$-time algorithm that computes weighted string edit distance exactly, thus bridging a fundamental decades-old gap between our understanding of unweighted and weighted edit distance. Bringing in several new techniques, we then generalize our deterministic $\mathcal{O}(n+k^5)$-time procedure for weighted string edit distance to an $\mathcal{O}(n+k^7\log k)$-time algorithm for weighted tree edit distance and an $\mathcal{O}(n+k^{12})$-time solution for weighted Dyck edit distance. Notably, the former significantly improves upon our earlier result for unweighted tree edit distance and remains the state of the art even for unlabeled trees.

#### References

[1] S. Akmal and C. Jin. Faster algorithms for bounded tree edit distance. In N. Bansal, E. Merelli, and J. Worrell, eds., *48th International Colloquium on Automata, Languages, and Programming, ICALP 2021, July 12-16, 2021, Glasgow, Scotland (Virtual Conference)*, 2021, LIPIcs 198, pp. 12:1–12:15. Schloss Dagstuhl - Leibniz-Zentrum für Informatik.

- [2] D. Das, J. Gilbert, M. Hajiaghayi, T. Kociumaka, and B. Saha. Weighted edit distance computation: Strings, trees, and Dyck. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing (STOC 2023)*, Orlando, FL, USA, 2023. ACM. Accepted.
- [3] D. Das, J. Gilbert, M. Hajiaghayi, T. Kociumaka, B. Saha, and H. Saleh. $\tilde{O}(n + \text{poly}(k))$-time algorithm for bounded tree edit distance. In *FOCS 2022, IEEE 63rd Annual Symposium on Foundations of Computer Science*, Denver, CO, USA, 2022, pp. 686–697. IEEE.

[4] D. Fried, S. Golan, T. Kociumaka, T. Kopelowitz, E. Porat, and T. Starikovskaya. An improved algorithm for the k-dyck edit distance problem. In J. S. Naor and N. Buchbinder, eds., *Proceedings of the 2022 ACM-SIAM Symposium on Discrete Algorithms, SODA 2022, Virtual Conference / Alexandria, VA, USA, January 9 - 12, 2022*, 2022, pp. 3650–3669. SIAM.

[5] G. M. Landau and U. Vishkin. Fast string matching with k differences. *J. Comput. Syst. Sci.*, 37(1):63–78, 1988.

[6] X. Mao. Breaking the cubic barrier for (unweighted) tree edit distance. In *62nd IEEE Annual Symposium on Foundations of Computer Science, FOCS 2021, Denver, CO, USA, February 7-10, 2022*, 2021, pp. 792–803. IEEE.

[7] E. W. Myers. An O(ND) difference algorithm and its variations. *Algorithmica*, 1(2):251–266, 1986.


### Approximation Algorithms for Dyck Edit Distance and RNA Folding

*Investigators: Tomasz Kociumaka, in cooperation with Debarati Das (Pennsylvania State University) and Barna Saha (University of California, San Diego)*

The Dyck language, which consists of well-balanced sequences of parentheses, is one of the most fundamental context-free languages. The Dyck edit distance quantifies the number of edits (character insertions, deletions, and substitutions) required to make a given length-$n$ parenthesis sequence well-balanced. The RNA FOLDING problem is a similar problem, where a closing parenthesis can match an opening parenthesis of the same type irrespective of their ordering. For example, in RNA FOLDING, both `()` and `)(` are valid matches, whereas the Dyck language only allows `()` as a match. Both of these problems have been studied extensively in the literature. Using fast matrix multiplication, it is possible to compute their exact solutions in time $\mathcal{O}(n^{2.687})$ [1], and a $(1+\varepsilon)$-multiplicative approximation is known with a running time of $\Omega(n^{2.372})$.

The impracticality of fast matrix multiplication often makes combinatorial algorithms much more desirable. Unfortunately, it is known that the problems of computing the Dyck edit distance and the folding distance exactly are at least as hard as Boolean matrix multiplication. Thereby, they are unlikely to admit truly subcubic-time combinatorial algorithms. In terms of fast approximation algorithms that are combinatorial in nature, the state of the art for both problems includes an $\varepsilon n$-additive approximation in $\widetilde{\mathcal{O}}(\varepsilon^{-1}n^2)$ time [4]. Moreover, Dyck edit distance admits an $\mathcal{O}(\log n)$-factor near-linear-time approximation algorithm [3].

In [2], we make substantial improvements for Dyck edit distance (with any number of parenthesis types). We design a constant-factor approximation algorithm that runs in $\mathcal{O}(n^{1.971})$ time (the first constant-factor approximation in subquadratic time). Moreover, we develop a $(1+\varepsilon)$-factor approximation algorithm running in $\widetilde{\mathcal{O}}(\varepsilon^{-1}n^2)$ time, which improves upon the earlier additive approximation. Finally, we design a $(3+\varepsilon)$-approximation that takes $\widetilde{\mathcal{O}}(\varepsilon^{-1}nk)$ time, where $k \geq 1$ is an upper bound on the sought distance.

As for Rna Folding, for parameter any $s \geq 1$, we design a factor-$s$ approximation algorithm that runs in $\mathcal{O}(n+(n/s)^3)$ time. To the best of our knowledge, this is the first nontrivial approximation algorithm for Rna Folding that can go below the $n^2$ barrier. All our algorithms are combinatorial in nature.

### References


[1] S. Chi, R. Duan, T. Xie, and T. Zhang. Faster min-plus product for monotone instances. In S. Leonardi and A. Gupta, eds., *STOC '22: 54th Annual ACM SIGACT Symposium on Theory of Computing, Rome, Italy, June 20 - 24, 2022*, 2022, pp. 1529–1542. ACM.

[2] D. Das, T. Kociumaka, and B. Saha. Improved approximation algorithms for Dyck edit distance and RNA folding. In M. Bojańczyk, E. Merelli, and D. P. Woodruff, eds., *49th EATCS International Conference on Automata, Languages, and Programming (ICALP 2022)*, Paris, France, 2022, Leibniz International Proceedings in Informatics 229, Article 49. Schloss Dagstuhl.

[3] B. Saha. The Dyck language edit distance problem in near-linear time. In *55th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2014, Philadelphia, PA, USA, October 18-21, 2014*, 2014, pp. 611–620. IEEE Computer Society.

[4] B. Saha. Fast & space-efficient approximations of language edit distance and RNA folding: An amnesic dynamic programming approach. In C. Umans, ed., *58th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2017, Berkeley, CA, USA, October 15-17, 2017*, 2017, pp. 295–306. IEEE Computer Society.


#### An Algorithmic Bridge Between Hamming and Edit Distances

*Investigators: Tomasz Kociumaka, in cooperation with Elazar Goldenberg (Academic College of Tel Aviv-Yafo), Barna Saha (University of California, San Diego), and Robert Krauthgamer (Weizmann Institute of Science)*

The edit distance between strings classically assigns the same cost to every character insertion, deletion, and substitution, whereas the Hamming distance only allows substitutions. In many real-life scenarios, insertions and deletions (abbreviated *indels*) appear frequently but significantly less so than substitutions. To model this, we consider substitutions being cheaper than indels, with cost $1/a$ for an integer parameter $a \geq 1$. This basic variant, denoted $\mathtt{ED}_a$, bridges classical edit distance ($a = 1$) with Hamming distance ($a$ ). So far, $\mathtt{ED}_a$ has only been studied for $a = \Theta(1)$, where its complexity matches that of the edit distance [1]. Interpreting $a$ as a parameter leads to interesting algorithmic challenges: Does the time complexity of computing $\mathtt{ED}_a$ interpolate between that of Hamming distance (linear time) and edit distance (quadratic time)? What about approximating $\mathtt{ED}_a$?

In [2], we first present a simple deterministic $\mathcal{O}(n+\min(nk, ak^2))$-time exact algorithm for $\mathtt{ED}_a$ and further prove that this runtime is near-optimal (as a function of the string length $n$, the parameter $a$, and an upper bound $k$ on $\mathtt{ED}_a$) assuming the Orthogonal Vectors Hypothesis. Our main result is a randomized algorithm computing a $(1+\varepsilon)$-approximation of $\mathtt{ED}_a$. For simplicity, let us focus on $k \geq 1$ and a constant $\varepsilon > 0$; then, our algorithm takes $\widetilde{\mathcal{O}}((n/a) + ak^3)$ time. Unless $a = \widetilde{\mathcal{O}}(1)$, in which case $\mathtt{ED}_a$ resembles the standard edit distance, and for the most interesting low-distance regime, this running time is sublinear in $n$.

We also consider a very natural version that asks to find a $(k_I, k_S)$-*alignment*, that is, an alignment with at most $k_I$ indels and $k_S$ substitutions. In this setting, we give an exact algorithm and, more importantly, an $\widetilde{\mathcal{O}}((nk_I/k_S) + k_S k_I^3)$-time $(1, 1+\varepsilon)$-bicriteria approximation. The latter solution is based on the techniques we develop for $\mathtt{ED}_a$ for $a = \Theta(k_S/k_I)$, and its running time is again sublinear in $n$ whenever $k_I \ll k_S$ and the overall distance is small.

These bounds are in stark contrast to unit-cost edit distance, where state-of-the-art algorithms are far from achieving $(1+\varepsilon)$-approximation in sublinear time, even for a favorable choice of $k$.

## References


[1] K. Bringmann and M. Künnemann. Quadratic conditional lower bounds for string problems and dynamic time warping. In *FOCS 2015, IEEE 56th Annual Symposium on Foundations of Computer Science*, Berkeley, CA, USA, 2015, pp. 79–97. IEEE.

[2] E. Goldenberg, T. Kociumaka, R. Krauthgamer, and B. Saha. An algorithmic bridge between Hamming and Levenshtein distances. In Y. Tauman Kalai, ed., *14th Innovations in Theoretical Computer Science Conference (ITCS 2023)*, Cambridge, MA, USA, 2023, Leibniz International Proceedings in Informatics, Article 58. Schloss Dagstuhl.


### Approximating Longest Common Subsequence in Linear Time

*Investigators: Karl Bringmann in cooperation with Vincent Cohen-Addad (Sorbonne Université, CNRS) and Debarati Das (University of Copenhagen, BARC)*

The longest common subsequence (LCS) of two strings $X$ and $Y$ is the longest string $Z$ that appears as a subsequence of $X$ and of $Y$. The length of the LCS is a measure of similarity that is in some sense complementary to the classic edit distance. Similar as for edit distance, it is known that the 40-year-old quadratic-time algorithm for computing an LCS is near-optimal unless the Strong Exponential Time Hypothesis fails [3].

Therefore, in parallel to the long line of work on approximating edit distance discussed earlier in this section, recent work approached the problem of approximating LCS in linear time: A simple linear-time algorithm yields an approximation ratio of $\mathcal{O}(\sqrt{n})$. This bound was broken by Hajiaghayi et al. [4], who improved the approximation ratio to $\mathcal{O}(n^{0.498})$. In [2, 1], we contributed to this line of research by optimizing the previous approach, obtaining an approximation ratio of $\widetilde{\mathcal{O}}(n^{0.4})$. Our approximation ratio was further improved in [5] to $n^{o(1)}$.

## References


[1] K. Bringmann, V. Cohen-Addad, and D. Das. A linear-time $n^{0.4}$-approximation for longest common subsequence. *ACM Transactions on Algorithms*, 19(1), Article 9, 2023.

[2] K. Bringmann and D. Das. A linear-time $n^{0.4}$-approximation for longest common subsequence. In N. Bansal, E. Merelli, and J. Worrell, eds., *48th International Colloquium on Automata, Languages, and Programming (ICALP 2021)*, Glasgow, UK (Virtual Conference), 2021, Leibniz International Proceedings in Informatics 198, Article 39. Schloss Dagstuhl.

[3] K. Bringmann and M. Künnemann. Quadratic conditional lower bounds for string problems and dynamic time warping. In *FOCS 2015, IEEE 56th Annual Symposium on Foundations of Computer Science*, Berkeley, CA, USA, 2015, pp. 79–97. IEEE.

[4] M. Hajiaghayi, M. Seddighin, S. Seddighin, and X. Sun. Approximating LCS in linear time: Beating the $\sqrt{n}$ barrier. In *SODA*, 2019, pp. 1181–1200. SIAM.

[5] N. S. Nosatzki. Approximating the longest common subsequence problem within a sub-polynomial factor in linear time. *CoRR*, arxiv abs/2112.08454, 2021.

### 27.7.3 Towards a Definitive Compressibility Measure for Repetitive Sequences

*Investigators: Tomasz Kociumaka, in cooperation with Nicola Prezza (Ca' Foscari University of Venice), Gonzalo Navarro, and Francisco Olivares (University of Chile)*

While the $k$-th order empirical entropy is an accepted measure of the compressibility of individual sequences on classical text collections, it is useful only for small values of $k$ and thus fails to capture the compressibility of globally repetitive sequences. In the absence of an established way of quantifying the latter, ad-hoc measures like the size $z$ of the Lempel–Ziv parse [5] are frequently used to estimate repetitiveness. Recently, a more principled measure, the size $\gamma$ of the smallest *string attractor*, was introduced [2]. The measure $\gamma \leq z$ lower-bounds all the previous relevant ones, while length-$n$ strings can be represented and efficiently indexed within space $\mathcal{O}(\gamma \log(n/\gamma))$ [1], which also upper-bounds many measures, including $z$. Although $\gamma$ is arguably a better measure of repetitiveness than $z$, it is NP-complete to compute and not monotone, and it is unknown if one can represent all strings in $o(\gamma \log n)$ space.

In [4], we study an even smaller measure, $\delta \leq \gamma$, which can be computed in linear time, is monotone, and allows encoding every string in $z = \mathcal{O}(\delta \log(n/\delta))$ space. We further argue that $\delta$ better captures the compressibility of repetitive strings. Concretely, we show that:

1. $\delta$ can be strictly smaller than $\gamma$, by up to a logarithmic factor;
2. there are string families needing $\Omega(\delta \log(n/\delta))$ space to be encoded, so this space is optimal for every $n$ and $\delta$;
3. one can build run-length context-free grammars of size $\mathcal{O}(\delta \log(n/\delta))$, whereas the smallest (non-run-length) grammar can be up to $\Theta(\log n / \log \log n)$ times larger;
4. within $\mathcal{O}(\delta \log(n/\delta))$ space, we also offer logarithmic-time access to the symbols of the string $S$ and to the Karp–Rabin fingerprints of the substrings of $S$.
5. the $\mathcal{O}(\delta \log(n/\delta))$ space also allows for efficient pattern matching queries: one can find all the *occ* occurrences in $S$ of any pattern of length $m$ in time $\mathcal{O}(m \log n + occ \log^{\varepsilon} n)$ for any constant $\varepsilon > 0$.

In contrast, using $\mathcal{O}(\gamma \log(n/\gamma))$ space allows for an improved search time of $\mathcal{O}(m + (occ + 1) \log^{\varepsilon} n)$ [1]. In a follow-up paper [3], we combine the techniques of [1, 4] to achieve the same search time in $\mathcal{O}(\delta \log(n/\delta))$ space. We also show that an extra sublogarithmic factor on top of this space enables $\mathcal{O}(m + occ)$ search time, as well as efficient counting of the pattern occurrences.

In [4] and in the forthcoming the journal version of [3], we further refine the above results to account for the alphabet size $\sigma$ of the string, showing that $\Theta(\delta \log(n \log \sigma / \delta \log n))$ space is necessary and sufficient to represent the string and to efficiently support access, fingerprinting, and pattern matching queries.

## References

[1] A. R. Christiansen, M. B. Ettienne, T. Kociumaka, G. Navarro, and N. Prezza. Optimal-time dictionary-compressed indexes. *ACM Trans. Algorithms*, 17(1):8:1–8:39, 2021.

[2] D. Kempa and N. Prezza. At the roots of dictionary compression: String attractors. In I. Diakonikolas, D. Kempe, and M. Henzinger, eds., *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2018, Los Angeles, CA, USA, June 25-29, 2018*, 2018, pp. 827–840. ACM.

• [3] T. Kociumaka, G. Navarro, and F. Olivares. Near-optimal search time in $\delta$-optimal space. In A. Castañeda and F. Rodríguez-Henríquez, eds., *LATIN 2022: Theoretical Informatics*, Guanajuato, Mexico, 2022, LNCS 13568, pp. 88–103. Springer.

• [4] T. Kociumaka, G. Navarro, and N. Prezza. Toward a definitive compressibility measure for repetitive sequences. *IEEE Transactions on Information Theory*, 69(4):2074–2092, 2022.

[5] J. Ziv and A. Lempel. A universal algorithm for sequential data compression. *IEEE Trans. Inf. Theory*, 23(3):337–343, 1977.

### 27.7.4 Breaking the $\mathcal{O}(n)$-Barrier in the Construction of Compressed Suffix Arrays

*Investigators: Tomasz Kociumaka, in cooperation with Dominik Kempa (Stony Brook University)*

The suffix array, describing the lexicographical order of suffixes of a given text, and the suffix tree, a path-compressed trie of all suffixes, are the two most fundamental data structures for string processing, with a plethora of applications in data compression, bioinformatics, and information retrieval. For a length-$n$ text, however, they use $\Theta(n \log n)$ bits of space, which is often too costly. To address this, Grossi and Vitter [2] and, independently, Ferragina and Manzini [1] introduced space-efficient versions of the suffix array, known as the *compressed suffix array* (CSA) and the FM-*index*. Sadakane [6] then showed how to augment them to obtain the *compressed suffix tree* (CST). For a length-$n$ text over an alphabet of size $\sigma$, these structures use only $\mathcal{O}(n \log \sigma)$ bits. Nowadays, these structures are part of the standard toolbox: modern textbooks spend dozens of pages describing their applications, and they almost completely replaced suffix arrays and suffix trees in space-critical applications. The biggest remaining open question is how efficiently they can be constructed. After two decades, the fastest algorithms still run in $\mathcal{O}(n)$ time [3, 5], which is $\Theta(\log_\sigma n)$ factor away from the lower bound of $\Omega(n/\log_\sigma n)$ (following from the necessity to read the input).

In [4], we make the first in 20 years improvement in $n$ for this problem by proposing a new compressed suffix array and a new compressed suffix tree that admit $o(n)$-time construction algorithms while matching the space bounds and the query times of the original CSA/CST and the FM-index. More precisely, our structures take $\mathcal{O}(n \log \sigma)$ bits, support suffix array queries and full suffix tree functionality in $\mathcal{O}(\log^{\varepsilon} n)$ time per operation (for any constant $\varepsilon > 0$), and can be constructed in $\mathcal{O}(n \cdot \min(1, \log\sigma/\sqrt{\log n}))$ time using $\mathcal{O}(n\log\sigma)$ bits of working space. (For example, if $\sigma$ is constant, the construction time is $\mathcal{O}(n/\sqrt{\log n}) = o(n)$.) We derive this result as a corollary from a much more general reduction: We prove that all parameters of a compressed suffix array/tree (query time, space, construction time, and construction working space) can essentially be reduced to those of a data structure answering new query types that we call *prefix rank* and *prefix selection*. Using the novel techniques, we also develop a new compact index for exact pattern matching. It counts the occurrences of any length-$m$ pattern in $\mathcal{O}(m/\log_{\sigma} n + \log^{\varepsilon} n)$ time, whereas its construction algorithm shares the time and space complexity with our compressed suffix array.

### References


[1] P. Ferragina and G. Manzini. Indexing compressed text. *J. ACM*, 52(4):552–581, 2005.

[2] R. Grossi and J. S. Vitter. Compressed suffix arrays and suffix trees with applications to text indexing and string matching. *SIAM J. Comput.*, 35(2):378–407, 2005.

[3] W. Hon, K. Sadakane, and W. Sung. Breaking a time-and-space barrier in constructing full-text indices. In *44th Symposium on Foundations of Computer Science (FOCS 2003), 11-14 October 2003, Cambridge, MA, USA, Proceedings*, 2003, pp. 251–260. IEEE Computer Society.

• [4] D. Kempa and T. Kociumaka. Breaking the $O(n)$-Barrier in the construction of compressed suffix arrays. In N. Bansal and V. Nagarajan, eds., *Proceedings of the 2023 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA 2023)*, Florence, Italy, 2023, pp. 5122–5202. SIAM.

[5] J. I. Munro, G. Navarro, and Y. Nekrich. Space-efficient construction of compressed indexes in deterministic linear time. In P. N. Klein, ed., *Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2017, Barcelona, Spain, Hotel Porta Fira, January 16-19*, 2017, pp. 408–424. SIAM.

[6] K. Sadakane. Succinct representations of LCP information and improvements in the compressed suffix arrays. In D. Eppstein, ed., *Proceedings of the Thirteenth Annual ACM-SIAM Symposium on Discrete Algorithms, January 6-8, 2002, San Francisco, CA, USA*, 2002, pp. 225–232. ACM/SIAM.


## 27.8 Optimization

Optimization problems are ubiquitious. Whenever there is a choice between several alternatives, we are faced with an optimization problem. Depending on the domain of the decision variables, such problems can be characterized as continuous or discrete optimization problems. The former often appear in the area of Machine Learning, for instance, when continuous parameters of Neural Networks are optimized, while the latter are central to classical combinatorial optimization problems, for instance, as in the well-known Knapsack problem.

In the last reporting period, we contributed to both areas by developing novel techniques to create more efficient algorithms, obtain more precise analyses, and strengthen upper and lower bounds on the computational complexity of various optimization problems.

**Optimization and Machine Learning.** Optimization plays a fundamental role in machine learning, where the objective is to learn patterns and make predictions from data. Optimization techniques are used to train models by minimizing a loss function, which measures the discrepancy between the predicted outputs and the actual ones. This typically leads to high-dimensional optimization problems, which require efficient algorithms. So-called kernel functions are used to compute the similarity between input items. We published several results that make their evaluation more efficient.

For more details on these results, consult Section 27.8.1.

**Learning-augmented Algorithms.** Machine Learning can be used to predict good solutions for unknown instances of optimization problems. Hoever, they typically do not provide a guarantee on the quality of the solution or sometimes even not for the feasibility of their proposal. Learning-augmented algorithms add a safety net by blending the prediction and a solution with a provable guarantee so that the resulting approximation factor is the better the smaller the prediction error is, but remains bounded even when the prediction error is large.

For more details on these results, consult Section 27.8.2.

**Optimization in Practice.** We are also motivated by interdisciplinary and application-driven research.

For more details on these results, consult Section 27.8.3.

### 27.8.1 Optimization and Machine Learning

#### KDEformer: Accelerating Transformers via Kernel Density Estimation

*Investigators: Amir Zandieh, in cooperation with and Majid Daliri (New York University), Amin Karbasi and Insu Han (Yale University)*

Dot-product attention mechanism plays a crucial role in modern deep architectures (e.g., Transformer) for sequence modeling, however, naïve exact computation of this model incurs quadratic time and memory complexities in sequence length, hindering the training of longsequence models. Critical bottlenecks are due to the computation of partition functions in the denominator of softmax function as well as the multiplication of the softmax matrix with the matrix of values. Our key observation is that the former can be reduced to a variant of the kernel density estimation (KDE) problem, and an efficient KDE solver can be further utilized to accelerate the latter via subsampling-based fast matrix products. Our proposed KDEformer [2] can approximate the attention in sub-quadratic time with provable spectral norm bounds, while all prior results merely provide entry-wise error bounds.

Empirically, we verify that KDEformer outperforms other attention approximations in terms of accuracy, memory, and runtime on various pre-trained models. Particularly, we apply our method to image generation with BigGAN [1] and observe that our images, shown in Fig. 27.5, look more natural than others and our generative score is even better than the exact attention. Furthermore, we get over $4\times$ speedup compared to exact computation.

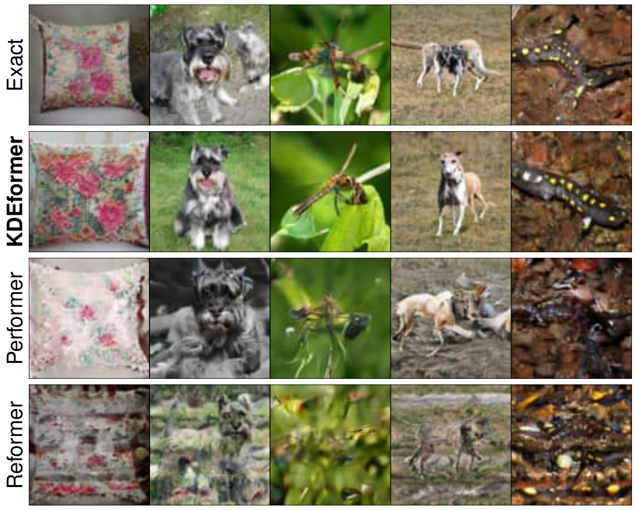



Figure 27.5: Image generations by the pre-trained BigGAN using exact and approximate attention without fine-tuning.

## References

[1] A. Brock, J. Donahue, and K. Simonyan. Large scale GAN training for high fidelity natural image synthesis. In *International Conference on Learning Representations*, 2019. https://openreview.net/forum?id=B1xsqj09Fm.

• [2] A. Zandieh, I. Han, M. Daliri, and A. Karbasi. *KDEformer: Accelerating Transformers via Kernel Density Estimation*, 2023. arXiv: 2302.02451.

### Gegenbauer polynomials for scaling up kernel methods and near optimal reconstruction of Spherical Harmonic expansions

*Investigators: Amir Zandieh, in cooperation with Haim Avron (Tel Aviv University), and Insu Han (Yale University)*

We propose efficient random features for approximating a new and rich class of kernel functions that we refer to as Generalized Zonal Kernels (GZK) [1]. Our proposed GZK family, generalizes the zonal kernels (that is, dot-product kernels on the unit sphere) by introducing radial factors in the Gegenbauer series expansion of these kernel functions. The GZK class of kernels includes a wide range of ubiquitous kernel functions such as the entirety of dot-product kernels as well as the Gaussian and the recently introduced Neural Tangent kernels. Interestingly, by exploiting the reproducing property of the Gegenbauer (Zonal) Harmonics, we can construct efficient random features for the GZK family based on randomly oriented Gegenbauer harmonics. We prove subspace embedding guarantees for our Gegenbauer features which ensures that our features can be used for approximately solving learning problems such as Kernel $k$-Means Clustering, Kernel Ridge Regression, etc.

We additionally propose an algorithm for robust recovery of the spherical harmonic expansion of functions defined on the $d$-dimensional unit sphere $\mathbb{S}^{d-1}$ using a near-optimal number of function evaluations [2]. We show that for any square-integrable function $f$, the number of evaluations of $f$ needed to recover its degree-$q$ spherical harmonic expansion equals the dimension of the space of spherical harmonics of degree at most $q$ up to a logarithmic factor. Moreover, we develop a simple yet efficient algorithm to recover degree-$q$ expansion of $f$ by only evaluating the function on uniformly sampled points on $\mathbb{S}^{d-1}$. Our algorithm is based on the connections between spherical harmonics and Gegenbauer polynomials and leverage score sampling methods. Unlike the prior results on fast spherical harmonic transform, our proposed algorithm works efficiently using a nearly optimal number of samples in any dimension $d$.

## References


- [1] I. Han, A. Zandieh, and H. Avron. Random Gegenbauer features for scalable kernel methods. In K. Chaudhuri, S. Jegelka, S. Le, S. Csaba, N. Gang, and S. Sabato, eds., *Proceedings of the 39th International Conference on Machine Learning (ICML 2022)*, Baltimore, MA, USA, 2022, Proceedings of the Machine Learning Research 162, pp. 8330–8358. https://proceedings.mlr.press/v162/han22g.html.
- [2] A. Zandieh, I. Han, and H. Avron. *Near Optimal Reconstruction of Spherical Harmonic Expansions*, 2022. arXiv: 2202.12995.


### Subspace embedding for tensor product matrices in input sparsity time

*Investigators: Amir Zandieh, in cooperation with David Woodruff (Carnegie Mellon University),*

We propose an input sparsity time sampling algorithm that can spectrally approximate the Gram matrix corresponding to the $q$-fold column-wise tensor product of $q$ matrices using a nearly optimal number of samples, improving upon all previously known methods by $\texttt{poly}(q)$ factors [1]. Furthermore, for the important special case of the $q$-fold self-tensoring of a dataset, which is the feature matrix of the degree-$q$ polynomial kernel, the leading term of our method's runtime is proportional to the size of the dataset and has no dependence on $q$. Previous techniques either incur a $\texttt{poly}(q)$ factor slowdown in their runtime or remove the dependence on $q$ at the expense of having sub-optimal target dimension, and depend quadratically on the number of data-points in their runtime. Our sampling technique relies on a collection of $q$ partially correlated random projections which can be simultaneously applied to a dataset $X$ in total time that only depends on the size of $X$, and at the same time their $q$-fold Kronecker product acts as a near-isometry for any fixed vector in the column span of $X^{\otimes q}$. We also show that our sampling methods generalize to other classes of kernels beyond polynomial, such as Gaussian and Neural Tangent kernels.

## References


- [1] D. Woodruff and A. Zandieh. Leverage score sampling for tensor product matrices in input sparsity time. In K. Chaudhuri, S. Jegelka, S. Le, S. Csaba, N. Gang, and S. Sabato, eds., *Proceedings of the 39th International Conference on Machine Learning (ICML 2022)*, Baltimore, MA, USA, 2022, Proceedings of the Machine Learning Research 162, pp. 23933–23964. https://proceedings.mlr.press/v162/woodruff22a.html.

### Fast Neural Kernel Embeddings via Sketching and Random Features

*Investigators: Amir Zandieh, in cooperation with Jaehoon Lee, Roman Novak, and Lechao Xiao (Google Research), Insu Han and Amin Karbasi (Yale University)*

The Neural Tangent Kernel (NTK) characterizes the behavior of infinitely-wide neural networks trained under least squares loss by gradient descent. Recent works also report that NTK regression can outperform finitely-wide neural networks trained on small-scale datasets. However, the computational complexity of kernel methods has limited its use in large-scale learning tasks. To accelerate learning with NTK, we design a near input-sparsity time approximation algorithm for NTK, by sketching the polynomial expansions of arc-cosine kernels: our sketch for the convolutional counterpart of NTK (CNTK) can transform any image using a linear runtime in the number of pixels [2]. Furthermore, we prove a spectral approximation guarantee for the NTK matrix, by combining random features (based on leverage score sampling) of the arc-cosine kernels with a sketching algorithm.

In a follow up work [1], we provide methods to work with general activations. First, we compile and expand the list of activation functions admitting exact dual activation expressions to compute neural kernels. When the exact computation is unknown, we present methods to effectively approximate them. We propose a fast sketching method that approximates any multi-layered Neural Network Gaussian Process (NNGP) kernel and Neural Tangent Kernel (NTK) matrices for a wide range of activation functions, going beyond the commonly analyzed ReLU activation. This is done by showing how to approximate the neural kernels using the truncated Hermite expansion of any desired activation functions. While most prior works require data points on the unit sphere, our methods do not suffer from such limitations and are applicable to any dataset of points in $\mathbb{R}^d$. Furthermore, we provide a subspace embedding for NNGP and NTK matrices with near input-sparsity runtime and near-optimal target dimension which applies to any homogeneous dual activation functions with rapidly convergent Taylor expansion.

We benchmark our methods on various large-scale regression and classification tasks and show that a linear regressor trained on our CNTK features matches the accuracy of an exact 5-layer Myrtle CNTK network on CIFAR-10 dataset while achieving 150x speedup [2, 1].

### References


- [1] I. Han, A. Zandieh, J. Lee, R. Novak, L. Xiao, and A. Karbasi. Fast neural kernel embeddings for general activations. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, eds., *Advances in Neural Information Processing Systems 35 (NeurIPS 2022)*, New Orleans, LA, USA, 2022, pp. 35657–35671. Curran Associates, Inc.

- [2] A. Zandieh, I. Han, H. Avron, N. Shoham, C. Kim, and J. Shin. Scaling neural tangent kernels via sketching and random features. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P. S. Liang, and J. Wortman Vaughan, eds., *Advances in Neural Information Processing Systems 34 (NeurIPS 2021)*, Virtual, 2021, pp. 1062–1073. Curran Assoicates, Inc.

### Sample Complexity

*Investigators: Pieter Kleer, and Hans U. Simon in cooperation with Adish Singla (MPI for Software Systems), David Kirkpatrick (University of British Columbia), Shaun Fallat, Abolghasem Soltani, Sandra Zilles (University of Regina), and Farnam Mansouri (University of Toronto)*

The *sample complexity* of a function-learning task is the smallest number of correctly labeled examples needed to successfully learn a target function. In case of random examples and binary functions, the sample complexity can be characterized by the VC-dimension of the underlying (binary) hypothesis class. As for multi-valued functions (used in multi-class classification or regression problems), the role of the VC-dimension is taken by combinatorial parameters like the pseudo- or the fat-shattering dimension. If the examples are not drawn at random but cleverly chosen by a teacher, then so-called teaching dimensions come into play. Among the most popular teaching models are the Preference-Based and the No-Clash model. The corresponding teaching dimensions are referred to as PB- and NC-dimension, respectively.

The purpose of the work in [3] is to understand the relation between the primal and dual of various combinatorial dimensions for multi-valued function classes. After having improved on one of the classical upper bounds, we present a couple of tightness results. Moreover, we analyze how Assouad's bound (a well known bound on the dual in terms of the primal VC-dimension) generalizes to classes of multi-valued functions.

The collusion-freeness condition of Goldman and Mathias [2] imposes a restriction on teaching models which is supposed to rule out "dirty coding tricks". The NC- and the PB-model, for instance, satisfy this condition. In [1], we show that the NC-teaching dimension is less than or equal to the teaching dimension induced by any collusion-free model. A similar result (dealing with "strong collusion-freeness") is shown for the PB-dimension.

The NC-dimension is always upper-bounded by the PB-dimension. In [5], we present a family $(H_n)$ of hypotheses classes (induced by so-called "Quadratic-Residue tournaments") such that the NC-dimension of $H_n$ equals 1 while the PB-dimension of $H_n$ asymptotically equals log $H_n$ . This the largest possible separation of these two dimensions. We also present an improved "bound of the Sauer-type" for the NC-dimension. The verification of the new bound makes use of Johnson graphs and maximum subgraphs not containing large narrow cliques.

It is not known whether the PB- or the NC-dimension is linearly bounded by the VC-dimension. In [4], we present a new teaching model whose dimension does provably not exceed the VC-dimension. The new model is *not* collusion-free but is shown to have a couple of other desirable properties.

### References

[1] S. Fallat, D. Kirkpatrick, H. U. Simon, A. Soltani, and S. Zilles. On batch teaching without collusion. *Journal of Machine Learning Research*, 24:1–33, 2023.

[2] S. A. Goldman and H. Mathias. Teaching a smarter learner. *Journal of Computer and System Sciences*, 52(2):255–267, 1996.

- [3] P. Kleer and H. U. Simon. Primal and dual combinatorial dimensions. *Discrete Applied Mathematics*, 327:185–196, 2023.
- [4] F. Mansouri, H. U. Simon, A. Singla, and S. Zilles. On batch teaching with sample complexity bounded by VCD. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, eds., *Advances in Neural Information Processing Systems 35 (NeurIPS 2022)*, New Orleans, LA, USA, 2022, pp. 15732–15742. Curran Associates, Inc.
- [5] H. U. Simon. Tournaments, Johnson graphs, and NC-teaching. In S. Agrawal and F. Orabona, eds., *Proceedings of the 34th International Conference on Algorithmic Learning Theory (ALT 2023)*, Singapore, Singapore, 2023, Proceedings of the Machine Learning Research 201, pp. 1411–1428. PMLR.


### 27.8.2 Learning-augmented Algorithms

#### A Novel Prediction Setup for Online Speed-Scaling

*Investigators: Golnoosh Shahkarami, in cooperation with Peyman Jabbarzade (University of Maryland), and Antonios Antoniadis (University of Twente)*

Given the rapid rise in energy demand by data centers and computing systems in general, it is fundamental to incorporate energy considerations when designing (scheduling) algorithms. Machine learning can be a useful approach in practice by predicting the future load of the system based on, for example, historical data. However, the effectiveness of such an approach highly depends on the quality of the predictions and can be quite far from optimal when predictions are sub-par. On the other hand, while providing a worst-case guarantee, classical online algorithms can be pessimistic for large classes of inputs arising in practice.

In [1], we study online, deadline-based *speed-scaling* scheduling, augmented with machine-learned predictions. More specifically, a set of jobs $\mathcal{J}$, each job $j$ $\mathcal{J}$ with an associated release time $r_j$, deadline $d_j$ and processing requirement $w_j$, arrives online and has to be scheduled on a single speed-scalable processor. A scheduling algorithm needs to decide for each time point $t$ on: (i) the processor speed $s(t)$ and (ii) which job $j$ $\mathcal{J}$ to execute at $t$ ($j(t)$). Both decisions have to be made by the algorithm at any time point $t$ while only having knowledge of the jobs with a release time equal to or less than $t$. The energy consumption of a schedule, which we seek to minimize over all feasible schedules, is given by $\int_0^\infty s(t)^\alpha dt$, where $\alpha > 1$ is a constant.

In the spirit of the new area of machine learning augmented algorithms, we attempt to obtain the best of both worlds for the classical, deadline based, online speed-scaling problem: Based on the introduction of a novel prediction setup, we develop algorithms that (i) obtain provably low energy-consumption in the presence of adequate predictions, and (ii) are robust against inadequate predictions, and (iii) are smooth, that is, their performance gradually degrades as the prediction error increases.

#### References


- [1] A. Antoniadis, P. Jabbarzade, and G. Shahkarami. A novel prediction setup for online speed-scaling. In A. Czumai and Q. Xin, eds., *18th Scandinavian Symposium and Workshops on Algorithm Theory (SWAT 2022)*, Tórshavn, Faroe Islands, 2022, Leibniz International Proceedings in Informatics 227, Article 9. Schloss Dagstuhl.

### Learning-Augmented Algorithms for Online TSP on the Line

*Investigators: Golnoosh Shahkarami, in cooperation with Konstantinos Lakis (National and Kapodistrian University of Athens), and Themis Gouleakis (National University of Singapore)*

The Traveling Salesperson Problem (tsp) is one of the most fundamental and widely studied problems in computer science, both in the offline version, where the input is known in advance, and the online version where input arrives sequentially. In [1], we consider the Online tsp on the real line augmented with machine-learned predictions. In the classical problem, there is a stream of requests released over time along the real line. The goal is to minimize the makespan of the algorithm. We distinguish between the *open* variant and the *closed* one, in which we additionally require the algorithm to return to the origin after serving all requests. The state of the art is a 1.64-competitive algorithm for the closed variant and a 2.04-competitive algorithm for the open variant. In both cases, a tight lower bound is known.

In both variants, our primary prediction model involves predicted *positions* of the requests. We introduce algorithms that (i) obtain a tight 1.5 competitive ratio for the closed variant and a 1.66 competitive ratio for the open variant in the case of perfect predictions, (ii) are robust against unbounded prediction error, and (iii) are smooth, that is, their performance degrades gracefully as the prediction error increases.

In addition, we further investigate the learning-augmented setting by including a prediction for the last request served by the optimal offline algorithm in the *open* variant. In this enhanced setting, our algorithm achieves a 1.33 competitive ratio with perfect predictions while also being smooth and robust, beating our lower bound of 1.44 for the original prediction setting. Moreover, we provide a lower bound of 1.25 for this enhanced setting.

### References

• [1] T. Gouleakis, K. Lakis, and G. Shahkarami. Learning-augmented algorithms for online TSP on the line. In *Proceedings of the 37th AAAI Conference on Artificial Intelligence*, Washington, DC, USA, 2023. AAAI. Accepted.

## 27.8.3 Optimization in Practice

### Computational Interaction

*Investigators: Andreas Karrenbauer, in cooperation with Niraj Ramesh Dayama, Morteza Sharipour, Antti Oulasvirta (Aalto University, Helsinki, Finland) and Evgeny Ivanko (Ural Federal University, Yekaterinburg, Russia)*

The problem of computational design for menu systems has been addressed in some specific cases such as the linear menu (list). The classical approach has been to model this problem as an assignment task, where commands are assigned to menu positions while optimizing for users' selection performance and grouping of associated items. However, we show that this approach fails with larger, hierarchically organized menus because it does not take into account the ways in which users navigate hierarchical structures.

In [1], we address the computational menu design problem by presenting a novel integer programming formulation that yields usable, well-ordered command hierarchies from a single model. First, it introduces a novel objective function based on information foraging theory, which minimizes navigation time in a hierarchical structure. Second, it models the hierarchical menu design problem as a combination of the exact set covering problem and the assignment problem, organizing commands into ordered groups of ordered groups. The approach is efficient for large, representative instances of the problem. In a controlled usability evaluation, the performance of computationally designed menus was faster to use than existing commercial designs. We also discuss applications of this approach for personalization and adaptation.

## References

• [1] N. R. Dayama, M. Shiripour, A. Oulasvirta, E. Ivanko, and A. Karrenbauer. Foraging-based optimization of menu systems. *International Journal of Human-Computer Studies*, 151, Article 102624, 2021.

### Computational Design and Optimization of Electro-physiological Sensors

*Investigators: Andreas Karrenbauer, in cooperation with Tobias Kraus (INM – Leibnitz Institute for New Materials, Saarbrücken, Germany), Aditya Shekhar Nittala, Arshad Khan, Jürgen Steimle (Saarland University, Saarbrücken, Germany)*

Electro-physiological sensing devices are becoming increasingly common in diverse applications. However, designing such sensors in compact form factors and for high-quality signal acquisition is a challenging task even for experts, is typically done using heuristics, and requires extensive training. Our work [1] proposes a computational approach for designing multi-modal electro-physiological sensors. By employing an optimization-based approach alongside an integrated predictive model for multiple modalities, compact sensors can be created which offer an optimal trade-off between high signal quality and small device size. The task is assisted by a graphical tool that allows to easily specify design preferences and to visually analyze the generated designs in real-time, enabling designer-in-the-loop optimization. Experimental results show high quantitative agreement between the prediction of the optimizer and experimentally collected physiological data. They demonstrate that generated designs can achieve an optimal balance between the size of the sensor and its signal acquisition capability, outperforming expert generated solutions.

## References

• [1] A. S. Nittala, A. Karrenbauer, A. Khan, T. Kraus, and J. Steimle. Computational design and optimization of electro-physiological sensors. *Nature Communications*, 12(1), Article 6351, 2021.

### Car (Re-)Sequencing in Mixed-Model Assembly Lines

*Investigators: Andreas Karrenbauer, Leonie Krull, Kurt Mehlhorn, Paolo Luigi Rinaldi, and Anna Twelsiek, in cooperation with Pranabendu Misra (Chennai Mathematical Institute (CMI)),*

Cars are manufactured on assembly lines, that is, they are attached to conveyors that move them through the plant at a certain speed so that they leave the factory one after the other at a fixed rate (between half a minute and a minute). Since customers of new cars can configure a wide variety of features (color, engine, tow-bar, GPS, etc.) individually, there are rarely two identical *orders* on the same production day.

The build sequence for a production day is typically generated a few days in advance to allow the suppliers to deliver the corresponding parts just in time. However, a sequence that allows for large batches of the same color in the paint shop, might cause a line stop in the final assembly when the workload for a certain feature of back to back cars exceeds the capacity of the corresponding station, for instance, it might be only possible to mount a tow-bar to every other car on time.

Therefore, a plant is partitioned into three main stages: welding of the body, painting of the body, assembly of the car (engine, seats, wheels, and so on). These stages are separated by buffers so that a line stop in one area does not immediately affect the production in an other area. These buffers can consist of multiple parallel lanes so that they can not only be used as a storage, but also allow for a limited reorganization of the build sequence. This is particularly helpful when the initial sequence is distorted due to extra work on some cars. In particular, paint is not perfect and hence a certain fraction of the cars needs additional work, the original sequence is disturbed after paint and needs to be restored as much as possible.

In the first stage, bodies have no paint and no particular features. Therefore, at this point in the process, a body is compatible with many possible *orders*. The *order* is the set of all the features that a car is expected to have at the end of the process. Each *order* is identified with a *Vehicle Identification Number* (VIN). Since a body can be paired with several possible VINs, we aim to exploit this flexibility to produce a sequence that results in minimal costs for the plant. As these bodies proceed with the next steps, more features are applied to them, and the degrees of freedom decrease. Perfect sequencing could result in an efficiency gain of several percent.

In a project together with the Ford Motor Company[1], we developed two algorithms: one for computing an optimal initial sequence and one for restoring order in a buffer with multiple parallel lanes. The latter algorithm has been integrated in the production process in a plant of our industrial partner in Q1 of 2023.

The development of our effective algorithms for practice would not have been possible without a profound understanding of the limited sorting capabilities of such lane buffers. To this end, we addressed this sorting problem in an ideal and formal framework [1]. More precisely, given a sequence of $n$ numbers, and a collection of $k$ parallel queues, can we sort this sequence with these $k$ queues? In general, it is not always possible in a realistic scenario because $k$ is too small. Therefore, we are interested in reducing the disorder. We attribute

[1] www.modigpro.saarland

the disorder to two possible definitions:

- Longest Decreasing Subsequence (LDS): Given a sequence of length $n$ with LDS of $L$, we show that with $k$ queues one can always obtain a resulting sequence of LDS at most $L - k + 1$. Furthermore, we have also shown that this is also a lower bound, in the sense that there exists a sequence of LDS $L$ such that, with $k$ queues it's impossible to reduce the LDS to a value lower than $L - k + 1$,
- Number of downsteps: A downstep in a sequence is an item immediately followed by a smaller item. We give an algorithm that for any number $k$ of queues, outputs a sequence with the least number of downsteps attainable with $k$ queues. This is in fact an online algorithm, but nevertheless, we proved that it is optimal.

## References


• [1] A. Karrenbauer, L. Krull, K. Mehlhorn, P. Misra, P. L. Rinaldi, and A. Twelsiek. *Improving Order with Queues*, 2022. arXiv: 2207.02476.


### 27.8.4 Miscellaneous Topics in Optimization

#### Physarum

*Investigators: Yuan Gao, Andreas Karrenbauer, Pavel Kolev, Kurt Mehlhorn, and Golnoosh Shahkarami, in cooperation with Quentin Vermande (École Normale Supérieure), Frederic Folz, Giovanna Morigi (Saarland University), Enrico Facca (Scuola Normale Superiore), Mohammadamin Sharifi (Sharif University of Technology) Vincenzo Bonifaci (Università degli Studi Roma Tre), and Hamidreza Kamkari (University of Toronto),*

The slime mold Physarum polycephalum was experimentally demonstrated to be able to solve several algorithmic problems on graphs, including shortest path problem [6, 7]. A mathematical model defined via a system of differential equations for the dynamic behavior of the Physarum was proposed [9] and shown to converge to shortest path for all graphs [2]. In this report period, we extended our previous work in three directions and published [1, 4, 3]

**Multi-Commodity Flow.** The shortest path problem can be viewed as a min-cost flow problem: the shortest path between two nodes is equal to the cost of a min-cost flow between the two nodes for uniform edge costs. In a multi-commodity flow problem, we have $k$ source-sink pairs $(s_i, t_i)$, $1 \leq i \leq k$, and want to send for each $i$ a unit of flow from source to sink. We aimed at a model where sharing edges is beneficial and therefore used for each edge the two-norm of the individual flows as the effective flow across the edge. In [1], we studied the resulting dynamics experimentally and theoretically. The computer experiments show that the dynamics is able to construct efficient networks, for example for the rail-road system in the greater Tokyo area, which resemble networks resulting from human ingenuity. In the theoretical part of our work, we show that the dynamics converges to a minimum cost solution.

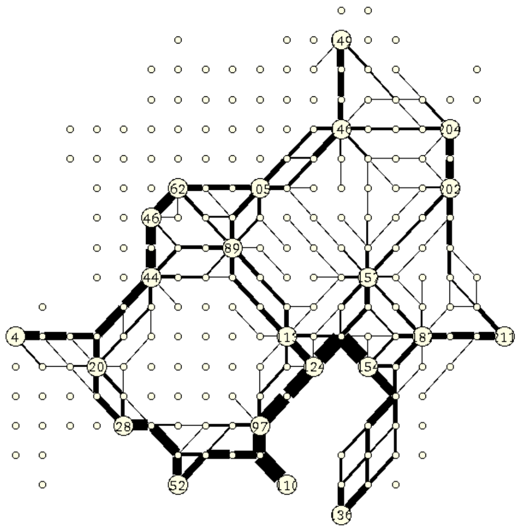


Figure 27.6: The network constructed for the railroad system in the Greater Tokyo area. There are 40 cities and hence $40 \cdot 39/2$ commidities to be routed. The flows into Tokyo (node 24) are five times the other flows.

**Semi-Definite Programming.** The Physarum dynamics can be described by the following system of ordinary differential equations:

$$\dot{x} = -(I - GA^T(AGA^T)^{\dagger}A)Gc \tag{27.1}$$

where $A$ is a arc-node incidence matrix of the underlying graph, $G = C^{-1}X$ is the conductance while $C$ and $X$ are diagonal matrices defined by the arc-costs $c$ and flow-vector $x$, respectively. Efforts have been made to generalize the Physarum dynamics beyond just the shortest path problem on graphs, for instance, convergence to the optimum of $\min \{c^T x : Ax = b, x \geq 0\}$ was shown for the case when $c$ is positive [5, 8].

In [4], we go a step further and we generalize the Physarum dynamics (27.1) to positive semi-definite programs (SDP) of the form

$$\min \; tr(C^T X) : tr(A_\ell^T X) = b_\ell \;\; \ell \quad [m],\, X \succeq 0 \quad \text{where } C \succ 0. \tag{27.2}$$

For an arbitrary positive definite conductance matrix $G$, we study a general recipe for SDP Physarum dynamics

$$\underline{\dot{X}} = -(I - G\mathcal{A}^T(\mathcal{A}G\mathcal{A}^T)^{\dagger}\mathcal{A})G\underline{C}\,, \tag{27.3}$$

and we prove that the dynamic reduces the objective value of the SDP over time. By introducing a correction term, the dynamic with a linearly infeasible starting point approaches feasibility exponentially fast. Moreover, we study two concrete choices of the conductance matrix. When $G := {}^1\!/_2(C^{-1} \otimes X + X \otimes C^{-1})$, the SDP dynamic (27.3) reduces to the LP dynamic (27.1) when SDP reduces to LP. For this choice of conductance, we prove, under the condition that $C^{-1}$ is feasible, the dynamic stays within the positive definite cone, and converges to an equilibrium where $\dot{X} = 0$. We conjecture that the condition can be removed and the equilibrium lies in the optimal face of the SDP. We complement the conjecture with empirical evidences. When $G := X \otimes X$, we prove that for any positive definite starting point, the Physarum dynamic converges to the optimum of SDP unconditionally, by showing the equivalence between the trajectory of the dynamic and some (infeasible) central path the SDP.

**Noise and Convergence.** Our previous work on the Physarum dynamics investigated the *ideal* dynamics in an environment without noise. In a cooperation [3] with Giovanna Morigi and Frederic Folz from the physics department of Saarland University, we started to investigate the influence of noise on the convergence behavior. More precisely, we studied the interplay of a nonlinear interaction function and Gaussian, additive noise. We varied the amplitude of the noise and investigated its influence on the convergence rate and the resulting network topology. We show that different network topologies emerge. Remarkably, the system converges to the most robust configuration at a non-zero noise amplitude. Interestingly, this configuration is not found by the deterministic dynamics and is reached with the maximum convergence rate. Our results suggest that stochastic dynamics can boost transport in a nonlinear network.

## References


- [1] V. Bonifaci, E. Facca, F. Folz, A. Karrenbauer, P. Kolev, K. Mehlhorn, G. Morigi, G. Shahkarami, and Q. Vermande. Physarum-inspired multi-commodity flow dynamics. *Theoretical Computer Science*, 920:1–20, 2022.

[2] V. Bonifaci, K. Mehlhorn, and G. Varma. Physarum can compute shortest paths. In *Proceedings of the Twenty-Third Annual ACM-SIAM Symposium on Discrete Algorithms*, USA, 2012, SODA '12, pp. 233–240. Society for Industrial and Applied Mathematics.

- [3] F. Folz, K. Mehlhorn, and G. Morigi. Interplay of periodic dynamics and noise: Insights from a simple adaptive system. *Physical Review E*, 104(5), Article 054215, 2021.

- [4] Y. Gao, H. Kamkari, A. Karrenbauer, K. Mehlhorn, and M. Sharifi. *Physarum Inspired Dynamics to Solve Semi-Definite Programs*, 2022. arXiv: 2111.02291.

[5] A. Johannson and J. Zou. A slime mold solver for linear programming problems. In S. B. Cooper, A. Dawar, and B. Löwe, eds., *How the World Computes*, Berlin, Heidelberg, 2012, pp. 344–354. Springer Berlin Heidelberg.

[6] T. Nakagaki, M. Iima, T. Ueda, Y. Nishiura, T. Saigusa, A. Tero, R. Kobayashi, and K. Showalter. Minimum-risk path finding by an adaptive amoebal network. *Phys. Rev. Lett.*, 99:068104, 2007.

[7] T. Nakagaki, H. Yamada, and Á. Tóth. Maze-solving by an amoeboid organism. *Nature*, 407(6803):470–470, 2000.

[8] D. Straszak and N. K. Vishnoi. On a natural dynamics for linear programming. In M. Sudan, ed., *Proceedings of the 2016 ACM Conference on Innovations in Theoretical Computer Science, Cambridge, MA, USA, January 14-16, 2016*, 2016, p. 291. ACM.

[9] A. Tero, R. Kobayashi, and T. Nakagaki. A mathematical model for adaptive transport network in path finding by true slime mold. *Journal of Theoretical Biology*, 244(4):553–564, 2007.

### Approximation Algorithms for Center-Based Clustering Problems

*Investigators: Martin Herold, Evangelos Kipouridis, Joachim Spoerhase in cooperation with Aristides Gionis (KTH Royal Institute of Technology), Bruno Ordozgoiti (Queen Mary University of London), Kamyar Khodamoradi (University of British Columbia), Moritz Beck, Sabine Storandt (University of Konstanz), Benedikt Riegel (University of Stuttgart)*

In Center-based Clustering problems, we ask to find a set of centers in a metric space and assign each input point to one of the selected centers so as to optimize a certain objective function based on the distances between points and centers. A prominent example is the $k$-Median problem where we open $k$ centers so as to minimize the *sum* of distances between points and their respective center. Another such problem is the $k$-Center problem where we aim at minimizing the *maximum* distance of a point to its center. We investigate approximation algorithms for variants of classic center-based clustering problems.

One such variant is the Reconciliation $k$-Median problem, which asks to cluster a set of data points by picking $k$ cluster centers so as to minimize the sum of distances of the data points to their cluster centers plus the sum of pairwise distances between the centers. The problem aims to find a set of cluster centers that are not too far from each other, and it has applications, for example, when selecting a committee to deliberate on a controversial topic. This problem was introduced recently in [3] where it was shown that a local-search-based algorithm is always within a factor $\mathcal{O}(k)$ of an optimum solution and performs well in practice.

In [2], we demonstrate a close connection of Reconciliation $k$-median to a variant of the $k$-Facility Location problem, in which each potential cluster center has an individual opening cost and we aim at minimizing the sum of client-center distances and the opening costs. This connection enables us to provide a new algorithm for Reconciliation $k$-Median that yields a *constant-factor approximation* (independent of $k$). We also provide a *sparsification* scheme that reduces the number of potential cluster centers to $\mathcal{O}(k)$ in order to substantially speed up approximation algorithms. We empirically compare our new algorithms with the previous approach.

In [1], we consider a variant of $k$-Center. As in $k$-Center, the metric space is specified by shortest-path distances in an edge-weighted graph. However, we study an *edge* variant where pick $k$ edges (rather than nodes) so as to minimize the maximum distance from any node to the closest center edge. The problem can be motivated by applications in urban planning, network analysis, and data visualization. An obstacle in handling this variant is that the underlying distance measure does not satisfy the triangle inequality, which is key in the known approximation algorithms for vanilla $k$-Center. We are able to get around this obstacle and obtain constant-factor upper and lower bounds on the approximability of this problem and some variants. We complement the theoretical results by an experimental evaluation on real-world instances.

In the above-mentioned clustering problems, the difficulty lies in selecting the centers. Once this is achieved, finding an optimal assignment is easier. For example, in $k$-Median (and $k$-Means, $k$-Center) it is optimal to assign each point to the closest center. In Capacitated $k$-Median (with or without outliers) solving a min-cost max-flow problem is sufficient. However, in some settings finding an optimal assignment is non-trivial. One such example is the MIN-SUM RADII problem, whose objective is to minimize the sum of distances from each center to its furthest assigned point. In an ongoing project, we study such problems, with the goal of approximating the optimal assignment, given a set of centers. Ideally, we would like to find simple combinatorial algorithms that can give insights to the structure of approximately optimal solutions.

## References


- [1] M. Beck, J. Spoerhase, and S. Storandt. Mind the gap: Edge facility location problems in theory and practice. In A. Bagchi and R. Muthu, eds., *Algorithms and Discrete Applied Mathematics (CALDAM 2023)*, Gandhinagar, India, 2023, LNCS 13947, pp. 321–334. Springer.
- [2] A. Gionis, K. Khodamoradi, B. Ordozgoiti, B. Riegel, and J. Spoerhase. A constant-factor approximation algorithm for reconciliation $k$-median. In *Proceedings of the 26th International Conference on Artificial Intelligence and Statistics (AISTATS 2023)*, Valencia, Spain, 2023, Proceedings of the Machine Learning Research. PMLR. Accepted.

[3] B. Ordozgoiti and A. Gionis. Reconciliation $k$-median: Clustering with non-polarized representatives. In *Proceedings of the The World Wide Web Conference (WWW 2019)*, 2019, pp. 1387–1397. ACM.


### Partially Dynamic Packing/Covering Linear Programs

*Investigators: Peter Kiss in cooperation with Thathaphol Saranurak (University of Michigan), and Sayan Bhattacharya (University of Warwick)*

Packing/covering linear programs are powerful generalizations of numerous optimization problems such as BIPARTITE MATCHING and VERTEX COVER. In the partially dynamic setting, our goal is to maintain a solution to the linear program as its solution polytope is either expanding or restricting due to updates. In the context of graph optimization problems these updates usually translate to edge or vertex insertions/deletions, but not both insertions and deletions.

In [1], we have presented the first algorithm for maintaining $(1+\varepsilon)$-approximate solutions to partially dynamic packing/covering linear programs in polylogarithmic update time, yielding new algorithmic implications for specific partially dynamic LPs such as WEIGHTED DECREMENTAL BIPARTITE MATCHING. Previous algorithms were either restricted to specific applications of packing/covering LPs or were restricted to LPs with a small number of variables.

## References


- [1] S. Bhattacharya, P. Kiss, and T. Saranurak. Dynamic algorithms for packing-covering LPs via multiplicative weight. In N. Bansal and V. Nagarajan, eds., *Proceedings of the 2023 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA 2023)*, Florence, Italy, 2023, pp. 1–47. SIAM.

## 27.9 Academic Activities

### 27.9.1 Journal Positions

*Danupon Nanongkai:*

- Editorial board of *TheoretiCS* (since 2022).

*Tomasz Kociumaka:*

- Editorial board of *Information Processing Letters* (since 2020).

*Karl Bringmann:*

- Editorial board of *SIAM Journal on Computing* (since 2023).

### 27.9.2 Conference and Workshop Positions

#### Membership in Program and Organization Committees

*Danupon Nanongkai:*

- *64th IEEE Symposium on Foundations of Computer Science (FOCS)*, Santa Cruz, CA, USA, November 2023 (PC member).

*Kurt Mehlhorn:*

- *6th Symposium on Simplicity in Algorithms, SOSA@SODA 2023*, Florence, Italy, January 2023 (PC co-chair with Kavitha Telikepalli).

*Philip Wellnitz:*

- *30th Annual European Symposium on Algorithms (ESA), track A*, Potsdam, Germany, September 2022 (PC member).

*Karol Węgrzycki:*

- *6th Symposium on Simplicity in Algorithms, SOSA@SODA 2023*, Florence, Italy, January 2023 (PC member).
- *17th International Symposium on Parameterized and Exact Computation (IPEC)*, Potsdam, Germany, September 2022 (PC member).

*Roohani Sharma:*

- *49th International Workshop on Graph-Theoretic Concepts in Computer Science (WG)*, Fribourg, Switzerland, June 2023 (PC member).
- *17th International Symposium on Parameterized and Exact Computation (IPEC)*, Potsdam, Germany, September 2022 (PC member).

*Adam Polak:*

- *31st Annual European Symposium on Algorithms (ESA), track A*, Amsterdam, Netherlands, September 2023 (PC member).

*André Nusser:*

- *Computational Geometry: Young Researchers Forum (CG:YRF)*, Berlin, Germany, June 2022 (PC member).

*Christoph Lenzen:*

- *27th IEEE International Symposium on Asynchronous Circuits and Systems (ASYNC)*, online, September 2021 (PC member).

*Marvin Künnemann:*

- *47th International Workshop on Graph-Theoretic Concepts in Computer Science (WG)*, Warsaw, Poland (virtual), June 2021 (PC member).

*Tomasz Kociumaka:*

- *34th International Workshop on Combinatorial Algorithms (IWOCA)*, Tainan, Taiwan, June 2023 (PC member),
- *34th Annual Symposium on Combinatorial Pattern Matching (CPM)*, Marne-la-Vallée, France, June 2023 (PC member),
- *30th Annual European Symposium on Algorithms (ESA), track S*, Potsdam, Germany, September 2022 (PC member),
- *18th Scandinavian Symposium and Workshops on Algorithm Theory (SWAT)*, Thorshavn, Faroe Islands, June 2022 (PC member),
- *33rd Annual Symposium on Combinatorial Pattern Matching (CPM)*, Prague, Czech Republic, June 2022 (PC member).

*Corinna Coupette:*

- *ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*, Long Beach, CA, USA, August 2023 (PC member),
- *Learning on Graphs Conference (LoG)*, Virtual Conference, December 2022 (PC member),
- *Natural Legal Language Processing Workshop (NLLP) at the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Abu Dhabi, December 2022 (PC member),
- *Conference on Neural Information Processing Systems, Datasets and Benchmarks Track (NeurIPS D&B)*, Washington, DC, USA, November 2022 (PC member).

*Karl Bringmann:*

- *50th International Colloquium on Automata, Languages, and Programming (ICALP)*, Paderborn, Germany, July 2023 (PC member),
- *54th Annual ACM SIGACT Symposium on Theory of Computing (STOC)*, Rome, Italy, June 2022 (PC member),
- *5th Symposium on Simplicity in Algorithms, SOSA@SODA 2022*, Virtual Conference, January 2022 (PC co-chair with Timothy Chan),

– *16th International Symposium on Parameterized and Exact Computation (IPEC)*, Lisbon, Portugal, September 2021 (PC member),

### 27.9.3 Invited Talks and Tutorials

*Kurt Mehlhorn:*

– Invited Talk at ACM India Symposium, 2022.

*Roohani Sharma:*

– *Designing FPT algorithms for problems with con ict-free constraints*, Invited Talk, Sabarmati Young Researcher Seminar Series at IIT Gandhinagar, India, October 2021.

*Tomasz Kociumaka:*

– *Dynamic string algorithms*, Keynote Talk, 2nd IGAFIT Workshop for Algorithms Postdocs in Europe, Będlewo, Poland, August 2022.
– *Small space and streaming pattern matching with k edits*, Highlights of CPM (a special session for presenting the highlights of recent developments in combinatorial pattern matching published in other venues.), 33rd Annual Symposium on Combinatorial Pattern Matching (CPM), Prague, Czech Republic, June 2022.
– *Recent advances in dynamic string algorithms*, Dynamic Algorithms Recent Advances and Applications, satellite workshop of STOC 2022, Rome, Italy, June 2022.

*Corinna Coupette:*

– *Exploring Evolving Hypergraphs: Methods and Applications*, Invited Talk, Complexity Science Hub Vienna, Vienna, March 2023.
– *Purposes of Legal Network Analysis*, Invited Talk, University of Maastricht: Workshop Series "Legal Network Analysis", Virtual, February 2023.
– *Law as Data*, Invited Talk, Max Planck Institute for Tax Law and Public Finance: Workshop "Method and Cognitive Interest in Private Law", Munich, September 2022.
– *Panel Discussion on Long-Range Graph Representation Learning*, Invited Panelist, International Conference on Learning Representations (ICLR): Workshop on Geometric and Topological Representation Learning, Virtual, April 2022.
– *Legal Network Science*, Guest Lecture, Bucerius Law School: Bucerius Legal Tech Essentials, Virtual, Summer 2022.
– *Legal Network Analysis*, Guest Lecture, Bucerius Law School: Bucerius Summer Program in Legal Technology and Operations, Hamburg, Summer 2022.
– *Legal Network Analysis: A Whirlwind Tour*, Guest Lecture, Techical University Munich: Legal Data Science & Informatics, Virtual, Summer 2021.
– *Quantitative Legal Studies: Technological Foundations*, Guest Lecture, Free University Berlin and Humboldt University Berlin: Quantitative Legal Studies, Virtual, Summer 2021.

*Karl Bringmann:*

- *Hardness of Approximation in P: Fine-Grained Complexity of Graph Distance Approximation*, Invited Talk, GI-Theorietag 2022, 83rd Workshop on Algorithms and Complexity, Bonn, November 2022.

### 27.9.4 Other Academic Activities

- Kurt Mehlhorn is a member of the Research and Innovation Advisory Board (RIAB) of Tata Consultancy Services (2019 – present).
- Kurt Mehlhorn is a member of the Joint Advisory Board of Carnegie Mellon University, Qatar (2009 – present).
- Danupon Nanongkai was interviewed by Quanta Magazine for the article "Finally, a Fast Algorithm for Shortest Paths on Negative Graphs", 2023.
- Kurt Mehlhorn was a member of the Review Committee of the Computer Science Department of EPFL (2023).
- Kurt Mehlhorn was a member of the ERC Scientific Council and chairman of its working group on Innovation (2016 – 2022).
- Kurt Mehlhorn was a member of the Review Committee of the Computer Science of KAUST (2023).
- Kurt Mehlhorn was the chair of the Review Committee of the Computer Science Department of Weizmann Institute (2022).
- Karl Bringmann was a Committee Member of the National Research Assessment of Computer Science in the Netherlands (2022).
- Karl Bringmann was a Panel Member of the final round of "Bundeswettbewerb Informatik". (2022)

## 27.10 Teaching Activities

Unless mentioned otherwise, all courses were taught at Saarland University.

### Winter Semester 2022/2023

#### Courses

- Introduction to Algorithms and Data Structures (Karl Bringmann)
- Randomized Algorithms and Probabilistic Analysis of Algorithms (Philip Wellnitz)
- Approximation Algorithms (Joachim Spoerhase)
- Ideas of Informatics (Corinna Coupette, Kurt Mehlhorn)

### Summer Semester 2022

#### Courses

- Optimization (Andreas Karrenbauer)

- Competitive Programming (Markus Bläser (Saarland University), Karl Bringmann, Martin Bromberger, Christoph Weidenbach)
- Programming for Lawyers, University of Heidelberg (Corinna Coupette)

#### Seminars

- Reading Group: String Algorithms (Karl Bringmann, Tomasz Kociumaka)

### Winter Semester 2021/2022

#### Courses

- Algorithms and Data Structures (Karl Bringmann, Raimund Seidel (Saarland University))
- Fine-Grained Complexity Theory (Karl Bringmann)
- Parameterized Algorithms (Dániel Marx, Roohani Sharma)
- Ideas of Informatics (Corinna Coupette, Kurt Mehlhorn)
- Introduction to Informatics, Bucerius Law School (Corinna Coupette, Dirk Hartung (Bucerius Law School))

#### Seminars

- Reading Group Algorithms: Continuous Methods for Combinatorial Problems (Kurt Mehlhorn, Roohani Sharma, Hans Simon, Philip Wellnitz)

### Summer Semester 2021

#### Courses

- Optimization (Andreas Karrenbauer)
- Geometric algorithms with limited resources (Themistoklis Gouleakis, Sándor Kisfaludi-Bak)
- Clock Synchronization and Adversarial Fault Tolerance (Christoph Lenzen, Danny Dolev (The Hebrew University of Jerusalem))
- Programming for Lawyers, University of Heidelberg (Corinna Coupette)

### Master Theses

- Rishabh Sharma, A system for conducting and evaluating user studies in the context of a graphical user interface for modeling optimization problems, 2022.
- Zahra Parsaeian, Approximate Near-Neighbor Problem for Curves Using Discrete Fréchet Distance under Translation and Rotation, 2022.
- Sami Shalayel, Non-destructive testing of concrete by interpreting ultrasound signals via linear optimization, 2022.
- Baris Can Esmer, On $(1+\varepsilon)$-Approximate Block Sparse Recovery, 2021.

### Bachelor Theses

- Sam Jonas Niro, Optimization under Privacy, 2022.
- Jannik Kudla: Approximation Algorithms for the Subset Sum Ratio Problem, 2021.

## 27.11 Dissertations, Habilitations, Awards

### 27.11.1 Dissertations

- André Nusser, *Fine-grained complexity and algorithm engineering of geometric similarity measures*, 2022,
- Ben Wiederhake, *Pulse Propagation, Graph Cover, and Packet Forwarding*, 2022.
- Attila Kinali-Dogan, *On Time, Time Synchronization and Noise in Time Measurement Systems*, 2022.
- Philip Wellnitz, *Counting Patterns in Strings and Graphs*, 2021,
- Bhaskar Ray Chaudhury, *Finding Fair and Efficient Allocations*, 2021,
- Anurag Pandey, *Variety Membership Testing in Algebraic Complexity Theory*, 2021.

### 27.11.2 Offers for Faculty Positions

- Joachim Spoerhase: University of Sheffield, UK, Lecturer.
- Adam Polak: Bocconi University, Italy, tenure-track assistant professor (offered before joining MPI-INF).
- Christoph Lenzen: CISPA, Germany, fast tenure track faculty position.

### 27.11.3 Best Paper Awards

- Roohani Sharma received the best paper award for the paper titled “Balanced Substructures in Bicolored Graphs” at the 48th International Conference on Current Trends in Theory and Practice of Computer Science, SOFSEM 2023. The paper was co-authored with P.S. Ardra, R. Krithika, and Saket Saurabh, 2023.
- Danupon Nanongkai received the best paper award for the paper titled “Negative-Weight Single-Source Shortest Paths in Near-Linear Time” at the 63rd IEEE Symposium on Foundations of Computer Science, FOCS 2022. The paper was co-authored with Aaron Bernstein and Christian Wulff-Nilsen, 2022.
- Tomasz Kociumaka received the Best Paper Award (33rd International Workshop on Combinatorial Algorithms (IWOCA), Trier, Germany, June 2022), for the paper “Computing Longest (Common) Lyndon Subsequence” co-authored with H. Bannai, T. I, D. Köppl, S. Puglisi, 2022.
- Tomasz Kociumaka received the Alberto Apostolico Best Paper Award (33rd Annual Symposium on Combinatorial Pattern Matching (CPM), Prague, Czech Republic, June 2022), for the paper “The Dynamic k-Mismatch Problem” co-authored with R. Clifford, P. Gawrychowski, D. Martin, and P. Uznański, 2022.

### 27.11.4 Awards

- Kurt Mehlhorn obtained an Honorary Doctorate Degree from Aalto University, 2023.
- Karol Węgrzycki received the Open Mind Prize, PCC, 2022.
- Karol Węgrzycki received the Prime Minister of Poland Award for PhD Thesis, Prime Minister of Poland, 2022.
- Philip Wellnitz received the Dieter-Rampacher-Prize for the youngest Ph.D. student of the Max Planck Society that graduated in 2021.
- Corinna Coupette received the Caroline von Humboldt Prize for outstanding female junior researchers awarded by Humboldt University Berlin, 2022.
- Corinna Coupette received the Best Reviewer Award at the Learning on Graphs Conference, 2022.
- Karol Węgrzycki received the Witold Lipski Prize, Foundation for Information Technology Development, 2021.
- Karol Węgrzycki received a Distinction for his PhD thesis, University of Warsaw, 2021.

## 27.12 Grants

- Joachim Spoerhase, *Approximation Algorithms for Combinatorial Optimization Problems with Packing Constraints*, DFG Research Grant, 2022–2024.
- Andreas Karrenbauer, Leonie Krull, and Paolo Luigi Rinaldi in collaboration with Ford-Werke GmbH and Saarland University, *Modernisierung und Digitalisierung der Produktionssteuerung des Ford Werkes in Saarlouis zur störungsfreien kontinuierlichen Produktion durch multidimensionale Algorithmen und künstliche Intelligenz (MoDigPro)*, Zentrales Technologieprogramm Saar (ZTS) cofinanced by the European Regional Development Fund (ERDF), 2020–2023.
- Christoph Lenzen, *A Theory of Reliable Hardware (ToRH)*, ERC Starting Grant, European Research Council, 2017–2022.

## 27.13 Publications

### Journal Articles

[1] A. Abboud, K. Bringmann, D. Hermelin, and D. Shabtay. Scheduling lower bounds via AND subset sum. *Journal of Computer and System Sciences*, 127:29–40, 2022.

[2] A. Abboud, K. Bringmann, D. Hermelin, and D. Shabtay. SETH-based lower bounds for Subset Sum and bicriteria path. *ACM Transactions on Algorithms*, 18(1), Article 6, 2022.

[3] A. Agrawal, D. Lokshtanov, P. Misra, S. Saurabh, and M. Zehavi. Erdős-Pósa property of obstructions to interval graphs. *Journal of Graph Theory*, 102(4):702–727, 2022.

[4] G. Amanatidis and P. Kleer. Rapid mixing of the switch Markov chain for 2-class joint degree matrices. *SIAM Journal on Discrete Mathematics*, 36(1):118–146, 2022.

[5] S. A. Amiri and B. Wiederhake. Distributed distance-$r$ dominating set on sparse high-girth graphs. *Theoretical Computer Science*, 906:18–31, 2022.

[6] H. An, M. Gurumukhani, R. Impagliazzo, M. Jaber, M. Künnemann, and M. P. Parga Nina. The fine-grained complexity of multi-dimensional ordering properties. *Algorithmica*, 84:3156–3191, 2022.

[7] R. Becker, S. Forster, A. Karrenbauer, and C. Lenzen. Near-optimal approximate shortest paths and transshipment in distributed and streaming models. *SIAM Journal on Computing*, 50(3):815–856, 2021.

[8] B. A. Berendsohn, L. Kozma, and D. Marx. Finding and counting permutations via CSPs. *Algorithmica*, 83:2552–2577, 2021.

[9] V. Bonifaci, E. Facca, F. Folz, A. Karrenbauer, P. Kolev, K. Mehlhorn, G. Morigi, G. Shahkarami, and Q. Vermande. Physarum-inspired multi-commodity flow dynamics. *Theoretical Computer Science*, 920:1–20, 2022.

[10] M. Briański, G. Joret, K. Majewski, P. Micek, M. T. Seweryn, and R. Sharma. Treedepth vs circumference. *Combinatorica*. Accepted 2023.

[11] K. Bringmann, V. Cohen-Addad, and D. Das. A linear-time $n^{0.4}$-approximation for longest common subsequence. *ACM Transactions on Algorithms*, 19(1), Article 9, 2023.

[12] K. Bringmann, N. Fischer, D. Hermelin, D. Shabtay, and P. Wellnitz. Faster minimization of tardy processing time on a single machine. *Algorithmica*, 84:1341–1356, 2022.

[13] K. Bringmann, R. Keusch, J. Lengler, Y. Maus, and A. R. Molla. Greedy routing and the algorithmic small-world phenomenon. *Journal of Computer and System Sciences*, 125:59–105, 2022.

[14] K. Bringmann, M. Künnemann, and A. Nusser. Discrete Fréchet distance under translation: Conditional hardness and an improved algorithm. *ACM Transactions on Algorithms*, 17(3), Article 25, 2021.

[15] K. Bringmann, M. Künnemann, and A. Nusser. Walking the dog fast in practice: Algorithm engineering of the Fréchet distance. *Journal of Computational Geometry*, 12(1):70–108, 2021.

[16] K. Bringmann and A. Nusser. Translating Hausdorff is hard: Fine-grained lower bounds for Hausdorff distance under translation. *Journal of Computational Geometry*, 13(2):30–50, 2021.

[17] D. Coudert, A. Nusser, and L. Viennot. Enumeration of far-apart pairs by decreasing distance for faster hyperbolicity computation. *ACM Journal of Experimental Algorithmics*, 27, Article 1.15, 2022.

[18] C. Coupette, J. Beckedorf, D. Hartung, M. Bommarito, and D. M. Katz. Measuring law over time: A network analytical framework with an application to statutes and regulations in the United States and Germany. *Frontiers in Physics*, 9, Article 658463, 2021.

[19] C. Coupette and D. Hartung. Rechtsstrukturvergleichung. *Rabels Zeitschrift für ausländisches und internationales Privatrecht*, 86(4):935–975, 2022.

[20] C. Coupette, D. Hartung, J. Beckedorf, M. Bother, and D. M. Katz. Law smells – defining and detecting problematic patterns in legal drafting. *Artificial Intelligence and Law*, 2022.

[21] E. Cruciani, E. Natale, A. Nusser, and G. Scornavacca. Phase transition of the 2-choices dynamics on core-periphery networks. *Distributed Computing*, 34:207–225, 2021.

[22] N. R. Dayama, M. Shiripour, A. Oulasvirta, E. Ivanko, and A. Karrenbauer. Foraging-based optimization of menu systems. *International Journal of Human-Computer Studies*, 151, Article 102624, 2021.

[23] J. Dörfler, M. Roth, J. Schmitt, and P. Wellnitz. Counting induced subgraphs: An algebraic approach to #W[1]-hardness. *Algorithmica*, 2021.

[24] A. Driemel, A. Nusser, J. M. Phillips, and I. Psarros. The VC dimension of metric balls under Fréchet and Hausdorff distances. *Discrete & Computational Geometry*, 2021.

[25] M. Dyer, C. Greenhill, P. Kleer, J. Ross, and L. Stougie. Sampling hypergraphs with given degrees. *Discrete Mathematics*, 344(11), Article 112566, 2021.

[26] S. Fallat, D. Kirkpatrick, H. U. Simon, A. Soltani, and S. Zilles. On batch teaching without collusion. *Journal of Machine Learning Research*, 24:1–33, 2023.

[27] A. M. Feit, M. Nancel, M. John, A. Karrenbauer, D. Weir, and A. Oulasvirta. AZERTY amélioré: Computational design on a national scale. *Communications of the ACM*, 64(2):48–58, 2021.

[28] F. Folz, K. Mehlhorn, and G. Morigi. Interplay of periodic dynamics and noise: Insights from a simple adaptive system. *Physical Review E*, 104(5), Article 054215, 2021.

[29] F. V. Fomin, P. A. Golovach, W. Lochet, P. Misra, S. Saurabh, and R. Sharma. Parameterized complexity of directed spanner problems. *Algorithmica*, 84:2292–2308, 2022.

[30] M. Függer, A. Kinali, C. Lenzen, and B. Wiederhake. Fast all-digital clock frequency adaptation circuit for voltage droop tolerance. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 41(8):2518–2531, 2022.

[31] J. Giliberti and A. Karrenbauer. Improved online algorithm for fractional knapsack in the random order model. In J. Koenemann and B. Preis, eds., *Approximation and Online Algorithms*, Lisbon, Portugal, 2021, LNCS 12982, pp. 188–205. Springer.

[32] D. Halperin, S. Har-Peled, K. Mehlhorn, E. Oh, and M. Sharir. The maximum-level vertex in an arrangement of lines. *Discrete & Computational Geometry*, 67:439–461, 2022.

[33] D. Kirkpatrick, H. U. Simon, and S. Zilles. Optimal collusion-free teaching. *Journal of Machine Learning Research*. arXiv: 1903.04012. Accepted 2022.

[34] P. Kleer and H. U. Simon. Primal and dual combinatorial dimensions. *Discrete Applied Mathematics*, 327:185–196, 2023.

[35] T. Kociumaka, G. Navarro, and N. Prezza. Toward a definitive compressibility measure for repetitive sequences. *IEEE Transactions on Information Theory*, 69(4):2074–2092, 2022.

[36] D. Lokshtanov, P. Misra, J. Mukherjee, F. Panolan, G. Philip, and S. Saurabh. 2-approximating feedback vertex set in tournaments. *ACM Transactions on Algorithms*, 17(2), Article 11, 2021.

[37] J. Madathil, R. Sharma, and M. Zehavi. A sub-exponential FPT algorithm and a polynomial kernel for minimum directed bisection on semicomplete digraphs. *Algorithmica*, 83:1861–1884, 2021.

[38] P. Misra, S. Saket, R. Sharma, and M. Zehavi. Sub-exponential time parameterized algorithms for graph layout problems on digraphs with bounded independence number. *Algorithmica*, 2023.

[39] A. S. Nittala, A. Karrenbauer, A. Khan, T. Kraus, and J. Steimle. Computational design and optimization of electro-physiological sensors. *Nature Communications*, 12(1), Article 6351, 2021.

[40] B. Ray Chaudhury, Y. K. Cheung, J. Garg, N. Garg, M. Hoefer, and K. Mehlhorn. Fair division of indivisible goods for a class of concave valuations. *Journal of Artificial Intelligence Research*, 74:111–142, 2022.

[41] B. Ray Chaudhury, T. Kavitha, K. Mehlhorn, and A. Sgouritsa. A little charity guarantees almost envy-freeness. *SIAM Journal on Computing*, 50(4):1336–1358, 2021.

## Conference Articles

[1] A. Abboud, K. Bringmann, and N. Fischer. Stronger 3-SUM lower bounds for approximate distance oracles via additive combinatorics. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing (STOC 2023)*, Orlando, FL, USA, 2023. ACM. Accepted.

[2] A. Abboud, K. Bringmann, S. Khoury, and O. Zamir. Hardness of approximation in P via short cycle removal: Cycle detection, distance oracles, and beyond. In S. Leonardi and A. Gupta, eds., *STOC '22, 54th Annual ACM Symposium on Theory of Computing*, Rome, Italy, 2022, pp. 1487–1500. ACM.

[3] H. Akrami, B. Ray Chaudhury, M. Hoefer, K. Mehlhorn, M. Schmalhofer, G. Shahkarami, G. Varricchio, Q. Vermande, and E. van Wijland. Maximizing Nash social welfare in 2-value instances. In *Proceedings of the 36th AAAI Conference on Artificial Intelligence*, Virtual Conference, 2022, pp. 4760–4767. AAAI.

[4] H. Akrami, R. Rezvan, and M. Seddighin. An EF2X allocation protocol for restricted additive valuations. In L. de Raedt, ed., *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence (IJCAI 2022)*, Vienna, Austria, 2022, pp. 17–23. IJCAI.

[5] H. An, M. Gurumukhani, R. Impagliazzo, M. Jaber, M. Künnemann, and M. P. Parga Nina. The fine-grained complexity of multi-dimensional ordering properties. In P. A. Golovach and M. Zehavi, eds., *16th International Symposium on Parameterized and Exact Computation (IPEC 2021)*, Lisbon, Portugal, 2021, Leibniz International Proceedings in Informatics 214, Article 3. Schloss Dagstuhl.

[6] I. Anagnostides, C. Lenzen, B. Haeupler, G. Zuzic, and T. Gouleakis. Almost universally optimal distributed Laplacian solvers via low-congestion shortcuts. In C. Scheideler, ed., *36th International Symposium on Distributed Computing (DISC 2022)*, Augusta, GA, USA, 2022, Leibniz International Proceedings in Informatics 246, Article 6. Schloss Dagstuhl.

[7] A. Antoniadis, P. Jabbarzade, and G. Shahkarami. A novel prediction setup for online speed-scaling. In A. Czumai and Q. Xin, eds., *18th Scandinavian Symposium and Workshops on Algorithm Theory (SWAT 2022)*, Tórshavn, Faroe Islands, 2022, Leibniz International Proceedings in Informatics 227, Article 9. Schloss Dagstuhl.

[8] S. Apers, Y. Efron, P. Gawrychowski, T. Lee, S. Mukhopadhyay, and D. Nanongkai. Cut query algorithms with star contraction. In *FOCS 2022, IEEE 63rd Annual Symposium on Foundations of Computer Science*, Denver, CO, USA, 2022, pp. 507–518. IEEE.

[9] P. S. Ardra, R. Krithika, S. Saurabh, and R. Sharma. Balanced substructures in bicolored graphs. In L. Gąsieniec, ed., *SOFSEM 2023: Theory and Practice of Computer Science*, Nový Smokovec, Slovakia, 2023, LNCS 13878, pp. 177–191. Springer.

[10] B. Banyassady, M. de Berg, K. Bringmann, K. Buchin, H. Fernau, D. Halperin, I. Kostitsyna, Y. Okamoto, and S. Slot. Unlabeled multi-robot motion planning with tighter separation bounds. In X. Goaoc and M. Kerber, eds., *38th International Symposium on Computational Geometry (SoCG 2022)*, Berlin, Germany, 2022, Leibniz International Proceedings in Informatics 224, Article 12. Schloss Dagstuhl.

[11] M. Beck, J. Spoerhase, and S. Storandt. Mind the gap: Edge facility location problems in theory and practice. In A. Bagchi and R. Muthu, eds., *Algorithms and Discrete Applied Mathematics (CALDAM 2023)*, Gandhinagar, India, 2023, LNCS 13947, pp. 321–334. Springer.

[12] M. de Berg, S. Kisfaludi-Bak, M. Monemizadeh, and L. Theocharous. Clique-based separators for geometric intersection graphs. In H.-K. Ahn and K. Sadakane, eds., *32nd International Symposium on Algorithms and Computation (ISAAC 2021)*, Fukuoka, Japan, 2021, Leibniz International Proceedings in Informatics 212, Article 22. Schloss Dagstuhl.

[13] A. Bernstein, D. Nanongkai, and C. Wulff-Nilsen. Negative-weight single-source shortest paths in near-linear time. In *FOCS 2022, IEEE 63rd Annual Symposium on Foundations of Computer Science*, Denver, CO, USA, 2022, pp. 600–611. IEEE.

[14] S. Bhattacharya, P. Kiss, and T. Saranurak. Dynamic algorithms for packing-covering LPs via multiplicative weight. In N. Bansal and V. Nagarajan, eds., *Proceedings of the 2023 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA 2023)*, Florence, Italy, 2023, pp. 1–47. SIAM.

[15] S. Bhattacharya, P. Kiss, and T. Saranurak. Sublinear algorithms for $(1.5+\epsilon)$-approximate matching. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing (STOC 2023)*, Orlando, FL, USA, 2023. ACM. Accepted.

[16] S. Bhattacharya, P. Kiss, T. Saranurak, and D. Wajc. Dynamic matching with better-than-2 approximation in polylogarithmic update time. In N. Bansal and V. Nagarajan, eds., *Proceedings of the 2023 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA 2023)*, Florence, Italy, 2023, pp. 100–128. SIAM.

[17] J. Blikstad, T.-W. Tu, D. Nanongkai, and S. Mukhopadhyay. Fast algorithms via dynamic-oracle matroids. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing (STOC 2023)*, Orlando, FL, USA, 2023. ACM. Accepted.

[18] S. Boodaghians, B. Ray Chaudhury, and R. Mehta. Polynomial time algorithms to find an approximate competitive equilibrium for chores. In S. Naor and N. Buchbinder, eds., *Proceedings of the Annual ACM-SIAM Symposium on Discrete Algorithms (SODA 2022)*, Virtual, 2022, pp. 2285–2302. SIAM.

[19] K. Bringmann. Fine-grained complexity theory: Conditional lower bounds for computational geometry. In L. De Mol, A. Weiermann, F. Manea, and D. Fernández-Duque, eds., *Connecting with Computability (CiE 2021)*, Ghent, Belgium (Virtual Event), 2021, LNCS 12813, pp. 60–70. Springer.

[20] K. Bringmann, N. Carmeli, and S. Mengel. Tight fine-grained bounds for direct access on join queries. In L. Libkin, P. Barceló, and A. Grez, eds., *PODS '22, 41st ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems*, Philadelphia, PA, USA, 2022, pp. 427–436. ACM.

[21] K. Bringmann and A. Cassis. Faster Knapsack algorithms via bounded monotone Min-Plus-Convolution. In M. Bojańczyk, E. Merelli, and D. P. Woodruff, eds., *49th EATCS International Conference on Automata, Languages, and Programming (ICALP 2022)*, Paris, France, 2022, Leibniz International Proceedings in Informatics 229, Article 31. Schloss Dagstuhl.

[22] K. Bringmann, A. Cassis, N. Fischer, and M. Künnemann. Fine-grained completeness for optimization in P. In M. Wootters and L. Sanità, eds., *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2021)*, Seattle, WA, USA, 2021, Leibniz International Proceedings in Informatics 207, Article 9. Schloss Dagstuhl.

[23] K. Bringmann, A. Cassis, N. Fischer, and M. Künnemann. A structural investigation of the approximability of polynomial-time problems. In M. Bojańczyk, E. Merelli, and D. P. Woodruff, eds., *49th EATCS International Conference on Automata, Languages, and Programming*

*(ICALP 2022)*, Paris, France, 2022, Leibniz International Proceedings in Informatics 229, Article 30. Schloss Dagstuhl.

[24] K. Bringmann, A. Cassis, N. Fischer, and V. Nakos. Almost-optimal sublinear-time edit distance in the low distance regime. In S. Leonardi and A. Gupta, eds., *STOC '22, 54th Annual ACM Symposium on Theory of Computing*, Rome, Italy, 2022, pp. 1102–1115. ACM.

[25] K. Bringmann, A. Cassis, N. Fischer, and V. Nakos. Improved sublinear-time edit distance for preprocessed strings. In M. Bojańczyk, E. Merelli, and D. P. Woodruff, eds., *49th EATCS International Conference on Automata, Languages, and Programming (ICALP 2022)*, Paris, France, 2022, Leibniz International Proceedings in Informatics 229, Article 32. Schloss Dagstuhl.

[26] K. Bringmann and D. Das. A linear-time $n^{0.4}$-approximation for longest common subsequence. In N. Bansal, E. Merelli, and J. Worrell, eds., *48th International Colloquium on Automata, Languages, and Programming (ICALP 2021)*, Glasgow, UK (Virtual Conference), 2021, Leibniz International Proceedings in Informatics 198, Article 39. Schloss Dagstuhl.

[27] K. Bringmann, A. Driemel, A. Nusser, and I. Psarros. Tight bounds for approximate near neighbor searching for time series under the Fréchet distance. In S. Naor and N. Buchbinder, eds., *Proceedings of the Annual ACM-SIAM Symposium on Discrete Algorithms (SODA 2022)*, Virtual, 2022, pp. 517–550. SIAM.

[28] K. Bringmann, N. Fischer, and V. Nakos. Sparse nonnegative convolution is equivalent to dense nonnegative convolution. In S. Khuller and V. Vassilevska Williams, eds., *STOC '21, 53rd Annual ACM SIGACT Symposium on Theory of Computing*, Virtual, Italy, 2021, pp. 1711–1724. ACM.

[29] K. Bringmann, N. Fischer, and V. Nakos. Deterministic and Las Vegas algorithms for sparse nonnegative convolution. In S. Naor and N. Buchbinder, eds., *Proceedings of the Annual ACM-SIAM Symposium on Discrete Algorithms (SODA 2022)*, Virtual, 2022, pp. 3069–3090. SIAM.

[30] K. Bringmann, M. Kapralov, M. Makarov, V. Nakos, A. Yagudin, and A. Zandieh. Traversing the FFT computation tree for dimension-independent sparse Fourier transforms. In N. Bansal and V. Nagarajan, eds., *Proceedings of the 2023 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA 2023)*, Florence, Italy, 2023, pp. 4768–4845. SIAM.

[31] K. Bringmann, S. Kisfaludi-Bak, M. Künnemann, D. Marx, and A. Nusser. Dynamic time warping under translation: Approximation guided by space-filling curves. In X. Goaoc and M. Kerber, eds., *38th International Symposium on Computational Geometry (SoCG 2022)*, Berlin, Germany, 2022, Leibniz International Proceedings in Informatics 224, Article 20. Schloss Dagstuhl.

[32] K. Bringmann, S. Kisfaludi-Bak, M. Künnemann, A. Nusser, and Z. Parsaeian. Towards sub-quadratic diameter computation in geometric intersection graphs. In X. Goaoc and M. Kerber, eds., *38th International Symposium on Computational Geometry (SoCG 2022)*, Berlin, Germany, 2022, Leibniz International Proceedings in Informatics 224, Article 21. Schloss Dagstuhl.

[33] K. Bringmann and V. Nakos. Fast $n$-fold Boolean convolution via additive combinatorics. In N. Bansal, E. Merelli, and J. Worrell, eds., *48th International Colloquium on Automata, Languages, and Programming (ICALP 2021)*, Glasgow, UK (Virtual Conference), 2021, Leibniz International Proceedings in Informatics 198, Article 41. Schloss Dagstuhl.

[34] K. Bringmann and A. Nusser. Translating Hausdorff is hard: Fine-grained lower bounds for Hausdorff distance under translation. In K. Buchin and É. Colin de Verdière, eds., *37th*

*International Symposium on Computational Geometry (SoCG 2021)*, Buffalo, NY, USA (Virtual Conference), 2021, Leibniz International Proceedings in Informatics 189, Article 18. Schloss Dagstuhl.

[35] K. Bringmann and J. Slusallek. Current algorithms for detecting subgraphs of bounded treewidth are probably optimal. In N. Bansal, E. Merelli, and J. Worrell, eds., *48th International Colloquium on Automata, Languages, and Programming (ICALP 2021)*, Glasgow, UK (Virtual Conference), 2021, Leibniz International Proceedings in Informatics 198, Article 40. Schloss Dagstuhl.

[36] K. Buchin, A. Nusser, and S. Wong. Computing continuous dynamic time warping of time series in polynomial time. In X. Goaoc and M. Kerber, eds., *38th International Symposium on Computational Geometry (SoCG 2022)*, Berlin, Germany, 2022, Leibniz International Proceedings in Informatics 224, Article 22. Schloss Dagstuhl.

[37] P. Charalampopoulos, T. Kociumaka, S. P. Pissis, J. Radoszewski, W. Rytter, T. Waleń, and W. Zuba. Approximate circular pattern matching. In S. Chechik, G. Navarro, E. Rotenberg, and G. Herman, eds., *30th Annual European Symposium on Algorithms (ESA 2022)*, Berlin/Potsdam, Germany, 2022, Leibniz International Proceedings in Informatics 244, Article 35. Schloss Dagstuhl.

[38] P. Charalampopoulos, T. Kociumaka, and P. Wellnitz. Faster pattern matching under edit distance: A reduction to dynamic puzzle matching and the seaweed monoid of permutation matrices. In *FOCS 2022, IEEE 63rd Annual Symposium on Foundations of Computer Science*, Denver, CO, USA, 2022, pp. 698–707. IEEE.

[39] B. R. Chaudhury, J. Garg, K. Mehlhorn, R. Mehta, and P. Misra. Improving EFX guarantees through rainbow cycle number. In P. Biró, S. Chawla, F. Echenique, and E. Sodomka, eds., *EC '21, 22nd ACM Conference on Economics and Computation*, Budapest, Hungary (Virtual), 2021, pp. 310–311. ACM.

[40] D. Coudert, A. Nusser, and L. Viennot. Computing graph hyperbolicity using dominating sets. In C. A. Phillips and B. Speckman, eds., *Proceedings of the Symposium on Algorithm Engineering and Experiments (ALENEX 2022)*, Alexandria, VA, USA, 2022, pp. 78–90. SIAM.

[41] C. Coupette, S. Dalleiger, and B. Rieck. Ollivier-Ricci curvature for hypergraphs: A unified framework. In *Eleventh International Conference on Learning Representations (ICLR 2023)*, Kigali, Rwanda, 2023. OpenReview.net. Accepted.

[42] C. Coupette, S. Dalleiger, and J. Vreeken. Differentially describing groups of graphs. In *Proceedings of the 36th AAAI Conference on Artificial Intelligence*, Virtual Conference, 2022, pp. 3959–3967. AAAI.

[43] C. Coupette, J. Singh, and H. Spamann. Simplify your law: Using information theory to deduplicate legal documents. In B. Xue, M. Pechenizkiy, and Y. S. Koh, eds., *21st IEEE International Conference on Data Mining Workshops (ICDMW 2021)*, Virtual Conference, 2021, pp. 631–638. IEEE.

[44] C. Coupette and J. Vreeken. Graph similarity description: How are these graphs similar? In F. Zhu, C. Ooi, Beng, C. Miao, G. Cong, J. Tang, and T. Derr, eds., *KDD '21, 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, Virtual Event, Singapore, 2021, pp. 185–195. ACM.

[45] D. Das, J. Gilbert, M. Hajiaghayi, T. Kociumaka, and B. Saha. Weighted edit distance computation: Strings, trees, and Dyck. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing (STOC 2023)*, Orlando, FL, USA, 2023. ACM. Accepted.

[46] D. Das, J. Gilbert, M. Hajiaghayi, T. Kociumaka, B. Saha, and H. Saleh. $\tilde{O}(n + \text{poly}(k))$-time algorithm for bounded tree edit distance. In *FOCS 2022, IEEE 63rd Annual Symposium on Foundations of Computer Science*, Denver, CO, USA, 2022, pp. 686–697. IEEE.

[47] D. Das, T. Kociumaka, and B. Saha. Improved approximation algorithms for Dyck edit distance and RNA folding. In M. Bojańczyk, E. Merelli, and D. P. Woodruff, eds., *49th EATCS International Conference on Automata, Languages, and Programming (ICALP 2022)*, Paris, France, 2022, Leibniz International Proceedings in Informatics 229, Article 49. Schloss Dagstuhl.

[48] F. Dross, K. Fleszar, K. Węgrzycki, and A. Zych-Pawlewicz. Gap-ETH-tight approximation schemes for red-green-blue separation and bicolored noncrossing Euclidean travelling salesman tours. In N. Bansal and V. Nagarajan, eds., *Proceedings of the 2023 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA 2023)*, Florence, Italy, 2023, pp. 1433–1463. SIAM.

[49] B. C. Esmer, A. Kulik, D. Marx, D. Neuen, and R. Sharma. Faster exponential-time approximation algorithms using approximate monotone local search. In S. Chechik, G. Navarro, E. Rotenberg, and G. Herman, eds., *30th Annual European Symposium on Algorithms (ESA 2022)*, Berlin/Potsdam, Germany, 2022, Leibniz International Proceedings in Informatics 244, Article 50. Schloss Dagstuhl.

[50] B. C. Esmer, A. Kulik, D. Marx, P. Schepper, and K. Węgrzycki. Computing generalized convolutions faster than brute force. In H. Dell and J. Nederlof, eds., *17th International Symposium on Parameterized and Exact Computation (IPEC 2022)*, Potsdam, Germany, 2022, Leibniz International Proceedings in Informatics 249, Article 12. Schloss Dagstuhl.

[51] O. Firman and J. Spoerhase. Hypergraph representation via axis-aligned point-subspace cover. In P. Mutzel, M. S. Rahman, and Slamin, eds., *WALCOM: Algorithms and Computation*, Jember, Indonesia, 2022, LNCS 13174, pp. 328–339. Springer.

[52] J. Focke, D. Marx, F. M. Inerney, D. Neuen, G. S. Sankar, P. Schepper, and P. Wellnitz. Tight complexity bounds for counting generalized dominating sets in bounded-treewidth graphs. In N. Bansal and V. Nagarajan, eds., *Proceedings of the 2023 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA 2023)*, Florence, Italy, 2023, pp. 3664–3683. SIAM.

[53] E. Galby, L. Khazaliya, F. Mc Inerney, R. Sharma, and P. Tale. Metric dimension parameterized by feedback vertex set and other structural parameters. In S. Szeider, R. Ganian, and A. Silva, eds., *47th International Symposium on Mathematical Foundations of Computer Science (MFCS 2022)*, Vienna, Austria, 2022, Leibniz International Proceedings in Informatics 241, Article 51. Schloss Dagstuhl.

[54] E. Galby, D. Marx, P. Schepper, R. Sharma, and P. Tale. Domination and cut problems on chordal graphs with bounded leafage. In H. Dell and J. Nederlof, eds., *17th International Symposium on Parameterized and Exact Computation (IPEC 2022)*, Potsdam, Germany, 2022, Leibniz International Proceedings in Informatics 249, Article 14. Schloss Dagstuhl.

[55] E. Galby, D. Marx, P. Schepper, R. Sharma, and P. Tale. Parameterized complexity of weighted multicut in trees. In M. A. Bekos and M. Kaufmann, eds., *Graph-Theoretic Concepts in Computer Science (WG 2022)*, Tübingen, Germany, 2022, LNCS 13453, pp. 257–270. Springer.

[56] J. Giliberti and A. Karrenbauer. Improved online algorithm for fractional knapsack in the random order model. In J. Koenemann and B. Preis, eds., *Approximation and Online Algorithms*, Lisbon, Portugal, 2021, LNCS 12982, pp. 188–205. Springer.

[57] A. Gionis, K. Khodamoradi, B. Ordozgoiti, B. Riegel, and J. Spoerhase. A constant-factor approximation algorithm for reconciliation $k$-median. In *Proceedings of the 26th International*

*Conference on Artificial Intelligence and Statistics (AISTATS 2023)*, Valencia, Spain, 2023, Proceedings of the Machine Learning Research. PMLR. Accepted.

[58] E. Goldenberg, T. Kociumaka, R. Krauthgamer, and B. Saha. Gap edit distance via non-adaptive queries: Simple and optimal. In *FOCS 2022, IEEE 63rd Annual Symposium on Foundations of Computer Science*, Denver, CO, USA, 2022, pp. 674–685. IEEE.

[59] E. Goldenberg, T. Kociumaka, R. Krauthgamer, and B. Saha. An algorithmic bridge between Hamming and Levenshtein distances. In Y. Tauman Kalai, ed., *14th Innovations in Theoretical Computer Science Conference (ITCS 2023)*, Cambridge, MA, USA, 2023, Leibniz International Proceedings in Informatics, Article 58. Schloss Dagstuhl.

[60] G. Goranci, M. Henzinger, D. Nanongkai, T. Saranurak, M. Thorup, and C. Wulff. Fully dynamic exact edge connectivity in sublinear time. In N. Bansal and V. Nagarajan, eds., *Proceedings of the 2023 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA 2023)*, Florence, Italy, 2023, pp. 70–86. SIAM.

[61] T. Gouleakis, K. Lakis, and G. Shahkarami. Learning-augmented algorithms for online TSP on the line. In *Proceedings of the 37th AAAI Conference on Artificial Intelligence*, Washington, DC, USA, 2023. AAAI. Accepted.

[62] G. Gutowski, F. Mittelstädt, I. Rutter, J. Spoerhase, A. Wolff, and J. Zink. Coloring mixed and directional interval graphs. In P. Angelini and R. von Hanxleden, eds., *Graph Drawing and Network Visualization (GD 2022)*, Tokyo, Japan, 2022, LNCS 13764, pp. 418–431. Springer.

[63] I. Han, A. Zandieh, and H. Avron. Random Gegenbauer features for scalable kernel methods. In K. Chaudhuri, S. Jegelka, S. Le, S. Csaba, N. Gang, and S. Sabato, eds., *Proceedings of the 39th International Conference on Machine Learning (ICML 2022)*, Baltimore, MA, USA, 2022, Proceedings of the Machine Learning Research 162, pp. 8330–8358. https://proceedings.mlr.press/v162/han22g.html.

[64] I. Han, A. Zandieh, J. Lee, R. Novak, L. Xiao, and A. Karbasi. Fast neural kernel embeddings for general activations. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, eds., *Advances in Neural Information Processing Systems 35 (NeurIPS 2022)*, New Orleans, LA, USA, 2022, pp. 35657–35671. Curran Associates, Inc.

[65] M. Hatzel, L. Jaffke, P. T. Lima, T. Masařík, M. Pilipczuk, R. Sharma, and M. Sorge. Fixed-parameter tractability of DIRECTED MULTICUT with three terminal pairs parameterized by the size of the cutset: Twin-width meets flow-augmentation. In N. Bansal and V. Nagarajan, eds., *Proceedings of the 2023 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA 2023)*, Florence, Italy, 2023, pp. 3229–3244. SIAM.

[66] Y. Jiang and S. Mukhopadhyay. Finding a small vertex cut on distributed networks. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing (STOC 2023)*, Orlando, FL, USA, 2023. ACM. Accepted.

[67] Y. Jiang and C. Zheng. Robust and optimal contention resolution without collision detection. In K. Agrawal and I.-T. A. Lee, eds., *SPAA '22, 34th ACM Symposium on Parallelism in Algorithms and Architectures*, Philadelphia, PA, USA, 2022, pp. 107–118. ACM.

[68] D. Kempa and T. Kociumaka. Breaking the $O(n)$-Barrier in the construction of compressed suffix arrays. In N. Bansal and V. Nagarajan, eds., *Proceedings of the 2023 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA 2023)*, Florence, Italy, 2023, pp. 5122–5202. SIAM.

[69] S. Kisfaludi-Bak, J. Nederlof, and K. Węgrzycki. A Gap-ETH-tight approximation scheme for Euclidean TSP. In *FOCS 2021, IEEE 62nd Annual Symposium on Foundations of Computer Science*, Denver, CO, USA (Virtual Conference), 2022, pp. 351–362. IEEE.

[70] P. Kleer. Sampling from the Gibbs distribution in congestion games. In P. Biró, S. Chawla, F. Echenique, and E. Sodomka, eds., *EC '21, 22nd ACM Conference on Economics and Computation*, Budapest, Hungary (Virtual), 2021, pp. 679–680. ACM.

[71] T. Kociumaka, G. Navarro, and F. Olivares. Near-optimal search time in $\delta$-optimal space. In A. Castañeda and F. Rodríguez-Henríquez, eds., *LATIN 2022: Theoretical Informatics*, Guanajuato, Mexico, 2022, LNCS 13568, pp. 88–103. Springer.

[72] R. Krithika, R. Sharma, and P. Tale. The complexity of contracting bipartite graphs into small cycles. In M. A. Bekos and M. Kaufmann, eds., *Graph-Theoretic Concepts in Computer Science (WG 2022)*, Tübingen, Germany, 2022, LNCS 13453, pp. 356–369. Springer.

[73] M. Künnemann and A. Nusser. Polygon placement revisited: (degree of freedom + 1)-SUM hardness and an improvement via offline dynamic rectangle union. In S. Naor and N. Buchbinder, eds., *Proceedings of the Annual ACM-SIAM Symposium on Discrete Algorithms (SODA 2022)*, Virtual, 2022, pp. 3181–3201. SIAM.

[74] C. Lenzen and H. Vahidi. Approximate minimum directed spanning trees under congestion. In T. Jurdziński and S. Schmid, eds., *Structural Information and Communication Complexity (SIROCCO 2021)*, Wrocław, Poland (Online), 2021, LNCS 12810, pp. 352–369. Springer.

[75] J. Li, D. Nanongkai, D. Panigrahi, and T. Saranurak. Near-linear time approximations for cut problems via fair cuts. In N. Bansal and V. Nagarajan, eds., *Proceedings of the 2023 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA 2023)*, Florence, Italy, 2023, pp. 240–257. SIAM.

[76] F. Mansouri, H. U. Simon, A. Singla, and S. Zilles. On batch teaching with sample complexity bounded by VCD. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, eds., *Advances in Neural Information Processing Systems 35 (NeurIPS 2022)*, New Orleans, LA, USA, 2022, pp. 15732–15742. Curran Associates, Inc.

[77] F. Mazowiecki, H. Sinclair-Banks, and K. Węgrzycki. Coverability in 2-VASS with one unary counter is in NP. In O. Kupferman and P. Sobocinski, eds., *Foundations of Software Science and Computation Structures (FoSSaCS 2023)*, Paris, France, 2023, LNCS 13992, pp. 196–217. Springer.

[78] J. Nederlof, M. Pilipczuk, C. M. F. Swennenhuis, and K. Węgrzycki. Isolation schemes for problems on decomposable graphs. In P. Berenbrink and B. Monmege, eds., *39th International Symposium on Theoretical Aspects of Computer Science (STACS 2022)*, Marseille, France (Virtual Conference), 2022, Leibniz International Proceedings in Informatics 219, Article 50. Schloss Dagstuhl.

[79] J. Nederlof and K. Węgrzycki. Improving Schroeppel and Shamir's algorithm for subset sum via orthogonal vectors. In S. Khuller and V. Vassilevska Williams, eds., *STOC '21, 53rd Annual ACM SIGACT Symposium on Theory of Computing*, Virtual, Italy, 2021, pp. 1670–1683. ACM.

[80] J. Olkowski, M. Pilipczuk, M. Rychlicki, K. Węgrzycki, and A. Zych-Pawlewicz. Dynamic data structures for parameterized string problems. In P. Berenbrink, P. Bouyer, A. Dawar, and M. M. Kanté, eds., *40th International Symposium on Theoretical Aspects of Computer Science (STACS 2023)*, Hamburg, Germany, 2023, Leibniz International Proceedings in Informatics 254, Article 50. Schloss Dagstuhl.

[81] A. Polak, L. Rohwedder, and K. Węgrzycki. Knapsack and subset sum with small items. In N. Bansal, E. Merelli, and J. Worrell, eds., *48th International Colloquium on Automata, Languages, and Programming (ICALP 2021)*, Glasgow, UK (Virtual Conference), 2021, Leibniz International Proceedings in Informatics 198, Article 106. Schloss Dagstuhl.

[82] H. U. Simon. Tournaments, Johnson graphs, and NC-teaching. In S. Agrawal and F. Orabona, eds., *Proceedings of the 34th International Conference on Algorithmic Learning Theory (ALT 2023)*, Singapore, Singapore, 2023, Proceedings of the Machine Learning Research 201, pp. 1411–1428. PMLR.

[83] D. Woodruff and A. Zandieh. Leverage score sampling for tensor product matrices in input sparsity time. In K. Chaudhuri, S. Jegelka, S. Le, S. Csaba, N. Gang, and S. Sabato, eds., *Proceedings of the 39th International Conference on Machine Learning (ICML 2022)*, Baltimore, MA, USA, 2022, Proceedings of the Machine Learning Research 162, pp. 23933–23964. https://proceedings.mlr.press/v162/woodruff22a.html.

[84] A. Zandieh, I. Han, H. Avron, N. Shoham, C. Kim, and J. Shin. Scaling neural tangent kernels via sketching and random features. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P. S. Liang, and J. Wortman Vaughan, eds., *Advances in Neural Information Processing Systems 34 (NeurIPS 2021)*, Virtual, 2021, pp. 1062–1073. Curran Assoicates, Inc.

## Technical Reports and Preprints

[1] H. Akrami, N. Alon, B. Ray Chaudhury, J. Garg, K. Mehlhorn, and R. Mehta. *EFX Allocations: Simplifications and Improvements*, 2022. arXiv: 2205.07638.

[2] H. Akrami, B. Ray Chaudhury, M. Hoefer, K. Mehlhorn, M. Schmalhofer, G. Shahkarami, G. Varricchio, Q. Vermande, and E. van Wijland. *Maximizing Nash Social Welfare in 2-Value Instances: The Half-Integer Case*, 2022. arXiv: 2207.10949.

[3] G. Amanatidis and P. Kleer. *Approximate Sampling and Counting of Graphs with Near-Regular Degree Intervals*, 2021. arXiv: 2110.09068.

[4] I. Anagnostides, C. Lenzen, B. Haeupler, G. Zuzic, and T. Gouleakis. *Almost Universally Optimal Distributed Laplacian Solvers via Low-Congestion Shortcuts*, 2021. arXiv: 2109.05151.

[5] A. Antoniadis, R. Hoeksma, S. Kisfaludi-Bak, and K. Schewior. *Online Search for a Hyperplane in High-Dimensional Euclidean Space*, 2021. arXiv: 2109.04340.

[6] S. Bhattacharya, P. Kiss, and T. Saranurak. *Dynamic* $(1+\epsilon)$*-Approximate Matching Size in Truly Sublinear Update Time*, 2023. arXiv: 2302.05030.

[7] J. Blikstad and P. Kiss. *Incremental* $(1-\epsilon)$*-approximate dynamic matching in* $O(poly(1/\epsilon))$ *update time*, 2023. arXiv: 2302.08432.

[8] M. Caoduro, J. Cslovjecsek, M. Pilipczuk, and K. Węgrzycki. *Independence Number of Intersection Graphs of Axis-Parallel Segments*, 2022. arXiv: 2205.15189.

[9] C. Coupette and D. Hartung. *Sharing and Caring: Creating a Culture of Constructive Criticism in Computational Legal Studies*, 2022. arXiv: 2205.01071.

[10] C. Coupette, J. Vreeken, and B. Rieck. *All the World's a (Hyper)Graph: A Data Drama*, 2022. arXiv: 2206.08225.

[11] J. Cslovjecsek, M. Pilipczuk, and K. Węgrzycki. *Parameterized Approximation for Maximum Weight Independent Set of Rectangles and Segments*, 2022. arXiv: 2212.01620.

[12] E. Galby, D. Marx, P. Schepper, R. Sharma, and P. Tale. *Domination and Cut Problems on Chordal Graphs with Bounded Leafage*, 2022. arXiv: 2208.02850.

[13] Y. Gao, H. Kamkari, A. Karrenbauer, K. Mehlhorn, and M. Sharifi. *Physarum Inspired Dynamics to Solve Semi-Definite Programs*, 2022. arXiv: 2111.02291.

[14] Y. Gao, R. Kyng, and D. A. Spielman. *Robust and Practical Solution of Laplacian Equations by Approximate Elimination*, 2023. arXiv: 2303.00709.

[15] L. Jaffke, P. T. Lima, and R. Sharma. *b-Coloring Parameterized by Pathwidth is XNLP-complete*, 2022. arXiv: 2209.07772.

[16] A. Karrenbauer, L. Krull, K. Mehlhorn, P. Misra, P. L. Rinaldi, and A. Twelsiek. *Improving Order with Queues*, 2022. arXiv: 2207.02476.

[17] E. J. Kim, T. Masařík, M. Pilipczuk, R. Sharma, and M. Wahlström. *On Weighted Graph Separation Problems and Flow-Augmentation*, 2022. arXiv: 2208.14841.

[18] J. Nederlof, M. Pilipczuk, and K. Węgrzycki. *Bounding Generalized Coloring Numbers of Planar Graphs Using Coin Models*, 2022. arXiv: 2201.09340.

[19] J. Nederlof, C. M. F. Swennenhuis, and K. Węgrzycki. *Makespan Scheduling of Unit Jobs with Precedence Constraints in $O(1.995^n)$ time*, 2022. arXiv: 2208.02664.

[20] A. Zandieh, I. Han, and H. Avron. *Near Optimal Reconstruction of Spherical Harmonic Expansions*, 2022. arXiv: 2202.12995.

[21] A. Zandieh, I. Han, M. Daliri, and A. Karbasi. *KDEformer: Accelerating Transformers via Kernel Density Estimation*, 2023. arXiv: 2302.02451.

## Theses

[1] B. R. Chaudhury. *Finding Fair and Efficient Allocations.* PhD thesis, Universität des Saarlandes, Saarbrücken, 2021.

[2] A. Kinali-Dogan. *On Time, Time Synchronization and Noise in Time Measurement Systems.* PhD thesis, Universität des Saarlandes, Saarbrücken, 2022.

[3] A. Nusser. *Fine-Grained Complexity and Algorithm Engineering of Geometric Similarity Measures.* PhD thesis, Universität des Saarlandes, Saarbrücken, 2022.

[4] A. Pandey. *Variety Membership Testing in Algebraic Complexity Theory.* PhD thesis, Universität des Saarlandes, Saarbrücken, 2021.

[5] P. Wellnitz. *Counting Patterns in Strings and Graphs.* PhD thesis, Universität des Saarlandes, Saarbrücken, 2021.

[6] B. Wiederhake. *Pulse Propagation, Graph Cover, and Packet Forwarding.* PhD thesis, Universität des Saarlandes, Saarbrücken, 2022.

# 28 D2: Computer Vision and Machine Learning

## 28.1 Personnel

**Head of Group**

Prof. Dr. Bernt Schiele

**Senior Researchers and Research Group Leaders**

Prof. Dr. Zeynep Akata (affiliated)
Dr. Dengxin Dai (Until January 2023)
Prof. Dr. Margret Keuper (affiliated since November 2021)
Dr. Jan Eric Lenssen
Prof Dr. Gerard Pons-Moll (affiliated)
Dr. Paul Swoboda (until August 2022)

**Researchers**

Dr. Jiangxin Dong (until September 2021)
Dr. Xinting Hu (since June 2022)
Dr. Li Jiang (since October 2021)
Dr. Shaoshuai Shi (since October 2021)
Dr. Yongqin Xian (until July 2021)
Dr. Guo Yong (since July 2021)

**Ph.D. Students**

Ahmed Abbas
Stephan Alaniz (until September 2021)
Bharat Lal Bhatnagar (until July 2021)
Apratim Bhattacharyya (until June 2021)
Moritz Böhle
Julian Chibane
Anurag Das (since October 2021)
Yue Fan
Siddhartha Gairola (since September 2022)
Vladimir Guzov
Xudong Hong (until December 2022)
Andrea Hornakova (until April 2022)
Jiang Jiaxi (until November 2022)
Steffen Jung
Anna Kukleva

Verica Lazova (until August 2021)
Zhi Li (since June 2022)
Yaoyao Liu (until January 2023)
Max Losch
Jovita Lukasik (since March 2022)
Aymen Mir (until February 2022)
Mohamed Omran (until August 2022)
Farzaneh Rezaeianaran (until September 2021)
Sukrut Rao (since September 2022)
Edgar Schönfeld
Mattia Segu (since September 2022)
David Stutz (until March 2022)
Garvita Tiwari
Haoran Wang (since December 2022)
Christopher Wewer (since July 2022)
Ning Yu (until September 2021)
Raza Yunus (since December 2022)
Xiaohan Zhang
Keyang Zhou

**Secretaries**
Connie Balzert

## 28.2 Visitors

From March 2021 to February 2023, the following researchers visited our group:

| | | |
|---|---|---|
| Alkis Gkotovos | 31.03.2021 | MIT Computer Science & Artificial Intelligence Lab, remote |
| Radek Mackowiak | 08.07.2021 | Heidelberg University, remote |
| Lynton Ardizzone | 08.07.2021 | Heidelberg University, remote |
| Ullrich Köthe | 08.07.2021 | Heidelberg University, remote |
| Jiang Li | 20.07.2021 | Chinese University of Hong Kong, remote |
| Haoran Wang | 28.07.2021 | ETH Zurich, SenseTime, remote |
| Shaoshuai Shi | 04.08.2021 | Chinese University of Hong Kong, remote |
| Jiaxi Jiang | 07.10.2021 | RWTH Aachen University, remote |
| Alexey Dosovitskiy | 02.11.2021 | Google AI Brain, remote |
| Hanwen Liang | 16.12.2021 | Huawei Noah's Ark Lab, remote |
| Ankit Singh | 26.01.2022 | Indian Institute of Technology, Madras, remote |
| Saman Motamed | 27.01.2022 | Carnegie Mellon's Robotics Institute, remote |

| | | |
|---|---|---|
| Michael Dorkenwald | 27.01.2022 | Heidelberg Collaboratory for Image Processing (HCI), remote |
| Siddhartha Gairola | 28.01.2022 | Microsoft Research India, remote |
| Christopher Wewer | 03.02.2022 | RWTH Aachen University, remote |
| Vedant Nanda | 24.02.2022 | MPI-SWS and University of Maryland, remote |
| Kai Xu | 09.03.2022 | National University of Singapore, remote |
| Yutong Chen | 24.03.2022 | Tsinghua University, remote |
| Jiaming Han | 24.03.2022 | Wuhan University, remote |
| Rowan Zellers | 11.04.2022 | University of Washington |
| Yulun Zhang | 31.05.2022 | ETH Zurich |
| Nina Shvetsova | 02.06.2022 | Goethe-Universität Frankfurt |
| Chenyang Si | 11.08.2022 | Sea AI Lab, Singapore, remote |
| Francesco Locatello | 03.-04.11.2022<br>09.-10.03.2023 | Amazon Web Services, Tübingen |
| Zihang Lai | 01.12.2022 | Carnegie Mellon University, remote |
| Almut-Sophia Köpke | 06.12.2022 | University of Tübingen |
| Christian Rupprecht | 20.01.2023 | University of Oxford |
| Iro Laina | 20.01.2023 | University of Oxford |
| Jiahao Xie | 24.03.2023 | Nanyang Technological University, remote |

## 28.3 Group Organization

The group currently consists of the following seven research groups, led by a research group leader:

- Computer Vision and Machine Learning (Bernt Schiele)
- Vision for Autonomous Systems (Dengxin Dai)
- Real Virtual Humans (Gerard Pons-Moll)
- Multimodal Deep Learning (Zeynep Akata)
- Robust Visual Learning (Margret Keuper)
- Combinatorial Computer Vision (Paul Swoboda)
- Geometric Representation Learning (Jan Eric Lenssen)

The entire group meets once per week for scientific talks and discussions. Any organizational matters are also discussed at these meetings. During the pandemic, meetings were mostly online, however recently we have shifted back to office or hybrid meetings. In fact the vast majority of researchers are back in the office every day. Senior researchers and research group leaders regularly meet with their respective research group to discuss more specific topics related to the research of their group. PhD-students and PostDocs have a weekly or bi-weekly scientific meeting with their respective supervisors. For many of the PhD-students there are also other collaborating researchers present during the discussions.

Following the tradition from previous research groups of Bernt Schiele (TU Darmstadt and ETH Zurich) the group normally organizes two retreats per year. These retreats were stalled during the pandemic and happened again in the second half of the reporting period (July, 2022 and January, 2023). They are important to foster scientific and social interactions between all group members. During these retreats every PhD-student and researcher gives a presentation about his or her current scientific work with an emphasis on the current and future work for the next six to twelve months. An important component during the retreats is that the entire group gives feedback and discusses openly. There is a general rule that the presentations should be between 15-20 min, to allow ample time for discussions. About a third and up to half of the time of every retreat is dedicated to social activities.

The selection of PhD-students and PostDocs includes discussions, interviews and demos with a significant number of the current students, researchers, and research group leaders. Decisions are taken by the research group leaders and Bernt Schiele.

## 28.4 Computer Vision & Machine Learning

*Coordinator: Bernt Schiele*

Since the establishment of D2 at the Max Planck Institute for Informatics, the sub-group headed by Bernt Schiele has been working at the intersection of computer vision and machine learning. As a result of the increased importance of machine learning, in particular for our research but also overall in computer science, D2 itself and the sub-group have been renamed in *computer vision and machine learning* during the last reporting period. Therefore, in the following, we will discuss research in the area of computer vision, in the area of machine learning, and at the intersection of these two areas.

### 28.4.1 Interpretable Machine Learning

*Investigators: Moritz Böhle, Amin Parchami, Sukrut Rao, and Bernt Schiele, in cooperation with Mario Fritz (CISPA Helmholtz Center for Information Security)*

Despite their success for a wide variety of different tasks, it is still difficult to understand the 'decision making process' within deep neural networks and how information is aggregated and processed. Given how ubiquitously these models are employed in our everyday lives, however, it is of paramount importance to gain a better understanding of their inner workings. Especially in safety-critical tasks—such as autonomous driving, health care, or in the judicial system—one needs to ensure that decisions are made for the right reasons.

In our work on Interpretable Machine Learning, we approached this problem setting from a variety of different angles. On the one hand, we performed an in-depth evaluation of the state-of-the art post-hoc attribution methods to understand their promises and shortcomings. On the other hand, we developed *inherently interpretable* deep neural networks (DNNs) by including the goal of interpretability in the optimisation process, thereby foregoing the need for post-hoc explanations. In this context, we evaluated and developed *model guidance* techniques that allow us to ensure that the models are indeed right for the right reasons. Finally, we explored how to increase the inherent interpretability of *conventional* models: for this, we propose to fine-tune conventional DNNs to use more human-interpretable concepts by inserting *semantic bottlenecks* into the model architectures.

### Evaluating post-hoc explanation methods

*Investigators: Sukrut Rao, Moritz Böhle, and Bernt Schiele*

In order to make neural network models more interpretable, there has been a growing interest in recent years in deriving explanations of model predictions. Typically, the goal of such explanation methods is to better understand the output of fixed pre-trained models—i.e., the explanations are derived *post-hoc*. However, given the lack of a 'ground truth' explanation, it is unclear whether these explanations can in fact be trusted.

To better understand differences between such attribution methods and to identify potential shortcomings, in [6, 8] we perform an in-depth evaluation of a wide range of attribution methods in controlled settings. Specifically, by carefully controlling the flow of information within the network, we are able to make definitive statements regarding *provably incorrect attributions*. By doing so, we are able to convincingly show that some attribution methods do not faithfully describe the model they are supposed to explain. Additionally, we develop visualisation and analysis tools that allow for a systematic and fair comparison between attribution methods. This enables us (1) to show that when compared fairly many attribution methods perform more similarly than previously reported, (2) to make recommendations regarding the usage of specific attribution methods, and (3) to succinctly visualise the full spectrum of attribution maps across a dataset, thereby avoiding the problem of relying on just a few qualitative examples when analysing a given attribution method.

### Developing inherently interpretable models

*Investigators: Moritz Böhle and Bernt Schiele, in cooperation with Mario Fritz (CISPA Helmholtz Center for Information Security)*

As discussed above, post-hoc explanation methods might not faithfully describe the underlying model and sometimes even attribute importance to features that *provably* (i.e., by construction of the problem setting) cannot have influenced the model output.

In [2, 3, 1, 4], we therefore take a different approach to address this problem and develop novel neural network architectures that are inherently much more interpretable than standard architectures. For this, we take inspiration from simple linear models, in which the contributions from individual input dimensions to the final output are trivial to obtain. We generalise this idea to *dynamic linear models* in which the linear function applied to the input is itself *input-dependent*—importantly, we show that these *dynamic linear functions* can be constructed in such a way that the model weights align with task-relevant patterns in the input during optimisation, i.e., the model weights exhibit a strong correlation with relevant patterns in the input. As a result, the linear weights can be used to visualise which parts of the inputs the model was able to 'reconstruct'.

Based on this idea, we developed two variants of 'dynamic linear models': the Convolutional Dynamic Alignment Networks [2, 3] and, as a more efficient instantiation of the same underlying principle, the B-cos Networks [1, 4]. Importantly, we not only show that such

Figure 28.1: Qualitative examples of the model-inherent explanations (bottom row) derived from a B-cos Network for samples from the ImageNet dataset (top row).

networks exhibit a high degree of inherent interpretability (see e.g. Fig. 28.1) and constitute performant classifiers, but also that (in the case of B-cos Networks) this approach is compatible with a wide range of convolutional neural networks [1], as well as vision transformers [4].

### Using Explanations to Guide Models

*Investigators: Sukrut Rao, Moritz Böhle, Amin Parchami, and Bernt Schiele*

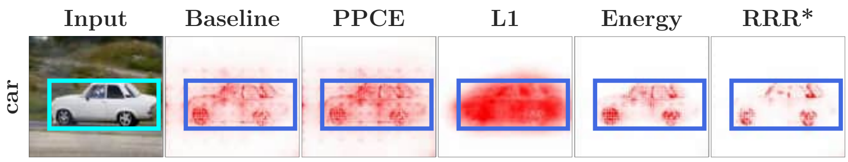



Figure 28.2: Qualitative comparison between various loss functions for model guidance. We find that the energy loss we propose (second to last column) not only yields visually convincing results, but also provides the highest quantitative gains in object guiding models to not focus on background features.

In order to not only understand whether models use the right features to arrive at their conclusions (e.g., via inherently interpretable models, see previous section), but rather to *ensure* it, in [7], we investigate various techniques for 'guiding' DNN models. In particular, we propose a new loss function that encourages the models' explanations to agree with human-annotated bounding and compare it to previously proposed approaches for model guidance on large scale datasets. We find that by not imposing a uniformity prior within the bounding box as is done in many other approaches, we are able to obtain more fine-grained explanations that focus on object-specific features, see e.g. Fig. 28.2 and 28.3.

### Semantic Bottlenecks: Quantifying & Improving Inspectability of Deep Representations

Apart from post-hoc explanations and designing inherently interpretable models from scratch, we evaluate yet another dimension for increasing the interpretability of DNN-based decision systems: modifying already trained models to allow for a higher degree of inspectability. Specifically, in our work on Semantic Bottlenecks [5], we aim to address the problem that the concepts on which neural networks base their decisions are difficult to inspect and evaluate.

In particular, we argue that there are at least two reasons making the inspectability of learnt concepts challenging: (i) representations are distributed across hundreds of channels


Figure 28.3: Effect of model guidance. For various examples from the PASCAL VOC dataset (first row), we show the attributions of a B-cos Network before (second row) and after applying our proposed model guidance loss (third row).

and (ii) a unifying metric quantifying such inspectability is lacking. In order to address these problems, we introduce Semantic Bottlenecks (SB) into neural networks and propose the model-agnostic AUiC metric to quantify the alignment between channel activations and human-understandable visual concepts. In details, we show that the SBs can be integrated into pre-trained networks without much impact on model performance, whilst drastically reducing the number of feature dimensions. We present a case study on semantic segmentation to demonstrate that SBs improve the AUiC up to four-fold over regular network outputs. For this, we explore two types of SB-layers: concept-supervised SB-layers (SSB), for which channel activations are explicitly trained to align with pre-defined visual concepts, and unsupervised SBs (USB) without such explicit supervision. While the SSBs offer the greatest inspectability, we show that the USBs can match the SSBs by producing one-hot encodings. Importantly, for both SB types, we can recover state of the art segmentation performance despite a drastic dimensionality reduction from 1000s of non aligned channels to 10s of semantically aligned channels that all downstream results are based on. For a qualitative visualisation of the improved coherence in activation patterns, see Fig. 28.4.

## References


- [1] M. Böhle, M. Fritz, and B. Schiele. B-cos networks: Alignment is all we need for interpretability. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2022)*, New Orleans, LA, USA, 2022, pp. 10319–10328. IEEE.
- [2] M. D. Böhle, M. Fritz, and B. Schiele. Convolutional dynamic alignment networks for interpretable classifications. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2021)*, Nashville, TN, USA (Virtual), 2021, pp. 10029–10038. IEEE.
- [3] M. D. Böhle, M. Fritz, and B. Schiele. Optimising for interpretability: Convolutional dynamic alignment networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022.

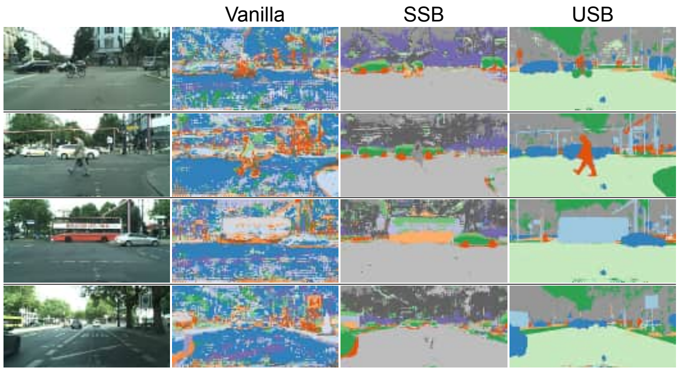



Figure 28.4: Qualitative comparison of the activations of the 20 best semantically aligned channels in the pyramid layer of a PSPNet for different input images and three different settings: the vanilla PSPNet, the PSPNet fine-tuned with concept-supervised Semantic Bottlenecks (SSBs) and with unsupervised Semantic Bottlenecks (USBs). We observe much greater coherence in channel activations for both SB types compared to the vanilla PSPNet.


- [4] M. D. Böhle, M. Fritz, and B. Schiele. Holistically Explainable Vision Transformers. *arXiv*, 2023. arxiv.org/abs/2301.08669.
- [5] M. Losch, M. Fritz, and B. Schiele. Semantic bottlenecks: Quantifying and improving inspectability of deep representations. *International Journal of Computer Vision*, 129:3136–3153, 2021.
- [6] S. Rao, M. Böhle, and B. Schiele. Towards better understanding attribution methods. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2022)*, New Orleans, LA, USA, 2022, pp. 10213–10222. IEEE.
- [7] S. Rao, M. D. Böhle, A. Parchami, and B. Schiele. Using Explanations to Guide Models. *arXiv*, 2023. Under submission at ICCV 2023, IEEE/CVF International Conference on Computer Vision).
- [8] S. Rao, M. D. Böhle, and B. Schiele. Better Understanding Differences in Attribution Methods via Systematic Evaluations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023. Under submission. Preprint available at arxiv.org/abs/2303.11884.


### 28.4.2 3D Scene Understanding

*Investigators: Li Jiang, Shaoshuai Shi, Dengxin Dai, Vladislav Golyanik, Bernt Schiele, in cooperation with Hongsheng Li, Jiaya Jia (Chinese University of Hong Kong), Xiaojuan Qi (University of Hong Kong), and Haiyang Wang, Liwei Wang (Peking University)*

3D scene understanding aims to enable machines to perceive, interpret, and reason about the 3D world in a way that mimics human perception. As the world we live in is inherently three-dimensional, understanding the spatial structures and relationships between 3D objects is essential for a wide range of applications, such as robotics, autonomous driving, augmented and virtual reality, and smart city planning. In this section, we present our contributions for more accurate and efficient 3D scene understanding. Our works encompass various aspects,

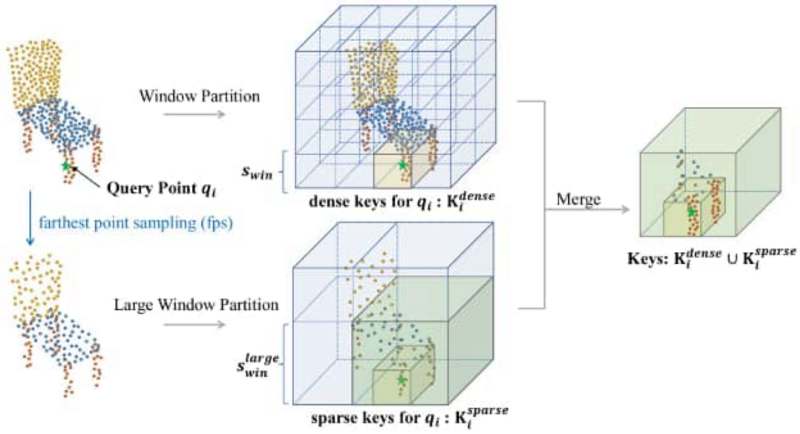



Figure 28.5: Illustration of the stratified strategy for key sampling.

including the development of effective backbone architectures, efficient representation learning in point clouds, object-level understanding for 3D indoor scenarios, and perception and prediction for autonomous driving systems.

### Backbone Architectures for 3D Point Cloud Understanding

Inspired by the success of transformer structure in natural language processing and image recognition, we have developed a series of 3D transformer-based backbone architectures. The self-attention mechanism, which is the core of the transformer structure, provides a powerful means to process input elements in a permutation-invariant manner, and thus is particularly suitable for feature extraction for unordered point clouds. In our Stratified Transformer [3], we demonstrate that the standard dot-product attention layers can also be applied to 3D point clouds with excellent performance. Instead of performing global attention on the whole point cloud, Stratified Transformer partitions the 3D scene into local windows and performs self-attention operations in each window. Also, to capture long-range contextual information, for each query point, Stratified Transformer samples key points in a stratified manner, with nearby points densely sampled and distant points sparsely sampled, as shown in Fig. 28.5. In this way, the model is equipped with a large effective receptive field at a low computational cost. Furthermore, we introduce a memory-efficient implementation to overcome the issue of varying point numbers in each window. Extensive experiments on object-level and indoor datasets, including S3DIS, ScanNetv2 and ShapeNetPart, demonstrate the effectiveness and superiority of Stratified Transformer.

In DSVT [7], we focus on designing an efficient transformer architecture for outdoor autonomous driving scenarios. To support sparse feature learning from raw point clouds, previous methods mainly apply customized sparse operations, such as PointNet++ and sparse convolution. These specifically-designed operations generally can not be implemented with well-optimized deep learning tools (*e.g.*, PyTorch and TensorFlow) and require writing customized CUDA codes, which greatly limits their deployment in real-world applications. To overcome this limitation, we propose Dynamic Sparse Window Attention, a window-based

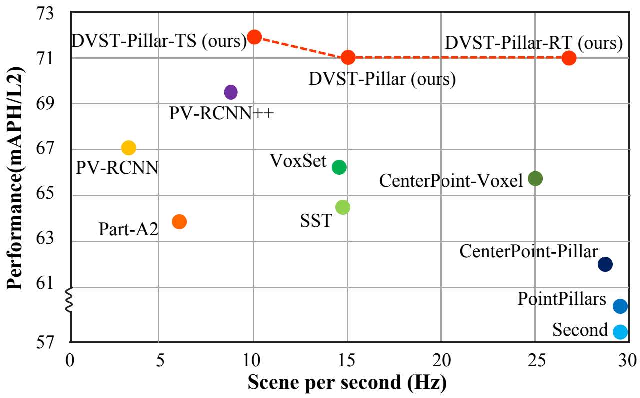



Figure 28.6: Detection performance (mAPH) vs speed (Hz) of different methods on Waymo 3D perception dataset.

attention strategy that can efficiently handle sparse 3D voxels in parallel. Our approach reformulates the window-based attention as parallel computing self-attention within a series of window-bounded and size-equivalent subsets, which enables local feature aggregation in parallel for the entire point clouds, without requiring any self-designed CUDA operations. The extensive experiments demonstrate that our approach outperforms previous state-of-the-art methods on the large-scale Waymo and nuScenes datasets across various 3D perception tasks with faster running speed. In Fig. 28.6, we also provide performance and speed comparisons of our approach against previous methods.

Additionally, our research delves into the development of a unified backbone architecture for a diverse range of 3D scenarios and tasks. Distinct 3D scenarios, such as outdoor and indoor scenes or single shapes, exhibit considerable variation in point distributions and ranges. Also, different task heads usually have their unique requirements on the output feature maps of the backbone network. As a result, data representations (*e.g.*, voxels or points) and 3D backbone networks are typically tailored to address specific scenarios and tasks. For example, voxel-based networks are commonly employed for outdoor detection tasks, where they extract coarse bird's-eye view feature maps for large-range autonomous driving scenes. On the other hand, point-based networks are often utilized in indoor object detection to extract features for subsampled keypoints, as fine-grained details and the information at the height dimension play significant roles for indoor scenarios. To unify various 3D architectures and enable a more flexible task head design, we propose the Embedding-Querying Paradigm (EQ-Paradigm) [10]. Equipped with our transformer-based querying network, a EQ-Paradigm model is able to take any positions in the 3D space as query positions and generate their features based on the support features extracted by any embedding network. Hence, EQ-Paradigm is a unified query-based paradigm enabling combination of arbitrary point- or voxel-based networks with different task heads (Fig. 28.7), which achieves top performance on diverse 3D understanding tasks, *e.g.*, semantic segmentation, object detection, and shape classification.

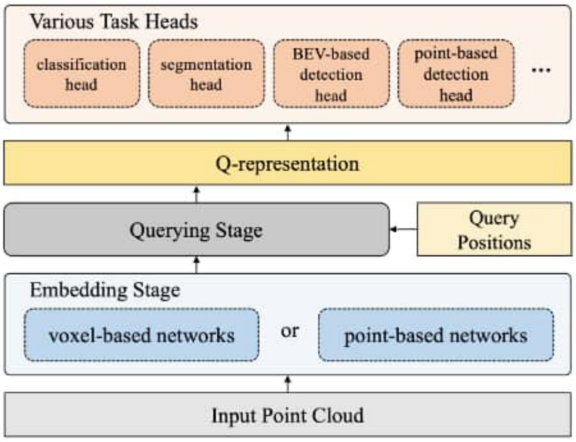



Figure 28.7: Illustration of the unified query-based EQ-Paradigm. The query position can be randomly designated in the 3D scene, thus making it possible to combine any backbone embedding networks with different task heads.

### Efficient Representation Learning in Point Cloud

We further investigate representation learning in point clouds to extract semantically richer and more generalizable features, aiming to enhance data and model efficiency with improved performance. Our exploration encompasses two perspectives: self-supervised pre-training and knowledge distillation.

Self-supervised pre-training is important for improving the efficiency in 3D scene understanding, since it can speed up the downstream training process and boost the performance with limited labeled data. We propose a 3D scene-level self-supervised pre-training framework, Masked Shape Prediction (MSP) [2], to conduct masked signal modeling in 3D scenes. Masked signal modeling is a pretext task that has greatly advanced self-supervised pre-training for language and 2D images, which first partially masks out the input and then reconstructs the masked part given the remaining content, forcing the network to learn semantic knowledge for completing the missing part. To perform masked signal modeling in 3D scenes, MSP uses the essential 3D semantic cue, *i.e.*, geometric shape, as the prediction target for masked points. The context-enhanced shape target consisting of explicit shape context and implicit deep shape feature is proposed to facilitate exploiting contextual cues in shape prediction. Fig. 28.8 shows the overall pipeline of our MSP and illustrates the context-enhanced shape target. Moreover, the pre-training architecture in MSP is carefully designed to alleviate the masked shape leakage from point coordinates. Experiments on multiple 3D understanding tasks on both indoor and outdoor datasets demonstrate the effectiveness of MSP in learning good feature representations to consistently boost downstream performance.

Despite substantial progress in 3D scene understanding, advanced 3D perception models often suffer from heavy computation overheads. To this end, we explore the potential of knowledge distillation (KD) in [9] for developing efficient 3D object detectors. In the absence of well-developed teacher-student pairs for 3D detection task, we first study how to obtain student models with good tradeoffs between accuracy and efficiency from the perspectives of model compression and input resolution reduction. Then, we build a benchmark to

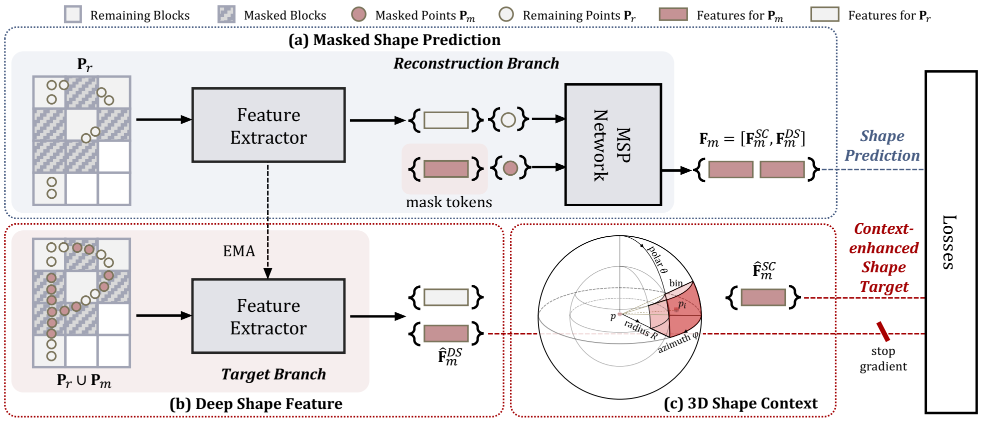



Figure 28.8: Illustration of the masked shape prediction (MSP) pipeline and the context-enhanced shape target.

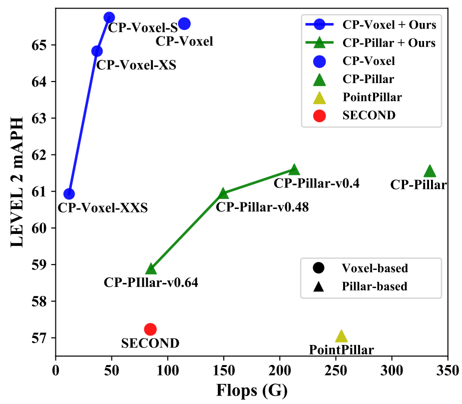



Figure 28.9: Performance and flops comparison of single-stage 3D detectors on Waymo dataset.

assess existing KD methods developed in the 2D domain for 3D object detection upon six wellconstructed teacher-student pairs. Further, we propose an improved KD pipeline incorporating an enhanced logit KD method that performs KD on only a few pivotal positions determined by teacher classification response, and a teacher-guided student model initialization to facilitate transferring teacher model's feature extraction ability to students through weight inheritance. As shown in Fig. 28.9, our best performing student model surpasses its teacher model by requiring only 44% of its teacher flops. Moreover, our most efficient model runs 51 FPS on an NVIDIA A100, which is 2.2× faster than the famous PointPillar with even higher accuracy on Waymo dataset.

### Object-level Understanding for 3D Indoor Scenes

Object-level understanding for indoor scenes is a critical task in 3D scene understanding, as accurate and robust 3D object detection is essential for a wide range of applications such as robotics, augmented reality, and smart homes. Existing approaches typically handle

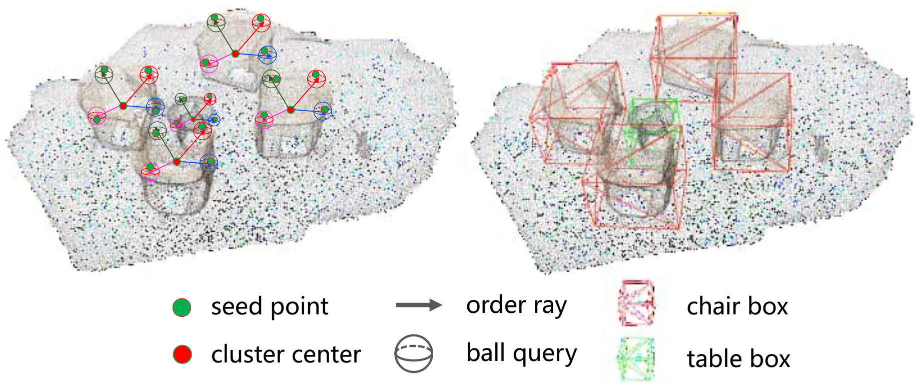



Figure 28.10: 3D object detection in point clouds with a raybased feature grouping module. Given the point clouds of a 3D scene, our model aggregates the point-wise features on object surface by a group of ordered rays to boost the performance of 3D object detection.

this task through a bottom-up scheme, which extracts point-wise features from input point clouds and groups the point features into their respective instances to generate a set of 3D bounding boxes. However, these feature grouping strategies have not well explored the fine-grained surface geometry to help improve the performance of 3D box generation. Hence, we propose the RBGNet [8] framework, which leverages surface geometry of foreground objects to enhance feature grouping and 3D box generation. Specifically, we formulated a ray-based mechanism to capture the object surface points, where a number of rays are uniformly emitted from the clustered object center with determined angles (see Fig. 28.10). A number of anchor points are densely sampled on each ray, where the aggregated local features of each anchor point are utilized to predict whether they are on the object surface to learn the geometry shape. Our experiments on the challenging ScanNetv2 and SUN RGB-D datasets show that our ray-based feature grouping strategy can effectively encode the surface geometry of foreground objects and significantly improves 3D detection performance.

While RBGNet [8] achieves impressive average performance, it may fail in cluttered indoor scenes where various objects are close but belong to different categories. Additionally, object sizes are diverse for different categories, making class-agnostic local grouping partial to covering the boundary points of large objects and involving more noise outliers for small objects. To address these challenges, we introduce CAGroup3D [6], a two-stage fully convolutional 3D object detection framework. CAGroup3D consists of two novel components: a class-aware 3D proposal generation module that generates reliable proposals by utilizing class-specific local group strategy on the object surface voxels with the same semantic predictions, and an efficient fully sparse convolutional RoI pooling module that recovers the features of the missed surface voxels due to semantic segmentation errors, improving the quality of predicted boxes. Our proposed CAGroup3D considers both the voxel-wise semantic consistency within the same local group and the object-level shape diversity among different categories, and it further improves the performance of RBGNet with remarkable gains on the challenging ScanNetv2 and SUN RGB-D datasets, achieving new state-of-the-art performance.

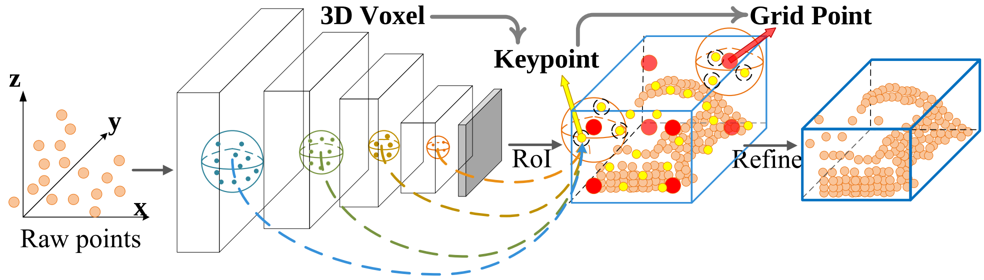



Figure 28.11: The proposed PV-RCNN frameworks deeply integrate both the voxel-based and the point-based representations via a twostep strategy including the voxel-to-keypoint 3D scene encoding and the keypoint-to-grid RoI feature abstraction for improving the performance of 3D object detection.

### 3D Perception and Prediction for Autonomous Driving

3D perception and prediction for autonomous driving is another critical area of 3D scene understanding, as it is essential for autonomous vehicles to navigate safely and efficiently by perceiving and predicting the surrounding environment using LiDAR sensors. Unlike point clouds in 3D indoor scenes that represent small regions with fine-grained points, the LiDAR-generated point clouds in driving scenes cover large-scale regions (*e.g.*, 150m × 150m) with sparse points. To perform 3D object detection on these large-scale regions, we proposed the PV-RCNN framework in [5]. This framework introduced a novel keypoint scheme to deeply integrate the feature learning of two different representations into one framework (see Fig. 28.11), where the first point-based representation maintains accurate point location information while the second voxel-based representation enables efficient multi-scale feature extraction from large-scale point clouds. Building upon PV-RCNN, we further proposed the PV-RCNN++ framework for more efficient and accurate 3D object detection. This framework consists of two major improvements: sectorized proposal-centric sampling for efficiently producing more representative keypoints, and VectorPool aggregation for better aggregating local point features with much less resource consumption. With these two strategies, our PV-RCNN++ is about 3x faster than PV-RCNN, while also achieving better performance. Our experiments demonstrate that the PV-RCNN++ framework achieves state-of-the-art 3D detection performance on the large-scale and highly-competitive Waymo Open Dataset with 10 FPS inference speed on the detection range of 150m × 150m.

Although our existing 3D detection frameworks have achieved impressive performance in either indoor scenes [8, 6] or outdoor scenes [5], they lack the ability to leverage temporal information. As we live in a dynamic world, point cloud videos can offer much more information than a single-frame point cloud, helping to better understand the 3D world and detect objects more accurately. For example, in self-driving cars, LiDAR sensors can continuously generate point clouds over time as the car moves, and this point cloud video can capture multiple views of an object in the scene, allowing for a comprehensive shape understanding of the object and improving detection accuracy (see Fig.28.12). Therefore, we proposed MPPNet [1], which effectively models temporal information from point cloud videos to boost 3D object

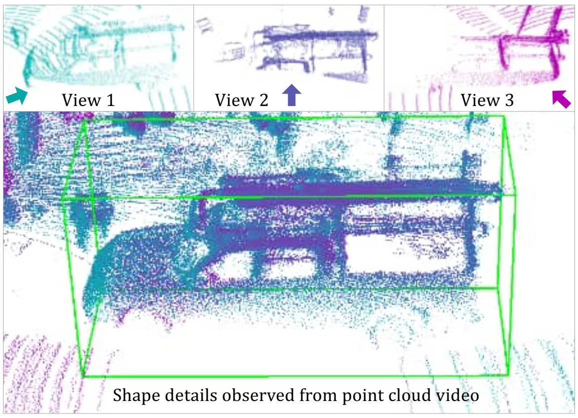
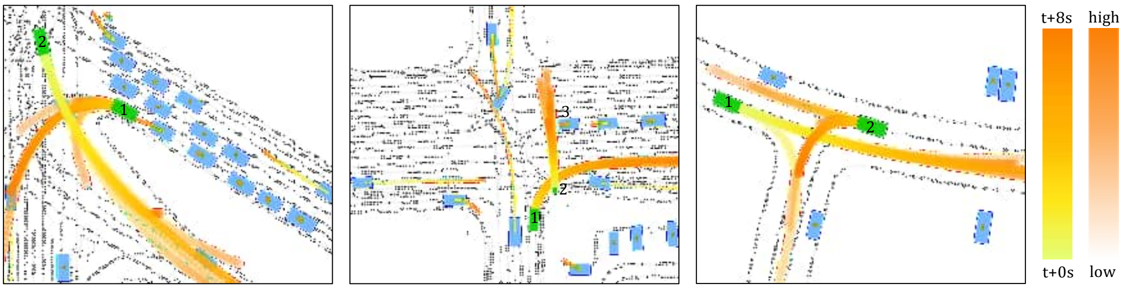



Figure 28.12: Point cloud video can provide more object details from multiple views.

detection performance. The key idea is to use 3D trajectory to align temporal information: 3D bounding boxes within the same trajectory probably contain different surface parts of the same 3D object, thus the feature aggregation of these sequentially-captured object parts can depict more details of the object. By taking a long-term point cloud video (*e.g.*, 16 frames) as input, our model achieved significantly better performance than state-of-the-art single-frame approaches.

In addition to detecting objects in the 3D world, predicting the future behaviors of agents is critical for autonomous driving. Motion prediction is the problem that aims to predict the potential future trajectories of each agent in the 3D world. However, this is a very challenging task since the agents' behaviors are highly multimodal, and their exact behaviors are unknown until they happen. Besides that, we also need to model the agents' historical states and the complex scene environments to forecast their future behaviors. To address these challenges, we developed the Motion Transformer framework [4] that models motion prediction as the joint optimization of global intention localization and local movement refinement. Our approach models different motion intentions with different learnable motion queries, allowing us to predict the occurrence probability of each potential intention and estimate the future trajectory of each intention using their specific learnable parameters. Moreover, we model multimodal future trajectories with a Gaussian Mixture Model, which forms an occurrence probability map that covers all potential moving trajectories of the agent, facilitating a comprehensive understanding of the scene in the future few seconds. Our approach has remained state-of-the-art on the Waymo motion prediction benchmark for over eight months, outperforming existing approaches with remarkable margins. Furthermore, our approach won the championship in the highly competitive 2022 Waymo Open Dataset Motion Prediction Challenge, and we have also shown some qualitative results in Fig. 28.13. By modeling the global intention and local movement of agents, our Motion Transformer framework provides a more comprehensive understanding of future behaviors, facilitating the planning of future actions of ego-vehicles in 3D worlds.




(a) **V2** is passing the intersection to turn left with high speed. Our model predicts multimodal behaviors for **V1:** turn left or make a U-turn. In any case, **V1** is predicted to yield for **V2**.

(b) **P2** is passing the road through the crosswalk while **V1** is on the right-turn lane to turn right. Both **V1** and **V3** are predicted to yield for **P2**.

(c) Our model predicts multimodal behaviors for **V1**: go straight and turn right, since it still has a distance to the intersection. **V2** is predicted to yield for **V1** when turning left, since **V1** is moving fast towards the intersection.

Figure 28.13: Qualitative results of MTR framework. There are two interested agents in each scene (green rectangle), where our model predicts 6 multimodal future trajectories for each of them. For other agents (blue rectangle), a single trajectory is predicted by dense future prediction module. We use gradient color to visualize the trajectory waypoints at different future time step, and trajectory confidence is visualized by setting different transparent. Abbreviation: Vehicle (V), Pedestrian (P).

## References


- [1] X. Chen, S. Shi, B. Zhu, K. C. Cheung, H. Xu, and H. Li. MPPNet: Multi-frame feature intertwining with proxy points for 3D temporal object detection. In S. Avidan, G. Brostow, M. Cissé, G. M. Farinella, and T. Hassner, eds., *Computer Vision – ECCV 2022*, Tel Aviv, Israel, 2022, LNCS 13668, pp. 680–697. Springer.
- [2] L. Jiang, Z. Yang, S. Shi, V. Golyanik, D. Dai, and B. Schiele. Self-supervised pre-training with masked shape prediction for 3D scene understanding. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2023)*, Vancouver, Canada, 2023. IEEE. Accepted.
- [3] X. Lai, J. Liu, L. Jiang, L. Wang, H. Zhao, S. Liu, X. Qi, and J. Jia. Stratified transformer for 3D point cloud segmentation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2022)*, New Orleans, LA, USA, 2022, pp. 8490–8499. IEEE.
- [4] S. Shi, L. Jiang, D. Dai, and B. Schiele. Motion transformer with global intention localization and local movement refinement. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, eds., *Advances in Neural Information Processing Systems 35 (NeurIPS 2022)*, New Orleans, LA, USA, 2022, pp. 6531–6543. Curran Associates, Inc.
- [5] S. Shi, L. Jiang, J. Deng, Z. Wang, C. Guo, J. Shi, X. Wang, and H. Li. PV-RCNN++: Point-voxel feature set abstraction with local vector representation for 3D object detection. *International Journal of Computer Vision*, 131:531–551, 2022.
- [6] H. Wang, L. Ding, S. Dong, S. Shi, A. Li, J. Li, Z. Li, and L. Wang. CAGroup3D: Class-aware grouping for 3D object detection on point clouds. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, eds., *Advances in Neural Information Processing Systems 35 (NeurIPS 2022)*, New Orleans, LA, USA, 2022, pp. 29975–29988. Curran Associates, Inc.
- [7] H. Wang, C. Shi, S. Shi, M. Lei, S. Wang, D. He, B. Schiele, and L. Wang. DSVT: Dynamic sparse voxel transformer with rotated sets. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2023)*, Vancouver, Canada, 2023. IEEE. Accepted.

- [8] H. Wang, S. Shi, Z. Yang, R. Fang, Q. Qian, H. Li, B. Schiele, and L. Wang. RBGNet: Ray-based grouping for 3D object detection. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2022)*, New Orleans, LA, USA, 2022, pp. 1100–1109. IEEE.
- [9] J. Yang, S. Shi, R. Ding, Z. Wang, and X. Qi. Towards efficient 3D object detection with knowledge distillation. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, eds., *Advances in Neural Information Processing Systems 35 (NeurIPS 2022)*, New Orleans, LA, USA, 2022, pp. 21300–21313. Curran Associates, Inc.
- [10] Z. Yang, L. Jiang, Y. Sun, B. Schiele, and J. Jia. A unified query-based paradigm for point cloud understanding. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2022)*, New Orleans, LA, USA, 2022, pp. 8531–8541. IEEE.


### 28.4.3 Robustness

*Investigators: David Stutz, Yong Guo, Max Losch, Bernt Schiele, in cooperation with Matthias Hein (University of Tübingen), Nandhini Chandramoorthy (IBM T. J. Watson Research Center), Huio Po Wang, Ning Yu and Mario Fritz (CISPA Helmholtz Center for Information Security), as well as Krishnamurthy, Dvijotham, Ali Taylan Cemgil and Arnaud Doucet (DeepMind).*

This subsection presents our contributions discussing the performance of machine learning models, e.g., (convolutional) neural networks or transformers, in high-stakes applications where robustness, proper uncertainty calibration and understanding mis-use become essential for deployment. Here, robustness studies how the performance of models is impacted by various types of perturbed or corrupted images. This includes imperceptibly or visibly manipulated images, so-called adversarial images or patches, as well as random corruptions due to noise or other effects that commonly simulate various compression or image processing effects. While adversarially manipulated images contribute to a “worst-case” assessment of model performance, random corruptions are typically used to assess various degrees of distribution shifts expected upon deployment. Besides perturbations of the models’ inputs, recent work also highlights several scenarios where the models’ parameters, i.e., weights, are subject to malicious or random errors. For example, when deploying machine learning models on special-purpose hardware, so-called accelerators, weights need to be quantized. This introduces quantization errors and makes parameters vulnerable to bit-level errors and attacks. For many safety-critical applications, however, empirically improving robustness might not be sufficient. In many applications, we require *guarantees* on performance alongside meaningful uncertainty estimates, e.g., indicating when to trust the models’ predictions. Moreover, when deploying or publicly releasing models, we need to make sure we understand and potentially counter-act potential mis-use of these systems by malicious actors.

#### Improving Robustness by Enhancing Weak Subnets

Recently, deep convolutional networks have achieved great success but still remained highly susceptible to image perturbations, often causing significant drops in accuracy. In [1], we explicitly investigate model robustness on perturbed inputs by studying the performance of internal sub-networks (subnets). Interestingly, we observe that most subnets show particularly poor robustness against perturbations. More importantly, these weak subnets are correlated with the overall lack of robustness. Tackling this phenomenon, we propose a new training procedure that identifies and enhances weak subnets (EWS) to improve robustness. Specifically, we develop a search algorithm to find particularly weak subnets and explicitly strengthen them via knowledge distillation from the full network. We show that EWS greatly improves both robustness against corrupted images as well as accuracy on clean data. Being complementary to popular data augmentation methods, EWS consistently improves robustness when combined with these approaches. To highlight the flexibility of our approach, we combine EWS also with popular adversarial training methods resulting in improved adversarial robustness.

#### Relating Adversarially Robust Generalization to Flat Minima

While adversarial training, e.g., in combination with EWW [1], has become the de-factor standard to obtain models robust against adversarial examples, it commonly exhibits severe robust overfitting: cross-entropy loss on adversarial examples, so-called robust loss, decreases continuously on training examples, while eventually increasing on test examples In practice, this leads to poor robust generalization, i.e., adversarial robustness does not generalize well to new examples. To better understand and potentially avoid this problem, in [6], we empirically study the relationship between robust generalization and flatness of the robust loss landscape in weight space, i.e., whether robust loss changes significantly when perturbing weights. Using newly developed metrics to measure “robust flatness”, we show a clear correlation between good robust generlaization and flatness. Specifically, flatness tends to reduce significantly during overfitting and adversarial training variants achieving higher adversarial obustness also find flatter minimna. Supported by evidence from many popular adversarial training variants (e.g., AT-AWP, TRADES, MART, AT with self-supervision or additional unlabeled examples), this suggests that avoiding overly sharp minima is a promising direction in further improving adversarial robustness.

#### Improving Robustness of Vision Transformers by Reducing Sensitivity to Patch Corruptions

Besides convolutional neural networks, we further study and improve the robustness of recent vision transformers (ViTs) models that achieve state-of-the-art performance in many computer vision tasks. In practice, ViTs are also very vulnerable to common corruptions and adversarial perturbations. In [2], we find that the vulnerability mainly stems from the unstable self-attention mechanism, which is inherently built upon patch-based inputs and often becomes overly sensitive to the corruptions across patches. For example, as illustrated in Figure 28.14, when we only occlude a small number of patches with random noise (e.g., 10%), these patch corruptions would lead to severe accuracy drops and greatly distract intermediate attention layers. To address this, we propose a new training method that improves the robustness of transformers from a new perspective, i.e., reducing sensitivity to patch corruptions (RSPC). Specifically, we first identify and occlude/corrupt the most vulnerable patches and then explicitly reduce sensitivity to them by aligning the intermediate features between clean and corrupted examples. We highlight that the construction of patch corruptions is learned


Figure 28.14: Sensitivity to patch perturbations/corruptions in terms of confidence score of the ground-truth class. We randomly select 10% patches to be perturbed/corrupted for a robust vision transformer. In practice, adversarial patch perturbations (often invisible) significantly reduce the confidence, indicating the high sensitivity of transformers to patches. However, directly adding random noise only yields marginal degradation. By contrast, occluding patches with noise greatly reduces the confidence and can be used as a good proxy of adversarial patch perturbations to reveal the patch sensitivity issue.

adversarially to the following feature alignment process, which is particularly effective and essentially different from existing methods. In experiments, our RSPC greatly improves the stability of attention layers and consistently yields better robustness on various benchmarks, including CIFAR-10/100-C, ImageNet-A, ImageNet-C, and ImageNet-P.

#### Certified Robust Models with Slack Control and Large Lipschitz Constants

Beyond empirical robustness, in [3], we also work towards certified robustness against adversarial examples. Here, recent work considers Lipschitz-based regularizers or constraints while at the same time increasing prediction margin. Unfortunately, this comes at the cost of significantly decreased accuracy. Instead, we propose a Calibrated Lipschitz Margin loss (CLip) that addresses this issue and improves certified robustness by tackling two problems: Firstly, commonly used margin losses do not adjust the penalties to the shrinking output distribution; caused by minimizing the Lipschitz constant $K$. Secondly, and most importantly, we observe that minimization of the Lipschitz constant can lead to overly smooth decision functions which limits the model's expressiveness. We address these issues by explicitly calibrating the loss wrt. margin and Lipschitz constant, thereby establishing full control over slack. On various datasets, we demonstrate that this approach yields greater Lipschitz constants, thereby improving clean accuracy, while also obtaining a new state-of-the-art certified $L_2$ robust accuracy.

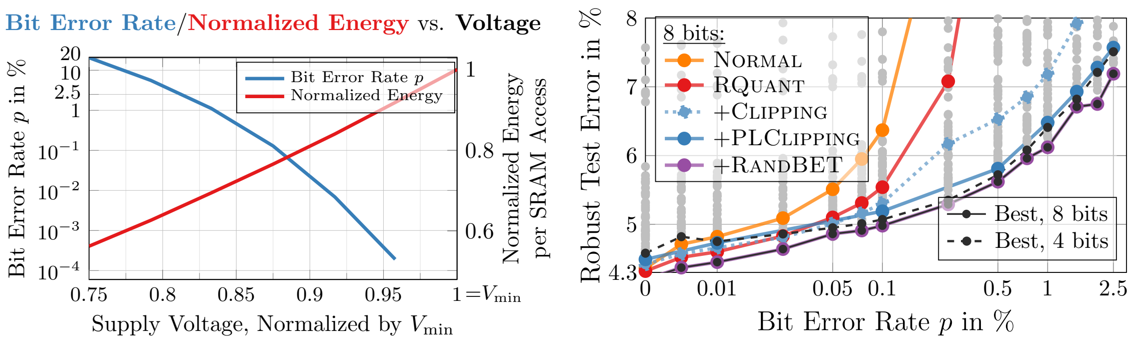



Figure 28.15: Left: Average bit error rate $p$ (blue, left y-axis) from 32 14nm SRAM arrays of size 512×64 and energy (red, right y-axis) vs. voltage (x-axis). Voltage and energy are normalized by the minimal measured voltage for error-free operation and the energy per SRAM access that voltage, respectively. Reducing voltage leads to exponentially increasing bit error rates. Right: Robust test error after injecting *random* bit errors (lower is better, y-axis) plotted against bit error rate $p$ (x-axis). For 8 bit, robust quantization (red), additionally weight clipping (blue) and finally adding random bit error training (violet) robustness improves significantly. Robustness to higher bit error rates allows more energy efficient operation. The Pareto optimal frontier is shown for 8 bit (black solid) and 4 bit (dashed) quantization

### Random and Adversarial Bit Error Robustness: Energy-Efficient and Secure DNN Accelerators

Deep neural networks are in practice commonly deployed on accelerators, special-purpose hardware for fast inference. Besides requiring weight quantization, these accelerators commonly operate in a low-voltage regime in order to improve energy-efficiency. However, this causes bit-level failures in the memory storing the quantized weights with the bit error rate increasing exponentially with reduced voltage, see Figure 28.15 (left). Furthermore, such accelerators are vulnerable to adversarial attacks on voltage controllers or individual bits. Thus, in [4], we show that a combination of robust fixed-point quantization, weight clipping, as well as bit error training improves robustness against random or adversarial bit errors significantly. This is highlighted in Figure 28.15 (right) showing that the increase in test error due to high bit error rates can be reduced substantially. Apart from leading to high energy savings from low-voltage operation and low-precision quantization, with also improves security. We empirically evaluate our methods using a novel adversarial bit error attack as well as on bit error patterns obtained from real profiled accelerator chips. Thereby, we provide benchmarks for the community in order to make progress on a very practical and challenging robustness problem.

### Learning Optimal Conformal Classifiers

Even with improved robustness, deep learning based classifiers are commonly evaluated in terms of (robust) accuracy on held-out test data. However, for many applications, this does not provide sufficient guarantees for safe deployment. Furthermore, such models do usually not provide reliable uncertainty estimates. Recent work on conformal prediction addresses these issues by using the classifier's predictions, e.g., its probability estimates, to predict confidence sets containing the true class with a user-specified probability. This is usually achieved in a separate post-processing step after training and thus prevents the underlying model from adapting to the prediction of these confidence sets. In [5], we explore strategies to differentiate through the conformal calibration during training with the goal of training model with the conformal wrapper end-to-end. Our approach, called conformal training, reduces the average confidence set size (inefficiency) of state-of-the-art conformal methods applied after training. Moreover, it allows to guide the composition of confidence sets in terms of the included classes, while retaining the guarantees obtained from conformal prediction.

### Hijack-GAN: Unintended-Use of Pretrained, Black-Box GANs

Besides robustness and uncertainty estimation, we also conducted research on other safety issues in state-of-the-art machine learning. In [7], we consider Generative Adversarial Networks (GANs) which recently achieved incredible performance in image generation. Commonly, the level of realism is becoming indistinguishable from natural images. With more of these models becoming publicly available, we show that state-of-the-art GAN models can be used for a range of applications beyond unconditional image generation. We achieve this by an iterative scheme that also allows gaining control over the image generation process despite the highly non-linear latent spaces of the latest GAN models. We demonstrate that this opens up the possibility to re-use state-of-the-art, difficult to train, pre-trained GANs with a high level of control even if only black-box access is granted. Our work also raises concerns and awareness that the use cases of a published GAN model may well reach beyond the creators' intention, which needs to be taken into account before a full public release.

### References


- [1] Y. Guo, D. Stutz, and B. Schiele. Improving robustness by enhancing weak subnets. In S. Avidan, G. Brostow, M. Cissé, G. Farinella, and T. Hassner, eds., *Computer Vision – ECCV 2022*, Tel Aviv, Israel, 2022, LNCS 13684, pp. 320–338. Springer.
- [2] Y. Guo, D. Stutz, and B. Schiele. Improving robustness of vision transformers by reducing sensitivity to patch corruptions. In *Conference on Computer Vision and Pattern Recognition 2023*, 2023. https://openreview.net/forum?id=rjM8U3DhlF.

[3] M. M. Losch, D. Stutz, B. Schiele, and M. Fritz. Certified robust models with slack control and large lipschitz constants. In *Under Submission*, 2022.

- [4] D. Stutz, N. Chandramoorthy, M. Hein, and B. Schiele. Random and adversarial bit error robustness: Energy-efficient and secure DNN accelerators. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(3):3632–3647, 2023.
- [5] D. Stutz, K. Dvijotham, A. T. Cemgil, and A. Doucet. Learning optimal conformal classifiers. In *The Tenth International Conference on Learning Representations, ICLR 2022, Virtual Event, April 25-29, 2022*, 2022. OpenReview.net.

- [6] D. Stutz, M. Hein, and B. Schiele. Relating adversarially robust generalization to flat minima. In *ICCV 2021, IEEE/CVF International Conference on Computer Vision*, Virtual Event, 2021, pp. 7787–7797. IEEE.
- [7] H.-P. Wang, N. Yu, and M. Fritz. Hijack-GAN: Unintended-use of pretrained, black-box GANs. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2021)*, Virtual Conference, 2021, pp. 7868–7877. IEEE.


### 28.4.4 Learning with Less Supervision

*Investigators: Anna Kukleva, Moritz Böhle, Enea Duka, Bernt Schiele, in cooperation with Hilde Kuehne (Goethe University Frankfurt, MIT-IBM Watson AI Lab), Christian Rupprecht (University of Oxford), Wei Lin, Kunyang Sun, Horst Possegger, Horst Bischof (Institute of Computer Graphics and Vision, Graz University of Technology)*

Deep learning models rely on supervision signals that are usually available in the dataset, e.g. in the form of class labels for each image. However, collecting human annotations can be time and cost intensive, especially for the currently popular large-scale datasets. Therefore, there is an increasing interest in reducing annotation costs by the utilization of very few or no human-annotated samples. In this subsection, we present our contributions on learning-with-less-supervision tasks, specifically few-shot learning and self-supervised learning.

#### Generalized and Incremental Few-Shot Learning

Few-shot learning addresses the task when only few annotated samples are available during training. In [3], we study generalized few-shot learning, where the model is pretrained with abundant amount of data first, and then novel few-shot classes are introduced. Additionally, we investigate incremental few-shot learning with the same framework, where novel classes with few annotated samples are continuously introduced in consecutive sessions. Both generalized and incremental few-shot learning have to deal with three major challenges: learning novel classes from only few samples per class, preventing catastrophic forgetting of base classes, and classifier calibration across novel and base classes. In [3], we propose a three-stage framework that allows to explicitly and effectively address these challenges, see Fig. 28.16. While the first phase learns base classes with many samples, the second phase learns a calibrated classifier for novel classes from few samples while also preventing catastrophic forgetting. In the final phase, calibration is achieved across all classes. We evaluate the proposed framework on four challenging benchmark datasets for image and video few-shot classification and obtain state-of-the-art results for both generalized and incremental few-shot learning.

#### Temperature Schedules for Self-Supervised Contrastive Methods on Long-Tail Data

In [2], we study self-supervised learning on long-tail data that requires no human annotations for training. Most approaches for self-supervised learning (SSL) are optimised on curated balanced datasets, e.g. ImageNet, despite the fact that natural data usually exhibits long-tail distributions. In [2], we analyse the behaviour of one of the most popular variants of SSL,

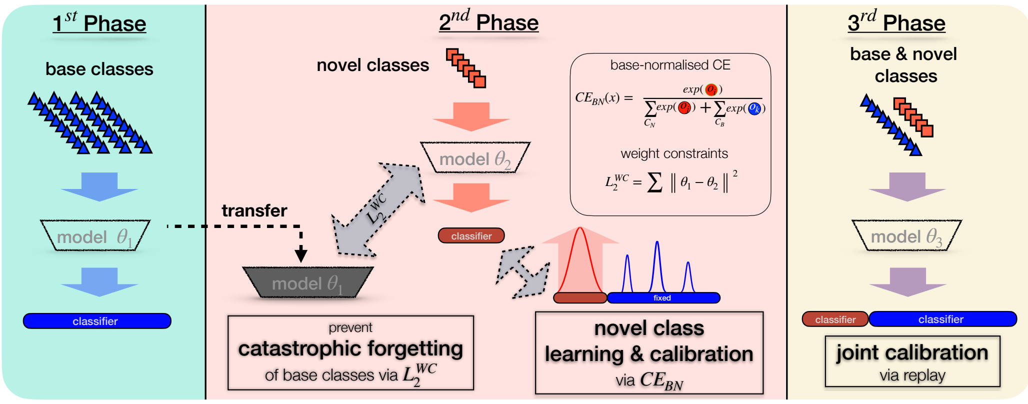



Figure 28.16: Overview of generalized few-shot learning framework. To achieve balanced performance on base and novel classes we dead with three problems learning novel classes, catastrophic forgetting, and calibration that we address in different phases of our framework. In the $1^{st}$ phase we pretrain model on base classes with abundant data. During the $2^{nd}$ phase we employ $L_2^{WC}$ weight constraints to preserve knowledge and base-normalized cross entropy $(CE_{BN})$ to calibrate learning of novel classes in the joint space with base classes. In the $3^{rd}$ phase we calibrate the performance with the balanced replay of novel and base samples.

i.e. contrastive methods, on imbalanced data. In particular, we investigate the role of the temperature parameter $\tau$ in the contrastive loss, by analysing the loss through the lens of average distance maximisation, and find that a large $\tau$ emphasises group-wise discrimination, whereas a small $\tau$ leads to a higher degree of instance discriminationi, see Fig. 28.17. While $\tau$ has thus far been treated exclusively as a constant hyperparameter, we propose to employ a dynamic and show that a simple cosine schedule can yield significant improvements in the learnt representations. Such a schedule results in a constant 'task switching' between an emphasis on instance discrimination and group-wise discrimination and thereby ensures that the model learns both group-wise features, as well as instance-specific details. Since frequent classes benefit from the former, while infrequent classes require the latter, we find this method to consistently improve separation between the classes in long-tail data without any additional computational cost.

### Leveraging Self-Supervised Training for Unintentional Action Recognition

In [1], we utilize self-supervised learning by creating self-supervised temporal modelling tasks to enhance specific features of the input unintentional action videos. Unintentional actions are rare occurrences that are difficult to define precisely and that are highly dependent on the temporal context of the action. In [1], we explore such actions and seek to identify the points in videos where the actions transition from intentional to unintentional by using self-supervised representation learning. We propose a multi-stage framework, that exploits in self-supervised way inherent biases such as motion speed, motion direction, and order, see Fig. 28.18. We propose these temporal transformations, called Temporal Transformations of Inherent Biases of Unintentional Actions ($T^2$IBUA) to enhance temporal representations for

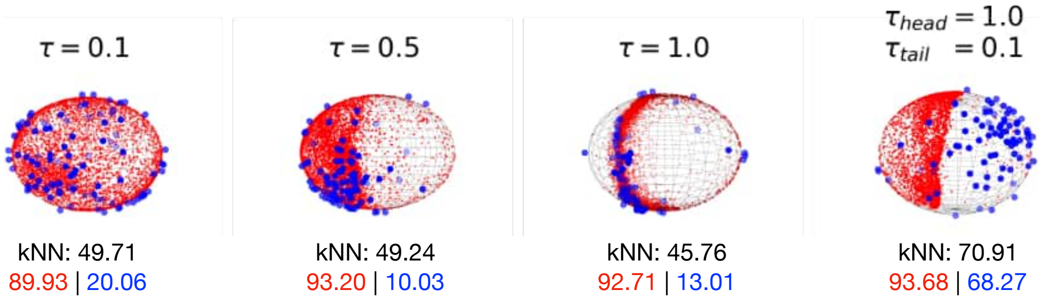



Figure 28.17: Visualization of the influence of $\tau$ on representation of two semantically close classes. Red denotes single head class and blue denotes single tail class from CIFAR10-LT. Small $\tau = 0.1$ promotes uniformity, while large $\tau = 1.0$ creates dense clusters. With $\tau_{head/tail}$ we refer to coarse supervision which separates tail from head classes that confirms our claim. In black / red / blue, we respectively show the average kNN accuracy over all classes / the head class / the tail class.

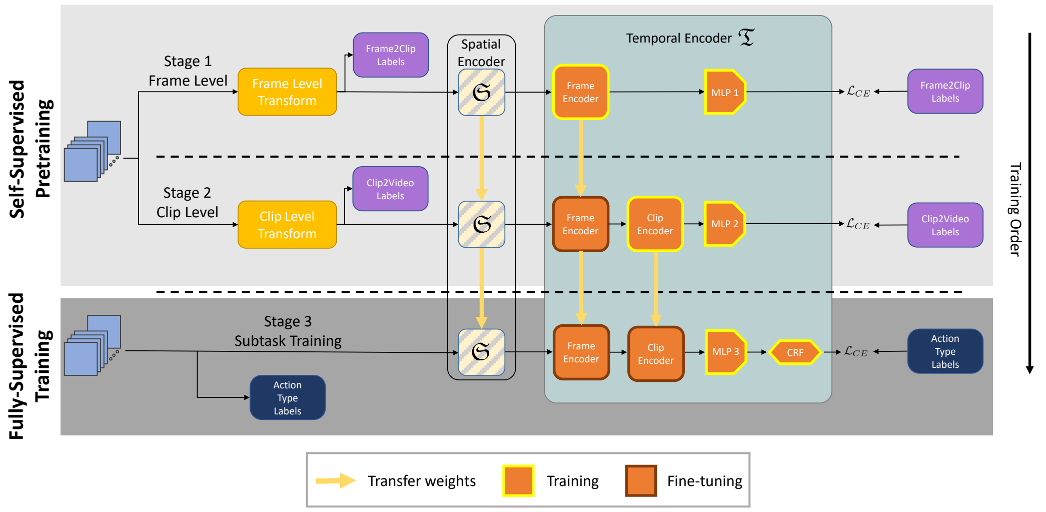



Figure 28.18: Unintentional action recognition framework overview. In the first and second stages, we use self-supervised feature enhancement by predicting $T^2$IBUA. During the last stage, we train in a fully-supervised way for downstream tasks.

specifically videos of unintentional actions. The multi-stage approach models the temporal information of individual frames and full clips by applying self-supervised T2IBUA on different levels. These enhanced representations show strong performance for unintentional action recognition tasks. We provide an extensive ablation study of our framework and report results that significantly improve over the state-of-the-art.

### CycDA: Unsupervised Cycle Domain Adaptation from Image to Video

Although action recognition has achieved impressive results over recent years, both collection and annotation of video training data are still time-consuming and cost intensive. Therefore,

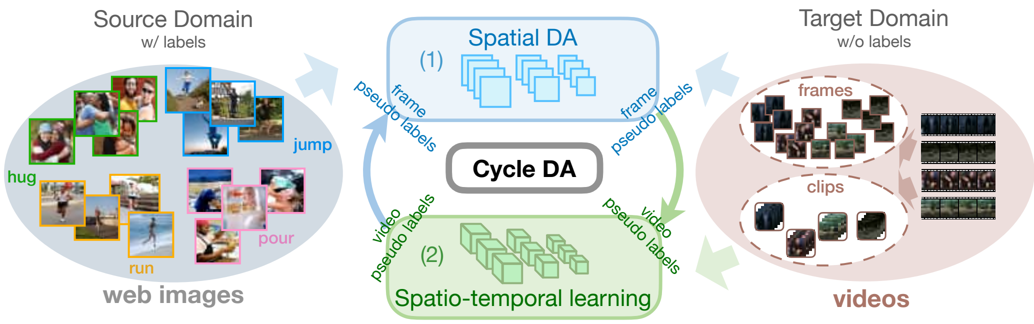



Figure 28.19: Cycle Domain Adaptation (CycDA) pipeline: we address image-to-video adaptation by training a spatial model and a spatio-temporal model alternately, passing pseudo labels to supervise each other in a cycle. The two alternating steps are: (1) domain alignment on the spatial model with pseudo labels from the spatio-temporal model, and (2) training the spatio-temporal model with updated pseudo labels from the spatial model.

image-to-video adaptation has been proposed to exploit labeling-free web image source for adapting on unlabeled target videos. This poses two major challenges: (1) spatial domain shift between web images and video frames; (2) modality gap between image and video data. To address these challenges, in [4] we propose Cycle Domain Adaptation (CycDA), a cycle-based approach for unsupervised image-to-video domain adaptation. We leverage the joint spatial information in images and videos on the one hand and, on the other hand, train an independent spatio-temporal model to bridge the modality gap. We alternate between the spatial and spatio-temporal learning with knowledge transfer between the two in each cycle, see Fig. 28.19. We evaluate our approach on benchmark datasets for image-to-video as well as for mixed-source domain adaptation achieving state-of-the-art results and demonstrating the benefits of our cyclic adaptation.

## References


- [1] E. Duka, A. Kukleva, and B. Schiele. *Leveraging Self-Supervised Training for Unintentional Action Recognition*, 2022. arXiv: 2209.11870.
- [2] A. Kukleva, M. Boehle, B. Schiele, H. Kuehne, and C. Rupprecht. Temperature schedules for self-supervised contrastive methods on long-tail data. In *International Conference on Learning Representations*, 2023.
- [3] A. Kukleva, H. Kuehne, and B. Schiele. Generalized and incremental few-shot learning by explicit learning and calibration without forgetting. In *ICCV 2021, IEEE/CVF International Conference on Computer Vision*, Virtual Event, 2021, pp. 9000–9009. IEEE.
- [4] W. Lin, A. Kukleva, K. Sun, H. Possegger, H. Kuehne, and H. Bischof. CycDA: Unsupervised cycle domain adaptation to learn from image to video. In S. Avidan, G. Brostow, M. Cissé, G. Farinella, and T. Hassner, eds., *Computer Vision – ECCV 2022*, Tel Aviv, Israel, 2022, LNCS 13663, pp. 698–715. Springer.

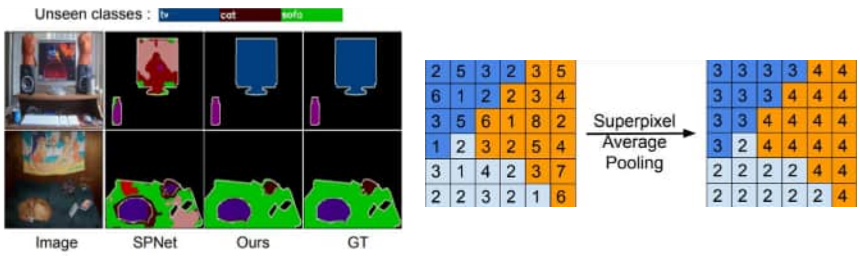



Figure 28.20: Our model can predict unseen classes (tv in row 1, and sofa + cat in row 2) correctly compared to the baseline SPNet model. We integrate a novel superpixel pooling module (left) in the segmentation network along with a bias reduction loss resulting in better generalization on the unseen classes as seen from the example.

### 28.4.5 Label Efficient Semantic Segmentation

*Investigators: Anurag Das, Li Jiang, Dengxin Dai, Zeynep Akata, Bernt Schiele, in cooperation with Yang He (Amazon), Yongqin Xian (Google), and Zhuotao Tian, Jiaya Jia (Chinese University of Hong Kong)*

#### Generalized Zero Label Semantic Segmentation

The data distribution in a real world setting is naturally long tailed. It is a difficult problem to semantically segment the rare classes having no or few labels. The task of generalized zero-label semantic segmentation aims to segment these rare (also called novel) classes without any labeled training samples along with the seen classes. It is a challenging problem as the networks trained on seen classes are biased and do not generalize well to the novel classes. In [3] we propose to utilize a class agnostic segmentation prior provided by superpixels and introduce a superpixel pooling module, that groups pixel features belonging to the same superpixels. Further to reduce the seen class biases in network predictions, we additionally propose a bias reduction loss term which reduces the seen class prediction probability in the ignored regions during training. The superpixel pooling module along with the bias reduction loss is effective in improving the segmentation performance on both seen and novel classes (see Fig. 28.20), and achieves state of the art on different data splits of PASCAL VOC 2012 and PASCAL-Context benchmark datasets.

#### Domain Adaptative Semantic Segmentation with Weak Labels

Semantic segmentation requires expensive pixel annotation. For urban scenes, it is even more challenging as there are too many objects to annotate in an image. Unsupervised Domain Adaptation tries to reduce this annotation effort by learning from cheaper synthetic images. However, the performance gap compared to supervised learning is still huge. In [2], we propose to utilize cheaper coarse annotation as weak labels to reduce the performance gap with supervised learning. Specifically we propose a coarse-to-fine self training framework that improves the predictions along the boundaries on coarse data by learning the fine boundaries from the synthetic data, and improves data diversity via cross-domain augmentation. In [1],

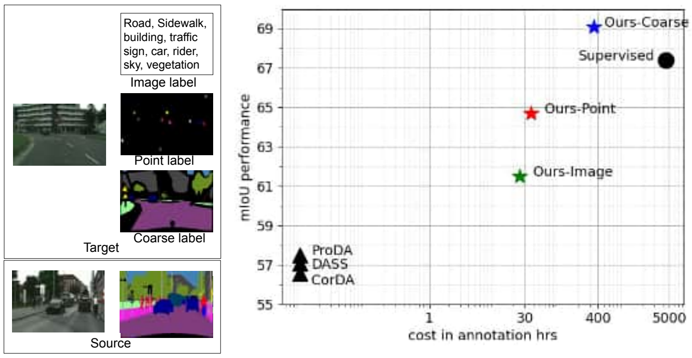



Figure 28.21: Our common framework for different weak label (image, point and coarse labels) for the task of Weakly Supervised Domain Adaptative Semantic Segmentation (WDASS) utilises cheaper weak labels and bridges the gap between UDA and supervised learning.

we extend [2] to other weak labels such as image and point label, and propose prototype based contrastive learning based on weak labels to further improve the performance. Notably, we reduce the performance gap wrt. supervised learning by using point and image labels and further, we even outperform supervised learning with coarse labels showing the importance of the weak labels (see Fig. 28.21).

### Generalized Few-shot Semantic Segmentation

Training semantic segmentation models requires a large amount of finely annotated data, making it hard to quickly adapt to novel classes not satisfying this condition. Few-Shot Segmentation (FS-Seg) tackles this problem with many constraints. In [4], we introduce a new benchmark, called Generalized Few-Shot Semantic Segmentation (GFS-Seg), to analyze the generalization ability of simultaneously segmenting the novel categories with very few examples and the base categories with sufficient examples. It is the first study showing that previous representative state-of-the-art FS-Seg methods fall short in GFS-Seg and the performance discrepancy mainly comes from the constrained setting of FS-Seg. Fig. 28.22 compares the pipelines of GFS-Seg and FS-Seg. To make GFS-Seg tractable, we set up a GFS-Seg baseline that achieves decent performance without structural change on the original model. Then, since context is essential for semantic segmentation, we propose the Context-Aware Prototype Learning (CAPL) that significantly improves performance by leveraging the co-occurrence prior knowledge from support samples, and dynamically enriching contextual information to the classifier, conditioned on the content of each query image. Both two contributions are experimentally shown to have substantial practical merit. Extensive experiments on PascalVOC and COCO manifest the effectiveness of CAPL, and CAPL generalizes well to FS-Seg by achieving competitive performance.

## References


- [1] A. Das, Y. Xian, D. Dai, and B. Schiele. Weakly-supervised domain adaptive semantic segmentation with prototypical contrastive learning. In *IEEE/CVF Conference on Computer Vision and*

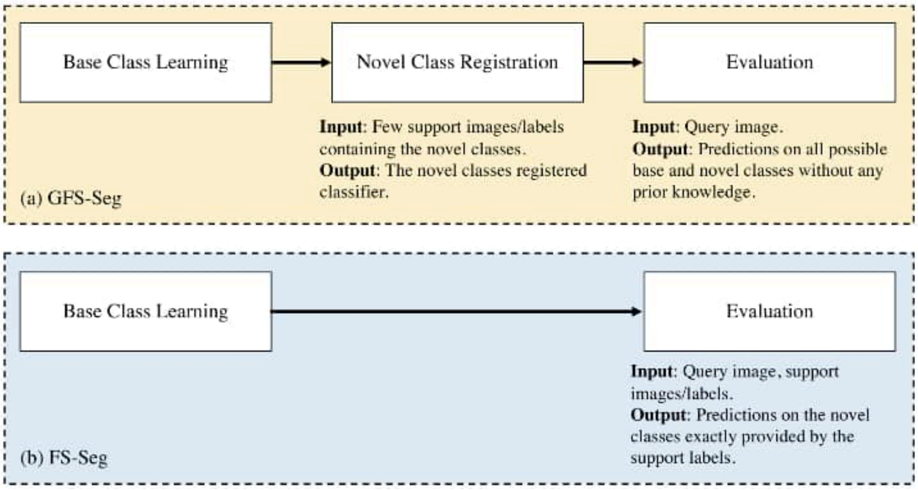



Figure 28.22: Pipeline illustrations of GFS-Seg and FS-Seg.

*Pattern Recognition (CVPR 2023)*, Vancouver, Canada, 2023. IEEE. Accepted.

- [2] A. Das, Y. Xian, Y. He, Z. Akata, and B. Schiele. Urban scene semantic segmentation with low-cost coarse annotation. In *2023 IEEE Winter Conference on Applications of Computer Vision (WACV 2023)*, Waikoloa Village, HI, USA, 2023, pp. 5967–5976. IEEE.
- [3] A. Das, Y. Xian, Y. He, B. Schiele, and Z. Akata. $SP^2$Net for generalized zero-label semantic segmentation. In C. Bauckhage, J. Gall, and A. Schwing, eds., *Pattern Recognition (GCPR 2021)*, Bonn, Germany, 2021, LNCS 13024, pp. 235–249. Springer.
- [4] Z. Tian, X. Lai, L. Jiang, S. Liu, M. Shu, H. Zhao, and J. Jia. Generalized few-shot semantic segmentation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2022)*, New Orleans, LA, USA, 2022, pp. 11553–11562. IEEE.

## 28.4.6 Semi-Supervised Learning

*Investigators: Yue Fan, Anna Kukleva, Dengxin Dai, Bernt Schiele, in cooperation with Hao Chen, Ran Tao, Marios Savvides, Bhiksha Raj (Carnegie Mellon University), Yidong Wang, Jindong Wang, Xing Xie (Microsoft Research Asia), Wang Sun (Tsinghua University), Wenxin Hou (Microsoft STCA), Renjie Wang, Zhi Zhou, Lan-Zhe Guo, Zhen Wu, Yu-Feng Li (Nanjing University), Linyi Yang, Yue Zhang (Westlake University), Heli Qi (Nara Institute of Science and Technology), Wei Ye (Peking University), Satoshi Nakamura, Takahiro Shinozaki (Tokyo Institute of Technology)*

### Revisiting Consistency Regularization for Semi-supervised Learning

Consistency regularization is one of the most widely-used techniques for semi-supervised learning (SSL). Generally, the aim is to train a model that is invariant to various data augmentations. In [3, 5], we revisit this idea and find that enforcing invariance by decreasing distances between features from differently augmented images leads to improved performance. However, encouraging equivariance instead, by increasing the feature distance, further improves performance. To this end, we propose an improved consistency regularization

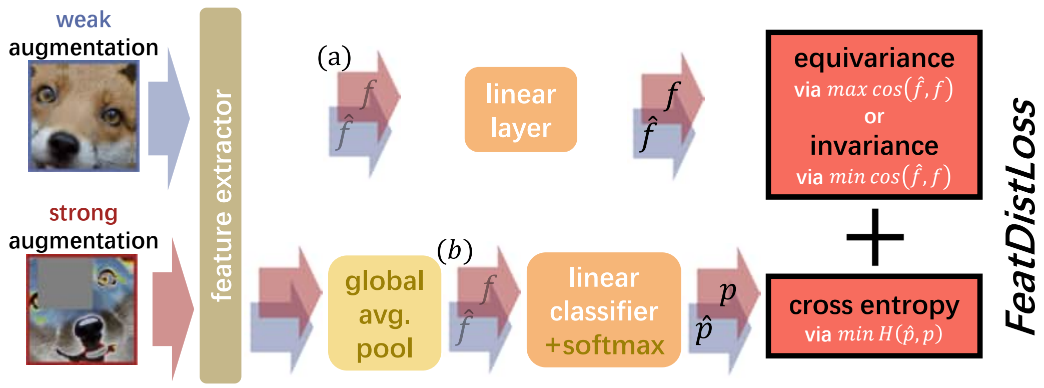



Figure 28.23: The proposed FeatDistLoss utilizes unlabeled images in two ways: On the classifier level, different versions of the same image should generate the same class label, whereas on the feature level, representations are encouraged to become either more equivariant (pushing away) or invariant (pulling together). $f$ and $\hat{f}$ denote strong and weak features; $p$ and $\hat{p}$ are predicted class distributions from strong and weak features; a) and b) denote features before and after the global average pooling layer. Our final model takes features from a) and encourages equivariance to differently augmented versions of the same image.

framework (see Fig. 28.23) by a simple yet effective technique, FeatDistLoss, that imposes consistency and equivariance on the classifier and the feature level, respectively. Experimental results show that our model defines a new state of the art across a variety of standard semi-supervised learning benchmarks as well as imbalanced semi-supervised learning benchmarks. Particularly, we outperform previous work by a significant margin in low data regimes and at large imbalance ratios.

### A Unified Semi-supervised Learning Benchmark for Classification

Semi-supervised learning (SSL) improves model generalization by leveraging massive unlabeled data to augment limited labeled samples. However, currently, popular SSL evaluation protocols are often constrained to computer vision (CV) tasks. In addition, previous work typically trains deep neural networks from scratch, which is time-consuming and environmentally unfriendly. To address the above issues, in [6], we construct a Unified SSL Benchmark (USB) for classification by selecting 15 diverse, challenging, and comprehensive tasks from CV, natural language processing (NLP), and audio processing (Audio), on which we systematically evaluate the dominant SSL methods, and also open-source a modular and extensible codebase for fair evaluation of these SSL methods. We further provide the pre-trained versions of the state-of-the-art neural models for CV tasks to make the cost affordable for further tuning. USB enables the evaluation of a single SSL algorithm on more tasks from multiple domains but with less cost (see Fig. 28.24). Specifically, on a single NVIDIA V100, only 39 GPU days are required to evaluate FixMatch on 15 tasks in USB while 335 GPU days (279 GPU days on 4 CV datasets except for ImageNet) are needed on 5 CV tasks with TorchSSL.

| Domain & Backbone | Dataset | Classification Task | Total GPU Hours |
|---|---|---|---|
| CV, ViTs | CIFAR-100<br>STL-10<br>EuroSAT<br>TissueMNIST<br>Semi-Aves | Natural Image<br>Natural Image<br>Satellite Image<br>Medical Image<br>Fine-grained, Long-tailed Natural Image | 924 GPU Hours<br>(39 GPU Days) |
| NLP, Bert | IMDB<br>AG News<br>Amazon Review<br>Yahoo! Answer<br>Yelp Review | Movie Review Sentiment<br>News Topic<br>Product Review Sentiment<br>QA Topic<br>Restaurant Review Sentiment | |
| Audio, Wave2Vec 2.0 and HuBert | GTZAN<br>UrtraSound8k<br>FSDnoisy18k<br>Keyword Spotting<br>ESC-50 | Music Genre<br>Urban Sound Event<br>Sound Event<br>Keyword<br>Environmental Sound Event | |

Figure 28.24: A summary of datasets and training cost used in USB. USB largely reduces the training cost while providing a diverse, challenging, and comprehensive benchmark covering a wide range of datasets from various domains. Training cost is estimated by using FixMatch on a single NVIDIA V100 GPU from Microsoft Azure Machine Learning platform, except for ImageNet where 4 V100s are used.

### Addressing the Quantity-Quality Tradeoff in Semi-supervised Learning

The critical challenge of Semi-Supervised Learning (SSL) is how to effectively leverage the limited labeled data and massive unlabeled data to improve the model's generalization performance. In [1], we first revisit the popular pseudo-labeling methods via a unified sample weighting formulation and demonstrate the inherent quantity-quality trade-off problem of pseudo-labeling with thresholding, which may prohibit learning. To this end, we propose SoftMatch to overcome the trade-off by maintaining both high quantity and high quality of pseudo-labels during training, effectively exploiting the unlabeled data. We derive a truncated Gaussian function to weight samples based on their confidence, which can be viewed as a soft version of the confidence threshold (see Fig. 28.25). We further enhance the utilization of weakly-learned classes by proposing a uniform alignment approach. In experiments, SoftMatch shows substantial improvements across a wide variety of benchmarks, including image, text, and imbalanced classification.

### Self-adaptive Thresholding for Semi-supervised Learning

Semi-supervised Learning (SSL) has witnessed great success owing to the impressive performances brought by various methods based on pseudo labeling and consistency regularization. However, we argue that existing methods might fail to utilize the unlabeled data more effectively since they either use a pre-defined / fixed threshold or an ad-hoc threshold adjusting scheme, resulting in inferior performance and slow convergence. In [7], we first analyze a motivating example to obtain intuitions on the relationship between the desirable threshold and model's learning status. Based on the analysis, we hence propose FreeMatch to adjust

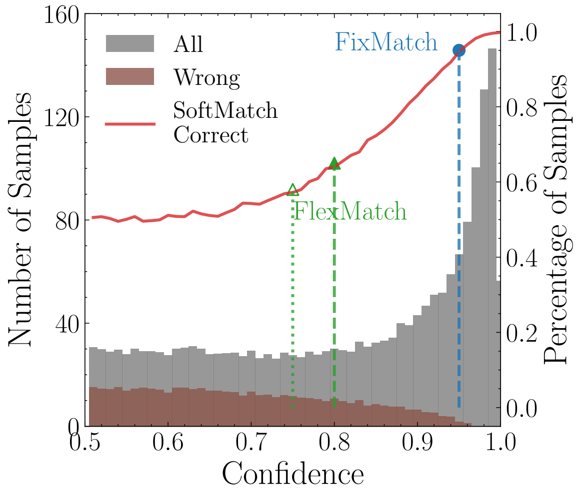



Figure 28.25: Confidence distribution, including all predictions and wrong predictions. The red line denotes the correct percentage of samples used by SoftMatch. The part of the line above scatter points denotes the correct percentage for FixMatch (blue) and FlexMatch (green).

the confidence threshold in a self-adaptive manner according to the model's learning status as in Fig. 28.26. We further introduce a self-adaptive class fairness regularization penalty to encourage the model for diverse predictions during the early training stage. Extensive experiments indicate the superiority of FreeMatch especially when the labeled data are extremely rare.

### Co-Learning of Representation and Classifier for Imbalanced Semi-Supervised Learning

In [2], we propose a novel co-learning framework (CoSSL) with decoupled representation learning and classifier learning for imbalanced SSL, see Fig. 28.27. To handle the data imbalance, we devise Tail-class Feature Enhancement (TFE) for classifier learning. Furthermore, the current evaluation protocol for imbalanced SSL focuses only on balanced test sets, which has limited practicality in real-world scenarios. Therefore, we further conduct a comprehensive evaluation under various shifted test distributions. In experiments, we show that our approach outperforms other methods over a large range of shifted distributions, achieving state-of-the-art performance on benchmark datasets ranging from CIFAR-10, CIFAR-100, ImageNet, to Food-101.

### Simple but Strong Baseline for Boosting Performance of Open-Set Semi-Supervised Learning

Semi-supervised learning (SSL) methods effectively leverage unlabeled data to improve model generalization. However, SSL models often underperform in open-set scenarios, where unlabeled data contain outliers from novel categories that do not appear in the labeled set. In [4], we study the challenging and realistico pen-set SSL setting, where the goal is to

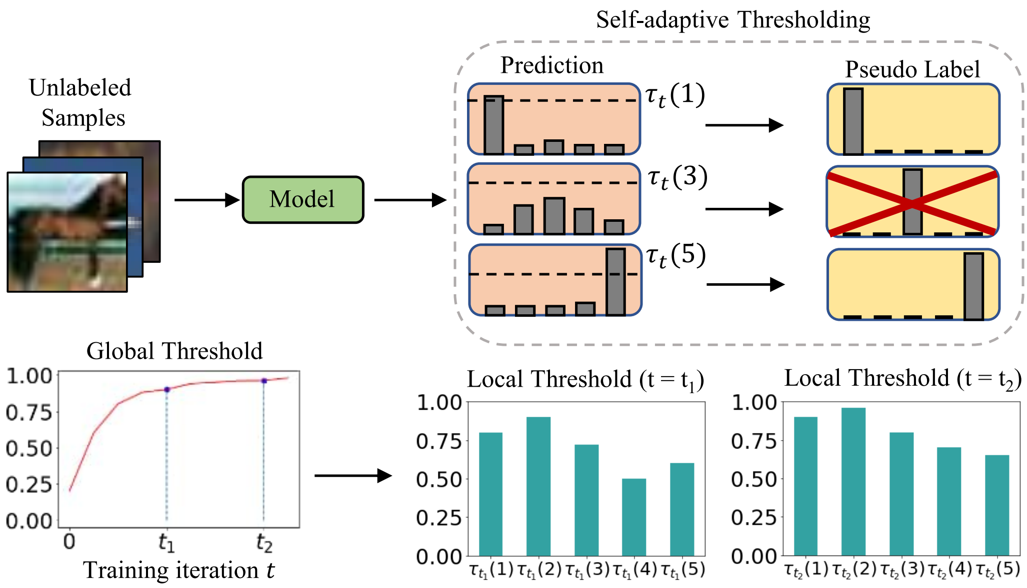



Figure 28.26: Illustration of Self-Adaptive Thresholding (SAT). FreeMatch adopts both global and local self-adaptive thresholds computed from the EMA of prediction statistics from unlabeled samples. Filtered (masked) samples are marked with red X.

both correctly classify inliers and to detect outliers. Intuitively, the inlier classifier should be trained on inlier data only. However, we find that inlier classification performance can be largely improved by incorporating high-confidence pseudo-labeled data, regardless of whether they are inliers or outliers. Also, we propose to utilize non-linear transformations to separate the features used for inlier classification and outlier detection in the multi-task learning framework, preventing adverse effects between them. Additionally, we introduce pseudo-negative mining, which further boosts outlier detection performance. The three ingredients lead to what we call **S**imple but **S**trong **B**aseline (SSB) for open-set SSL (see Fig. 28.28). In experiments, SSB greatly improves both inlier classification and outlier detection performance, outperforming existing methods by a large margin. Our code will be publicly available.

## References


- [1] H. Chen, R. Tao, Y. Fan, Y. Wang, M. Savvides, J. Wang, B. Raj, X. Xie, and B. Schiele. SoftMatch: Addressing the quantity-quality tradeoff in semi-supervised learning. In *Eleventh International Conference on Learning Representations (ICLR 2023)*, Kigali, Rwanda, 2023. OpenReview.net. Accepted.
- [2] Y. Fan, D. Dai, and B. Schiele. CoSSL: Co-learning of representation and classifier for imbalanced semi-supervised learning. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2022)*, New Orleans, LA, USA, 2022, pp. 14554–14564. IEEE.
- [3] Y. Fan, A. Kukleva, D. Dai, and B. Schiele. Revisiting consistency regularization for semi-supervised learning. *International Journal of Computer Vision*, 131:626–643, 2023.

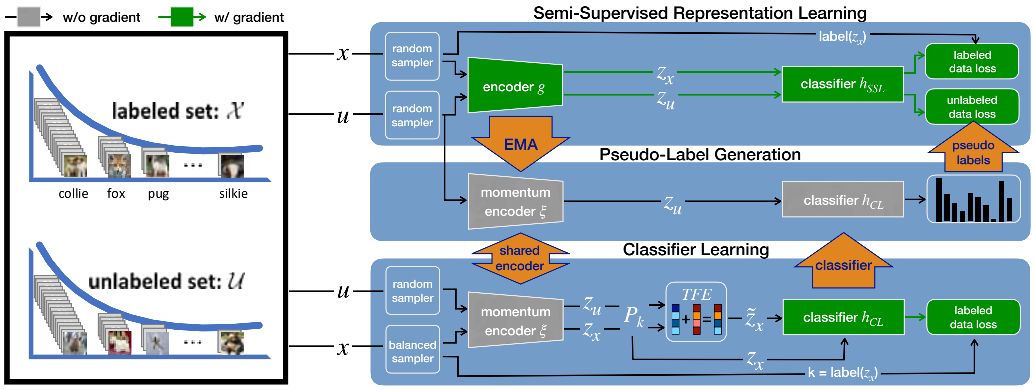



Figure 28.27: Our co-learning framework CoSSL decouples the training of representation and classifier while coupling them in a non-gradient manner. CoSSL consists of three modules: a semi-supervised representation learning module, a balanced classifier learning module, and a carefully designed pseudo-label generation module. The representation module provides a momentum encoder for feature extraction in the other two modules, and the classifier module produces a balanced classifier using our novel Tail-class Feature Enhancement (TFE). Then, pseudo-label module generates pseudo-labels for the representation module using the momentum encoder and the balanced classifier. The interplay between these modules enhances each other, leading to both a more powerful representation and a more balanced classifier. Additionally, our framework is flexible as it can accommodate any standard SSL methods and classifier learning methods.


- [4] Y. Fan, A. Kukleva, D. Dai, and B. Schiele. Ssb: Simple but strong baseline for boosting performance of open-set semi-supervised learning. In *Under Submission to the International Conference of Computer Vision (ICCV 2023)*, 2023.

- [5] Y. Fan, A. Kukleva, and B. Schiele. Revisiting consistency regularization for semi-supervised learning. In C. Bauckhage, J. Gall, and A. Schwing, eds., *Pattern Recognition (GCPR 2021)*, Bonn, Germany, 2021, LNCS 13024, pp. 63–78. Springer.

- [6] Y. Wang, H. Chen, Y. Fan, W. Sun, R. Tao, W. Hou, R. Wang, L. Yang, Z. Zhou, L.-Z. Guo, H. Qi, Z. Wu, Y.-F. Li, S. Nakamura, W. Ye, M. Savvides, B. Raj, T. Shinozaki, B. Schiele, J. Wang, X. Xie, and Y. Zhang. USB: A unified semi-supervised learning benchmark for classification. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, eds., *Advances in Neural Information Processing Systems 35 (NeurIPS 2022)*, New Orleans, LA, USA, 2022, pp. 3938–3961. Curran Associates, Inc.

- [7] Y. Wang, H. Chen, Q. Heng, W. Hou, Y. Fan, Z. Wu, J. Wang, M. Savvides, T. Shinozaki, B. Raj, B. Schiele, and X. Xie. FreeMatch: Self-adaptive thresholding for semi-supervised learning. In *Eleventh International Conference on Learning Representations (ICLR 2023)*, Kigali, Rwanda, 2023. OpenReview.net. Accepted.

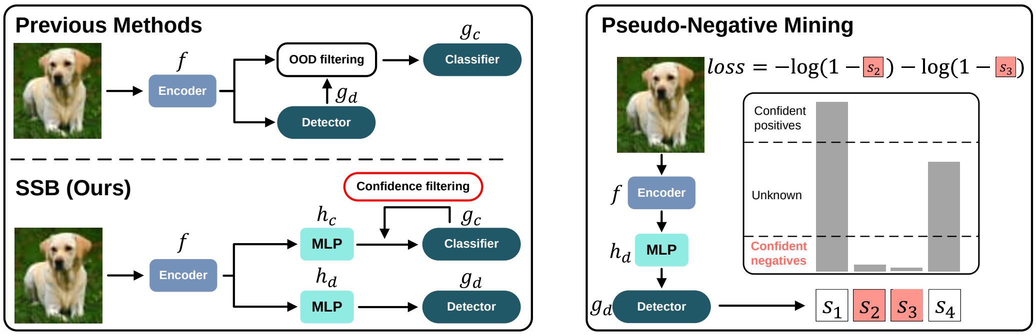



Figure 28.28: **Left:** Our baseline for open-set SSL consists of an inlier classifier $g_c$, an outlier detector $g_d$, and a shared feature encoder $f$ whose features are separated from the task-specific heads by two projection heads $h_c$ and $h_d$. Unlike the detector-based filtering, we adopt confidence-based pseudo-labeling by the inlier classifier to leverage useful OOD data for classifier training. For detector training, we train one-vs-all (OVA) classifiers as in OpenMatch. **Right:** Given the inlier scores ($s_1$ to $s_4$), pseudo-negative mining selects confident negatives ($s_2$ and $s_3$ in the figure), whose inlier scores are lower than a pre-defined threshold, as pseudo-outliers to help the outlier detector training.

### 28.4.7 Continual Learning and Meta Learning

*Investigators: Yaoyao Liu, Bernt Schiele, in cooperation with Qianru Sun (Singapore Management University), Zilin Luo (Singapore Management University), Christian Rupprecht (University of Oxford), Andrea Vedaldi (University of Oxford), Yingying Li (California Institute of Technology)*

This subsection presents our contributions on continual learning and meta learning.

The goal of continual learning is to develop a classification model that gradually learns from training data of different classes, phase-by-phase. In each phase, the classifier is trained using both the new class data and the memorized samples of old classes. The model is then evaluated on the test data of both old and new classes. Incremental learning allows the classification model to continually improve its performance as it receives additional data, making it suitable for tasks where new classes are continuously being introduced.

Meta-learning can be used to optimize hyperparameters of deep models. Technically, the meta-learning process can be formulated as a bilevel optimization program (BOP) where the network parameters are updated at one level, and the key hyperparameters are updated at another level.

#### Class-Incremental Exemplar Compression for Class-Incremental Learning

Exemplar-based class-incremental learning (CIL) finetunes the model with all samples of new classes but few-shot exemplars of old classes in each incremental phase, where the "few-shot" abides by the limited memory budget. In [6], we break this "few-shot" limit based on a simple yet surprisingly effective idea: compressing exemplars by downsampling non-discriminative

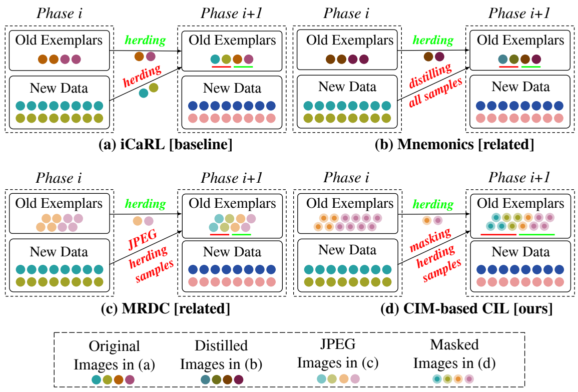



Figure 28.29: The phase-wise training data in different methods. (a) iCaRL is the baseline method using full new class data and few-shot old class exemplars. (b) Mnemonics distills all training samples into few-shot exemplars without increasing their quantity. (c) MRDC compresses each exemplar uniformly into a low-resolution image using JPEG. (d) Our approach [6] based on the proposed class-incremental masking (CIM) downsamples only non-discriminative pixels in the image.

pixels and saving "many-shot" compressed exemplars in the memory. Without needing any manual annotation, we achieve this compression by generating 0-1 masks on discriminative pixels from class activation maps (CAM). We propose an adaptive mask generation model called class-incremental masking (CIM) to explicitly resolve two difficulties of using CAM: 1) transforming the heatmaps of CAM to 0-1 masks with an arbitrary threshold leads to a trade-off between the coverage on discriminative pixels and the quantity of exemplars, as the total memory is fixed; and 2) optimal thresholds vary for different object classes, which is particularly obvious in the dynamic environment of CIL. We optimize the CIM model alternatively with the conventional CIL model through a bilevel optimization problem. We conduct extensive experiments on high-resolution CIL benchmarks including Food-101, ImageNet-100, and ImageNet-1000, and show that using the compressed exemplars by CIM can achieve a new state-of-the-art CIL accuracy, e.g., 4.8 percentage points higher than FOSTER on 10-Phase ImageNet-1000.

#### Online Hyperparameter Optimization for Class-Incremental Learning

Class-incremental learning (CIL) aims to train a classification model while the number of classes increases phase-by-phase. An inherent challenge of CIL is the stability-plasticity tradeoff, i.e., CIL models should keep stable to retain old knowledge and keep plastic to absorb new knowledge. However, none of the existing CIL models can achieve the optimal tradeoff in different data-receiving settings–where typically the training-from-half (TFH) setting needs more stability, but the training-from-scratch (TFS) needs more plasticity. To this end, in [2], we design an online learning method that can adaptively optimize the tradeoff without knowing the setting as a priori. Specifically, we first introduce the key hyperparameters that

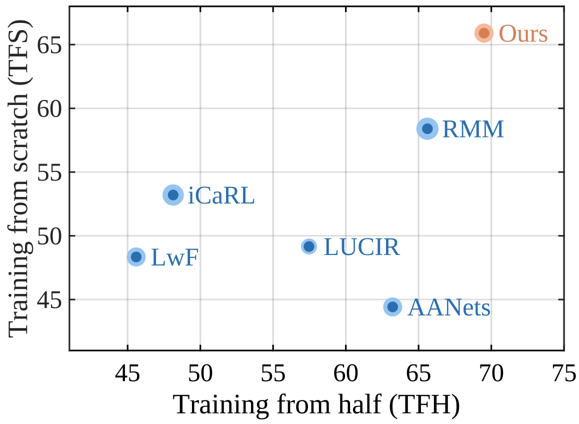



Figure 28.30: Average accuracy (%) on CIFAR-100 25-phase. Our method [2] uses an online learning algorithm to produce the key hyperparameters, e.g., the weights that control which KD losses are used. Thus, our method achieves the highest performance in both TFS and TFH.

influence the trade-off, e.g., knowledge distillation (KD) loss weights, learning rates, and classifier types. Then, we formulate the hyperparameter optimization process as an online Markov Decision Process (MDP) problem and propose a specific algorithm to solve it. We apply local estimated rewards and a classic bandit algorithm Exp3 to address the issues when applying online MDP methods to the CIL protocol. Our method consistently improves top-performing CIL methods in both TFH and TFS settings, e.g., boosting the average accuracy of TFH and TFS by 2.2 percentage points on ImageNet-Full, compared to the state-of-the-art.

### RMM: Reinforced Memory Management for Class-Incremental Learning

Class-Incremental Learning trains classifiers under a strict memory budget: in each incremental phase, learning is done for new data, most of which is abandoned to free space for the next phase. The preserved data are exemplars used for replaying. However, existing methods use a static and ad hoc strategy for memory allocation, which is often sub-optimal. In [4], we propose a dynamic memory management strategy that is optimized for the incremental phases and different object classes. We call our method reinforced memory management (RMM), leveraging reinforcement learning. RMM training is not naturally compatible with CIL as the past, and future data are strictly non-accessible during the incremental phases. We solve this by training the policy function of RMM on pseudo CIL tasks, e.g., the tasks built on the data of the zeroth phase, and then applying it to target tasks. RMM propagates two levels of actions: Level-1 determines how to split the memory between old and new classes, and Level-2 allocates memory for each specific class. In essence, it is an optimizable and general method for memory management that can be used in any replaying-based CIL method. For evaluation, we plug RMM into two top-performing baselines (LUCIR+AANets and POD+AANets and conduct experiments on three benchmarks (CIFAR-100, ImageNet-Subset, and ImageNet-Full).

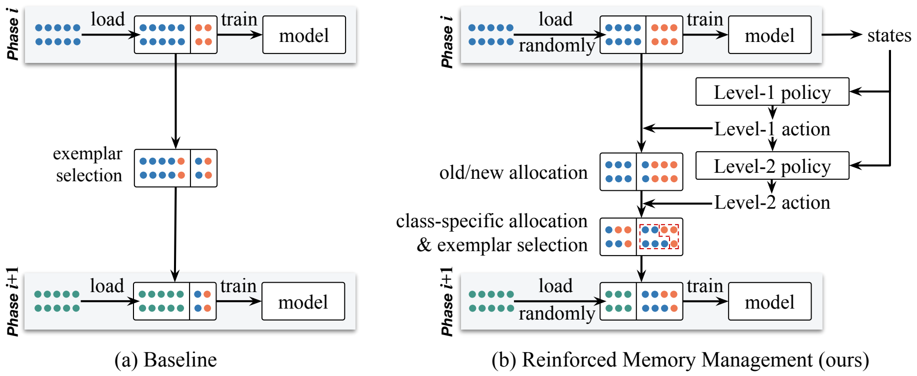



Figure 28.31: a) Existing CIL methods allocate memory between old and new classes in an arbitrary and frozen way, causing the data imbalance between old and new classes and exacerbating the catastrophic forgetting of old knowledge in the learned model. (b) Our proposed method—Reinforced Memory Management (RMM) [4]—is able to learn the optimal and class-specific memory sizes in different incremental phases. Please note we use orange, blue, and green dots to denote the samples observed in the ($i$-1)-th, $i$-th, and ($i$+1)-th phases, respectively.

### Adaptive Aggregation Networks for Class-Incremental Learning

Class-Incremental Learning (CIL) aims to learn a classification model with the number of classes increasing phase-by-phase. An inherent problem in CIL is the stability-plasticity dilemma between the learning of old and new classes, i.e., high-plasticity models easily forget old classes, but high-stability models are weak to learn new classes. We alleviate this issue by proposing a novel network architecture called Adaptive Aggregation Networks (AANets) [3], in which we explicitly build two types of residual blocks at each residual level (taking ResNet as the baseline architecture): a stable block and a plastic block. We aggregate the output feature maps from these two blocks and then feed the results to the next-level blocks. We adapt the aggregation weights in order to balance these two types of blocks, i.e., to balance stability and plasticity, dynamically. We conduct extensive experiments on three CIL benchmarks: CIFAR-100, ImageNet-Subset, and ImageNet, and show that many existing CIL methods can be straightforwardly incorporated into the architecture of AANets to boost their performances.

### Continual Detection Transformer for Incremental Object Detection

Incremental object detection (IOD) aims to train an object detector in phases, each with annotations for new object categories. As other incremental settings, IOD is subject to catastrophic forgetting, which is often addressed by techniques such as knowledge distillation (KD) and exemplar replay (ER). However, KD and ER do not work well if applied directly to state-of-the-art transformer-based object detectors such as Deformable DETR and UP-DETR. In [5], we solve these issues by proposing a ContinuaL DEtection TRansformer (CL-DETR), a new method for transformer-based IOD which enables effective usage of KD

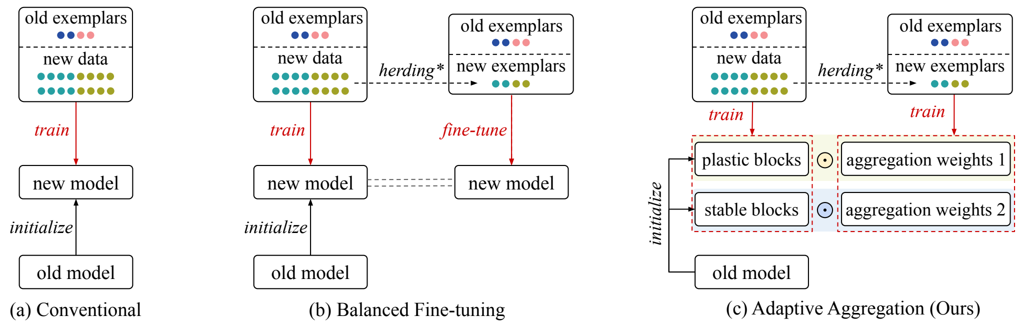



Figure 28.32: Conceptual illustrations of different CIL methods. (a) Conventional methods use all available data (which are imbalanced among classes) to train the model. (b) Recent methods follow this convention but add a fine-tuning step on a balanced subset of all classes. (c) The proposed Adaptive Aggregation Networks (AANets) [3] is a new architecture and it applies a different data strategy: using all available data to update the parameters of plastic and stable blocks, and the balanced set of exemplars to adapt the aggregation weights for these blocks. Our key lies in that adapted weights can balance the usage of the plastic and stable blocks, i.e., balance between plasticity and stability.

and ER in this context. First, we introduce a Detector Knowledge Distillation (DKD) loss, focusing on the most informative and reliable predictions from old versions of the model, ignoring redundant background predictions, and ensuring compatibility with the available ground-truth labels. We also improve ER by proposing a calibration strategy to preserve the label distribution of the training set, therefore better matching training and testing statistics. We conduct extensive experiments on COCO 2017 and demonstrate that CL-DETR achieves state-of-the-art results in the IOD setting.

### Meta-Transfer Learning through Hard Tasks

Meta-learning has been proposed as a framework to address the challenging few-shot learning setting. The key idea is to leverage a large number of similar few-shot tasks in order to learn how to adapt a base-learner to a new task for which only a few labeled samples are available. As deep neural networks (DNNs) tend to overfit using a few samples only, typical meta-learning models use shallow neural networks, thus limiting its effectiveness. In order to achieve top performance, some recent works tried to use the DNNs pre-trained on large-scale datasets but mostly in straight-forward manners, e.g., (1) taking their weights as a warm start of meta-training, and (2) freezing their convolutional layers as the feature extractor of base-learners. In [7], we propose a novel approach called meta-transfer learning (MTL), which learns to transfer the weights of a deep NN for few-shot learning tasks. Specifically, meta refers to training multiple tasks, and transfer is achieved by learning scaling and shifting functions of DNN weights (and biases) for each task. To further boost the learning efficiency of MTL, we introduce the hard task (HT) meta-batch scheme as an effective learning curriculum of few-shot classification tasks. We conduct experiments for five-class few-shot classification tasks on three challenging benchmarks, mini ImageNet, tiered ImageNet, and

Fewshot-CIFAR100 (FC100), in both supervised and semi-supervised settings. Extensive comparisons to related works validate that our MTL approach trained with the proposed HT meta-batch scheme achieves top performance. An ablation study also shows that both components contribute to fast convergence and high accuracy.

### Learning to Teach and Learn for Semi-supervised Few-shot Image Classification

In [1], we presents a novel semi-supervised few-shot image classification method named Learning to Teach and Learn (LTTL) to effectively leverage unlabeled samples in small-data regimes. Our method is based on self-training, which assigns pseudo labels to unlabeled data. However, the conventional pseudo-labeling operation heavily relies on the initial model trained by using a handful of labeled data and may produce many noisy labeled samples. We propose to solve the problem with three steps: firstly, cherry-picking searches valuable samples from pseudo-labeled data by using a soft weighting network; and then, cross-teaching allows the classifiers to teach mutually for rejecting more noisy labels. A feature synthesizing strategy is introduced for cross-teaching to avoid clean samples being rejected by mistake; finally, the classifiers are fine-tuned with a few labeled data to avoid gradient drifts. We use the meta-learning paradigm to optimize the parameters in the whole framework. The proposed LTTL combines the power of meta-learning and self-training, achieving superior performance compared with the baseline methods on two public benchmarks.

## References


- [1] X. Li, J. Huang, Y. Liu, Q. Zhou, S. Zheng, B. Schiele, and Q. Sun. Learning to teach and learn for semi-supervised few-shot image classification. *Computer Vision and Image Understanding*, 212, Article 103270, 2021.
- [2] Y. Liu, Y. Li, B. Schiele, and Q. Sun. Online hyperparameter optimization for class-incremental learning. In *Proceedings of the 37th AAAI Conference on Artificial Intelligence*, Washington, DC, USA, 2023. AAAI. Accepted.
- [3] Y. Liu, B. Schiele, and Q. Sun. Adaptive aggregation networks for class-incremental learning. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2021)*, Nashville, TN, USA (Virtual), 2021, pp. 2544–2553. IEEE.
- [4] Y. Liu, B. Schiele, and Q. Sun. RMM: Reinforced memory management for class-incremental learning. In M. Ranzato, A. Beygelzimer, P. S. Liang, J. W. Vaughan, and Y. Dauphin, eds., *Advances in Neural Information Processing Systems 34 (NeurIPS 2021)*, Virtual, 2021, pp. 3478–3490. Curran Associates, Inc.
- [5] Y. Liu, B. Schiele, A. Vedaldi, and C. Rupprecht. Continual detection transformer for incremental object detection. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR*, 2023.
- [6] Z. Luo, Y. Liu, B. Schiele, and Q. Sun. Class-incremental exemplar compression for class-incremental learning. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR*, 2023.
- [7] Q. Sun, Y. Liu, Z. Chen, T.-S. Chua, and B. Schiele. Meta-transfer learning through hard tasks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(3):1443–1456, 2022.

### 28.4.8 Multiple Object Tracking

*Investigators: Mattia Segu, Bernt Schiele, in cooperation with Tao Sun, Janis Postels, Yuxuan Wang, Luc Van Gool, Fisher Yu (ETH Zurich), Federico Tombari (Google)*

Multiple object tracking (MOT) represents a cornerstone of modern perception systems for challenging computer vision applications, such as autonomous driving, video surveillance, and augmented reality. Laying the ground for safety-critical downstream perception and planning tasks – e. g. obstacle avoidance, motion estimation, prediction of vehicles and pedestrians intentions, and the consequent path planning – the robustness of MOT to diverse conditions is of uttermost importance. However, domain shift could result in life-threatening failures of MOT pipelines, due to the perception system's inability to understand previously unseen environments and provide meaningful signals for downstream planning.

Despite the urge of addressing domain adaptation for MOT to enable safer driving and video analysis, no solution has ever been proposed. Moreover, most of the popular driving datasets provide annotations only for a limited variety of tasks and domains, hindering the progress in domain adaptation for multiple object tracking and other fundamental perception tasks. In [2], we introduce SHIFT, a synthetic driving dataset for continuous multi-task domain adaptation (28.4.8), enabling multiple tracking research under domain shift; in [1], we propose the first solution to the test-time domain adaptation problem for multiple object tracking (28.4.8).

#### SHIFT: A Synthetic Driving Dataset for Continuous Multi-Task Domain Adaptation

*Investigators: Mattia Segu and Bernt Schiele, in cooperation with Tao Sun, Janis Postels, Yuxuan Wang, Luc Van Gool, Fisher Yu (ETH Zurich), Federico Tombari (Google)*

Adapting to a continuously evolving environment is a safety-critical challenge inevitably faced by all autonomous driving systems. Existing image and video driving datasets, however, fall short of capturing the mutable nature of the real world.

In [2], we introduce the largest multi-task synthetic dataset for autonomous driving, SHIFT. It presents discrete and continuous shifts in cloudiness, rain and fog intensity, time of day, and vehicle and pedestrian density (28.33). Featuring a comprehensive sensor suite and annotations for several mainstream perception tasks, SHIFT allows investigating the degradation of a perception system performance at increasing levels of domain shift, fostering the development of continuous adaptation strategies to mitigate this problem and assess model robustness and generality. As discussed above, multiple object tracking is a safety-critical problem, and our dataset SHIFT provides the tools to assess the transferability and adaptation of MOT systems across domains. Our dataset and benchmark toolkit are publicly available at `www.vis.xyz/shift`.

#### DARTH: Holistic Test-time Adaptation for Multiple Object Tracking

*Investigators: Mattia Segu and Bernt Schiele, in cooperation with Fisher Yu (ETH Zurich)*

Despite the urge of safety in driving systems, no solution to the MOT adaptation problem to domain shift in test-time conditions has ever been proposed.

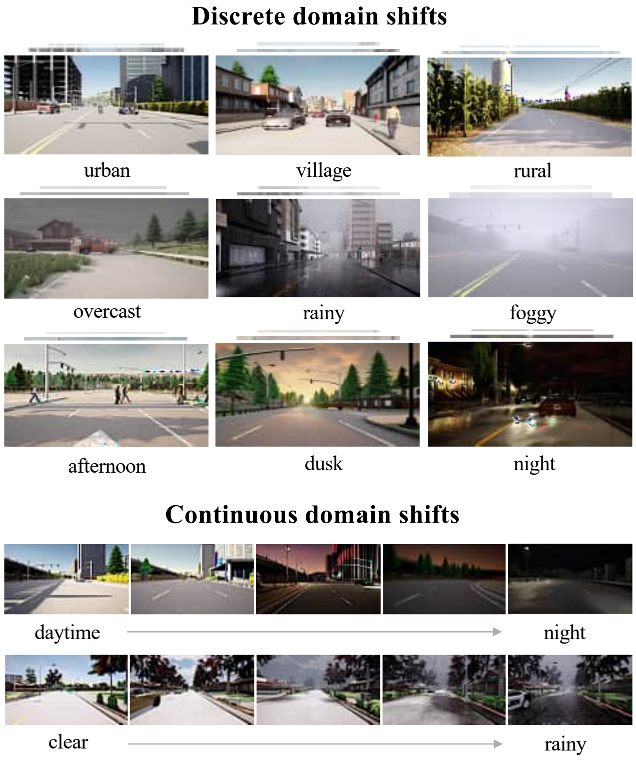



Figure 28.33: SHIFT provides: **discrete domain shifts**, a set of sequences each collected using different domain parameters and initial states; **continuous domain shifts**, a set of sequences where domain parameters change continuously during driving.

In [1], we analyze the effect of domain shift on multiple object tracking (28.34), and propose a test-time adaptation solution to counteract it. We focus on appearance-based tracking, which shows state-of-the-art performance across a variety of datasets, outperforms motion-based trackers in complex scenarios – i. e. BDD100K – and complements motion cues for superior tracking performance. Since appearance-based trackers associate detections through time based on the similarity of their learnable appearance embeddings, domain shift threatens the performance of both their detection and instance association stages.

In particular, we analyze the effect of domain shift on appearance-based trackers, and introduce DARTH, a holistic test-time adaptation framework for MOT (28.35). We propose a detection consistency formulation to adapt object detection in a self-supervised fashion, while adapting the instance appearance representations via our novel patch contrastive loss. We evaluate our method on a variety of domain shifts – including sim-to-real, outdoor-to-indoor, indoor-to-outdoor – and substantially improve the source model performance on all metrics.

## References


- [1] M. Segu, B. Schiele, and F. Yu. Darth: Holistic test-time adaptation for multiple object tracking. In *Under Submission to the International Conference of Computer Vision (ICCV 2023)*, 2023.

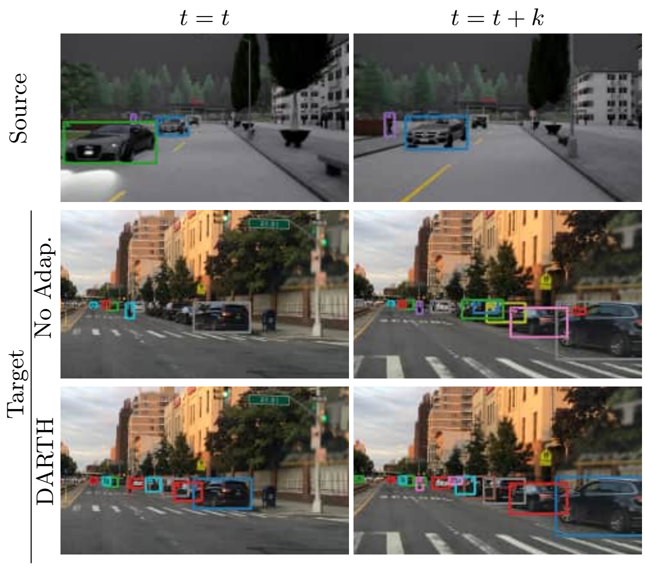



Figure 28.34: We illustrate the effect of domain shift on multiple object tracking, and how our test-time adaptation technique (DARTH) counteracts it. The top row shows the in-domain performance of a model trained on the synthetic dataset SHIFT (Source); the same model (No Adap.) suffers from domain shift when deployed on the real-world BDD100K (Target); the bottom row shows the benefits of DARTH. Each row shows two frames spaced by $k$=2 seconds; boxes of the same color correspond to the same tracking ID.

• [2] T. Sun, M. Segù, J. Postels, Y. Wang, L. Van Gool, B. Schiele, F. Tombari, and F. Yu. SHIFT: A synthetic driving dataset for continuous multi-task domain adaptation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2022)*, New Orleans, LA, USA, 2022, pp. 21339–21350. IEEE.

### 28.4.9 Human Performance Capture

*Investigators: Zhi Li, Bernt Schiele, Christian Theobalt, Soshi Shimada, Vladislav Golyanik, Weipeng Xu, in cooperation with Patrick Pérez (Valeo.ai)*

We aim at the very challenging yet highly practical problem of marker-less monocular 3D human motion capture (MoCap) with scene interactions.

Marker-less monocular 3D human motion capture (MoCap) with scene interactions is a challenging research topic relevant for extended reality, robotics and virtual avatar generation. Due to the inherent depth ambiguity of monocular settings, 3D motions captured with existing methods often contain severe artefacts such as incorrect body-scene inter-penetrations, jitter and body floating. To tackle these issues, we propose HULC in [2], a new approach for 3D human MoCap which is aware of the scene geometry. HULC estimates 3D poses and dense body-environment surface contacts for improved 3D localisations, as well as the absolute scale of the subject. Furthermore, we introduce a 3D pose trajectory optimisation based on a novel pose manifold sampling that resolves erroneous body-environment inter-penetrations.

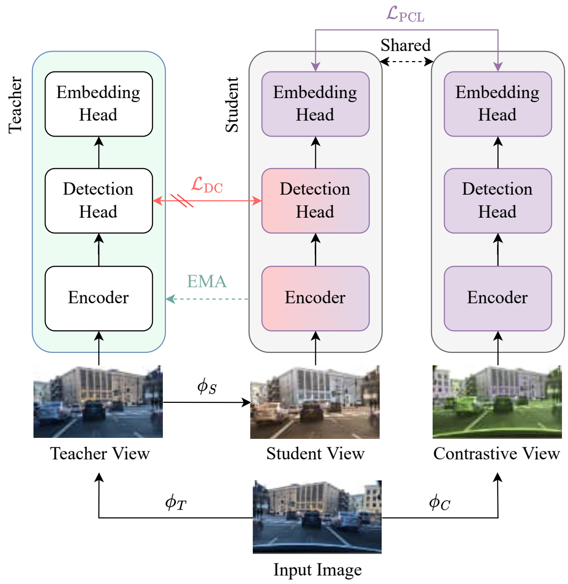



Figure 28.35: Schematic representation on the target domain of DARTH, our test-time adaptation method for MOT. Our patch contrastive loss $\mathcal{L}_{\text{PCL}}$ between the siamese student's instance embeddings adapts instance association. Our detection consistency loss $\mathcal{L}_{\text{DC}}$ enforces consistency to photometric changes. The EMA updates to the teacher gradually improve the detection targets for our consistency loss. $\phi_T$, $\phi_S$, and $\phi_C$ are the image transformations. '\\' = stop gradient.

Although the proposed method requires less structured inputs compared to existing scene-aware monocular MoCap algorithms, it produces more physically-plausible poses: HULC significantly and consistently outperforms the existing approaches in various experiments and on different metrics.

To take one step forward, in the scope of 3D human motion capture from monocular RGB images respecting scene interactions, we consider complex and possibly deformable environments. It is a very challenging, illposed and under-explored problem. Existing methods address it only weakly and do not model possible surface deformations often occurring when humans interact with scene surfaces. In contrast, in [1] we propose MoCapDeform, i.e., a new framework for monocular 3D human motion capture that is the first to explicitly model non-rigid deformations of a 3D scene for improved 3D human pose estimation and deformable environment reconstruction, shown in Fig. 28.36. MoCapDeform accepts a monocular RGB video and a 3D scene mesh aligned in the camera space. It first localises a subject in the input monocular video along with dense contact labels using a new raycasting based strategy. Next, our human-environment interaction constraints are leveraged to jointly optimise global 3D human poses and non-rigid surface deformations. MoCapDeform achieves superior accuracy than competing methods on several datasets, including our newly recorded one with deforming background scenes.

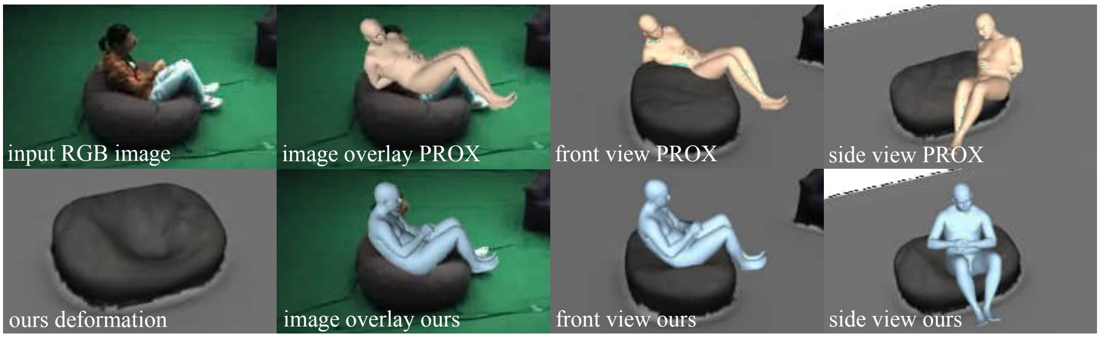



Figure 28.36: Existing monocular 3D human motion capture methods such as the previous SOTA method PROX ignore abundant scene deformation when penalising human-scene collisions, resulting in erroneous global poses (top). Our MoCapDeform algorithm is the first that models non-rigid scene deformations and finds the accurate global 3D poses of the subject by human-deformable scene interaction constraints, achieving increased accuracy with significantly fewer penetrations (bottom).

### References


- [1] Z. Li, S. Shimada, B. Schiele, C. Theobalt, and V. Golyanik. MoCapDeform: Monocular 3D human motion capture in deformable scenes. In *International Conference on 3D Vision*, Hybrid / Prague, Czechia, 2022, pp. 1–11. IEEE.
- [2] S. Shimada, V. Golyanik, Z. Li, P. Pérez, W. Xu, and C. Theobalt. HULC: 3D HUman motion capture with pose manifold sampLing and dense Contact guidance. In S. Avidan, G. Brostow, M. Cissé, G. M. Farinella, and T. Hassner, eds., *Computer Vision – ECCV 2022*, Tel Aviv, Israel, 2022, LNCS 13682, pp. 516–533. Springer.


## 28.4.10 People

*Investigators: Di Chen and Bernt Schiele, in cooperation with Shanshan Zhang and Jian Yang (Nanjing University of Science and Technology), Andreas Doering and Juergen Gall (University of Bonn), Qiuhong Ke (The University of Melbourne), Mario Fritz (CISPA Helmholtz Center for Information Security)*

Video-based person re-identification (re-ID) is an important technique in visual surveillance systems which aims to match video snippets of people captured by different cameras. Existing methods are mostly based on convolutional neural networks (CNNs), whose building blocks either process local neighbor pixels at a time, or, when 3D convolutions are used to model temporal information, suffer from the misalignment problem caused by person movement. In [1], we propose to overcome the limitations of normal convolutions with a human-oriented graph method (see Fig. 28.37). Specifically, features located at person joint keypoints are extracted and connected as a spatial-temporal graph. These keypoint features are then updated by message passing from their connected nodes with a graph convolutional network (GCN). During training, the GCN can be attached to any CNN-based person re-ID model to assist representation learning on feature maps, whilst it can be dropped after training for

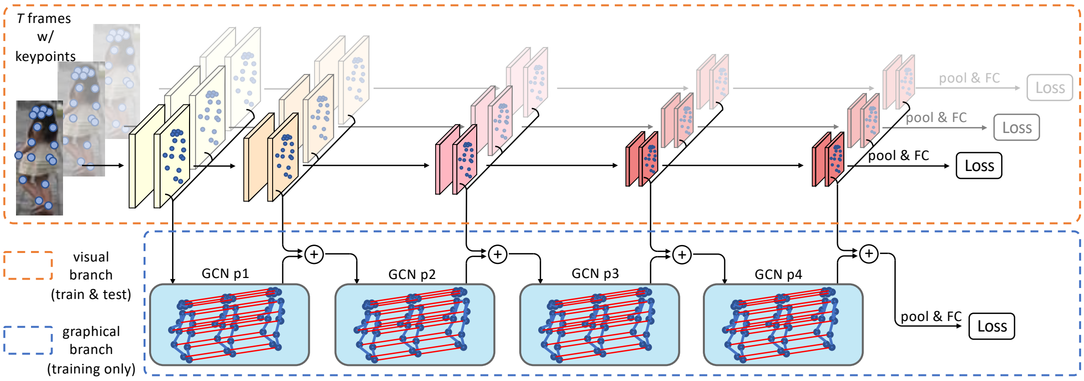



Figure 28.37: Overall pipeline for Keypoint Message Passing. The visual branch is a base CNN which is a typical 5-stage model and takes in a video with $T$ frames as input. The graphical branch is a GCN divided into 4 parts. Given the keypoint locations, we extract features according to the locations on the first CNN stage and use them as the inputs to the GCN. At the end of each stage, keypoint features are fused with the intermediate representations from GCN with an element-wise sum. The fused features serve as the new inputs to the subsequent GCN layers. Both the CNN and GCN are supervised with cross-entropy loss during training. Note, that the GCN branch including keypoint estimation can be dropped during inference.

better inference speed. Our method brings significant improvements over the CNN-based baseline model on the MARS dataset with generated person keypoints and a newly annotated dataset: PoseTrackReID. It also defines a new state-of-the-art method in terms of top-1 accuracy and mean average precision in comparison to prior works.

Pedestrian detection and re-identification have progressed significantly in the last few years. However, occluded people are notoriously hard to detect and recognize, as their appearance varies substantially depending on a wide range of occlusion patterns. In [5], we aim to propose a simple and compact method based on CNNs for occlusion handling. We start with interpreting CNN channel features of a pedestrian detector, and we find that different channels activate responses for different body parts respectively. These findings motivate us to employ an attention mechanism across channels to represent various occlusion patterns in one single model, as each occlusion pattern can be formulated as some specific combination of body parts. Therefore, an attention network with self or external guidances is proposed as an add-on to the baseline CNN method. Also, we propose an attention guided self-paced learning method to balance the optimization across different occlusion levels. Our proposed method shows significant improvements over the baseline methods for both pedestrian detection and re-identification tasks. For pedestrian detection, we achieve a considerable improvement of 8pp to the baseline FasterRCNN detector on the heavy occlusion subset of CityPersons and on Caltech we outperform the state-of-the-art method by 5pp. For pedestrian re-identification, our method surpasses the baseline and achieves state-of-the-art performance on multiple re-identification benchmarks.

Person detection and Re-identification are two welldefined support tasks for practically

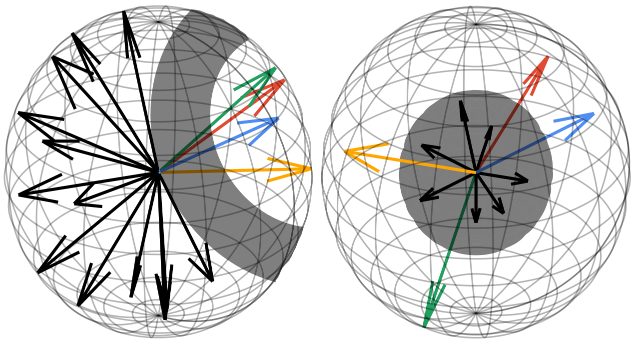


Figure 28.38: Illustration on how person and background representations are scattered in the embedding space. Black arrows denote background while colorful ones denote persons with different identities. Gray surfaces are the decision boundary for person and background. (Left) For $L_2$ normalized embeddings, the inter-class angle distances for different persons are squeezed by backgrounds. (Right) Norm-aware embeddings separate persons and background by norms and discriminate person identities by angles, thus the constrain on inter-class distances is relaxed.

relevant tasks such as Person Search and Multiple Person Tracking. Person Search aims to find and locate all instances with the same identity as the query person in a set of panoramic gallery images. Similarly, Multiple Person Tracking, especially when using the tracking-by-detection pipeline, requires to detect and associate all appeared persons in consecutive video frames. One major challenge shared by the two tasks comes from the contradictory goals of detection and re-identification, i.e., person detection focuses on finding the commonness of all persons while person re-ID handles the differences among multiple identities. Therefore, it is crucial to reconcile the relationship between the two support tasks in a joint model. In [2], we present a novel approach called Norm-Aware Embedding to disentangle the person embedding into norm and angle for detection and re-ID respectively, allowing for both effective and efficient multi-task training. An illustration on the concept of Norm-Aware Embedding is shown in Fig. 28.38. Our Norm-Aware Embedding achieves remarkable performance on both person search and multiple person tracking benchmarks, with the merit of being easy to train and resource-friendly.

Current research evaluates person search, multi-object tracking and multi-person pose estimation as separate tasks and on different datasets although these tasks are very akin to each other and comprise similar sub-tasks, e.g. person detection or appearance-based association of detected persons. Consequently, approaches on these respective tasks are eligible to complement each other. Therefore, in [3], we introduce PoseTrack21, a large-scale dataset for person search, multi-object tracking and multi-person pose tracking in real-world scenarios with a high diversity of poses. The dataset provides rich annotations like human pose annotations including annotations of joint occlusions, bounding box annotations even for small persons, and person-ids within and across video sequences. The dataset allows to evaluate multi-object tracking and multi-person pose tracking jointly with person re-identification or exploit structural knowledge of human poses to improve person search and tracking, particularly in the context of severe occlusions. With PoseTrack21, we want to encourage researchers to work on joint approaches that perform reasonably well on all three tasks.

The goal of Assessing Future Moment of an Action of Interest (AFM-AI) is to determine if an action of interest will happen or not as well as the starting moment of the action. In [4], we aim to assess starting moments at any time-horizon of the future. We tackle the regression task of the starting moments as a generation task using a Deterministic Residual Guided Variational Regression Module (DR-VRM), which is built on a Variational Regression Module (VRM) and a deterministic residual network. The VRM takes the uncertainty into account and is capable of generating diverse predictions for the starting moment. The deterministic network encourages the VRM to learn from deterministic residual information in order to generate more precise predictions for moment assessment. Experimental results on three datasets clearly show that the proposed method is capable of generating both diverse and precise predictions of starting moments for query actions.

### References


- [1] D. Chen, A. Doering, S. Zhang, J. Yang, J. Gall, and B. Schiele. Keypoint message passing for video-based person re-identification. In *Proceedings of the 36th AAAI Conference on Artificial Intelligence*, Virtual Conference, 2022, pp. 239–247. AAAI.
- [2] D. Chen, S. Zhang, J. Yang, and B. Schiele. Norm-aware embedding for efficient person search and tracking. *International Journal of Computer Vision*, 129:3154–3168, 2021.
- [3] A. Doering, D. Chen, S. Zhang, B. Schiele, and J. Gall. PoseTrack21: A dataset for person search, multi-object tracking and multi-person pose tracking. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2022)*, New Orleans, LA, USA, 2022, pp. 20931–20940. IEEE.
- [4] Q. Ke, M. Fritz, and B. Schiele. Future moment assessment for action query. In *IEEE Winter Conference on Applications of Computer Vision (WACV 2021)*, Virtual Event, 2021, pp. 3219–3227. IEEE.
- [5] S. Zhang, D. Chen, J. Yang, and B. Schiele. Guided attention in CNNs for occluded pedestrian detection and re-identification. *International Journal of Computer Vision*, 129:1875–1892, 2021.


### 28.4.11 Deblurring

*Investigators: Jiangxin Dong and Bernt Schiele in cooperation with Stefan Roth (TU Darmstadt)*

Image deblurring is a classical image restoration problem. Traditional methods usually separate this problem into two phases, blur kernel estimation and image restoration (i.e., non-blind image deblurring). The goal of non-blind image deblurring is to restore the clear image from its corrupted observation given the blur kernel. This subsection presents our contributions on non-blind image deblurring, combining classical techniques and deep learning.

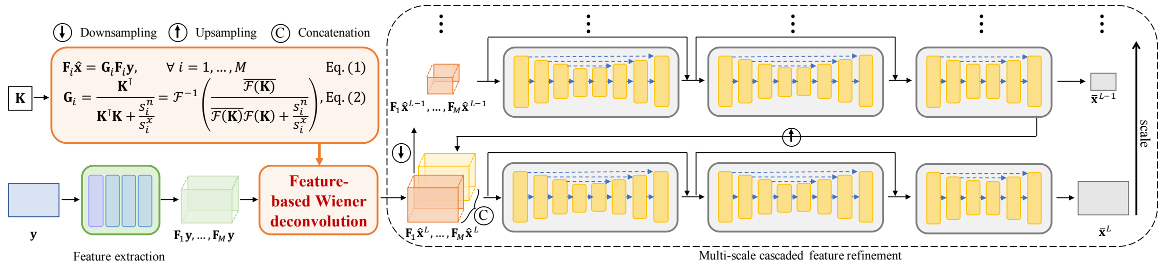



Figure 28.39: Deep Wiener deconvolution network. While previous work mostly relies on a deconvolution in the image space, our neural network first extracts useful feature information from the blurry input image and then conducts an explicit Wiener deconvolution in the (deep) feature space through Eqs. (1) and (2). A multi-scale cascaded encoder-decoder network progressively restores clear images, with fewer artifacts and finer detail. The whole network is trained in an end-to-end manner.

### Learning Spatially-Variant MAP Models for Non-blind Image Deblurring

The classical maximum a-posteriori (MAP) framework for non-blind image deblurring requires defining suitable data and regularization terms, whose interplay yields the desired clear image through optimization. The vast majority of prior work focuses on advancing one of these two crucial ingredients, while keeping the other one standard. Considering the indispensable roles and interplay of both data and regularization terms, in [1], we propose a simple and effective approach to jointly learn these two terms. We show that jointly learning both terms is more effective than learning only one term alone; the difference becomes even more striking in challenging application scenarios. To improve the goodness-of-fit and capture the properties of clear images, the neural networks not only yield suitable image-adaptive features for both terms, but actually predict per-pixel spatially-variant features instead of the commonly used spatially-uniform ones. We develop an end-to-end learning approach to better capture the spatially-variant properties, integrating the MAP-based optimization framework as a constraint for the deep neural network. The resulting spatially-variant data and regularization terms particularly improve the restoration of fine-scale structures and detail. Quantitative and qualitative results underline the effectiveness of our approach.

### DWDN: Deep Wiener Deconvolution Network for Non-Blind Image Deblurring

In [2], we present a simple and effective approach for non-blind image deblurring, combining classical techniques and deep learning. In contrast to existing methods that deblur the image directly in the standard image space, we propose to perform an explicit deconvolution process in a feature space by integrating a classical Wiener deconvolution framework with learned deep features. We show that this feature-space deconvolution is more effective in suppressing artifacts and recovering fine detail compared to previous methods that conduct the deconvolution in the image space. To reconstruct final clear images, a multi-scale cascaded feature refinement module then predicts the deblurred image from the deconvolved deep features, progressively recovering detail and small-scale structures. The proposed model is trained in an end-to-end manner and evaluated on scenarios with simulated Gaussian noise, saturated pixels, or JPEG compression artifacts as well as real-world images. Moreover, we present detailed analyses of the benefit of the feature-based Wiener deconvolution and of the multi-scale cascaded feature refinement as well as the robustness of the proposed approach. Our extensive experimental results show that the proposed deep Wiener deconvolution network facilitates deblurred results with visibly fewer artifacts and quantitatively outperforms state-of-the-art non-blind image deblurring methods by a wide margin.

### References


- [1] J. Dong, S. Roth, and B. Schiele. Learning spatially-variant MAP models for non-blind image deblurring. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2021)*, Nashville, TN, USA (Virtual), 2021, pp. 4886–4895. IEEE.
- [2] J. Dong, S. Roth, and B. Schiele. DWDN: Deep Wiener Deconvolution Network for non-blind image deblurring. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(12):9960–9976, 2022.


## 28.4.12 Image Synthesis

*Investigators: Edgar Schönfeld, Bernt Schiele, in cooperation with Vadim Sushko, Dan Zhang, Anna Khoreva (Bosch Center for Artificial Intelligence), and Juergen Gall (University of Bonn)*

Image synthesis models have the ability to generate strikingly realistic images, leading to their recent surge in popularity as tools for artists, but also for designers in the advertising and gaming industry. Generative adversarial networks (GANs) represent a popular and promising approach to image synthesis. The primary challenges involve achieving training stability and data efficiency, specifically in terms of attaining superior image quality with a fixed dataset size. Here, we concentrate on the latter aspect and explore strategies to enhance image quality by refining the network architecture of GANs.

#### Training semantic image synthesis models with only adversarial supervision

Next to unconditional and class-conditional image synthesis, GANs are also commonly used for the task of generating images from segmentation maps, known as semantic image synthesis. The standard GAN setup uses a single discriminator as a loss function for the generator. However, GANs for semantic image synthesis perform poorly under this standard setup, and therefore rely on ensembles of multiple discriminators and a VGG-based perceptual loss. This setup limits the progress of GAN models for semantic image synthesis. For example, in [1] we show that the perceptual loss strongly suppresses the diversity of the synthesized images. In [1], we propose a novel, simplified GAN model, which needs only the standard GAN setup of one generator and one discriminator to achieve high-quality results. We re-design the discriminator as a semantic segmentation network, directly using the given semantic label maps as the ground truth for training. By providing stronger supervision to the discriminator as well as to the generator through spatially- and semantically-aware discriminator feedback,


Figure 28.40: OASIS multi-modal synthesis results. The 3D noise can be sampled globally (first rows), changing the whole scene, or locally (second row), partially changing the image. For the latter, we sample different noise per region, like the bed segment (in red).

we are able to synthesize images of higher fidelity with better alignment to their input label maps, making the use of the perceptual loss superfluous. We name our model OASIS, since it Only Needs Adversarial Supervision for Semantic Image Synthesis. In contrast to classification-based discriminators, our proposed segmentation-based discriminator allows to directly balance the contribution of each semantic class to the loss, which improves performance on underrepresented classes. Since we no longer require the diversity-suppressing perceptual loss, we propose a technique to enable highly diverse multi-modal image synthesis. In particular, we propose global and local sampling of a 3D noise tensor injected into the generator, which allows complete or partial image change (see Fig. 28.40). We show that images synthesized by our model are more diverse and follow the color and texture distributions of real images more closely. We achieve an average improvement of 6 FID and 5 mIoU points over the state of the art across different datasets using only adversarial supervision. In [2], we investigate semantic image synthesis under severe class imbalance and sparse annotations, which are common aspects in practical applications but were overlooked in prior works. To this end, we evaluate our model on LVIS, a dataset originally introduced for long-tailed object recognition. We thereby demonstrate the high performance of our model in the sparse and unbalanced data regimes, achieved by means of the proposed 3D noise and the ability of our discriminator to balance class contributions directly in the loss function.

## References


- [1] E. Schönfeld, V. Sushko, D. Zhang, J. Gall, B. Schiele, and A. Khoreva. You only need adversarial supervision for semantic image synthesis. In *International Conference on Learning Representations (ICLR 2021)*, Vienna, Austria (Virtual), 2021. OpenReview.net.
- [2] V. Sushko, E. Schönfeld, D. Zhang, J. Gall, B. Schiele, and A. Khoreva. Oasis: Only Adversarial Supervision for Semantic Image Synthesis. *International Journal of Computer Vision*, 130:2903–2923, 2022.

### 28.4.13 Visually Grounded Story Generation

*Investigators: Xudong Hong, Bernt Schiele in cooperation with Asad Sayeed (University of Gothenburg), Vera Demberg, Khushboo Mehra, Dongqi Pu, Pin-Jie Lin, Ernie Chang, Qiankun Zheng (Saarland University)*

*Visually Grounded Story Generation* is a challenging task of generating a coherent story from a series of images. The generated stories must go beyond the literal description of a stand-alone image or scene, requiring not only observable elements such as events, characters, objects, or locations but also information that can be inferred via commonsense knowledge like states of previously mentioned entities, current actions, and subsequent events. Current state-of-the-art systems still struggle to capture the connections of these elements and maintain local coherence, which is crucial for generating high-quality stories. In addition, this task is low-resourced because collecting stories or annotations is expensive and time-consuming. This task also requires large pre-trained vision and language models to achieve synergy so parameter efficiency is important. To address these challenges, we made the following contributions: 1) a parameter-efficient two-stage method for movie script to story generation; 2) an image filtering pipeline and a crowdsourcing user interface to collect a high-quality dataset Visual Writing Prompts for character-based story generation; 3) a character matching metric of input-output relevance and a visual coherence loss function for generating more visually grounded stories.

#### Two-Stage Movie Script Summarization

To obtain more pair data for visually grounded story generation, one method is summarizing the scene-aligned movie scripts into short stories so that we can have scene-story pairs. However, efficiently summarizing movie scripts is challenging because the input is prohibitively long. To tackle this issue, we propose an innovative two-stage hierarchical architecture in [3] (see Figure 28.41). The first stage applies a heuristic extraction method to extract actions and essential dialogues, which not only reduces the average length of input movie scripts by 66% from about 24K to 8K tokens but also retains important information for story generation. This significant reduction in script length makes the summarization process more manageable and efficient. In the second stage, a state-of-the-art encoder-decoder model, Longformer-Encoder-Decoder (LED), is fine-tuned with BitFit and NoisyTune methods to generate summaries that effectively capture the essence of the movie scripts. Evaluations on the unseen test set indicate that our system outperforms both zero-shot LED baselines as well as other participants on various automatic metrics and ranks 1st in the Scriptbase track at COLING 2022, showcasing the effectiveness of our approach.

#### Visual Writing Prompts Dataset for Character-Based Story Generation

Existing image-based story generation datasets often suffer from a lack of coherent plots and limited diversity, which negatively affects the quality of generated stories. To address this limitation and promote more coherent and diverse visual story generation, we introduce a

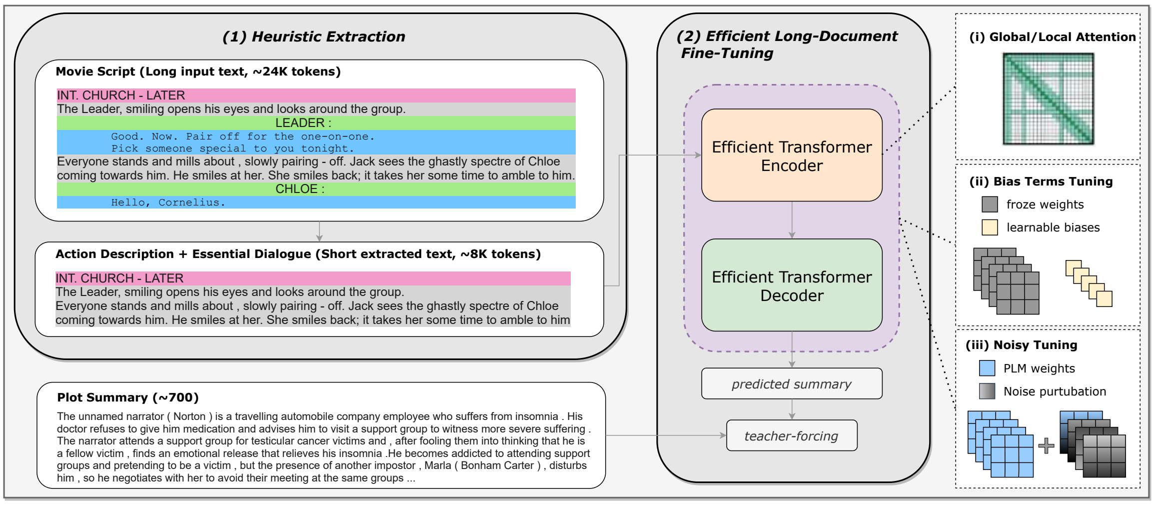



Figure 28.41: Diagram to depict the pipeline. Our system achieves efficient text summarization by (1) a heuristics extraction method to compress long input text into a relatively short input sequence, and (2) efficient long-document fine-tuning. In the first stage, the heuristics extraction method takes (a) long movie scripts with an average length of 24K and compresses them into 8K. (b) Following that, the transformer encoder with (i) global sliding winding attention takes movie scripts up to 8K tokens and is jointly fine-tuned with the decoder using (ii) bias-terms tuning or (iii) noisy tuning techniques to generate a long movie summarization (1K) in a computationally efficient manner.

novel image-grounded dataset called Visual Writing Prompts (VWP) in [2] (see Figure 28.42). VWP contains almost 2K selected sequences of movie shots, each including 5-10 images. These sequences were carefully curated to ensure the presence of coherent plots and meaningful interactions between characters. The image sequences are aligned with a total of 12K stories, which were collected via crowdsourcing given the image sequences and a set of grounded characters from the corresponding image sequence. Our new image sequence collection and filtering process have allowed us to obtain stories that are more coherent and more diverse compared to previous work. We also propose a character-based story generation model driven by coherence as a strong baseline model.

### Visual Coherence Loss Function and Character Matching Metric

Maintaining local coherence in visually grounded story generation is essential for generating high-quality stories. The recent large pre-trained language models still fail while tracking multiple entities. We focus on two key aspects: representing character re-occurrences within the image sequence to mention them correctly in the story and distinguishing different characters. In [1], we propose a loss function inspired by a linguistic theory of coherence for self-supervised learning of image sequence representations. We further propose combining features from an object and a face detector to construct stronger character features. To evaluate input-output relevance that current reference-based metrics don't measure, we propose a character


Figure 28.42: Comparison of Visual Writing Prompts dataset with Visual Storytelling and Travel Blogs datasets. Our dataset has recurring characters across all five images and sub-stories. Each occurrence of a character in a sub-story has a bounding box in the corresponding image, which grounds the textual appearance to visual input.

matching metric to check whether the models generate referring expressions correctly for characters in input image sequences. Experiments on a visual story generation dataset show that our proposed features and loss function are effective for generating more coherent and visually grounded stories.

## References

[1] X. Hong, V. Demberg, A. Sayeed, Q. Zheng, and B. Schiele. Visual coherence loss for coherent and visually grounded story generation. In *Under Submission*, 2023.

• [2] X. Hong, A. Sayeed, K. Mehra, V. Demberg, and B. Schiele. Visual writing prompts: Character-grounded story generation with curated image sequences. *Transactions of the Association for Computational Linguistics*, 11, 2023. E-Text: 2301.08571.

• [3] D. Pu, X. Hong, P.-J. Lin, E. Chang, and V. Demberg. Two-stage movie script summarization: An efficient method for low-resource long document summarization. In *Proceedings of The Workshop on Automatic Summarization for Creative Writing*, Gyeongju, Republic of Korea, 2022, pp. 57–66. Association for Computational Linguistics.

## 28.5 Vision for Autonomous Systems

*Coordinator: Dengxin Dai*

### 28.5.1 Adverse Weather

*Investigators: Dengxin Dai, in cooperation with Luc Van Gool (ETH Zurich and KU Lueven), Martin Hahner (ETH Zurich), and Christos Sakaridis (ETH Zurich)*

Adverse weather and illumination conditions (e.g. fog, rain, snow, low light, nighttime, glare and shadows) create visibility problems for the sensors that power automated systems. The aim of this line of work is to develop robust algorithms agains those conditions and to develop benchmarks to evaluate them.

In [2], we developed a novel method to address the challenging task of LiDAR-based 3D object detection in foggy weather. In [1], we address 3D object detection in snowfall. Collecting and annotating data in such scenarios is very time, labor and cost intensive. We tackle this problem by simulating physically accurate fog and snowfall into clear-weather scenes, so that the abundant existing real datasets captured in clear weather can be repurposed for our task. We are the first to provide strong 3D object detection baselines on the Seeing Through Fog dataset. The code is made publicly available and the 3D detection results are shown in Figure 28.43.

On top of the developed robust methods, we further proposed a new dataset ACDC [3] to evaluate the performance of semantic segmentation methods in adverse weather and lighting conditions. ACDC consists of a large set of 4006 images which are equally distributed between four common adverse conditions: fog, nighttime, rain, and snow. Each adverse-condition image comes with a high-quality fine pixel-level semantic annotation, a corresponding image of the same scene taken under normal conditions, and a binary mask that distinguishes between intra-image regions of clear and uncertain semantic content. Thus, ACDC supports both standard semantic segmentation and the newly introduced uncertainty-aware semantic segmentation. A detailed empirical study demonstrates the challenges that the adverse domains of ACDC pose to state-of-the-art supervised and unsupervised approaches and indicates the value of our dataset in steering future progress in the field. Our dataset and benchmark are publicly available.

#### References


- [1] M. Hahner, C. Sakaridis, M. Bijelic, F. Heide, F. Yu, D. Dai, and L. Van Gool. LiDAR snowfall simulation for robust 3D object detection. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2022)*, New Orleans, LA, USA, 2022, pp. 16343–16353. IEEE.
- [2] M. Hahner, C. Sakaridis, D. Dai, and L. Van Gool. Fog simulation on real LiDAR point clouds for 3D object detection in adverse weather. In *ICCV 2021, IEEE/CVF International Conference on Computer Vision*, Virtual Event, 2021, pp. 15263–15272. IEEE.
- [3] C. Sakaridis, D. Dai, and L. Van Gool. ACDC: The adverse conditions dataset with correspondences for semantic driving scene understanding. In *ICCV 2021, IEEE/CVF International Conference on Computer Vision*, Virtual Event, 2021, pp. 10745–10755. IEEE.

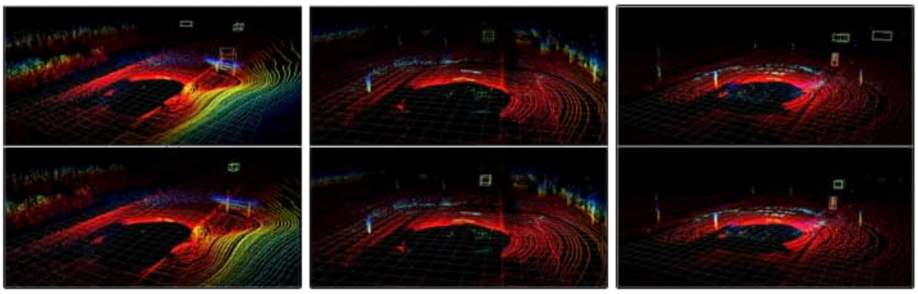


Figure 28.43: Our 3D object detection results: the (top) row shows predictions by PV-RCNN trained on the original clear weather data (first row in tables above), the (bottom) row shows predictions by PV-RCNN trained on a mix of clear weather and simulated foggy data (fourth row in tables above) on three example scenes from the STF dense fog test split. Ground truth boxes in color, predictions of the model in white. Best viewed on a screen (and zoomed in)..

### 28.5.2 Domain Adaptation

*Investigators: Dengxin Dai and Bernt Schiele, in cooperation with Luc Van Gool (ETH Zurich and KU Lueven), Rui Gong (ETH Zurich), Lukas Hoyer (ETH Zurich), Wen Li (ETH Zurich), Yuhua Chen (ETH Zurich), Martin Danelljan (ETH Zurich), Danda Pani Paudel (ETH Zurich), Ajad Chhatkuli (ETH Zurich), Fisher Yu (ETH Zurich), Shanshan Zhang (Nanjing University of Science and Technology), and Shenjian Gong (Nanjing University of Science and Technology)*

Humans are able to seamlessly adapt to changes in their environments. Current robotic perception systems, on the other hand, are typically trained in a rather fixed environment, allowing them to succeed in specific settings, but leading to failure in others. Therefore, one of the key challenge now is to build adaptive perception systems that can work well even when encountering domain changes. To this aim, we have made a few contributions for domain adaptation.

Given that the neural network architectures used in domain adaptation are rather outdated, in [7], we develop methods to improve both of them. We propose a novel method, DAFormer, which consists of a Transformer encoder and a multilevel context-aware feature fusion decoder. It is further enhanced by three simple but crucial training strategies. Since domain adaptation methods are usually GPU memory intensive, most previous methods operate only on downscaled images. Therefore, we further propose HRDA [8], a multi-resolution training approach for UDA, that combines the strengths of small high-resolution crops to preserve fine segmentation details and large low-resolution crops to capture long-range context dependencies, while maintaining a manageable GPU memory footprint. These two works significantly improve the state of the art.

To overcome large domain gap, we present a domain flow generation (DLOW) model [4] to bridge two different domains by generating a continuous sequence of intermediate domains. We have also made a few contributions to lift the constraints of common domain adaptation settings. For instance, we addressed how to lift the assumption that no domain change exists in the output space in [3]. This is important as in practice, different datasets are often labeled according to different semantic taxonomies. We further developed a novel method to combine and reuse existing datasets that may belong to different domains, have partial annotations, and/or have different data modalities [2]. In order to handle compound domain gaps, we have proposed a novel method disentangle multiple causal factors and then close the sub-domain gaps individually [9]. These techniques have significantly the application domain of existing methods.

The aforementioned methods are mainly designed for semantic image segmentation. We have also developed new domain adaptation methods for object detection [10] and object counting [5, 6]. In [10], we have introduced a simple but effective method to aggregate features at instance-level based on visual similarity before inducing group alignment via adversarial training. In [1], we have proposed a very simple yet effective method that introduces perturbation to batch normalization layers during training to increase model robustness for object detection. For object counting, a new adversarial learning based method is proposed to handle domain gaps for crowd counting [5] and a new counter robust to high intraclass diversity for class-agnostic object counting [6]. These four methods all outperform their competing methods.

## References


- [1] Q. Fan, M. Segu, Y.-W. Tai, F. Yu, C.-K. Tang, B. Schiele, and D. Dai. Normalization perturbation: A simple domain generalization method for real-world domain shifts. In *International Conference on Learning Representations*, 2023.
- [2] R. Gong, D. Dai, Y. Chen, W. Li, and L. Van Gool. mDALU: Multi-source domain adaptation and label unification with partial datasets. In *ICCV 2021, IEEE/CVF International Conference on Computer Vision*, Virtual Event, 2021, pp. 8856–8865. IEEE.
- [3] R. Gong, M. Danelljan, D. Dai, D. P. Paudel, A. Chhatkuli, F. Yu, and L. Van Gool. TACS: Taxonomy adaptive cross-domain semantic segmentation. In S. Avidan, G. Brostow, M. Cissé, G. M. Farinella, and T. Hassner, eds., *Computer Vision – ECCV 2022*, Tel Aviv, Israel, 2022, LNCS 13694, pp. 19–35. Springer.
- [4] R. Gong, W. Li, Y. Chen, D. Dai, and L. Van Gool. DLOW: Domain flow and applications. *International Journal of Computer Vision*, 129:2865–2888, 2021.
- [5] S. Gong, S. Zhang, J. Yang, D. Dai, and B. Schiele. Bi-level alignment for cross-domain crowd counting. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2022)*, New Orleans, LA, USA, 2022, pp. 7532–7540. IEEE.
- [6] S. Gong, S. Zhang, J. Yang, D. Dai, and B. Schiele. Class-agnostic object counting robust to intraclass diversity. In S. Avidan, G. Brostow, M. Cissé, G. M. Farinella, and T. Hassner, eds., *Computer Vision – ECCV 2022*, Tel Aviv, Israel, 2022, LNCS 13693, pp. 388–403. Springer.
- [7] L. Hoyer, D. Dai, and L. Van Gool. DAFormer: Improving network architectures and training strategies for domain-adaptive semantic segmentation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2022)*, New Orleans, LA, USA, 2022, pp. 9914–9925. IEEE.
- [8] L. Hoyer, D. Dai, and L. Van Gool. HRDA: Context-aware high-resolution domain-adaptive semantic segmentation. In S. Avidan, G. Brostow, M. Cissé, G. M. Farinella, and T. Hassner, eds., *Computer Vision – ECCV 2022*, Tel Aviv, Israel, 2022, LNCS 13690, pp. 372–391. Springer.

- [9] X. Ma, Z. Wang, Y. Zhan, Y. Zheng, Z. Wang, D. Dai, and C.-W. Lin. Both style and fog matter: Cumulative domain adaptation for semantic foggy scene understanding. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2022)*, New Orleans, LA, USA, 2022, pp. 18900–18909. IEEE.
- [10] F. Rezaeianaran, R. Shetty, R. Aljundi, D. O. Reino, S. Zhang, and B. Schiele. Seeking similarities over differences: Similarity-based domain alignment for adaptive object detection. In *ICCV 2021, IEEE/CVF International Conference on Computer Vision*, Virtual Event, 2021, pp. 9184–9193. IEEE.


### 28.5.3 Test-time Domain Adaptation

*Investigators: Zhi Li, Bernt Schiele, Shaoshuai Shi, Dengxin Dai, in cooperation with Luc Van Gool (ETH Zurich and KU Lueven), Qi Wang (ETH Zurich), and Olga Fink (ETH Zurich and EPFL)*

We aim at vision tasks in a new, highly practical setting: test-time domain adaptation. This problem setting enables us to adapt the source-trained models to the target domain by further updating the models on the unlabelled test set during inference, especially without the access to the source data (for data protection reasons). We aim at several popular and yet challenging vision tasks such as classification, semantic segmentation and depth estimation.

Test-time domain adaptation aims to adapt a source pretrained model to a target domain without using any source data. Existing works mainly consider the case where the target domain is static. However, real-world machine perception systems are running in non-stationary and continually changing environments where the target domain distribution can change over time. Existing methods, which are mostly based on self-training and entropy regularization, can suffer from these non-stationary environments. Due to the distribution shift over time in the target domain, pseudo-labels become unreliable. The noisy pseudolabels can further lead to error accumulation and catastrophic forgetting. To tackle these issues, we propose a continual test-time adaptation approach (CoTTA) in develop a test-time domain adaptation method for image classification and semantic segmentation tasks in [2] which comprises two parts. Firstly, we propose to reduce the error accumulation by using weight-averaged and augmentationaveraged predictions which are often more accurate. On the other hand, to avoid catastrophic forgetting, we propose to stochastically restore a small part of the neurons to the source pre-trained weights during each iteration to help preserve source knowledge in the long-term. The proposed method enables the long-term adaptation for all parameters in the network. CoTTA is easy to implement and can be readily incorporated in off-the-shelf pre-trained models. We demonstrate the effectiveness of our approach on four classification tasks and a segmentation task for continual testtime adaptation, on which we outperform existing methods.

In [1], we further develop a test-time domain adaptation method for a more challenging dense prediction task: depth estimation, shown in Fig. 28.44, which achieves both stability and adaptation performance by benefiting from both self-training of the supervised branch and pseudo labels from self-supervised branch, and is able to tackle the above problems: our scale alignment scheme aligns the input features between source and target data, correcting the absolute scale inference on the target domain; with pseudo label consistency check, we select confident pixels thus improve pseudo label quality; regularisation and self-training

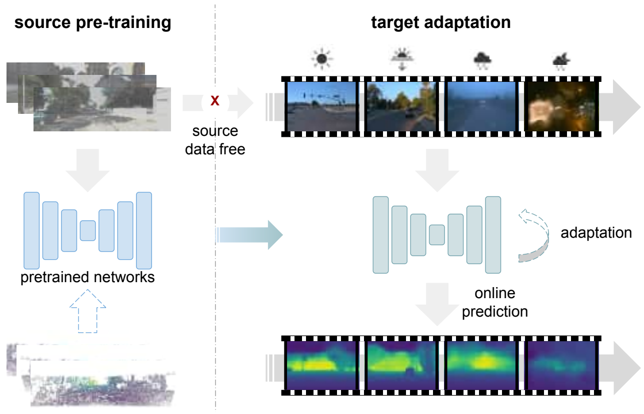



Figure 28.44: Overview of our test-time domain adaptation framework. We adapt our source-trained network to the changing target data during test time in an online fashion, without requiring the access of the source data anymore.

schemes are applied to help avoid catastrophic forgetting. Without requirement of further supervisions on the target domain, our method adapts the source-trained models to the test data with significant improvements over the direct inference results, providing scale-aware depth map outputs that outperform the state-of-the-arts.

## References


- [1] Z. Li, S. Shi, B. Schiele, and D. Dai. Test-time domain adaptation for monocular depth estimation. In *IEEE International Conference on Robotics and Automation (ICRA 2023)*, 2023. IEEE. Accepted.
- [2] Q. Wang, O. Fink, L. Van Gool, and D. Dai. Continual test-time domain adaptation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2022)*, New Orleans, LA, USA, 2022, pp. 7191–7201. IEEE.


### 28.5.4 Multi-Task Learning

*Investigators: Dengxin Dai, in cooperation with Luc Van Gool (ETH Zurich and KU Lueven), Arun Balajee Vasudevan (ETH Zurich), Jiri Matas ( Czech Technical University), Ke Li (ETH Zurich), Qin Wang (ETH Zurich), Lukas Hoyer (ETH Zurich), Guolei Sun (ETH Zurich), and Olga Fink (ETH Zurich)*

Despite the recent progress in deep learning, most approaches still go for a silo-like solution, focusing on learning each task in isolation: training a separate neural network for each individual task. Many real-world problems such as autonomous driving, however, call for a multi-tasking model aiming for lower memory footprint, higher overall inference speed.

In [1], we have developed an approach for multiple scene understanding tasks purely based on binaural sounds. The considered tasks include predicting the semantic masks of sound-making objects, the motion of sound-making objects, and the depth map of the scene, as illustrated in Figure 28.45. We formulate all the prediction tasks into one end-to-end


Figure 28.45: An illustration of our four tasks: (a) super-resolved binaural sounds – from azimuth angle 0 degree (red) to other angles such as 90 degree (blue), (b) auditory semantic prediction for three sound-making object classes, (c) auditory depth prediction and (d) auditory motion prediction.

trainable multi-tasking network aiming to boost the overall performance. Our method is the first of its kind. In [4], we have found that multitask pre-training is effective to pre-train a universal model such that it can be used for many different types of down-stream tasks. This way, we do not need to pre-train specific models for specific downstream tasks (e.g. classification vs. segmentation).

We have also developed Task Switching Networks [3], a task-conditioned architecture with a single unified encoder/decoder for efficient multi-task learning. Multiple tasks are performed by switching between them, performing one task at a time. TSNs have a constant number of parameters irrespective of the number of tasks. This scalable yet conceptually simple approach circumvents the overhead and intricacy of task-specific network components in existing works.

In addition to the setting of fully-supervised multitask learning, we have also developed novel algorithms for learning with auxiliary task, where auxiliary tasks are used to regularise the training of the primary task. In [5], we have developed a novel approach to successfully use self-supervised depth estimation to improve the adaptability of semantic segmentation to new domains. In [2], the RGB image super-resolution task is used as an auxiliary task to improve the performance of hyper-spectral image super-resolution. This way, the method is able to benefit from the large-scale training data of RGB image super-resolution. The two methods achieve the state-of-the-art results.

## References


- [1] D. Dai, A. B. Vasudevan, J. Matas, and L. Van Gool. Binaural SoundNet: Predicting semantics, depth and motion with binaural sounds. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(1):123–136, 2023.
- [2] K. Li, D. Dai, and L. van Gool. Hyperspectral image super-resolution with RGB image super-resolution as an auxiliary task. In *2022 IEEE Winter Conference on Applications of Computer Vision (WACV 2022)*, Waikoloa Village, HI, USA, 2022, pp. 4039–4048. IEEE.
- [3] G. Sun, T. Probst, D. P. Paudel, N. Popovic, M. Kanakis, J. Patel, D. Dai, and L. Van Gool. Task switching network for multi-task learning. In *ICCV 2021, IEEE/CVF International Conference on Computer Vision*, Virtual Event, 2021, pp. 8271–8280. IEEE.
- [4] A. B. Vasudevan, D. Dai, and L. Van Gool. Sound and visual representation learning with multiple

pretraining tasks. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2022)*, New Orleans, LA, USA, 2022, pp. 14596–14606. IEEE.

• [5] Q. Wang, D. Dai, L. Hoyer, L. Van Gool, and O. Fink. Domain adaptive semantic segmentation with self-supervised depth estimation. In *ICCV 2021, IEEE/CVF International Conference on Computer Vision*, Virtual Event, 2021, pp. 8495–8505. IEEE.

### 28.5.5 Scene Understanding

*Investigators: Dengxin Dai, Apratim Bhattacharyya, Mario Fritz, and Bernt Schiele, in cooperation with Luc Van Gool (ETH Zurich and KU Lueven), Ozan Unal (ETH Zurich), Vaishakh Patil (ETH Zurich), Alex Liniger (ETH Zurich), Niclas Vödisch (ETH Zurich & University of Freiburg), Ke Li (ETH Zurich), Shijie Li (University of Bonn), Cyrill Stachniss (University of Bonn), Juergen Gall (University of Bonn), and Daniel Olmeda Reino (Toyota Motor Europe)*

Semantic scene understanding is a very fundamental task in Computer Vision and it has many applications such as autonomous driving. We have made contributions on LiDAR-based semantic segmentation [4, 2], LiDAR-based 3D object detection [5], depth estimation [3] and pedestrian vehicle interactions understanding [1].

While great progress has been made for LiDAR semantic segmentation, densely annotating LiDAR point clouds remains too expensive and time-consuming to keep up with the ever growing volume of data. While current literature focuses on fully-supervised performance, developing efficient methods that take advantage of realistic weak supervision have yet to be explored. In [4], we have proposed using scribbles to annotate LiDAR point clouds and release ScribbleKITTI, the first scribble-annotated dataset for LiDAR semantic segmentation. Furthermore, we have presented a pipeline to reduce the performance gap that arises when using such weak annotations to up to 95.7% of the fully-supervised performance while using only 8% labeled points. A comparison of our scribble annotation with the full annotation can be found in Figure 28.46.

Real-time semantic segmentation of LiDAR data is crucial for autonomously driving vehicles and robots. Current projection-based methods either have a low accuracy or require millions of parameters. We therefore have proposed an efficient method which uses multiple paths with different scales and balances the computational resources between the scales [2]. The proposed network outperforms point-based, image-based, and projection-based methods in terms of accuracy, number of parameters, and runtime.

Besides annotation cost and running speed, we have developed method to reduce the cost of the sensor [5]. Existing learning methods for LiDAR-based applications use 3D points scanned under a pre-determined beam configuration. We have taken a new route to learn to optimize the LiDAR beam configuration for different applications like 3D object detection and localization. The method significantly reduces the cost of the LiDAR sensor while retaining the performance of the application tasks.

Accurate scene depth is fundamental for robot scene understanding as it adds spatial reasoning. However, accurate scene depth often comes at the cost of expensive additional depth sensors. In [3], we have proposed to use map-based depth data as an additional input instead of expensive depth sensors. Such an approach is especially appealing in autonomous

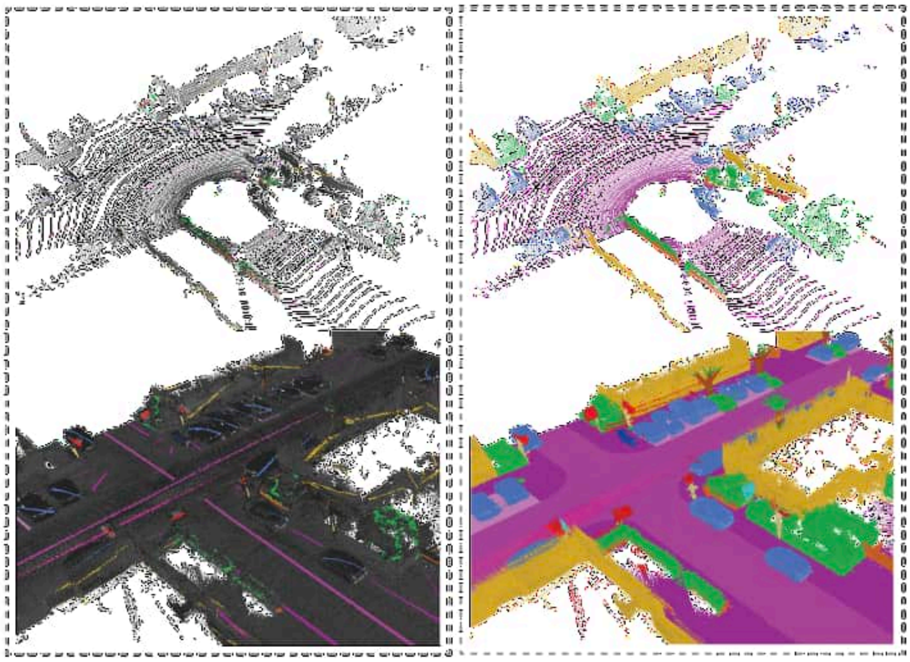


Figure 28.46: Example of scribble-annotated LiDAR point cloud scenes of a single frame (top) and superimposed frames (bottom). Compared are the proposed ScribbleKITTI (left) with the fully labeled counterpart from SemanticKITTI.

driving since map-based depth is commonly available from high-definition maps. Our method significantly outperforms other methods both in quantitative and qualitative results.

Accurate prediction of pedestrian and bicyclist paths is integral to the development of reliable autonomous vehicles in dense urban environments. We have proposed Euro-PVI [1], a dataset of pedestrian and bicyclist trajectories featuring more diverse and complex interactions. In this work, we further developed a joint inference model that learns an expressive multi-modal shared latent space across agents for accurate predictions.

## References


- [1] A. Bhattacharyya, D. O. Reino, M. Fritz, and B. Schiele. Euro-PVI: Pedestrian vehicle interactions in dense urban centers. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2021)*, Nashville, TN, USA (Virtual), 2021, pp. 6408–6417. IEEE.
- [2] S. Li, X. Chen, Y. Liu, D. Dai, C. Stachniss, and J. Gall. Multi-scale interaction for real-time LiDAR data segmentation on an embedded platform. *IEEE Robotics and Automation Letters*, 7(2):738–745, 2022.
- [3] V. Patil, A. Liniger, D. Dai, and L. Van Gool. Improving depth estimation using map-based depth priors. *IEEE Robotics and Automation Letters*, 7(2):3640–3647, 2022.
- [4] O. Unal, D. Dai, and L. Van Gool. Scribble-supervised LiDAR semantic segmentation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2022)*, New Orleans, LA, USA, 2022, pp. 2687–2697. IEEE.

- [5] N. Vödisch, O. Unal, K. Li, L. Van Gool, and D. Dai. End-to-end optimization of LiDAR beam configuration for 3D object detection and localization. *IEEE Robotics and Automation Letters*, 7(2):2242–2249, 2022.


### 28.5.6 Object Tracking

*Investigators: Dengxin Dai, in cooperation with Luc Van Gool (ETH Zurich and KU Lueven), Jan-Nico Zaech (ETH Zurich), Alex Liniger (ETH Zurich), and Martin Danelljan (ETH Zurich)*

Autonomous systems that operate in dynamic environments require robust object tracking in 3D as one of their key components. Most recent approaches for 3D multi-object tracking (MOT) from LIDAR use object dynamics together with a set of handcrafted features to match detections of objects across multiple frames. However, manually designing such features and heuristics is cumbersome and often leads to suboptimal performance. In [1], we have presented a unified and learning based approach to the 3D MOT problem. We have designed a graph structure to jointly process detection and track states in an online manner. Our approach provides a natural way for track initialization and handling of false positive detections, while significantly improving track stability. We demonstrated the merit of the proposed approach in the nuScenes tracking challenge 2021 with a state-of-the-art performance, resulting in the best LIDAR only submission.

MOT is most often approached in the tracking-by-detection paradigm, where object detections are associated through time. The association step naturally leads to discrete optimization problems. As these optimization problems are often NP-hard, they can only be solved exactly for small instances on current hardware. Adiabatic quantum computing (AQC) offers a solution for this, as it has the potential to provide a considerable speedup on a range of NP-hard optimization problems in the near future. However, current MOT formulations are unsuitable for quantum computing due to their scaling properties. In [2], we have proposed the first MOT formulation designed to be solved with AQC. We have employed an Ising model that represents the quantum mechanical system implemented on the AQC. We showed that our approach is competitive compared with state-of-the-art optimizationbased approaches, even when using of-the-shelf integer programming solvers. In this work, we have demonstrated that our MOT problem is already solvable on the current generation of real quantum computers for small examples, and analyze the properties of the measured solutions. Please see Figure 28.47 for an illustration of the method.

#### References


- [1] J.-N. Zaech, D. Dai, A. Liniger, M. Danelljan, and L. Van Gool. Learnable online graph representations for 3D multi-object tracking. *IEEE Robotics and Automation Letters*, 7(2):5103–5110, 2022.
- [2] J.-N. Zaech, A. Liniger, M. Danelljan, D. Dai, and L. Van Gool. Adiabatic quantum computing for multi object tracking. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2022)*, New Orleans, LA, USA, 2022, pp. 8801–8812. IEEE.

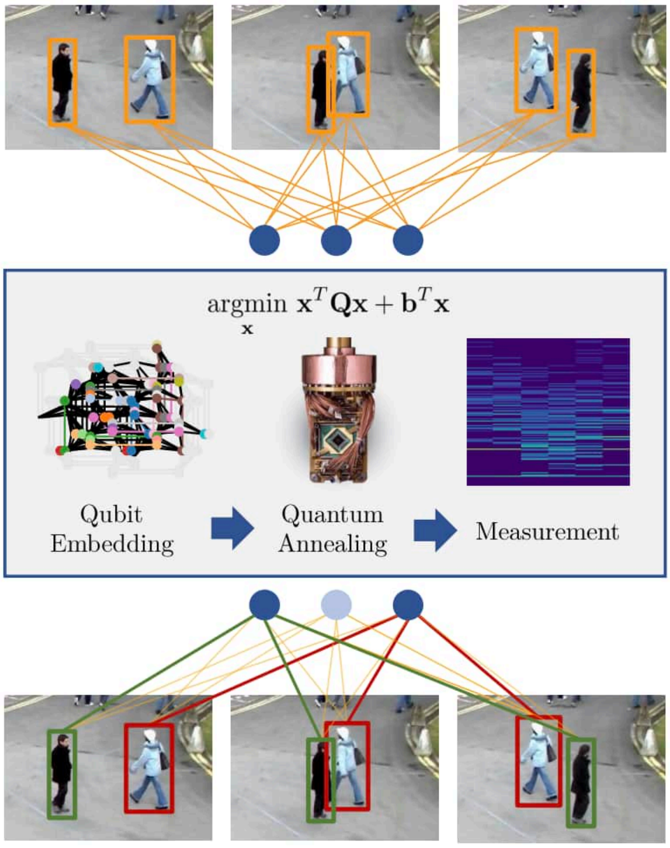



Figure 28.47: The proposed approach to MOT states the assignment problem between detections and a set of tracks as a quadratic unconstrained binary optimization task. We then represent the optimization problem as a quantum mechanical system that can be implemented on an AQC. Via quantum annealing, a minimum energy state is found that represent the best assignment..

## 28.6 Real Virtual Humans

*Coordinator: Gerard Pons-Moll*

### 28.6.1 People in Clothing

*Investigators: Gerard Pons-Moll (University of Tubingen), Garvita Tiwari (University of Tubingen), Marc Habermann, Christian Theobalt in cooperation with Nikolaos Sarafianos, Michael Zollhoefer, Weipeng Xu, Tony Tung (Reality Labs Research), Enric Corona, Albert Pumarola, Guillem Alenyà, Francesc Moreno-Noguer (Institut de Robotica i Informatica Industrial, CSIC-UPC)*

Modeling humans with realism and personalization requires modeling physically plausible clothing. This is crucial for AR/VR, content generation in 3D. Key challenges in modeling "people in clothing" are 1) lack of 3D data with corresponding SMPL parameters, 2) limited

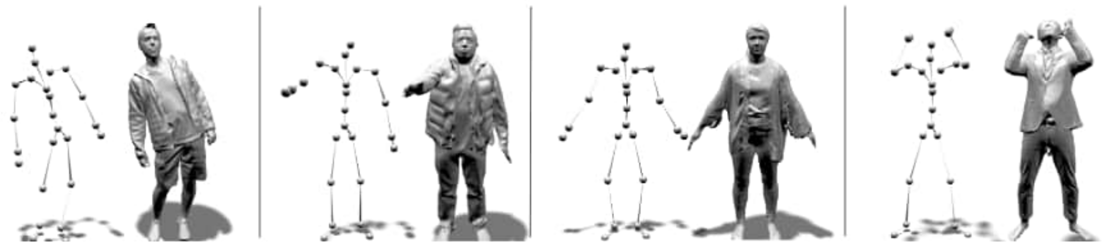


Figure 28.48: In [8], we present Neural Generalized Implicit Functions (Neural-GIF), to animate people in clothing as a function of body pose. Neural-GIF learns directly from scans, models complex clothing and produces pose-dependent details for realistic animation.

representation power of mesh based neural network, 3) difficulty to scale such mesh based representation methods, and 3) expensive and tedious 3D/4D capture setup. Here we discuss broad frontiers of research that we have been pushing forward in order to address the challenges mentioned above.

### "People in Clothing" using Neural Implicits

3D content generation for research, AR/VR, and animation industry requires controllable model of 3D humans, such as control over pose or clothing or shape. Prior work in this directions such as [5, 4, 7] either use synthetic data to learn the model or rely on parametric model based surface registration of real-world data to animate people. In later, a fixed mesh is used to represent people and clothing, which results in low resolution, limited topology and an expensive data processing step. With recent advances in neural implicit based surface representation such as [1], we propose a novel method, called Neural Generalized Implicit Functions (Neural-GIF) [8] to animate people in clothing using controllable neural implicits, which overcomes aforementioned limitations. We draw inspiration from template-based methods, which factorize motion into articulation and non-rigid deformation, but generalize this concept for implicit shape learning to obtain a more flexible model. Our formulation allows the learning of complex and non-rigid deformations of clothing and soft tissue, without computing a template registration as it is common with prior approaches, as shown in figure 28.48. Neural-GIF models person and clothing as a single layer of surface and can be used for separate clothing items/dressed body/SMPL shapes. Experimental evaluation shows that Neural-GIF show significant improvements than prior work in terms on fine details and pose generalization.

We also introduce SMPLicit [6], a neural implicit based model which incorporates different clothing layers along with pose-shape dependent deformation. Unlike prior template based method, SMPLicit represent all garment types using a unified model, while having controllable properties like garment style, size and shape. The representation flexibility of SMPLicit builds upon an implicit model conditioned with the SMPL human body parameters and a learnable latent space which is semantically interpretable and aligned with the clothing attributes. The proposed model is fully differentiable, allowing for its use into larger end-to-end trainable systems, as shown in figure 28.49.We demonstrate SMPLicit can be readily used for fitting

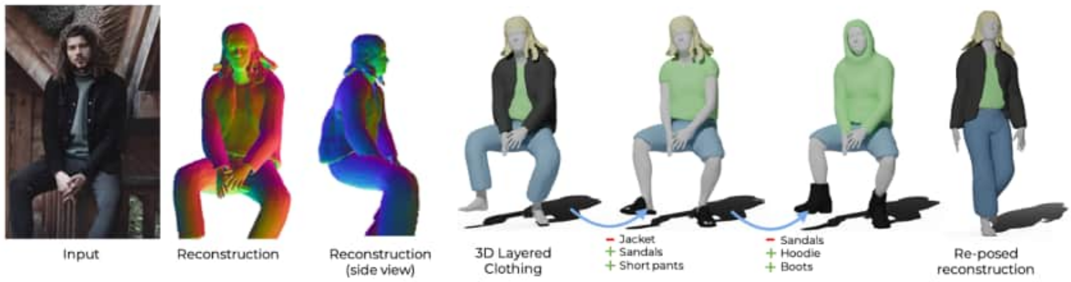



Figure 28.49: We introduce SMPLicit, a fully differentiable generative model for clothed bodies, capable of representing garments with different topology. Given an input image, we use SMPLicit to predict different models per cloth, even for multi-layer cases. The model can also be used for editing the outfits, removing/adding new garments and re-posing the body (right).

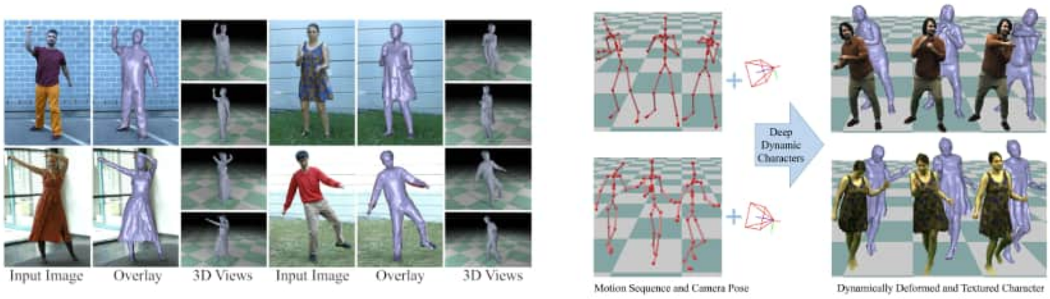



Figure 28.50: **Left**: In [3], we present the first learning-based approach for dense monocular human performance capture using weak multi-view supervision. Proposed model can recover varying types of apparel ranging from shirt-pants to loose clothing such as skirts. **Right**: In [2], motion- and view-dependent dynamic surface deformation and texture of a person-specific template is regressed to generate video realistic results.

3D scans and for 3D reconstruction in images of dressed people. In both cases we are able to go beyond state of the art, by retrieving complex garment geometries, handling situations with multiple clothing layers and providing a tool for easy outfit editing.

### Human performance capture

Capturing 3D data of people in clothing requires expensive camera setup, which are not usually portable and not easy to use for novice. This limits the scalability and diversity in data capture. On the other hand monocular human performance capture is a scalable way of capturing human data but is a highly unconstrained problem. In [3], we present the first learning-based approach for dense monocular human performance capture using weak multi-view supervision that not only predicts the pose but also the space-time coherent non-rigid deformations of the model surface. The key idea here is to disentangle pose estimation and non-rigid surface deformation. The proposed approach preserves the skeletal structure of the human body, produces a more accurate 3D reconstruction of pose and non-rigid deformation than existing methods, and handles loose clothes. We perform extensive qualitative and quantitative experiments in different scenarios like indoor, in-the-wild and lab setting with diverse set of subjects and varying types of apparel, as shown in figure refig:d2-gerard-human-perf (left). Our qualitative and quantitative results in different scenarios show that our method produces more accurate 3D reconstruction of pose and non-rigid deformation than existing methods.

### Real-time dynamic human model with realistic texture and deformation

Prior works have mainly focussed on improving geometry for human modeling [4, 5, 3]. In contrast to this, we also aim to recover realistic dynamic texture in [2]. We learn a model displaying highly realistic shape, motion, and dynamic appearance learned in a new weakly supervised way from multi-view images. The dynamic texture model also accounts for photo-realistic motion-dependent appearance details, as well as view-dependent lighting effects, which makes the final results video realistic. We use a novel graph convolutional network architecture to enable motion-dependent deformation learning of body and clothing, including dynamics, and a neural generative dynamic texture model creates corresponding dynamic texture maps. In figure 28.50 (right), we show that given skeletal motions as input, our model creates motion-dependent surface deformations, physically plausible dynamic clothing deformations, as well as video-realistic surface textures at a much higher level of detail than previous state of the art approaches, and even in real-time.

## References


[1] J. Chibane, T. Alldieck, and G. Pons-Moll. Implicit functions in feature space for 3D shape reconstruction and completion. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2020)*, Seattle, WA, USA (Virtual), 2020, pp. 6968–6979. IEEE.

• [2] M. Habermann, L. Liu, W. Xu, M. Zollhöfer, G. Pons-Moll, and C. Theobalt. Real-time deep dynamic characters. *ACM Transactions on Graphics (Proc. ACM SIGGRAPH)*, 40(4), Article 94, 2021.

• [3] M. Habermann, W. Xu, M. Zollhöfer, G. Pons-Moll, and C. Theobalt. A deeper look into DeepCap. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(4):4009–4002, 2023.

[4] Q. Ma, J. Yang, A. Ranjan, S. Pujades, G. Pons-Moll, S. Tang, and M. J. Black. Learning to dress 3D people in generative clothing. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2020)*, Seattle, WA, USA (Virtual), 2020, pp. 6468–6477. IEEE.

[5] C. Patel, Z. Liao, and G. Pons-Moll. Tailornet: Predicting clothing in 3D as a function of human pose, shape and garment style. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2020)*, Seattle, WA, USA (Virtual), 2020, pp. 7363–7373. IEEE.

• [6] G. Pons-Moll, F. Moreno-Noguer, E. Corona, A. Pumarola, and G. Alenyà. SMPLicit: Topology-aware generative model for clothed people. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2021)*, Virtual Conference, 2021, pp. 11870–11880. IEEE.

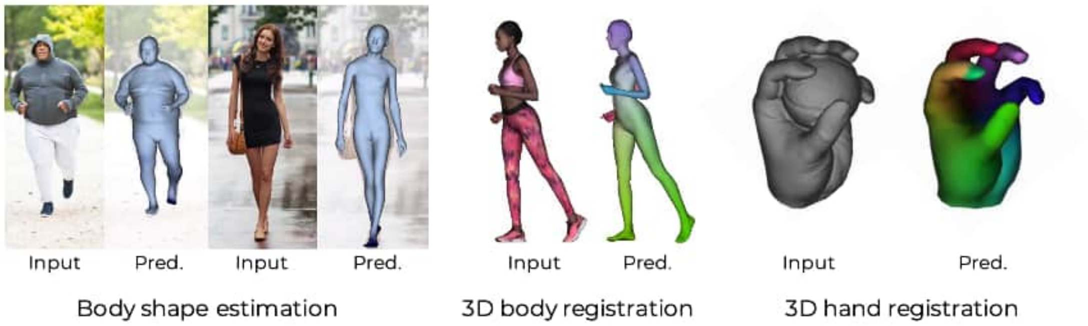



Figure 28.51: LVD [1] is a novel optimization strategy in which a network leverages local image or volumetric features to iteratively predict per-vertex directions towards an optimal body/hand surface. The proposed approach is directly applicable to different tasks with minimal changes on the network, and we show it can fit a much larger variability of body shapes than previous state-of-the-art. The figure depicts results on the three tasks where we have evaluated LVD: body shape reconstruction from a single image, and 3D fitting of body and hand scans

[7] G. Tiwari, B. L. Bhatnagar, T. Tung, and G. Pons-Moll. Sizer: A dataset and model for parsing 3D clothing and learning size sensitive 3D clothing. In A. Vedaldi, H. Bischof, T. Brox, and J.-M. Frahm, eds., *Computer Vision – ECCV 2020*, Glasgow, UK, 2020, LNCS 12348, pp. 1–18. Springer.

• [8] G. Tiwari, N. Sarafianos, T. Tung, and G. Pons-Moll. Neural-GIF: Neural generalized implicit functions for animating people in clothing. In *ICCV 2021, IEEE/CVF International Conference on Computer Vision*, Virtual Event, 2021, pp. 11688–11698. IEEE.

## 28.6.2 Pose and Shape

*Investigators: Gerard Pons-Moll (University of Tubingen), Garvita Tiwari (University of Tubingen), Zhouyingcheng Liao,, Jan Eric Lenssen in cooperation with Nikolaos Sarafianos, Tony Tung (Reality Labs Research), Jimei Yang, Jun Saito, Yang Zhou (Adobe Research), Dimitrije Antic (University of Tubingen), Enric Corona, Guillem Alenyà, Francesc Moreno-Noguer (Institut de Robotica i Informatica Industrial, CSIC-UPC)*

Pose and shape together are used to explain articulated/deformable shapes such as humans or animals or characters. This disentanglingemtn of shape into two components has motivated paramteric models like SMPL and based on it, most of the surface estimation problems is decoupled into pose and shape estimation. We propose two completely novel formulation along 3D shape fitting and 3D pose manifold using neural fields representation. At the end we also discuss a more application oriented research in the direction of pose transfer for content generation and animation.

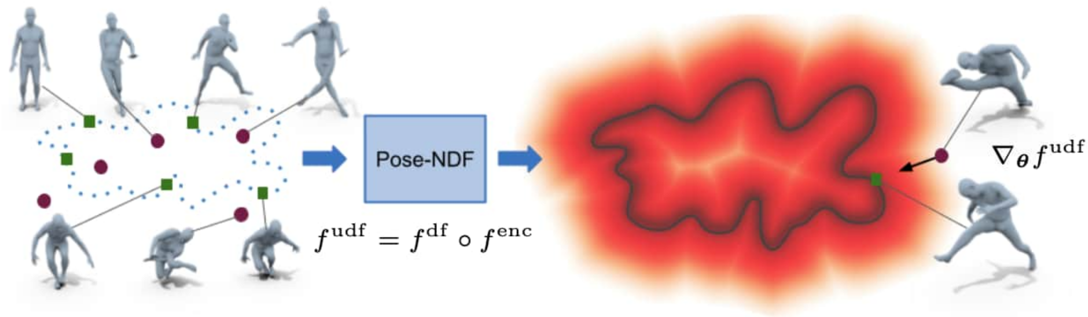



Figure 28.52: We present Pose-NDF, a neural unsigned distance field in $SO(3)^K$, which learns the manifold of plausible poses as zero level set. We learn the distance field representation from samples of plausible and unrealistic poses (**left**). We encode the input pose (given as a set of quaternions) using a structural MLP $f^{enc}$ and predict the distance from the joint representation using an MLP $f^{df}$. The gradient $_{\theta} f^{udf}$ and distance value $f^{udf}(\theta)$ are used to project implausible poses onto the manifold (**right**)

### Neural Distance Fields in Model Fitting

Fitting human pose and shape to 3D data/images/videos is an ambiguous problem. Prior work use parameters of a low-dimensional statistical body model (e.g. SMPL) in optimization or neural network based paradigm to estimate pose and shape from input of different modalities. However optimization based methods have tendency to get stuck in local minima. Moreover these parameter prediction based formulation struggle in capturing detailed body shape, specially for morphotypes departing from the mean shape or when the person is wearing loose clothing. To tackle such problem, we propose a novel optimization-based paradigm for 3D human model fitting on images and scans. We train an ensemble of per vertex neural fields network. The network predicts, in a distributed manner, the vertex descent direction towards the ground truth, based on neural features extracted at the current vertex projection. During inference LVD is used in a gradient-descent optimization pipeline and it convergences within fraction of seconds. LVD based fitting significantly improves state-of-the-art and especially for extreme body shapes and loose clothing for SMPL estimation from images in-the-wild and 3D scans. LVD based optimization converges even for bad initialization, e.g. also when initializing all vertices into a single point. This formulation can be used easily on different parametric models, we show results for SMPL and MANO model in figure 28.51.

### Neural Distance Fields in Pose manifolds

Modeling manifold of realistic poses is crucial to constrain pose/shape/motion estimation tasks and generation of novel poses. Prior work in this direction has relied on Gaussian assumption on pose space (e.g. GMM) or on latent space (VPoser). We present Pose-NDF, a continuous model for plausible human poses based on neural distance fields (NDFs). Pose-NDF learns a manifold of plausible poses as the zero level set of a neural implicit

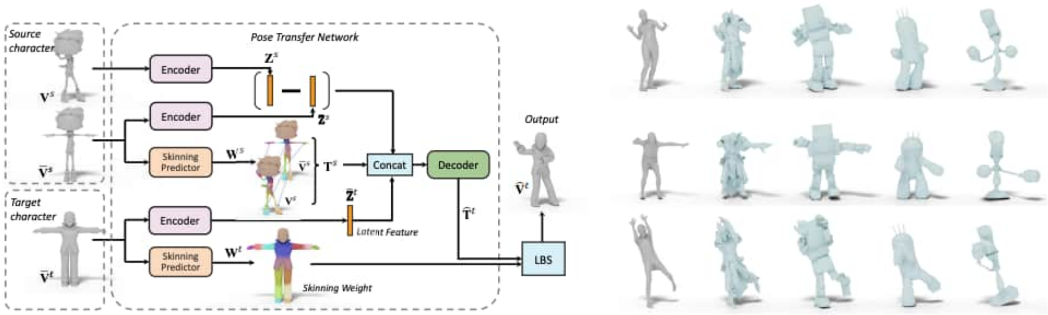



Figure 28.53: Given a posed source character and a target character as input, the pose transfer network estimates character skinning weights and part-wise transformations which articulate the target character through LBS to match the pose of the source (**Left**). Given source pose characters as input (grey), our model automatically transfers their poses to target subject characters with different body proportions and topologies (blue) (**Right**).

function, extending the idea of modeling implicit surfaces in 3D to the high-dimensional domain $SO(3)^K$, where a human pose is defined by a single data point, represented by K quaternions, as shown in figure 28.52. The resulting high-dimensional implicit function can be differentiated with respect to the input poses and thus can be used to project arbitrary poses onto the manifold by using gradient descent on the set of 3-dimensional hyperspheres. In contrast to previous VAE-based human pose priors, which transform the pose space into a Gaussian distribution, we model the actual pose manifold, preserving the distances between poses. We demonstrate that PoseNDF outperforms existing state-of-the-art methods as a prior in various downstream tasks, ranging from denoising real-world human mocap data, pose recovery from occluded data to 3D pose reconstruction from images.

We believe that extending neural fields formulation for novel tasks such as model fitting and pose manifold opens a new area of research and our works [1, 3], show promising success in these domain.

### Skeleton-free Pose Transfer

Posing and animating human/animals/animatronic-devices/cartoon characters is actively required in animation and digital industry. Instead of generating articulations from scratch, common and efficient practice is to transfer pose/motion from one subject to another. We present the first method, in [2] that automatically transfers poses between stylized 3D characters without skeletal rigging. In contrast to previous attempts to learn pose transformations on fixed or topology equivalent skeleton templates, our method focuses on a novel scenario to handle skeleton-free characters with diverse shapes, topologies, and mesh connectivities. The key idea of our method is to represent the characters in a unified articulation model so that the pose can be transferred through the correspondent parts. We first predict the character skinning weights and deformation transformations jointly to articulate the target character to match the desired pose, as shown in figure 28.53 (left). Our

Figure 28.54: We focus on modelling human-object interactions in 3D. To this end, our work BEHAVE [1] proposes the first dataset and method to capture the 3D human, object and the interactions between them in natural environments.

method is trained in a semi-supervised manner absorbing all existing character data with paired/unpaired poses and stylized shapes. As a result, we do not require neither annotations nor mesh correspondences for training and can make use of large amounts of static character It generalizes well to unseen stylized characters and inanimate objects. We conduct extensive experiments and demonstrate the effectiveness of our method on this novel task,as shown in figure 28.53 (right).

### References


- [1] E. Corona, G. Pons-Moll, G. Alenyà, and F. Moreno-Noguer. Learned vertex descent: A new direction for 3D human model fitting. In S. Avidan, G. Brostow, M. Cissé, G. Farinella, and T. Hassner, eds., *Computer Vision – ECCV 2022*, Tel Aviv, Israel, 2022, LNCS 13662, pp. 145–164. Springer.
- [2] Z. Liao, J. Yang, J. Saito, G. Pons-Moll, and Y. Zhou. Skeleton-free pose transfer for stylized 3D characters. In S. Avidan, G. Brostow, M. Cissé, G. Farinella, and T. Hassner, eds., *Computer Vision – ECCV 2022*, Tel Aviv, Israel, 2022, LNCS 13662, pp. 640–656. Springer.
- [3] G. Tiwari, D. Antic, J. E. Lenssen, N. Sarafianos, T. Tung, and G. Pons-Moll. Pose-NDF: Modeling human pose manifolds with neural distance fields. In S. Avidan, G. Brostow, M. Cissé, G. Farinella, and T. Hassner, eds., *Computer Vision – ECCV 2022*, Tel Aviv, Israel, 2022, LNCS 13665, pp. 572–589. Springer.


### 28.6.3 Human Object Interaction in 3D

*Investigators: Gerard Pons-Moll, Bharat Lal Bhatnagar, Xianghui Xie, Vladimir Guzov, Aymen Mir, Xioahan Zhang, Keyang Zhou, Julian Chibane, Jan Eric Lenssen, Christian Theobalt, in cooperation with Ilya Petrov (University of Tübingen), Riccardo Marin (University of Tübingen), Cristian Sminchisescu (Google Research), Torsten Sattler (Czech Technical University, Prague), Sebastian Starke (Meta)*

Humans interact with their surroundings in many different ways by making contacts at different ocations, creating a highly complex interaction space. One of the key challenges in building such a model is that there does not exist data that captures realistic human-object interactions in natural environments. And capturing the interactions between humans and their environment in 3D is important for many applications in robotics, graphics, and vision. To this end we propose several capture methods based on IMUs [2], Kinects [1] and hybrid sensors [6]. We then use this data to learn neural models for real world tasks like reconstructing humans and objects from images [4] and videos [5], tracking human-object interactions [1, 7], as well as synthesising controllable interactions [6, 3]. All the code and data collected for our projects is publicly available for further research.

### Capturing interactions between 3D humans and objects/scenes

Existing capture systems such as 4D scanners, marker based setups etc. are not suitable for recording human interactions as they have a limited recording volume and do not constitute a natural environment for interactions. We propose HPS [2], which uses IMUs attached at the body limbs and a head mounted camera looking outwards to record the human motion in a pre-recorded scene. HPS contains a dataset of humans interacting with large 3D scenes (300-1000 sq.m) consisting of 7 subjects and more than 3 hours of diverse motion.

Although IMU based HPS is quite powerful, it requires pre-scanning the 3D scene which is limiting. Moreover it does not capture the dynamics between the human and the object during interaction as the scene is static. Therefore we present the BEHAVE dataset [1], the first full body human object interaction dataset with multi-view RGBD frames and corresponding 3D SMPL and object fits along with the annotated contacts between them. We record $\sim$400k frames at 5 locations with 8 subjects performing a wide range of interactions with 20 common objects (see Fig. 28.54). We use this data to learn a model that can jointly track humans and objects in natural environments. Our key insight is to predict correspondences from the human and the object to a statistical body model to obtain human-object contacts during interactions. This allows us to record and track not just the humans and objects but also their interactions.

### Reconstructing human-object interaction from a monocular RGB camera

Most works in computer vision and learning have focused on perceiving 3D humans from single images in isolation. In reality, the problem is extremely challenging due to heavy occlusions between humans and objects, diverse interaction types and depth ambiguity. We introduce CHORE [4], a novel method that learns to jointly reconstruct humans and objects from a single image. We compute a neural reconstruction of human and object represented implicitly with two unsigned distance fields, and additionally predict a correspondence field to a parametric body as well as an object pose field. This allows us to robustly fit a parametric body model and a 3D object template, while reasoning about interactions.

Although quite powerful, CHORE does not explicitly reason about the occlusions between the human and the object during interaction and it does not take temporal information into account. This leads to significant performance drop in challenging scenarios. To address this, we propose a novel method to track the 3D human, object, contacts, and relative translation across frames from a single RGB camera, while being robust to heavy occlusions [5]. To this end we first fit a human model such as SMPL consistently to the video. We then predict the 3D object conditioned on the SMPL. This allows us to handle object occlusions and makes our human and object predictions consistent across video.

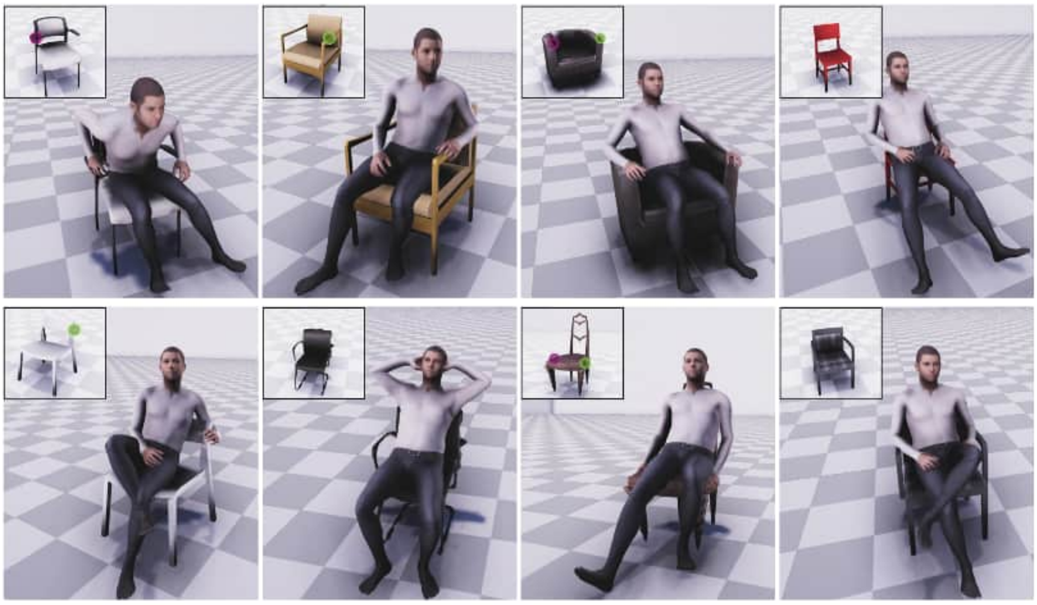


Figure 28.55: Our work COUCH [6] can synthesise controllable human-object interaction. Given an initial starting point for the human and the goal contacts on the chair, COUCH synthesizes realistic human motion that satisfies the user desired interaction.

### Synthesizing human-object interaction

Apart from capturing and reconstructing interactions, we study the problem of synthesizing interactions as well. We propose a novel synthesis framework COUCH [6] that can synthesise human motion conditioned on contact positions specified by the user, on the object (see Fig. 28.55). COUCH plans ahead the motion by predicting contact-aware control signals of the hands, which are then used to synthesise contact-conditioned interactions. Furthermore, we contribute a large human-chair interaction dataset with clean annotations, the COUCH Dataset. More importantly, our method enables control of the motion through user-specified or automatically predicted contacts.

We put special emphasis on modelling hand-object interactions as most of our interactions with our surroundings take place using our hands. However existing trackers for hands and objects are far from perfect and often do not take interaction into account. To this end we present TOCH [7], a method for refining incorrect 3D hand-object interaction sequences using a data prior. The core of our method are TOCH fields, a novel spatio-temporal representation for modelling correspondences between hands and objects during interaction, which allows us to outperform state-of-the-art (SOTA) 3D hand-object interaction models.

We further extend the idea of motion synthesis by synthesizing the 3D object that would correctly interact with a given human pose [3]. The relation between affordances and how we interact with an object is of large interest, for behavioural sciences, cognitive psychology, and computer vision. Existing work has looked into tasks like affordance modelling i.e. predict human pose conditioned on the object but the inverse task has not been explored. In this work, we infer 3D objects and their poses given 3D poses of the human which signals how the person would interact.

### References


- [1] B. L. Bhatnagar, X. Xie, I. Petrov, C. Sminchisescu, C. Theobalt, and G. Pons-Moll. BEHAVE: Dataset and method for tracking human object interactions. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2022)*, New Orleans, LA, USA, 2022, pp. 15914–15925. IEEE.
- [2] V. Guzov, A. Mir, T. Sattler, and G. Pons-Moll. Human POSEitioning system (HPS): 3D human pose estimation and self-localization in large scenes from body-mounted sensors. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2021)*, Nashville, TN, US (Virtual), 2021, pp. 4318–4329. IEEE.
- [3] I. A. Petrov, R. Marin, J. Chibane, and G. Pons-Moll. Object pop-up: Can we infer 3d objects and their poses from human interactions alone? In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023.
- [4] X. Xie, B. L. Bhatnagar, and G. Pons-Moll. CHORE: Contact, human and object reconstruction from a single RGB image. In S. Avidan, G. Brostow, M. Cissé, G. Farinella, and T. Hassner, eds., *Computer Vision – ECCV 2022*, Tel Aviv, Israel, 2022, LNCS 13662, pp. 125–145. Springer.
- [5] X. Xie, B. L. Bhatnagar, and G. Pons-Moll. Visibility aware human-object interaction tracking from single rgb camera. In *IEEE Conference on Computer Vision and PatternRecognition (CVPR)*, 2023.
- [6] X. Zhang, B. L. Bhatnagar, S. Starke, V. Guzov, and G. Pons-Moll. COUCH: Towards controllable human-chair interactions. In S. Avidan, G. Brostow, M. Cissé, G. Farinella, and T. Hassner, eds., *Computer Vision – ECCV 2022*, Tel Aviv, Israel, 2022, LNCS 13665, pp. 518–535. Springer.
- [7] K. Zhou, B. L. Bhatnagar, J. E. Lenssen, and G. Pons-Moll. TOCH: Spatio-temporal object correspondence to hand for motion refinement. In S. Avidan, G. Brostow, M. Cissé, G. Farinella, and T. Hassner, eds., *Computer Vision – ECCV 2022*, Tel Aviv, Israel, 2022, LNCS 13663, pp. 1–19. Springer.


### 28.6.4 3D Scenes and Objects

*Investigators: Gerard Pons-Moll, Julian Chibane, Verica Lazova, Tuan Anh Tran, Vladimir Guzov, Yongqin Xian, Bharat Lal Bhatnagar, Bernt Schiele, Zeynep Akata, Keyang Zhou in cooperation with Aayush Bansal (CMU), Francis Engelmann (ETH Zürich), Kyle Olszewski (Snap Inc.), Sergey Tulyakov (Snap Inc.), Albert Pumarola (Institut de Robotica i Informatica Industrial), Enric Corona (Institut de Robotica i Informatica Industrial), Francesc Moreno-Noguer (Institut de Robotica i Informatica Industrial)*

Crucial to learning about humans and representing them realistically in the 3D world is learning about their environment and the parts it comprises. There are several key problems to understanding the human environment: finding a good representation that can be learned from RGB input such as images or videos [1], allowing for the scene to change [3] and incorporating the dynamics of the real world [4]. Additionally studying individual objects can be equally as important. Separating [2] and understanding individual objects within the 3D scene and learning their shape [5, 6] can help us understand the virtual world. We have addressed all of these problems in our work. The code, models and data are publicly available to support further research in this topics.

### Scene representation, editing and dynamics

Recent works in the Neural Radiance field learn to represent environments as a radiance field network, that can predict color and density for each 3D location within the scene. Such representations have achieved impressive quality and realism at rendering, surpassing classical pipelines which rely on multi-view reconstruction. However they are usually scene specific and learned from a dense set of calibrated images. We have introduced Stereo Radiance Fields (SRF) [1], a neural view synthesis approach that is trained end-to-end, generalizes to new scenes, and requires only sparse views at test time. The core idea is a neural architecture inspired by classical multi-view stereo methods, which estimates surface points by finding similar image regions in stereo images. Experiments show that SRF learns structure instead of overfitting on a scene.

Additionally the human environment is not fixed, it can change as a result of human interaction. A good scene representation should allow for scene editing and manipulation. With this in mind we have presented Control-NeRF [3], a model that couples learnt scene-specific feature volumes with a scene agnostic neural rendering network. With this hybrid representation, we decouple neural rendering from scene-specific geometry and appearance. Since the feature volumes are independent of the rendering model, we can manipulate and combine scenes by editing their corresponding feature volumes. The edited volume can then be plugged into the rendering model to synthesize high-quality novel views.

Furthermore, the real world and our environment are dynamic, they consist of many nonstationary parts, animate objects or humans that move or change over time. We have introduced D-NeRF [4], a method that extends neural radiance fields to a dynamic domain, allowing us to reconstruct and render novel images of objects under rigid and non-rigid motions from a single camera moving around the scene. For this purpose we consider time as an additional input to the system, and split the learning process in two main stages: one that encodes the scene into a canonical space and another that maps this canonical representation into the deformed scene at a particular time. Both mappings are simultaneously learned using fully-connected networks.

### Object segmentation and reconstruction

While studying the human environment as a whole one should not forget about the individual parts that comprise it. Studying separate objects is just as important to understanding a scene and how a virtual human exists and interacts within it. In order to learn about individual objects one needs to learn how to separate them from the scene by solving a 3D segmentation problem, learn their shape or reconstruct them from a given input.

Current 3D segmentation methods heavily rely on large-scale point-cloud datasets, which are notoriously laborious to annotate. We have introduced Box2Mask [2], where the key idea is to leverage 3D bounding box labels which are easier and faster to annotate (See Figure 28.56). We showed that it is possible to train dense segmentation models using only bounding box labels. At the core of our method, lies a deep model, inspired by classical

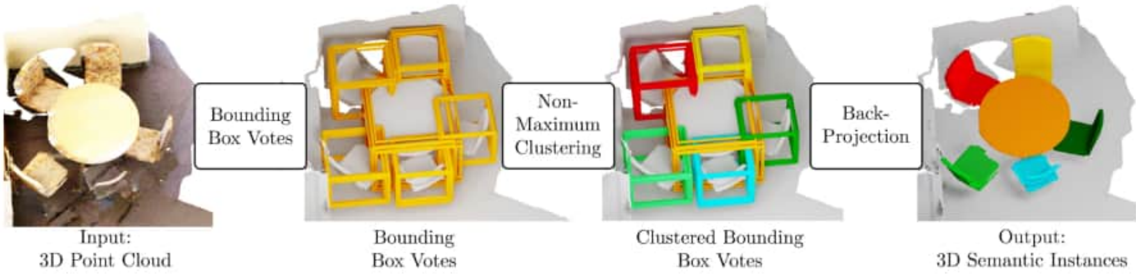



Figure 28.56: Box2Mask Overview. Given a colored 3D point cloud of a scene, Box2Mask predicts dense semantic instance masks. This is achieved with only coarse bounding box annotations. First, for each point in the input scene, our model votes for an instance representation, parameterized as a bounding box. Votes are clustered via non-maximum clustering. Probabilistic instance masks are obtained via back-projection to the original space.

Hough voting, that directly votes for bounding box parameters, and a clustering method specifically tailored to bounding box votes.

While studying object shape and reconstruction we have addressed the task of estimating the 3D shapes of novel shape classes from a single RGB image. Prior works are either limited to reconstructing known training classes or are unable to reconstruct high-quality shapes. To solve those issues, we proposed Generalizing Implicit Networks (GIN) [5] which decomposes 3D reconstruction into 1.) front-back depth estimation followed by differentiable depth voxelization, and 2.) implicit shape completion with 3D features (See Figure 28.57). The key insight is that the depth estimation network learns local class-agnostic shape priors, allowing us to generalize to novel classes, while our implicit shape completion network is able to predict accurate shapes with rich details by learning implicit surfaces in 3D voxel space.

Most learning methods for 3D data, including shape estimation, suffer significant performance drops when the data is not carefully aligned to a canonical orientation, which is usually the case for real-world in-the-wild data. Aligning real world data collected from different sources is non-trivial and requires manual intervention. We have proposed the Adjoint Rigid Transform (ART) Network [6], a neural module which can be integrated with a variety of 3D networks to significantly boost their performance. ART learns to rotate input shapes to a learned canonical orientation, which is crucial for a lot of tasks such as shape reconstruction, interpolation, non-rigid registration, and latent disentanglement. With only self-supervision, ART facilitates learning a canonical orientation for both rigid and nonrigid shapes, which leads to a notable boost in performance of aforementioned tasks.

## References


- [1] J. Chibane, A. Bansal, V. Lazova, and G. Pons-Moll. Stereo radiance fields (SRF): Learning view synthesis from sparse views of novel scenes. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2021)*, Nashville, TN, USA (Virtual), 2021, pp. 7907–7916. IEEE.
- [2] J. Chibane, F. Engelmann, A. T. Tran, and G. Pons-Moll. Box2Mask: Weakly supervised 3D semantic instance segmentation using bounding boxes. In S. Avidan, G. Brostow, M. Cissé,

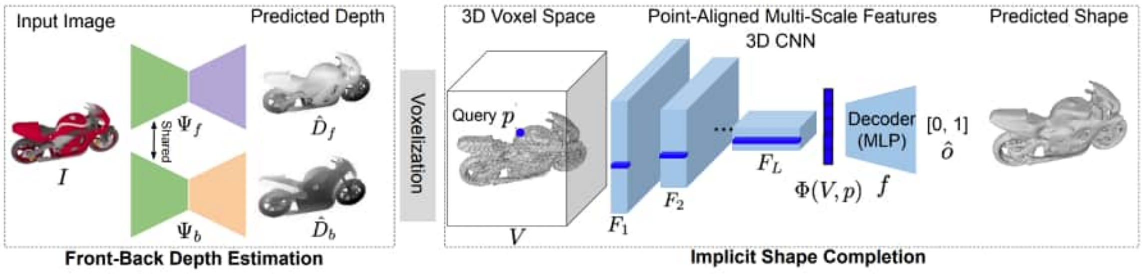



Figure 28.57: Overview of our proposed method GIN for single-image reconstruction of novel classes. Given an input image I, we first predict its front and back depth maps with depth estimation networks with a shared encoder. Then we unproject the depth maps into 3D points followed by voxelizing them into a 3D voxel grid V in viewer-centered coordinate system. Afterwards, we apply 3D convolutions and extract point-aligned multi-scale features by trilinearly interpolating intermediate feature maps at query point p. Finally, our implicit decoder predicts the occupancy of the query point p

G. Farinella, and T. Hassner, eds., *Computer Vision – ECCV 2022*, Tel Aviv, Israel, 2022, LNCS 13691, pp. 681–699. Springer.

- [3] V. Lazova, V. Guzov, K. Olszewski, S. Tulyakov, and G. Pons-Moll. *Control-NeRF: Editable Feature Volumes for Scene Rendering and Manipulation*, 2022. arXiv: 2204.10850.
- [4] A. Pumarola, E. Corona, G. Pons-Moll, and F. Moreno-Noguer. D-NeRF: Neural radiance fields for dynamic scenes. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2021)*, Nashville, TN, US (Virtual), 2021, pp. 10318–10327. IEEE.
- [5] Y. Xiang, J. Chibane, B. L. Bhatnagar, B. Schiele, Z. Akata, and G. Pons-Moll. Any-shot gin: Generalizing implicit networks for reconstructing novel classes. In *2022 International Conference on 3D Vision (3DV)*, 2022. IEEE.
- [6] K. Zhou, B. L. Bhatnagar, B. Schiele, and G. Pons-Moll. *Adjoint Rigid Transform Network: Task-conditioned Alignment of 3D Shapes*, 2021. arXiv: 2102.01161.

## 28.7 Multimodal Deep Learning

*Coordinator: Zeynep Akata*

### 28.7.1 Explainable AI

*Investigators: Stephan Alaniz, Bernt Schiele, Zeynep Akata, in cooperation with Diego Marcos (INRIA), Jae Myung Kim, Katrin Renz, Otniel-Bogdan Mercea, Yao Rong, Kashyap Chitta, A. Sophia Koepke, Seong Joon Oh, Andreas Geiger (University of Tübingen), Enkelejda Kasneci (Technical University of Munich), Wenjia Xu (Beijing University), and Junsuk Choe (Sogang University)*

Deep neural networks are considered black boxes which make it difficult to determine whether an AI system has made a mistake or biased decision. The goal of Explainable AI

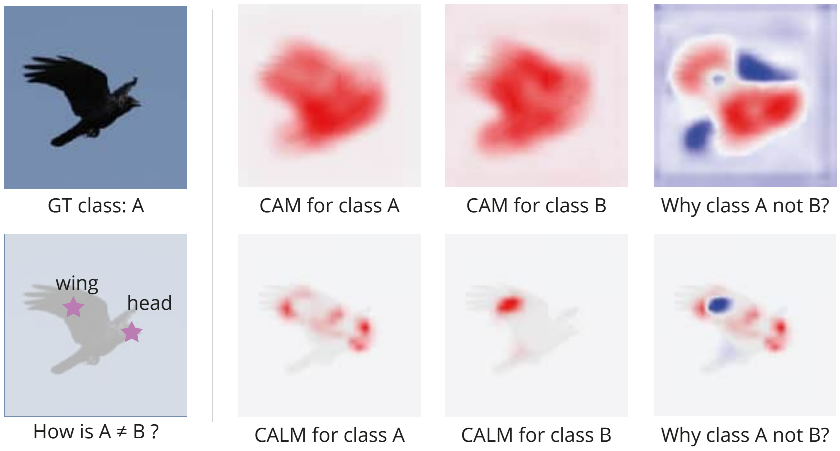



Figure 28.58: CAM vs CALM. CALM is better at locating the actual cues used for the recognition than CAM. Two bird classes A and B only differ in their head and wing attributes. Attributions for class A, B, and their difference are shown. While CAM fails to detect the head and wing, CALM captures them accurately.

(XAI) is to develop systems that can provide clear and understandable explanations of their decision-making processes and outcomes to the human. This way, XAI can establish trust and transparency in AI systems, which are increasingly being used in critical applications such as healthcare, finance, and transportation. Explainable systems help to identify and mitigate potential biases and errors, and provides a way to determine why an a particular decision was made.

Through four lines of work, we explored different topics in Explainable AI. In [2], we incorporate the class attributions into the training procedure to obtain better visual explanations of the vision model. On the other hand, [4] looks at how human attention differs from model explanations and how we can use human gaze data to improve our models. [1] explores a different direction of using a communication loop between two agents to solve a classification task making the decision process explainable in the process. Finally, [3] is an explainable planning network for driving that focuses on few important objects in the scene making its decision process easier to interpret.

### Keep CALM and Improve Visual Feature Attribution

The class activation mapping, or CAM, has been the cornerstone of feature attribution methods for multiple vision tasks. However, the computation of CAM attribution maps relies on ad-hoc calibration steps that are not part of the training computational graph, making it difficult for us to understand the real meaning of the attribution values. In paper [2], we improve CAM by explicitly incorporating a latent variable encoding the location of the cue for recognition in the formulation, thereby subsuming the attribution map into the training computational graph. The resulting model, *class activation latent mapping*, or *CALM*, identifies discriminative attributes for image classifiers more accurately than CAM and other visual attribution baselines. CALM also shows performance improvements over prior arts on the weakly-supervised object localization benchmarks. Figure 28.58 illustrates explanations from CALM.

#### Human Attention in Fine-grained Classification

The way humans attend to, process and classify a given image has the potential to vastly benefit the performance of deep learning models. Exploiting where humans are focusing can rectify models when they are deviating from essential features for correct decisions. To validate that human attention contains valuable information for decision-making processes such as fine-grained classification, we compare human attention and model explanations in discovering important features in [4]. Towards this goal, we collect human gaze data for the fine-grained classification dataset CUB and build a dataset named CUB-GHA (Gaze-based Human Attention). Furthermore, we propose the Gaze Augmentation Training (GAT) and Knowledge Fusion Network (KFN) to integrate human gaze knowledge into classification models. Our result reveals that integrating human attention knowledge benefits classification effectively, e.g. improving the baseline by 4.38% on CXR, a medical dataset.

#### Learning Decision Trees Recurrently Through Communication

Integrated interpretability without sacrificing the prediction accuracy of decision making algorithms has the potential of greatly improving their value to the user. Instead of assigning a label to an image directly, in [1] we propose to learn iterative binary sub-decisions, inducing sparsity and transparency in the decision making process. The key aspect of our model is its ability to build a decision tree whose structure is encoded into the memory representation of a Recurrent Neural Network jointly learned by two models communicating through message passing. In addition, our model assigns a semantic meaning to each decision in the form of binary attributes, providing concise, semantic and relevant rationalizations to the user. On three benchmark image classification datasets, including the large-scale ImageNet, our model generates human interpretable binary decision sequences explaining the predictions of the network while maintaining state-of-the-art accuracy.

#### PlanT: Explainable Planning Transformers via Object-Level Representations

Planning an optimal route in a complex environment requires efficient reasoning about the surrounding scene. While human drivers prioritize important objects and ignore details not relevant to the decision, learning-based planners typically extract features from dense, high-dimensional grid representations containing all vehicle and road context information. In paper [3], we propose PlanT, a novel approach for planning in the context of self-driving that uses a standard transformer architecture. PlanT is based on imitation learning with a compact object-level input representation. On the Longest6 benchmark for CARLA, PlanT outperforms all prior methods (matching the driving score of the expert) while being 5.3x faster than equivalent pixel-based planning baselines during inference. Furthermore, PlanT is explainable and our results indicate that PlanT can focus on the most relevant object in the scene, even when this object is geometrically distant.

## References


- [1] S. Alaniz, D. Marcos, B. Schiele, and Z. Akata. Learning decision trees recurrently through communication. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2021)*, Nashville, TN, USA (Virtual), 2021, pp. 13518–13527. IEEE.
- [2] J. M. Kim, J. Choe, Z. Akata, and S. J. Oh. Keep CALM and improve visual feature attribution. In *ICCV 2021, IEEE/CVF International Conference on Computer Vision*, Virtual Event, 2021, pp. 8330–8340. IEEE.
- [3] K. Renz, K. Chitta, O.-B. Mercea, A. S. Koepke, Z. Akata, and A. Geiger. PlanT: Explainable planning transformers via object-level representations. In K. Liu, D. Kulic, and J. Ichnowski, eds., *Proceedings of the 6th Annual Conference on Robot Learning (CoRL 2022)*, Auckland, New Zealand, 2022, Proceedings of the Machine Learning Research 205, pp. 459–470. MLResearchPress.
- [4] Y. Rong, W. Xu, Z. Akata, and E. Kasneci. Human attention in fine-grained classification. In *British Machine Vision Conference (BMVC)*, 2021.


## 28.7.2 Multimodal Learning

*Investigators: Stephan Alaniz, Yongqin Xian, Zeynep Akata, in cooperation with Otniel-Bogdan Mercea, Thomas Hummel, Shyamgopal Karthik, Uddeshya Upadhyay, Lukas Riesch, Andrei Neculai, Tobias Hepp, Sergios Gatidis, Massimiliano Mancini, A. Sophia Koepke (University of Tübingen), Andreea-Maria Oncescu, Joao F. Henriques, Samuel Albanie (University of Oxford), Abhra Chaudhuri (University of Exeter), Anjan Dutta (University of Surrey), Yanbei Chen (Amazon), Marco Federici (University of Amsterdam), and Ying Shan (Tencent)*

When employing multimodal learning, we train our deep networks on multiple sources of data, such as images, video, audio, and text. This allows the model to learn complex relationships between different types of data, leading to improved performance and better generalizability and robustness to unseen data compared to using a single modality.

Our research in multimodal learning can be roughly categorized into audio-visual learning [8, 7, 4, 10], combining vision with language [6, 11], compositionality through multi-modality [5, 1, 9] and working on diverse modalities [3, 2, 12]. We are describing each of these research directions in the following.

### Audio-Visual Learning

The audio modality can enrich representations of many visual tasks with supplementary information not otherwise available from images. We show the efficacy of combining audio and visual cues from video data in the challenging zero-shot learning setting in [8, 7]. In [8] we propose to learn multi-modal representations from audio-visual data using cross-modal attention and exploit textual label embeddings for transferring knowledge from seen classes to unseen classes. We further introduce a (generalised) zero-shot learning benchmark on three audio-visual datasets of varying sizes and difficulty, VGGSound, UCF, and ActivityNet. In [7], we propose a multi-modal and Temporal Cross-attention Framework (TCaF) for audio-visual generalised zero-shot learning. Our proposed framework that ingests temporal features yields state-of-the-art performance by encouraging it to focus on cross-modal correspondence across time instead of self-attention within the modalities.

We further show in [4] that distillation allows us to compose audio, image, and video representations uncover richer multi-modal knowledge. We establish a new multi-modal distillation benchmark on three video datasets: UCF101, ActivityNet, and VGGSound, and demonstrate that our model significantly outperforms a variety of existing knowledge distillation methods. Finally, we consider the task of retrieving audio using free-form natural language queries. In [10], we introduce challenging new benchmarks for text-based audio retrieval using text annotations sourced from the Audiocaps and Clotho datasets, and demonstrate the benefits of pre-training on diverse audio tasks.

### Combining Vision with Language

Semantics learned from language allow vision models to learn structured representations for a diverse set of tasks. Firstly with [11], we propose a language guidance objective for visual similarity learning in the Deep Metric Learning (DML) framework. Leveraging language embeddings of expert- and pseudo-classnames, we contextualize and realign visual representation spaces corresponding to meaningful language semantics for better semantic consistency. Extensive experiments and ablations show language guidance offering significant, model-agnostic improvements for DML, achieving competitive and state-of-the-art results on all benchmarks.

Secondly, we look at models capable of generating natural language explanations (NLEs) for their predictions on vision-language (VL) tasks in [6]. However, there is a lack of comparison between existing methods, which is due to a lack of re-usable evaluation frameworks and a scarcity of datasets. In this work, we introduce e-ViL and e-SNLI-VE, benchmarks for explainable vision-language tasks that establishe a unified evaluation framework. We also propose a new model that combines UNITER, which learns joint embeddings of images and text, and GPT-2, a pre-trained language model that is well-suited for text generation.

### Compositionality through Multi-Modality

Learning compositionality allows us to recognize compositions of states and objects in images, given only a subset of them during training and no prior on the unseen compositions.

In [5], we predict the primitives, i.e. states and objects, independently instead of jointly as employed by previous work. Moreover, we estimate the feasibility of each composition through external knowledge, using this prior to remove unfeasible compositions from the output space. Our model, Knowledge-Guided Simple Primitives (KG-SP), achieves state of the art in open-world compositional zero-shot learning surpassing most recent competitors even when coupled with semi-supervised learning techniques. Figure 28.59 illustrates the problem and our model.

We also present CoMix [1], a model that learns a shared Gaussian mixture representation imposing the compositionality of the text onto the visual domain without having explicit location supervision. We learn to split images into separately encoded patches to associate visual and textual representations in an interpretable manner and show that our model is able to perform weakly supervised object detection and extrapolates to unseen combination of objects.

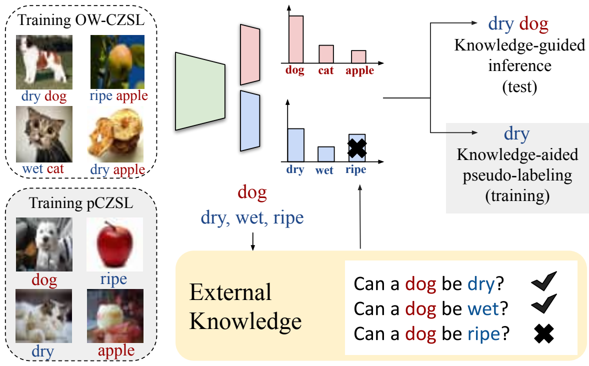



Figure 28.59: We consider the problems of open-world CZSL (OW-CZSL), where we lack priors on unseen compositions at test time, and CZSL under partial supervision (pCZSL) where we also lack compositional labels during training (left). We tackle them by independently predicting object (red) and state (blue) labels and by using external knowledge (bottom) to estimate the feasibility of compositions, reducing the search space during inference and improving pseudo-labeling during training in pCZSL.

Lastly, [9] takes compositional image retrieval. To this end, we propose a novel multimodal probabilistic composer (MPC) that models input images and texts as probabilistic embeddings, which can be further composed by a probabilistic composition rule to facilitate image retrieval with multiple multimodal queries. Our probabilistic model formulation outperforms existing related methods on multimodal image retrieval while generalizing well to query with different amounts of inputs.

### Diverse Modalities

Investigating a diverse set of modalities allows drawing different information sources together. Sketches are a popular modality for sketch-based image retrieval. We find in [3] that sketches and images can often provide complementary information describing the underlying concept. Thus, we propose a cross-attention framework that fuses modality-specific information instead of discarding them. Our framework first maps paired datapoints to fused representations that unify information from both modalities. We achieve state-of-the-art results on three fine-grained sketch-based image retrieval benchmarks.

Furthermore, we explore image-to-image translation tasks. In [2], we tackle semantic image synthesis and find that jointly learning the semantic conditioning and image latents significantly improves the modeling capabilities of the generative transformer model. Tying the two modalities at the autoencoding stage proves to be an important ingredient to improve autoregressive modeling performance. We show that our model improves semantic image synthesis using autoregressive models on popular semantic image datasets ADE20k, Cityscapes and COCO-Stuff. And in [12], we look at image-to-image translations for medical applications. We propose an uncertainty-guided progressive learning scheme to estimate the uncertainty in the predictions. By incorporating aleatoric uncertainty as attention maps for

GANs trained in a progressive manner, we generate images of increasing fidelity progressively. Our model generalizes well in three different tasks and improves performance over state of the art under full-supervision and weak-supervision with limited data.

## References


- [1] S. Alaniz, M. Federici, and Z. Akata. Compositional mixture representations for vision and text. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPR 2022)*, New Orleans, LA, USA, 2021, pp. 4201–4210. IEEE.
- [2] S. Alaniz, T. Hummel, and Z. Akata. Semantic image synthesis with semantically coupled VQ-model. In *ICLR Workshop on Deep Generative Models for Highly Structured Data (ICLR 2022 DGM4HSD)*, Virtual, 2022. OpenReview.net.
- [3] A. Chaudhuri, M. Mancini, Y. Chen, Z. Akata, and A. Dutta. Cross-modal fusion distillation for fine-grained sketch-based image retrieval. In *33rd British Machine Vision Conference (BMVC 2022)*, London, UK, 2022, Article 499. BMVA Press.
- [4] Y. Chen, Y. Xian, A. S. Koepke, and Z. Akata. Distilling audio-visual knowledge by compositional contrastive learning. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2021)*, Virtual Conference, 2021, pp. 7012–7021. IEEE.
- [5] S. Karthik, M. Mancini, and Z. Akata. KG-SP: knowledge guided simple primitives for open world compositional zero-shot learning. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022.
- [6] M. Kayser, O.-M. Camburu, L. Salewski, C. Emde, V. Do, Z. Akata, and T. Lukasiewicz. e-ViL: A dataset and benchmark for natural language explanations in vision-language tasks. In *ICCV 2021, IEEE/CVF International Conference on Computer Vision*, Virtual Event, 2021, pp. 1224–1234. IEEE.
- [7] O.-B. Mercea, T. Hummel, A. S. Koepke, and Z. Akata. Temporal and cross-modal attention for audio-visual zero-shot learning. In S. Avidan, G. Brostow, M. Cissé, G. M. Farinella, and T. Hassner, eds., *Computer Vision – ECCV 2022*, Tel Aviv, Israel, 2022, LNCS 13680, pp. 488–505. Springer.
- [8] O.-B. Mercea, L. Riesch, A. S. Koepke, and Z. Akata. Audio-visual generalised zero-shot learning with cross-modal attention and language. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2022)*, New Orleans, LA, USA, 2022, pp. 10543–10553. IEEE.
- [9] A. Neculai, Y. Chen, and Z. Akata. Probabilistic compositional embeddings for multimodal image retrieval. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPR 2022)*, New Orleans, LA, USA, 2021, pp. 4546–4556. IEEE.
- [10] A. Oncescu, A. S. Koepke, J. F. Henriques, Z. Akata, and S. Albanie. Audio retrieval with natural language queries. In *Interspeech*, 2021.
- [11] K. Roth, O. Vinyals, and Z. Akata. Integrating language guidance into vision-based deep metric learning. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022.
- [12] U. Upadhyay, Y. Chen, T. Hepp, S. Gatidis, and Z. Akata. Uncertainty-guided progressive gans for medical image translation. In *Medical Image Computing and Computer Assisted Intervention (MICCAI)*, 2021.

### 28.7.3 Learning with Less Supervision

*Investigators: Yongqin Xian, Stephan Alaniz, Wenjia Xu, Zeynep Akata and Bernt Schiele, in cooperation with Giuseppe Pastore (Politecnico di Torino), Fabio Cermelli (Politecnico di Torino), Massimiliano Mancini (University of Tübingen), Muhammad Ferjad Naeem (Technical University of Munich), Federico Tombari (Technical University of Munich), Jiuniu Wang (City University of Hong Kong), Sarkhan Badirli (Purdue University), George O. Mohler, Christine Picard (Purdue University), Ushasi Chaudhuri (Indian Institute of Technology), Ruchika Chavan (Indian Institute of Technology), Biplab Banerjee (Indian Institute of Technology), Anjan Dutta (University of Surrey), Bruno Korbar (University of Oxford), Matthijs Douze (Meta), Samarth Sinha (University of Toronto), Karsten Roth (University of Tübingen), Anirudh Goyal (University of Montreal), Marzyeh Ghassemi (MIT), Hugo Larochelle (MILA), Youngwook Kim (Seoul National University), Jae Myung Kim (University of Tübingen), Lennart Alexander der Goten Van (KTH), Tobias Hepp (University of Tübingen), Seong Joon Oh (University of Tübingen), Lorenzo Torresani (Dartmouth College), Animesh Garg (University of Toronto), Barbara Caputo (Politecnico di Torino), Murat Dundar (Purdue University), Hyunjung Shim (KAIST), Kevin Smith (KTH), Jungwoo Lee (Seoul National University), Diego Marcos (INRIA), Yuchen Ma (University of Tübingen), Yanbei Chen (Amazon)*

Deep learning with less supervision has gained significant attention in recent years due to its potential to reduce the need for expensive and time-consuming manual labeling of data. In this section, we show that compelling results can be achieved for visual recognition tasks i.e., object recognition, image retrieval and semantic segmentation with only a limited number of training examples or weak supervision signals. In weakly supervised learning, we introduce a new evaluation protocol for weakly-supervised object localization in [5] and propose to correct the large loss samples to tackle weakly supervised multi-label classification in [8]. In the area of few-shot learning, we present a new few-shot video classification method with video retrieval and video feature generation in [13] and introduce a new task called Incremental Few-Shot Segmentation as well as strong baselines in [3]. Moreover, we have made significant contributions in the zero-shot learning field. We propose a new attribute prototype networks that jointly learn discriminative global and local features using only class-level attributes in [14] and improve unsupervised semantic embedding for zero-shot learning in [15]. In [4], we address the challenges in zero-shot sketch-based image retrieval by performing a bi-level domain adaptation and utilizing a symmetric loss. On sketches, we further show in [1] how a self-supervised abstraction through simple primitives can help in downstream tasks under constrained information budgets. In [10], we propose a novel graph convolutional neural network for compositional zero-shot learning. In [11], we propose a consistency regularizer to improve pseudo-labeling for zero-shot semantic segmentation. In [2], we propose a new method that uses DNA as side information for the first time for fine-grained zero-shot classification of species. Moreover, we explored data efficient learning and show that models that learn uniformly distributed features perform better transfer learning at test-time in [12]. We also show in [6] that combining a discriminative and alignment learning objective for self-supervised learning consistently outperforms the existing state-of-the-art. With our Embedding Graph Alignment from [9], we can distill the knowledge from self-supervised

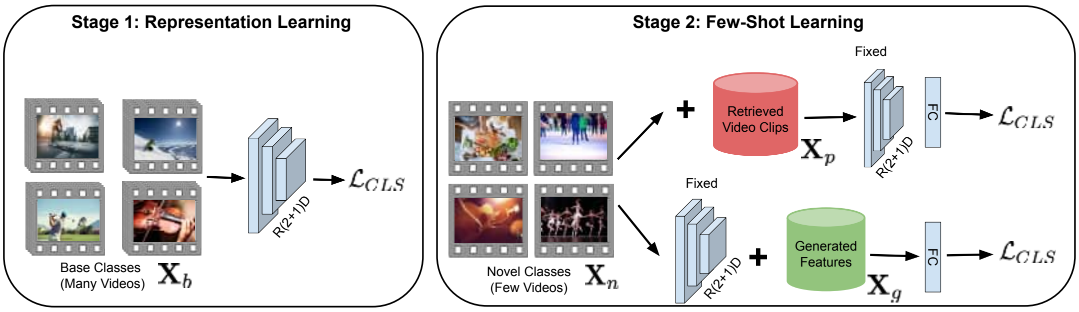



Figure 28.60: Our few-shot video classification approach in [13] comprises two learning stages: representation learning and few-shot learning. In representation learning, we train a R(2+1)D CNN using the base classes. In few-shot learning, we propose to expand the few-shot training set of novel classes by either retrieving videos from another tag-labeled video dataset and generating video features from semantic embeddings.

teacher to a student network with a different architecture. Finally, in [7], we introduce a new conditional multi-scale GAN architecture to de-identify diagnostic scans.

### Evaluation for Weakly Supervised Object Localization: Protocol, Metrics, and Datasets

*Investigators: Junsuk Choe, Seong Joon Oh, Sanghyuk Chun, Seungho Lee, Zeynep Akata and Hyunjung Shim*

Weakly-supervised object localization (WSOL) has gained popularity over the last years for its promise to train localization models with only image-level labels. Since the seminal WSOL work of class activation mapping (CAM), the field has focused on how to expand the attention regions to cover objects more broadly and localize them better. However, these strategies rely on full localization supervision for validating hyperparameters and model selection, which is in principle prohibited under the WSOL setup. In this paper, we argue that WSOL task is ill-posed with only image-level labels, and propose a new evaluation protocol where full supervision is limited to only a small held-out set not overlapping with the test set. We observe that, under our protocol, the five most recent WSOL methods have not made a major improvement over the CAM baseline. Moreover, we report that existing WSOL methods have not reached the few-shot learning baseline, where the full-supervision at validation time is used for model training instead. Based on our findings, we discuss some future directions for WSOL.

### Generalized Few-Shot Video Classification With Video Retrieval and Feature Generation

*Investigators: Yongqin Xian, Bruno Korbar, Matthijs Douze, Lorenzo Torresani, Bernt Schiele and Zeynep Akata*

Few-shot learning aims to recognize novel classes from a few examples. Although significant progress has been made in the image domain, few-shot video classification is relatively unexplored. We argue that previous methods underestimate the importance of video feature learning and propose to learn spatiotemporal features using a 3D CNN. Proposing a two-stage approach that learns video features on base classes followed by fine-tuning the classifiers on novel classes, we show that this simple baseline approach outperforms prior few-shot video classification methods by over 20 points on existing benchmarks. An overview of our method is shown in Figure 28.60. To circumvent the need of labeled examples, we present two novel approaches that yield further improvement. First, we leverage tag-labeled videos from a large dataset using tag retrieval followed by selecting the best clips with visual similarities. Second, we learn generative adversarial networks that generate video features of novel classes from their semantic embeddings. The experimental results show that our retrieval and feature generation approach significantly outperform the baseline approach.

### BDA-SketRet: Bi-level domain adaptation for zero-shot SBIR

*Investigators: Ushasi Chaudhuri, Ruchika Chavan, Biplab Banerjee, Anjan Dutta and Zeynep Akata*

The efficacy of zero-shot sketch-based image retrieval (ZS-SBIR) models is governed by two challenges. The immense distributions-gap between the sketches and the images requires a proper domain alignment. Moreover, the fine-grained nature of the task and the high intra-class variance of many categories necessitates a class-wise discriminative mapping among the sketch, image, and the semantic spaces. Under this premise, we propose BDA-SketRet, a novel ZS-SBIR framework performing a bi-level domain adaptation for aligning the spatial and semantic features of the visual data pairs progressively. In order to highlight the shared features and reduce the effects of any sketch or image-specific artifacts, we propose a novel symmetric loss function based on the notion of information bottleneck for aligning the semantic features while a cross-entropy-based adversarial loss is introduced to align the spatial feature maps. Finally, our CNN-based model confirms the discriminativeness of the shared latent space through a novel topology-preserving semantic projection network. Experimental results on the extended Sketchy, TU-Berlin, and QuickDraw datasets exhibit sharp improvements over the literature.

### Uniform Priors for Data-Efficient Learning

*Investigators: Samarth Sinha, Karsten Roth, Anirudh Goyal, Marzyeh Ghassemi, Zeynep Akata, Hugo Larochelle, Animesh Garg*

Few or zero-shot adaptation to novel tasks is important for the scalability and deployment of machine learning models. It is therefore crucial to find properties that encourage more transferable features in deep networks for generalization. In this paper, we show that models that learn uniformly distributed features from the training data, are able to perform better transfer learning at test-time. Motivated by this, we evaluate our method: uniformity regularization ($\mathcal{UR}$) on its ability to facilitate adaptation to unseen tasks and data on six distinct domains: Few-Learning with Images, Few-shot Learning with Language, Deep Metric Learning, Zero-Shot Domain Adaptation, Out-of-Distribution classification, and Neural Radiance Fields. Across all experiments, we show that using $\mathcal{UR}$, we are able to learn robust

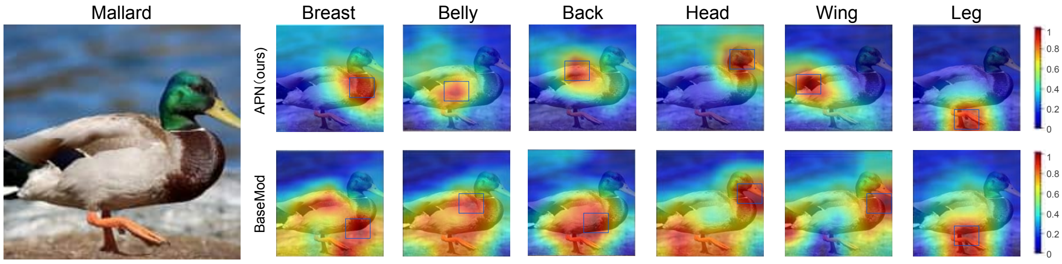



Figure 28.61: Part localization results in [14]: Attention maps for each body part of the bird Mallard generated by our `APN` (first row) and `BaseMod` (second row). Boxes mark out the area with the highest attention.

vision systems which consistently offer benefits over baselines trained without uniformity regularization and are able to achieve state-of-the-art performance in Deep Metric Learning, Few-shot learning with images and language.

### Attribute Prototype Network for Any-Shot Learning

*Investigators: Wenjia Xu, Yongqin Xian, Jiuniu Wang, Bernt Schiele and Zeynep Akata*

Any-shot image classification allows to recognize novel classes with only a few or even zero samples. To better transfer attribute-based knowledge from seen to unseen classes, we argue that an image representation with integrated attribute localization ability would be beneficial for any-shot image classification tasks. To this end, we propose a novel representation learning framework that jointly learns discriminative global and local features using only class-level attributes. While a visual-semantic embedding layer learns global features, local features are learned through an attribute prototype network that simultaneously regresses and decorrelates attributes from intermediate features. We show that our locality augmented image representations achieve a new state-of-the-art on challenging benchmarks, i.e. CUB, AWA2, and SUN. In Figure 28.61, we show that our model points to the visual evidence of the attributes in an image, confirming the improved attribute localization ability of our image representation.

### Large Loss Matters in Weakly Supervised Multi-Label Classification

*Investigators: Youngwook Kim, Jae Myung Kim, Zeynep Akata and Jungwoo Lee*

Weakly supervised multi-label classification (WSML) task, which is to learn a multi-label classification using partially observed labels per image, is becoming increasingly important due to its huge annotation cost. In this work, we first regard unobserved labels as negative labels, casting the WSML task into noisy multi-label classification. From this point of view, we empirically observe that memorization effect, which was first discovered in a noisy multi-class setting, also occurs in a multi-label setting. That is, the model first learns the representation of clean labels, and then starts memorizing noisy labels. Based on this finding, we propose

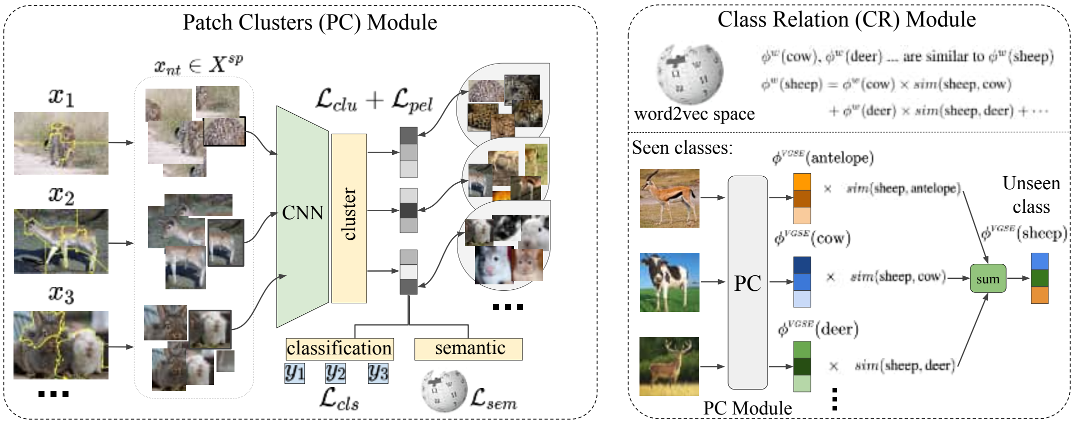



Figure 28.62: Our visually-grounded semantic embedding network consists of two modules. The Patch Clustering (PC) module learns clusters from patch images, and predicts semantic embeddings for seen classes with their images. The Class Relation (CR) module predicts the unseen class embeddings using unseen and seen class relations learned from external knowledge, e.g., word2vec.

novel methods for WSML which reject or correct the large loss samples to prevent model from memorizing the noisy label. Without heavy and complex components, our proposed methods outperform previous state-of-the-art WSML methods on several partial label settings including Pascal VOC 2012, MS COCO, NUSWIDE, CUB, and OpenImages V3 datasets. Various analysis also show that our methodology actually works well, validating that treating large loss properly matters in a weakly supervised multi-label classification.

### VGSE: Visually-Grounded Semantic Embeddings for Zero-Shot Learning

*Investigators: Wenjia Xu, Yongqin Xian, Jiuniu Wang, Bernt Schiele and Zeynep Akata*

Human-annotated attributes serve as powerful semantic embeddings in zero-shot learning. However, their annotation process is labor-intensive and needs expert supervision. Current unsupervised semantic embeddings, i.e., word embeddings, enable knowledge transfer between classes. However, word embeddings do not always reflect visual similarities and result in inferior zero-shot performance. We propose to discover semantic embeddings containing discriminative visual properties for zero-shot learning, without requiring any human annotation. Our model visually divides a set of images from seen classes into clusters of local image regions according to their visual similarity, and further imposes their class discrimination and semantic relatedness. To associate these clusters with previously unseen classes, we use external knowledge, e.g., word embeddings and propose a novel class relation discovery module. Through quantitative and qualitative evaluation, we demonstrate that our model discovers semantic embeddings that model the visual properties of both seen and unseen classes. Furthermore, we demonstrate on three benchmarks that our visually-grounded semantic embeddings further improve performance over word embeddings across various ZSL models by a large margin.

### Conditional De-Identification of 3D Magnetic Resonance Images

*Investigators: Lennart Alexander Van der Goten, Tobias Hepp, Zeynep Akata and Kevin Smith*

Privacy protection of medical image data is challenging. Even if metadata is removed, brain scans are vulnerable to attacks that match renderings of the face to facial image databases. Solutions have been developed to de-identify diagnostic scans by obfuscating or removing parts of the face. However, these solutions either fail to reliably hide the patient's identity or are so aggressive that they impair further analyses. We propose a new class of de-identification techniques that, instead of removing facial features, remodels them. Our solution relies on a conditional multi-scale GAN architecture. It takes a patient's MRI scan as input and generates a 3D volume conditioned on the patient's brain, which is preserved exactly, but where the face has been de-identified through remodeling. We demonstrate that our approach preserves privacy far better than existing techniques, without compromising downstream medical analyses. Analyses were run on the OASIS-3 and ADNI corpora

### Prototype-based Incremental Few-Shot Segmentation

*Investigators: Fabio Cermelli, Massimiliano Mancini, Yongqin Xian, Zeynep Akata and Barbara Caputo*

Semantic segmentation models have two fundamental weaknesses: i) they require large training sets with costly pixel-level annotations, and ii) they have a static output space, constrained to the classes of the training set. Toward addressing both problems, we introduce a new task, Incremental Few-Shot Segmentation (iFSS). The goal of iFSS is to extend a pretrained segmentation model with new classes from few annotated images and without access to old training data. To overcome the limitations of existing models iniFSS, we propose Prototype-based Incremental Few-Shot Segmentation (PIFS) that couples prototype learning and knowledge distillation. PIFS exploits prototypes to initialize the classifiers of new classes, fine-tuning the network to refine its features representation. We design a prototype-based distillation loss on the scores of both old and new class prototypes to avoid overfitting and forgetting, and batch-renormalization to cope with non-i.i.d.few-shot data. We create an extensive benchmark for iFSS showing that PIFS outperforms several few-shot and incremental learning methods in all scenarios.

### Learning Graph Embeddings for Compositional Zero-Shot Learning

*Investigators: Muhammad Ferjad Naeem, Yongqin Xian, Federico Tombari and Zeynep Akata*

In compositional zero-shot learning, the goal is to recognize unseen compositions (e.g. old dog) of observed visual primitives states (e.g. old, cute) and objects (e.g. car, dog) in the training set. This is challenging because the same state can for example alter the visual appearance of a dog drastically differently from a car. As a solution, we propose a novel graph formulation called Compositional Graph Embedding (CGE) that learns image features, compositional classifiers, and latent representations of visual primitives in an end-to-end

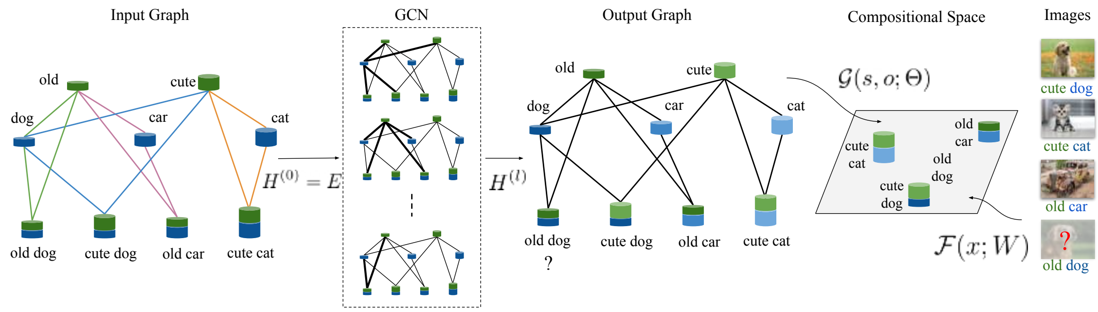



Figure 28.63: Compositional Graph Embed (CGE) [10] learns a globally consistent joint embedding space between image features and classes of seen and unseen compositions from a graph. In our novel graph formulation, nodes are connected if a dependency exists in form of a compositional label e.g. old, car and old car. We backpropagate the classification loss through the seen compositional nodes to the GCN $\mathcal{G}$ and the feature extractor $\mathcal{F}$. Hence, the representation of e.g. the dog is compatible with its different states and the representation of old dog aggregates the knowledge from old, dog, cute dog, old car etc

manner. The key to our approach is exploiting the dependency between states, objects, and their compositions within a graph structure to enforce the relevant knowledge transfer from seen to unseen compositions. By learning a joint compatibility that encodes semantics between concepts, our model allows for generalization to unseen compositions without relying on an external knowledge base like WordNet. Figure 28.63 illustrates an overview of our method. We show that in the challenging generalized compositional zero-shot setting our CGE significantly outperforms the state of the art on MIT-States and UT-Zappos. We also propose a new benchmark for this task based on the recent GQA dataset.

### A Closer Look at Self-Training for Zero-Label Semantic Segmentation

*Investigators: Giuseppe Pastore, Fabio Cermelli, Yongqin Xian, Massimiliano Mancini, Zeynep Akata and Barbara Caputo*

Being able to segment unseen classes not observed during training is an important technical challenge in deep learning, because of its potential to reduce the expensive annotation required for semantic segmentation. Prior zero-label semantic segmentation works approach this task by learning visual-semantic embeddings or generative models. However, they are prone to overfitting on the seen classes because there is no training signal for them. In this paper, we study the challenging generalized zero-label semantic segmentation task where the model has to segment both seen and unseen classes at test time. We assume that pixels of unseen classes could be present in the training images but without being annotated. Our idea is to capture the latent information on unseen classes by supervising the model with self-produced pseudo-labels for unlabeled pixels. We propose a consistency regularizer to filter out noisy pseudo-labels by taking the intersections of the pseudo-labels generated from different augmentations of the same image. Our framework generates pseudo-labels and then retrain the model with human-annotated and pseudo-labelled data. This procedure is repeated for several iterations. As a result, our approach achieves the new state-of-the-art on PascalVOC12 and COCO-stuff datasets in the challenging generalized zero-label semantic segmentation setting, surpassing other existing methods addressing this task with more complex strategies.

### Fine-Grained Zero-Shot Learning with DNA as Side Information

*Investigators: Sarkhan Badirli, Zeynep Akata, George O. Mohler, Christine Picard and Murat Dundar*

Fine-grained zero-shot learning task requires some form of side-information to transfer discriminative information from seen to unseen classes. As manually annotated visual attributes are extremely costly and often impractical to obtain for a large number of classes, in this study we use DNA as side information for the first time for fine-grained zero-shot classification of species. Mitochondrial DNA plays an important role as a genetic marker in evolutionary biology and has been used to achieve near-perfect accuracy in the species classification of living organisms. We implement a simple hierarchical Bayesian model that uses DNA information to establish the hierarchy in the image space and employs local priors to define surrogate classes for unseen ones. On the benchmark CUB dataset, we show that DNA can be equally promising yet in general a more accessible alternative than word vectors as a side information. This is especially important as obtaining robust word representations for fine-grained species names is not a practicable goal when information about these species in free-form text is limited. On a newly compiled fine-grained insect dataset that uses DNA information from over a thousand species, we show that the Bayesian approach outperforms state-of-the-art by a wide margin.

### Abstracting Sketches through Simple Primitives

*Investigators: Stephan Alaniz, Massimiliano Mancini, Anjan Dutta, Diego Marcos, Zeynep Akata*

Humans show high-level of abstraction capabilities in games that require quickly communicating object information. They decompose the message content into multiple parts and communicate them in an interpretable protocol. Toward equipping machines with such capabilities, we propose the Primitive-based Sketch Abstraction task where the goal is to represent sketches using a fixed set of drawing primitives under the influence of a budget. To solve this task, our Primitive-Matching Network (PMN), learns interpretable abstractions of a sketch in a self supervised manner. Specifically, PMN maps each stroke of a sketch to its most similar primitive in a given set, predicting an affine transformation that aligns the selected primitive to the target stroke. We learn this stroke-to-primitive mapping end-to-end with a distance-transform loss that is minimal when the original sketch is precisely reconstructed with the predicted primitives. Our PMN abstraction empirically achieves the highest performance on sketch recognition and sketch-based image retrieval given a communication budget, while at the same time being highly interpretable. This opens up new possibilities for sketch analysis, such as comparing sketches by extracting the most relevant primitives that define an object category.

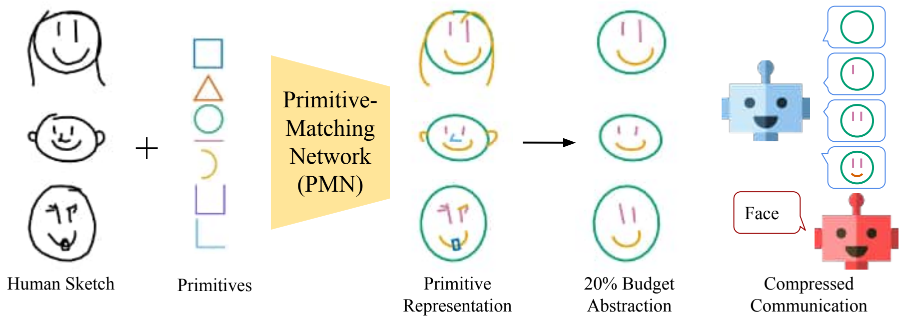



Figure 28.64: Our Primitive-Matching Network (PMN) takes human sketches and replaces their strokes with simple shapes from a set of 7 drawing primitives to create an abstract representation of the sketch. We can further compress the sketch by sub-selecting primitive strokes to meet a desired information budget. When communicating a sketch with a limited budget, our sketch abstractions retain the original semantics to perform well on downstream tasks.

### Concurrent Discrimination and Alignment for Self-Supervised Feature Learning

*Investigators: Anjan Dutta, Massimiliano Mancini, Zeynep Akata*

Existing self-supervised learning methods learn representation using pretext tasks that are either discriminative, where one explicitly specifies which features should be separated, or are alignment based, where one indicates which features should be close together. Prior works ignore the selection of features for these pretext tasks. In this work, we combine the positive aspects of the discriminative and alignment based approaches and design a hybrid method that addresses this issue. Our method explicitly uses the repulsion and attraction mechanism. Specifically, we use discriminative predictive task and maximize the mutual information between paired views that share redundant information. We qualitatively and quantitatively show that our proposed model learns better features that are more effective for the diverse downstream tasks ranging from classification to semantic segmentation. Our experiments on nine established benchmarks show that the proposed model consistently outperforms the existing state-of-the-art results on self-supervised and transfer learning settings.

### Distilling Knowledge from Self-Supervised Teacher by Embedding Graph Alignment

*Investigators: Yuchen Ma, Yanbei Chen, Zeynep Akata*

Recent advances have indicated the strengths of self-supervised pre-training for improving representation learning on downstream tasks. Existing works often utilize self-supervised pre-trained models by fine-tuning on downstream tasks. However, fine-tuning does not generalize to the case when one needs to build a customized model architecture different from the self-supervised model. In this work, we formulate a new knowledge distillation framework to transfer the knowledge from self-supervised pre-trained models to any other student network by a novel approach named Embedding Graph Alignment. Specifically, inspired by the spirit of instance discrimination in self-supervised learning, we model the instance-instance relations by a graph formulation in the feature embedding space and distill the self-supervised teacher knowledge to a student network by aligning the teacher graph and the student graph. Our distillation scheme can be flexibly applied to transfer the self-supervised knowledge to enhance representation learning on various student networks. We demonstrate that our model outperforms multiple representative knowledge distillation methods on three benchmark datasets, including CIFAR100, STL10, and TinyImageNet.

## References


- [1] S. Alaniz, M. Mancini, A. Dutta, D. Marcos, and Z. Akata. Abstracting sketches through simple primitives. In S. Avidan, G. Brostow, M. Cissé, G. M. Farinella, and T. Hassner, eds., *Computer Vision – ECCV 2022*, Tel Aviv, Israel, 2022, LNCS 13689, pp. 396–412. Springer.
- [2] S. Badirli, Z. Akata, G. O. Mohler, C. Picard, and M. Dundar. Fine-grained zero-shot learning with dna as side information. In *NeurIPS*, 2021.
- [3] F. Cermelli, M. Mancini, Y. Xian, Z. Akata, and B. Caputo. Prototype-based incremental few-shot segmentation. In *BMVC*, 2021.
- [4] U. Chaudhuri, R. Chavan, B. Banerjee, A. Dutta, and Z. Akata. Bda-sketret: Bi-level domain adaptation for zero-shot sbir. *Neurocomputing*, 2022.
- [5] J. Choe, S. J. Oh, S. Chun, S. Lee, Z. Akata, and H. Shim. Evaluation for weakly supervised object localization: Protocol, metrics, and datasets. *TPAMI*, 2023.
- [6] A. Dutta, M. Mancini, and Z. Akata. Concurrent discrimination and alignment for self-supervised feature learning. In *IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, 2021.
- [7] L. A. der Goten, Van, T. Hepp, Z. Akata, and K. Smith. Conditional de-identification of 3d magnetic resonance images. In *BMVC*, 2021.
- [8] Y. Kim, J. M. Kim, Z. Akata, and J. Lee. Large loss matters in weakly supervised multi-label classification. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2022)*, New Orleans, LA, USA, 2022, pp. 14136–14145. IEEE.
- [9] Y. Ma, Y. Chen, and Z. Akata. Distilling knowledge from self-supervised teacher by embedding graph alignment. In *33rd British Machine Vision Conference (BMVC 2022)*, London, UK, 2022, Article 973. BMVA Press.
- [10] M. F. Naeem, Y. Xian, F. Tombari, and Z. Akata. Learning graph embeddings for compositional zero-shot learning. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2021)*, Nashville, TN, US (Virtual), 2021, pp. 953–962. IEEE.
- [11] G. Pastore, F. Cermelli, Y. Xian, M. Mancini, Z. Akata, and B. Caputo. A closer look at self-training for zero-label semantic segmentation. In *CVPR*, 2021.
- [12] S. Sinha, K. Roth, A. Goyal, M. Ghassemi, Z. Akata, H. Larochelle, and A. Garg. Uniform priors for data-efficient learning. In *CVPR*, 2022.
- [13] Y. Xian, B. Korbar, M. Douze, L. Torresani, B. Schiele, and Z. Akata. Generalized few-shot video classification with video retrieval and feature generation. *TPAMI*, 2022.
- [14] W. Xu, Y. Xian, J. Wang, B. Schiele, and Z. Akata. Attribute prototype network for any-shot learning. *IJCV*, 2022.
- [15] W. Xu, Y. Xian, J. Wang, B. Schiele, and Z. Akata. Vgse: Visually-grounded semantic embeddings for zero-shot learning. In *CVPR*, 2022.

### 28.7.4 Representation Learning

*Investigators: Zeynep Akata, in cooperation with Karsten Roth, Mark Ibrahim (Meta), Pascal Vincent (Meta), Diane Boucancourt (Meta), Abhra Choudhury (University of Surrey), Anjan Dutta (University of Surrey), Ilke Cugu (Microsoft Turkey), Massimiliano Mancini (University of Tübingen), Yanbei Chen (Amazon), Oriol Vinyals (Deepmind), Michael Kirschhof (University of Tübingen), Enkelejda Kasneci (University of Tübingen), Shyamgopal Karthik, Uddeshya Upadhyay*

In representation learning, our aim is to learn robust and generalizable representations of images under various conditions, e.g. under domain shift, distribution shift and in low data regimes. We apply our methods in various applications, e.g. fine-grained classification, compositional zero-shot learning and deep metric learning.

Our research in multimodal learning can be roughly categorized into disentangled representation learning [3, 1], representation learning with uncertainty [6, 5], and data efficient learning [4, 2]. We are describing each of these research directions in the following.

#### Disentangled Representation Learning

*Investigators: Karsten Roth, Mark Ibrahim, Zeynep Akata, Pascal Vincent, Diane Bouchacourt, Abhra Choudhuri, Massimiliano Mancini, Anjan Dutta*

A grand goal in deep learning research is to learn representations capable of generalizing across distribution shifts. Disentanglement is one promising direction aimed at aligning a models representations with the underlying factors generating the data (e.g. color or background). Existing disentanglement methods, however, rely on an often unrealistic assumption – that factors are statistically independent. In reality, factors (like object color and shape) are correlated. To address this limitation, in [3], we propose a relaxed disentanglement criterion – the Hausdorff Factorized Support (HFS) criterion – that encourages a factorized support, rather than a factorial distribution, by minimizing a Hausdorff distance. This allows for arbitrary distributions of the factors over their support, including correlations between them. We show that the use of HFS consistently facilitates disentanglement and recovery of ground-truth factors across a variety of correlation settings and benchmarks, even under severe training correlations and correlation shifts, with in parts over +60% in relative improvement over existing disentanglement methods. In addition, we find that leveraging HFS for representation learning can even facilitate transfer to downstream tasks such as classification under distribution shifts. We hope our original approach and positive empirical results inspire further progress on the open problem of robust generalization.

Fine-grained categories that largely share the same set of parts cannot be discriminated based on part information alone, as they mostly differ in the way the local parts relate to the overall global structure of the object. In [1], We propose Relational Proxies, a novel approach that leverages the relational information between the global and local views of an object for encoding its semantic label. Starting with a rigorous formalization of the notion of distinguishability between fine-grained categories, we prove the necessary and sufficient conditions that a model must satisfy in order to learn the underlying decision boundaries in the fine-grained setting. We design Relational Proxies based on our theoretical findings and

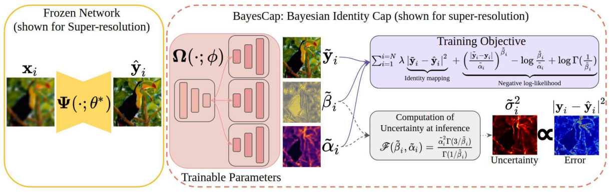



Figure 28.65: BayesCap in tandem with the pretrained network with frozen parameters. While the pretrained network can- not estimate the uncertainty, the proposed BayesCap feeds on the output of the pretrained network and maps it to the underlying probability distribution that allows computation of well calibrated uncertainty estimates.

evaluate it on seven challenging fine-grained benchmark datasets and achieve state-of-the-art results on all of them, surpassing the performance of all existing works with a margin exceeding 4% in some cases. We also experimentally validate our theory on fine-grained distinguishability and obtain consistent results across multiple benchmarks.

### Representation Learning with Uncertainty

*Investigators: Uddeshya Upadhyay, Yanbei Chen, Zeynep Akata, Shyamgopal Karthik, Massimiliano Mancini*

High-quality calibrated uncertainty estimates are crucial for numerous real-world applications, especially for deep learning-based deployed ML systems. While Bayesian deep learning techniques allow uncertainty estimation, training them with large-scale datasets is an expensive process that does not always yield models competitive with non-Bayesian counterparts. Moreover, many of the high-performing deep learning models that are already trained and deployed are non-Bayesian in nature and do not provide uncertainty estimates. To address these issues, in [6], we propose BayesCap (see Figure 28.65) that learns a Bayesian identity mapping for the frozen model, allowing uncertainty estimation. BayesCap is a memory-efficient method that can be trained on a small fraction of the original dataset, enhancing pretrained non-Bayesian computer vision models by providing calibrated uncertainty estimates for the predictions without (i) hampering the performance of the model and (ii) the need for expensive retraining the model from scratch. The proposed method is agnostic to various architectures and tasks. We show the efficacy of our method on a wide variety of tasks with a diverse set of architectures, including image super-resolution, deblurring, inpainting, and crucial application such as medical image translation. Moreover, we apply the derived uncertainty estimates to detect out-of-distribution samples in critical scenarios like depth estimation in autonomous driving.

Unpaired image-to-image translation refers to learning inter-image-domain mapping without corresponding image pairs. Existing methods learn deterministic mappings without explicitly modelling the robustness to outliers or predictive uncertainty, leading to performance degradation when encountering unseen perturbations at test time. To address this, in [5], we propose a novel probabilistic method based on Uncertainty-aware Generalized Adaptive Cycle Consistency (UGAC), which models the per-pixel residual by generalized Gaussian distribution, capable of modelling heavy-tailed distributions. We compare our model with a wide variety of state-of-the-art methods on various challenging tasks including unpaired image translation of natural images, using standard datasets, spanning autonomous driving, maps, facades, and also in medical imaging domain consisting of MRI. Experimental results demonstrate that our method exhibits stronger robustness towards unseen perturbations in test data.

### Data Efficient Learning

*Investigators: Karsten Roth, Oriol Vinyals, Zeynep Akata, Michael Kirchhof, Enkelejda Kasneci*

Deep Metric Learning (DML) aims to learn representation spaces on which semantic relations can simply be expressed through predefined distance metrics. Best performing approaches commonly leverage class proxies as sample stand-ins for better convergence and generalization. However, these proxy-methods solely optimize for sample-proxy distances. Given the inherent non-bijectiveness of used distance functions, this can induce locally isotropic sample distributions, leading to crucial semantic context being missed due to difficulties resolving local structures and intraclass relations between samples. To alleviate this problem, in [4], we propose non-isotropy regularization (NIR) for proxy-based Deep Metric Learning. By leveraging Normalizing Flows, we enforce unique translatability of samples from their respective class proxies. This allows us to explicitly induce a non-isotropic distribution of samples around a proxy to optimize for. In doing so, we equip proxy-based objectives to better learn local structures. Extensive experiments highlight consistent generalization benefits of NIR while achieving competitive and state-of-the-art performance on the standard benchmarks CUB200-2011, Cars196 and Stanford Online Products. In addition, we find the superior convergence properties of proxy-based methods to still be retained or even improved, making NIR very attractive for practical usage.

Proxy-based Deep Metric Learning (DML) learns deep representations by embedding images close to their class representatives (proxies), commonly with respect to the angle between them. However, this disregards the embedding norm, which can carry additional beneficial context such as class- or image-intrinsic uncertainty. In addition, proxy-based DML struggles to learn class-internal structures. To address both issues at once, in [2], we introduce non-isotropic probabilistic proxy-based DML. We model images as directional von Mises-Fisher (vMF) distributions on the hypersphere that can reflect image-intrinsic uncertainties. Further, we derive non-isotropic von Mises-Fisher (nivMF) distributions for class proxies to better represent complex class-specific variances. To measure the proxy-to-image distance between these models, we develop and investigate multiple distribution-to-point and distribution-to-distribution metrics. Each framework choice is motivated by a set of ablational studies, which showcase beneficial properties of our probabilistic approach to proxy-based DML, such as uncertainty-awareness, better-behaved gradients during training, and overall improved generalization performance. The latter is especially reflected in the competitive performance on the standard DML benchmarks, where our approach compares favorably, suggesting that existing proxy-based DML can significantly benefit from a more probabilistic treatment.

## References


- [1] A. Chaudhuri, M. Mancini, Z. Akata, and A. Dutta. Relational proxies: Emergent relationships as fine-grained discriminator. In *Neural Information Processing Systems (NeurIPS)*, 2022.
- [2] M. Kirchhof, K. Roth, Z. Akata, and E. Kasneci. A non-isotropic probabilistic take on proxy-based deep metric learning. In *European Conference on Computer Vision (ECCV)*, 2022.
- [3] K. Roth, M. Ibrahim, Z. Akata, P. Vincent, and D. Bouchacourt. Disentanglement of correlated factors via hausdorff factorized support. In *International Conference on Learning Representations (ICLR)*, 2023.
- [4] K. Roth, O. Vinyals, and Z. Akata. Non-isotropy regularization for proxy-based deep metric learning. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022.
- [5] U. Upadhyay, Y. Chen, and Z. Akata. Robustness via uncertainty-aware cycle consistency. In *Neural Information Processing Systems (NeurIPS)*, 2021.
- [6] U. Upadhyay, S. Karthik, Y. Chen, M. Mancini, and Z. Akata. Bayescap: Bayesian identity cap for calibrated uncertainty in frozen neural networks. In *European Conference on Computer Vision (ECCV)*, 2022.


## 28.8 Robust Visual Learning

*Coordinator: Margret Keuper*

### 28.8.1 Robust Generative Models

*Investigators: Margret Keuper in cooperation with Amrutha Saseendran, Stefan Falkner, and Kathrin Skubch (Bosch Center for Artificial Intelligence)*

Variational Autoencoders (VAEs) are powerful probabilistic models to learn representations of complex data distributions. One important limitation of VAEs is the strong prior assumption that latent representations learned by the model follow a simple uni-modal Gaussian distribution. The susceptibility of Variational Autoencoders (VAEs) to adversarial attacks indicates the necessity to evaluate the robustness of the learned representations along with the generation performance. The vulnerability of VAEs has been attributed to the limitations associated with their variational formulation. Deterministic autoencoders could overcome the practical limitations associated with VAEs and offer a promising alternative for image generation applications. Yet, since DAEs only provide weak control over the learned latent distribution, they require an ex-post density estimation step to generate samples comparable to those of VAEs.

In [1], we propose a simple and end-to-end trainable deterministic autoencoding framework, that efficiently shapes the latent space of the model during training and utilizes the capacity of expressive multi-modal latent distributions. The proposed training procedure is based on the one-point Kolmogorov–Smirnov test and faciliates to efficiently shape the latent space into a given, potentially multi-modal distribution. The proposed regularization objective, that shapes the latent space of a DAE into a Gaussian mixture model, is visualized in 28.66 (left). In [2], we build upon [1] and propose an adversarially robust deterministic autoencoder with superior performance in terms of both generation and robustness of

Figure 28.66: Left: Learned latent representations in a deterministic autoencoder regularized towards a Gaussian mixture model prior with two components (blue shaded regions). Consider a set of latent points $z_1, \ldots, z_N$ (blue dots), in a subspace (green shaded region) within a component and the corresponding adversarial examples $z_1^{\text{adv}}, \ldots, z_N^{\text{adv}}$ (red crosses). The adversarial examples tend to explore regions not covered by the input samples. If we assume that $z$ and $z^{\text{adv}}$ follow the same prior assumptions independently, the adversarial examples tend to move closer to the original samples (blue crosses). In the worst case scenario, an adversarial example might reside in a different component. Right: By establishing a strong coupling via the proposed 2-point Kolmogorov-Smirnov-distance regularization, the adversarial examples tend to move more closer to the original samples (orange crosses) after regularization.

the learned representations. First, we show that the regularized latent space proposed therein is already (slightly) more robust than the non-regularized version. Further, to leverage adversarial training, we introduce a further loss component based on the two-point Kolmogorov–Smirnov test between representations of benign and adversarial data. This regularization objective, that shapes the latent space of a DAE into a Gaussian mixture model, is visualized in 28.66 (right). This facilitates the regularization of adversarially perturbed data points in the training pipeline without increasing the computational complexity or compromising the generation fidelity. We conduct extensive experimental studies on popular image benchmark datasets to quantify the robustness of the proposed approach based on the adversarial attacks targeted at VAEs. Our empirical findings show that the proposed method achieves significant performance in both robustness and fidelity when compared to the robust VAE models.

## References


- [1] A. Saseendran, K. Skubch, S. Falkner, and M. Keuper. Shape your space: A Gaussian mixture regularization approach to deterministic autoencoders. In M. Ranzato, A. Beygelzimer, P. S. Liang, J. W. Vaughan, and Y. Dauphin, eds., *Advances in Neural Information Processing Systems 34 pre-proceedings (NeurIPS 2021)*, Virtual Event, 2021, pp. 7319–7332. Curran Associates, Inc.
- [2] A. Saseendran, K. Skubch, and M. Keuper. Trading off image quality for robustness is not necessary with regularized deterministic autoencoders. In S. Koyejo, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, eds., *Advances in Neural Information Processing Systems 35 (NeurIPS 2022)*, New Orleans, LA, USA, 2022, pp. 26751–26763. Curran Associates, Inc.

### 28.8.2 Image Synthesis

*Investigators: Steffen Jung, Ning Yu, and Margret Keuper, in cooperation with Yang He and Mario Fritz (CISPA Helmholtz Center for Information Security)*

#### Detection of Synthesized Images

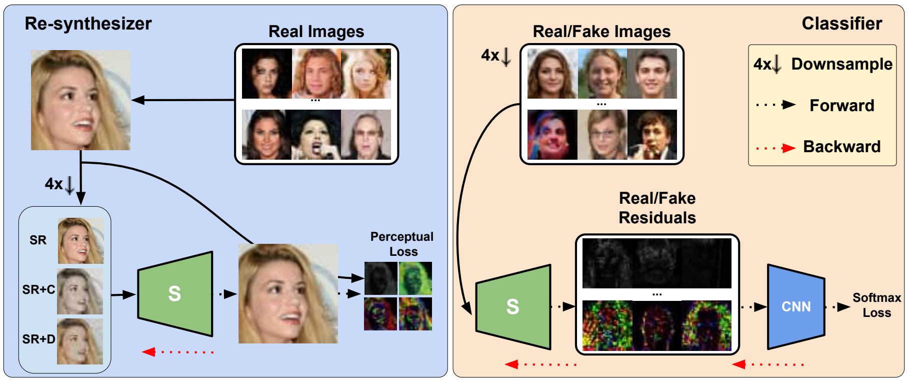



Figure 28.67: The diagram of our detection pipeline. Our end-to-end model has two components. A classifier is trained with real/synthesized images. We learn a re-synthesizer with real images only to help extract robust features and isolate synthesized images. The re-synthesizer takes different forms of inputs to capture various visual patterns from those tasks for robust representations, including super-resolution (SR), colorization (C) and denoising (D).

Advances in image synthesis with generative adversarial networks (GANs) has led to photorealistic results that are hardly distinguishable from real imagery with the human eye. Consequently, the interest in detecting GAN-generated images has increased concurrently. In [3] we show that (i) GAN-generated images containing high frequency artifacts, (ii) several detection methods rely on the presence of those artifacts, and (iii) those artifacts can be successfully removed by incorporating an additional discriminator network acting in the spectral domain. This spectral discriminator maps a given image to a one-dimensional spectral profile via azimuthal integration of its frequency components. Given these profiles, a logistic regression is trained to linearly separate real from synthesized images. By incorporating the output of the logistic regression as an additional training signal to the generator network, spectral artifacts are reduced substantially.

As shown above, relying on spectral artifacts cannot be a sustainable method for detecting synthesized images. Hence, in [1] we propose a detection framework based on extracting visual cues after re-synthesizing testing images. Here, we can add an increasing number of re-synthesization procedures on different visual tasks (e.g. super-resolution, denoising, and colorization). Our detection approach consists of a re-synthesizer and a spatial CNN-based classifier (cf. Figure 28.67). The purpose of the re-synthesizer is to produce features for the classifier in the form of residuals between its input and output images. To further improve the robustness of the classifier, we compare features at different layers of the CNN, hence, at different hierarchical levels of images. The effectiveness of this approach has been validated on high-resolution images synthesized by state-of-the-art GANs.

### Bias in Image Synthesis Metrics

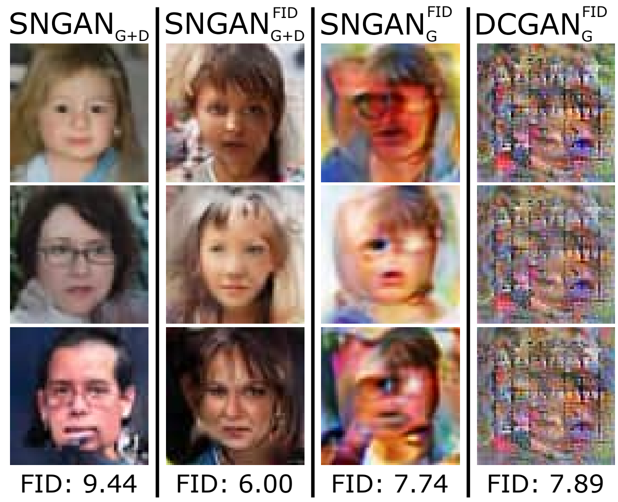



Figure 28.68: We show images generated by different GAN architectures (DCGAN, SNGAN) and training schemes. $G + D$ indicates that the generator $G$ is trained using discriminator $D$. $FID$ indicates that $G$ is trained by minimizing FID as a loss term. Minimizing FID directly exposes that it is not aligned with human judgment.

An important metric that is used to rank the image quality of different image synthesizing models is Fréchet inception distance (FID). FID compares the distribution of features between two image datasets that are estimated from training data and samples from the synthesizing model. In [2] we empirically investigate the biases of FID that are inherited by its underlying design decision of extracting image features using the Inception v3 image classification network. We showed that the way Inception v3 was trained enforced it to become robust to corruptions related to color, intensity, saturation, and horizontal translations, while being sensitive to corruptions of textures and to vertical translations. These underlying variances in sensitivity translate to certain biases towards image corruptions and influence the ranking when comparing generative models. Additionally, we showed that FID as a metric is not aligned with human perception by minimizing FID as a loss term. Here, we showed that generators trained with FID loss produce images with substantially improved FIDs, but worse visual appearance (cf. Figure 28.68). Finally, we showed that substituting Inception v3 with other image classification networks leads to different rankings and different biases towards certain corruptions. Hence, we conclude that simply exchanging the feature extraction backbone in FID can not solve its underlying biases.

## References


- [1] Y. He, N. Yu, M. Keuper, and M. Fritz. Beyond the spectrum: Detecting deepfakes via re-synthesis. In Z.-H. Zhou, ed., *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence (IJCAI 2021)*, Montreal, Canada, 2021, pp. 2534–2541. IJCAI.
- [2] S. Jung and M. Keuper. Internalized biases in Fréchet inception distance. In *NeurIPS 2021 Workshop on Distribution Shifts: Connecting Methods and Applications (NeurIPS 2021 Workshop DistShift)*, Virtual, 2021. OpenReview.net.
- [3] S. Jung and M. Keuper. Spectral distribution aware image generation. In *Thirty-Fifth AAAI Conference on Artificial Intelligence Technical Tracks 2*, Virtual Conference, 2021, pp. 1734–1742. AAAI.


## 28.8.3 Robust Discriminative Models

*Investigators: Margret Keuper in cooperation with Julia Grabinski and Janis Keuper (Offenburg University of Applied Sciences and Fraunhofer ITWM), Paul Gavrikov (Offenburg University of Applied Sciences), Yuxuan Zhou (University of Mannheim), Wangmeng Xiang and Lei Zhang (The Hong Kong Polytechnic University), Chao Li, Biao Wang, Xihan Wei and Xiansheng Hua (Alibaba Research), Yumeng Li, Dan Zhang and Anna Khoreva (Bosch Center for Artificial Intelligence), and Patrick Müller and Alexander Braun (Düsseldorf University of Applied Sciences)*

### Analyzing the Properties of Robust Discriminative Models

*Investigators: Margret Keuper in cooperation with Julia Grabinski and Janis Keuper (Offenburg University of Applied Sciences and Fraunhofer ITWM) and Paul Gavrikov (Offenburg University of Applied Sciences)*

Many commonly well-performing convolutional neural network models have shown to be susceptible to adversarial input data perturbations, indicating a low model robustness. Robustness against attacks can be gained for example by using adversarial examples during training, which effectively reduces the measurable model attackability. Here, we summarize three studies that investigate the properties of adversarially robust models.

In [3], we analyze adversarially trained, robust models in the context of a specifically suspicious network operation, the downsampling layer, and provide evidence that robust models have learned to downsample more accurately and suffer significantly less from aliasing than baseline models.

Then, in [2], we provide an in-depth analysis of the resulting robust models beyond adversarial robustness, specifically with respect to the distribution of their prediction confidences and calibration. We empirically analyze a variety of adversarially trained models that achieve high robust accuracies when facing state-of-the-art attacks and we show that adversarial training has an interesting side-effect: it leads to models that are significantly less overconfident with their decisions even on clean data than non-robust models – while they are not necessarily better calibrated, see Fig. 28.69.

Further, recent work has pointed out that the reasoning in neural networks is different from humans. Humans identify objects by shape, while neural nets mainly employ texture cues. Exemplarily, a model trained on photographs will likely fail to generalize to datasets

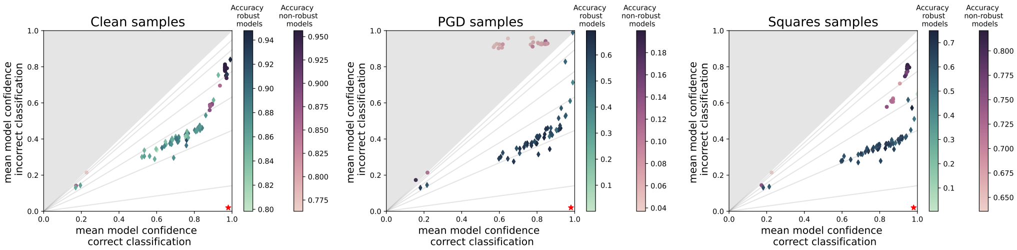



Figure 28.69: Mean model confidences on their correct (x-axis) and incorrect (y-axis) predictions over the full CIFAR10 dataset, clean (left) and perturbed with the attacks PGD (middle) and Squares (right). Each point represents a model. Circular points (purple color-map) represent non-robust models and diamond-shaped points (green color-map) represent robust models. The color of each point represents the models accuracy, darker signifies higher accuracy (better) on the given data samples. The star in the bottom right corner indicates the optimal model calibration and the gray area marks the area were the confidence distribution of the network is worse than random, i.e. more confident in incorrect predictions than in correct ones.

containing sketches. Interestingly, it was also shown that adversarial training seems to favorably increase the shift toward shape bias. In [1], we revisit this observation and provide an extensive analysis of this effect on various architectures, the common $\ell_2$- and $\ell_\infty$-training, and transformer-based models. Further, we provide a possible explanation for this phenomenon from a frequency perspective.

### Models and Model Hardening

*Investigators: Margret Keuper in cooperation with Yuxuan Zhou (University of Mannheim), Wangmeng Xiang and Lei Zhang (The Hong Kong Polytechnic University), Chao Li, Biao Wang, Xihan Wei and Xiansheng Hua (Alibaba Research), Yumeng Li, Dan Zhang and Anna Khoreva (Bosch Center for Artificial Intelligence), and Patrick Müller and Alexander Braun (Düsseldorf University of Applied Sciences)*

Because of their wide practical applicability in safety critical applications, Deep Neural Networks (DNNs) have to behave in a robust way to disturbances such as noise, image corruptions and other distribution shifts. In this subsubsection, we summarize three of our recent works that propose to train well-conditioned discriminative models in different ways. In [7], we propose a variant of self-attention that allows to train vision transformer models which perform well in the low data regime without overfitting, while in [6] and [4], we address two specific challenges for distribution shifts that models are expected to face in real world applications, i.e. optical aberrations and adverse wheather conditions. In the following, we provide more details on these works.

Recently, transformers have shown great potential in image classification and established state-of-the-art results on the ImageNet benchmark. However, compared to Convolutional Neural Networks, transformers converge slowly and are prone to overfitting in low-data regimes due to the lack of spatial inductive biases. Such spatial inductive biases can be especially beneficial since the 2D structure of an input image is not well preserved in transformers. In [7], we present Spatial Prior–enhanced Self-Attention (SP-SA), a variant of vanilla Self-Attention for vision transformers that enables to highlight certain groups of spatial relations. Unlike convolutional inductive biases, which are forced to focus exclusively on hard-coded local regions, our proposed SPs are learned by the model itself and take a variety of spatial relations into account, allowing for improved accuracy for example on ImageNet-1k without extra data, thus showing reduced overfitting behavior.

Blur impacts directly the performance of DNNs, often approximated as disk-shaped kernel to model defocus when evaluating the robustness. However, optics suggests that diverse kernel shapes exist in realistic scenarios, depending on wavelength and location, caused by optical aberrations. As the optical quality of a lens objective decreases, such aberrations increase in practice. In [5], we showed experimentally that such realistic optical effects strongly impact the behavior of exisitng models, even when trained to be robust to other kinds of corruptions. In [6], we propose *OpticsBench*, a benchmark to investigate robustness towards reaslistic, practically relevant optical blur effects. Experiments on ImageNet show for a multitude of different pretrained DNNs that the performance compared to disk-shaped kernels varies strongly, indicating the necessity of considering realistic image degradations. Additionally, we demonstrate on ImageNet-100 with *OpticsAugment* that robustness can be increased using optical kernels as data augmentation.

In [4], we further address the generalization of deep neural network models with respect to domain shifts, as they frequently appear in applications such as autonomous driving. Specifically, we propose an exemplar-based style synthesis pipeline to improve domain generalization in semantic segmentation. Our approach is based on a masked noise encoder for StyleGAN2 inversion. The model learns to faithfully reconstruct the image, preserving its semantic layout through noise prediction. Random masking of the estimated noise enables the style mixing capability of our model, i.e. it allows to alter the global appearance without affecting the semantic layout of an image. Using the proposed masked noise encoder to randomize style and content combinations in the training set, i.e., intra-source style augmentation (ISSA) effectively increases the diversity of training data and reduces spurious correlation. As a result, we achieve significant improvements on driving-scene semantic segmentation under different types of data shifts, i.e., changing geographic locations, adverse weather conditions, and day to night. ISSA is model-agnostic and straightforwardly applicable with CNNs and Transformers.

## References

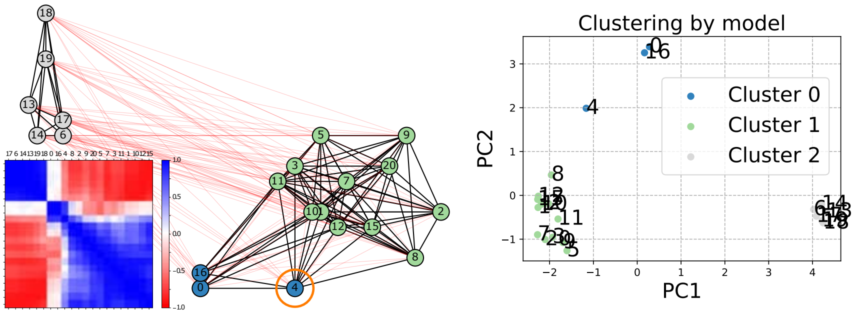

- [1] P. Gavrikov, J. Keuper, and M. Keuper. An extended study of human-like behavior under adversarial training. In *The 3rd Workshop of Adversarial Machine Learning on Computer Vision: Art of Robustness, CVPR workshops*, 2023.

- [2] J. Grabinski, P. Gavrikov, J. Keuper, and M. Keuper. Robust models are less over-confident. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, eds., *Advances in Neural Information Processing Systems 35 (NeurIPS 2022)*, New Orleans, LA, USA, 2022, pp. 39059–39075. Currran Assoicate, Inc.

- [3] J. Grabinski, J. Keuper, and M. Keuper. Aliasing and adversarial robust generalization of CNNs. *Machine Learning*, 111:3925–3951, 2022.
- [4] Y. Li, D. Zhang, M. Keuper, and A. Khoreva. Intra-source style augmentation for improved domain generalization. In *2023 IEEE Winter Conference on Applications of Computer Vision (WACV 2023)*, Waikoloa Village, HI, USA, 2023, pp. 509–519. IEEE.
- [5] P. Müller, A. Braun, and M. Keuper. Impact of realistic properties of the point spread function on classification tasks to reveal a possible distribution shift. In *NeurIPS 2022 Workshop on Distribution Shifts: Connecting Methods and Applications (NeurIPS 2022 Workshop DistShift)*, New Orelans, LA, USA, 2022. OpenReview.net.
- [6] P. Müller, A. Braun, and M. Keuper. Classification robustness to common optical aberrations. In *Under Submission to the International Conference of Computer Vision (ICCV 2023)*, 2023.
- [7] Y. Zhou, W. Xiang, C. Li, B. Wang, X. Wei, L. Zhang, M. Keuper, and X. Hua. SP-ViT: Learning 2D spatial priors for vision transformers. In *33rd British Machine Vision Conference (BMVC 2022)*, London, UK, 2022, Article 564. BMVA Press.


### 28.8.4 Graph Decomposition via Minimum Cost Multicuts

*Investigators: Steffen Jung and Margret Keuper, in cooperation with Amirhossein Kardoost, Sebastian Ziegler (University of Mannheim), Evgeny Levinkov (Bosch Center for Artificial Intelligence), and Bjoern Andres (Technical University Dresden)*




(a) Clustered nodes. (b) Node emb. PCA.

Figure 28.70: (a) Node clustering given by our approach on a complete graph and the ordered cosine similarity between all learned node embeddings. (b) The first two principal components for each node embedding of (a). Node 4 is part of the green cluster in the optimal solution. The closeness of both solutions is reflected in the embedding.

The minimum cost lifted multicut problem is a generalization of the multicut problem (also known as correlation clustering) and is a means to optimizing a decomposition of a graph w.r.t. both positive and negative edge costs. It has been shown to be useful in a large variety of applications in computer vision thanks to the fact that multicut-based formulations do not require the number of components given a priori; instead, it is deduced from the solution. However, the standard multicut cost function is limited to pairwise relationships between nodes, while several important applications either require or can benefit from a higher-order cost function, i.e., hyper-edges. In [3], we propose a pseudo-boolean formulation for a multiple model fitting problem. It is based on a formulation of any-order minimum cost lifted multicuts, which allows to partition an undirected graph with pairwise connectivity such as to minimize costs defined over any set of hyper-edges. As the proposed formulation is NP-hard and the branch-and-bound algorithm (as well as obtaining lower bounds) is too slow in practice, we propose an efficient local search algorithm for inference into resulting problems. We demonstrate versatility and effectiveness of our approach in several applications: 1) We define a geometric multiple model fitting, more specifically, a line fitting problem on all triplets of points and group points that belong to the same line, together. 2) We formulate homography and motion estimation as a geometric model fitting problem where the task is to find groups of points that can be explained by the same geometrical transformation. 3) In motion segmentation our model allows us to go from modeling translational motion to euclidean or affine transformations, which improves the segmentation quality in terms of F-measure.

While former methods, like [3], employ heuristics to solve multicut problems, in [1] we propose a learning-based framework. In this framework, we formulate solving the multicut problem via a binary edge-classification task, classifying edges to be cut or retained. To achieve this, we first reformulate message-passing neural networks in a way that allows processing negatively weighted edges. These networks then provide edge features for an MLP-based classification network (cf. Figure 28.70). Since this formulation can lead to infeasible solutions (removing edges not partitioning the graph), we formulate a soft cycle consistency loss that regularizes the network towards feasibility. Due to the lack of large-scale multicut problem datasets, we generate synthetic training data of different levels of complexity. While only training on synthetic data, we show that (i) our approach is generalizable to real-world datasets, (ii) provides improved scalability over existing heuristical solvers, and (iii) still maintains competitive solution quality.

In [2] we show that the constraints for feasible solutions of minimum cost multicut problems can be employed as additional training signal to improve the performance of learning-based edge detectors. For this, we formulate an adaptive CRF over multicut constraints that progressively considers more violated constraints, leading to edge maps that are less cluttered while accurately localizing the contours. Experiments on the BSDS500 benchmark for natural image segmentation as well as on electron microscopic recordings show that our approach yields more precise edge detection and image segmentation.

## References


- [1] S. Jung and M. Keuper. Learning to solve minimum cost multicuts efficiently using edge-weighted graph convolutional neural networks. In *Machine Learning and Knowledge Discovery in Databases (ECML PKDD 2022)*, Grenoble, France, 2022, Article 486. ecmlpkdd.org.

- [2] S. Jung, S. Ziegler, A. Kardoost, and M. Keuper. Optimizing edge detection for image segmentation with multicut penalties. In B. Andres, F. Bernard, D. Cremers, S. Frintrop, B. Goldlücke, and I. Ihrke, eds., *Pattern Recognition (DAGM GCPR 2022)*, Konstanz, Germany, 2022, LNCS 13485, pp. 182–197. Springer.

- [3] E. Levinkov, A. Kardoost, B. Andres, and M. Keuper. Higher-order multicuts for geometric model fitting and motion segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(1):608–622, 2023.

### 28.8.5 Neural Architecture Search

*Investigators: Jovita Lukasik, Steffen Jung and Margret Keuper, in cooperation with Jonas Geiping and Michael Moeller (University of Siegen)*

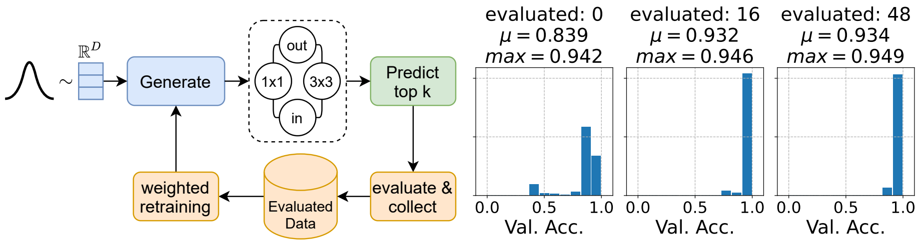



Figure 28.71: (**left**) Our search method generates architectures from points in an architecture representation space that is iteratively optimized. (**right**) The architecture representation space is biased towards better-performing architectures with each search iteration.

Advances in ever-improving neural architectures on computer vision tasks, such as image classification, come mostly with the cost of tedious trial-and-error processes and human expert knowledge for network design. To overcome these costs Neural Architecture Search (NAS) was introduced. NAS is the process of efficiently and automatically searching for well-performing neural architectures, which has drawn increasing attention in the recent past. Thereby, the predominant research objective is to reduce the necessity of costly evaluations of neural architectures while efficiently exploring large search spaces. One widely researched tool for the discovery of novel architectures is the differentiable one-shot architecture search (DAS). The main benefit of DAS is the effectiveness achieved through the weight-sharing one-shot paradigm, which allows efficient architecture search. In [1] we investigate the stability of DAS in a systematic case study of inverse problems, which allows us to analyze the potential benefits of DAS in a controlled manner. In [2] we further extend this analysis to stability and sensitivity to different hyperparameters and search spaces. Although DAS can be extended from image classification to signal reconstruction, in principle, we also expose fundamental difficulties in the evaluation of DAS-based methods, which challenge the common understanding of DAS as a one-shot method. In contrast to one-shot methods, surrogate and generative models as search methods, have the aim of facilitating query-efficient search in a well-structured latent space. Surrogate models embed architectures in a latent space and predict their performance, while generative models for neural architectures enable optimization-based search within the latent space the generator draws from. In [4], we improve the trade-off between query efficiency and promising architecture generation by leveraging advantages from both, efficient surrogate models and generative design. To this end, we propose a generative model, paired with a surrogate predictor, that iteratively learns to generate samples from increasingly promising latent subspaces (cf. Figure 28.71). This approach leads to a very effective and efficient architecture search while keeping the query amount low.

The increasing success of NAS in finding well-performing architectures is recently accompanied by the search for (more) robust architectures against perturbations. Just like the search for well-performing architectures in terms of clean accuracy, this usually involves a tedious trial-and-error process with one additional challenge: the evaluation of a network's robustness is significantly more expensive than its evaluation for clean accuracy. In [3], we facilitate better-streamlined research on architectural design choices w.r.t. their impact on robustness as well as the evaluation of surrogate measures for robustness. We, therefore, borrow one of the most commonly considered search spaces for NAS for image classification, NAS-Bench-201, which contains a manageable size of $6,466$ non-isomorphic network designs. We evaluate all these networks on a range of common adversarial attacks and corruption types and introduce a database on neural architecture design and robustness evaluations. We further present three exemplary use cases of this dataset, in which we (i) benchmark robustness measurements based on Jacobian and Hessian matrices for their robustness predictability, (ii) perform NAS on robust accuracies, and (iii) provide an initial analysis of how architectural design choices affect robustness.

### References


- [1] J. Geiping, J. Lukasik, M. Keuper, and M. Moeller. DARTS for inverse problems: A study on stability. In *NeurIPS 2021 Workshop on Deep Learning and Inverse Problems (NeurIPS 2021 Deep Inverse Workshop)*, Virtual, 2021. OpenReview.net.
- [2] J. Geiping, J. Lukasik, M. Keuper, and M. Moeller. Differentiable architecture search: a one-shot method? In *Under Submission to the AutoML Conference (AutoML23)*, 2023.
- [3] S. Jung, J. Lukasik, and M. Keuper. Neural architecture design and robustness: A dataset. In *Eleventh International Conference on Learning Representations (ICLR 2023)*, Kigali, Rwanda, 2023. OpenReview.net. Accepted.
- [4] J. Lukasik, S. Jung, and M. Keuper. Learning where to look – generative NAS is surprisingly efficient. In S. Avidan, G. Brostow, M. Cissé, G. Farinella, and T. Hassner, eds., *Computer Vision – ECCV 2022*, Tel Aviv, Israel, 2022, LNCS 13683, pp. 257–273. Springer.


## 28.8.6 Robust Architectures

*Investigators: Jovita Lukasik and Steffen Jung and Margret Keuper, in cooperation with Julia Grabinski and Janis Keuper (Offenburg University of Applied Sciences and Fraunhofer ITWM), Paul Gavrikov (Offenburg University of Applied Sciences), and Shashank Agnihotri (University of Siegen)*

Over the last few years, Convolutional Neural Networks (CNNs) have been the dominating neural architecture in a wide range of computer vision tasks. From an image and signal processing point of view, this success might be a bit surprising as the inherent spatial pyramid design of most CNNs is apparently violating basic signal processing laws, i.e., the Sampling Theorem in their down-sampling operations. However, since poor sampling appeared not to affect model accuracy, this issue has been broadly neglected until model

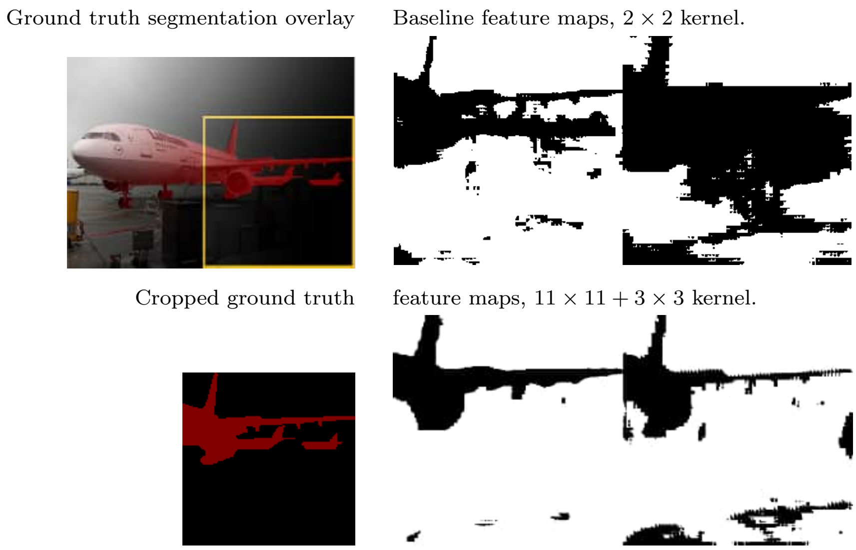



Figure 28.72: Comparing feature maps of the penultimate layer (before the class predictions) from a baseline model ($2 \times 2$ kernel for upsampling, top right) and our model with larger kernels ($11 \times 11$ with parallel $3 \times 3$, bottom right) for transposed convolutions. While the baseline feature maps miss out on many details and show severe grid artifacts, these artifacts are significantly reduced when larger kernels are used.

robustness started to receive more attention. In [2], we introduce an aliasing-free down-sampling operation that can easily be plugged into any CNN architecture, based on the finding of a strong correlation between the vulnerability of CNNs and aliasing artifacts induced by poor down-sampling operations. Our experiments show, that in combination with simple adversarial training, our hyperparameter-free operator substantially improves model robustness and avoids catastrophic overfitting. Adversarial training in general introduces worst-case perturbations into the training data and thus allows for better model generalization to some extent since neural networks tend to overfit the training distribution and perform poorly on out-of-distribution data. However, it is only one ingredient towards generally more robust models. In [3], we focus on the native robustness of CNNs, which can learn robust behavior directly from conventional training data. For that, we investigate the frequencies of learned network filters. Clean-trained models often prioritize high-frequency information, whereas adversarial training enforces models to shift the focus to low-frequency details during training. Based on this observation, we propose a frequency regularization in learned convolution weights. We show two possible realizations based on the discrete cosine transform and demonstrate that our proposed method increases native robustness while remaining compatible with adversarial training and favorable shifts decisions towards low-frequency information such as shapes.

Simultaneously, large convolution kernels have recently been re-gaining popularity as they either match or outperform state-of-the-art transformer-based models. The motivation for this success is two-fold: (i) larger kernels have a larger effective receptive field and thus provide more spatial context, similar to the concept of self-attention in vision transformers, and (ii) large convolution kernels help in reducing down-sampling artifacts in encoder networks, which are known to impair the network's adversarial robustness. However, the study of large kernels has so far been focusing on encoder (e.g. classification) networks. In [1], we hypothesize that both aspects are equally important for image decoding, e.g., in UNet-like models. For example, to predict the label of a pixel in semantic segmentation, the context of the object's shape in other image regions carries important information. Further, grid artifacts during upconvolution can decrease the quality of segmentation masks and make the models easier to attack (cf. Figure 28.72). Thus, we investigate the use of large kernels in normal and transposed convolutions in the CNN decoder. Our experiments show that increasing the kernel size during transposed convolutions significantly improves the adversarial robustness of the resulting models and reduces spectral artifacts during upsampling, across different architectures and tasks such as semantic segmentation or disparity estimation. At the same time, the resulting models' clean accuracy is on par with or better than the respective baselines.

### References


- [1] S. Agnihotri, J. Gabrinski, and M. Keuper. Context matters: the role of upsampling in pixel-wise prediction tasks. In *Under Submission to the International Conference of Computer Vision (ICCV 2023)*, 2023.
- [2] J. Grabinski, S. Jung, J. Keuper, and M. Keuper. FrequencyLowCut pooling – plug & play against catastrophic overfitting. In S. Avidan, G. Brostow, M. Cissé, G. Farinella, and T. Hassner, eds., *Computer Vision – ECCV 2022*, Tel Aviv, Israel, 2022, LNCS 13674, pp. 36–57. Springer.
- [3] J. Lukasik, P. Gavrikov, J. Keuper, and M. Keuper. Filter frequency regularization for improved native cnn robustness. In *Under Submission to the International Conference of Computer Vision (ICCV 2023)*, 2023.


## 28.9 Combinatorial Computer Vision

*Coordinator: Paul Swoboda*

### 28.9.1 Clustering

*Investigators: Ahmed Abbas and Paul Swoboda*

Several dense prediction tasks in computer vision can be modeled as graph clustering problems. One particular formulation is the multicut problem, also known as correlation clustering. We study practical aspects of this problem by proposing efficient algorithms and alleviate the need for manual model selection by enabling synergy with neural networks.

#### Efficient Algorithms

In [2] we propose a highly parallel primal-dual algorithm for the multicut problem (cf. Figure 28.73). Our dual algorithm finds conflicted cycles which correspond to violated inequatlities of the underlying multicut relaxation. Afterwards we perform message passing to

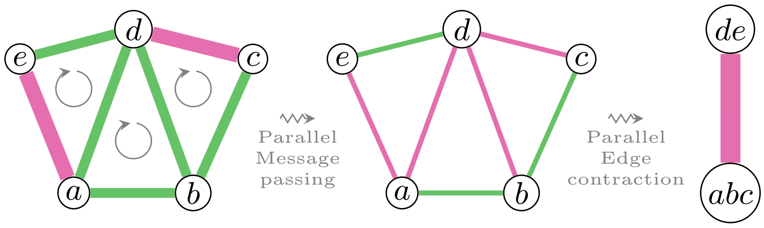



Figure 28.73: Example iteration of primal-dual multicut solver on a graph with attractive (green) and repulsive edges. Width of the edges indicate absolute cost. First we detect conflicted cycles and triangulate to get triangles. Next, parallel dual update reparametrizes edge costs which resolves the conflicted cycles. Lastly a primal update is done in parallel to contract attractive edges.

optimize the Lagrange relaxation producing reduced costs and lower bound on the objective value. Our primal algorithm takes the reduced costs and contracts edges with high reduced costs through matrix-matrix multiplications. We implement our algorithm on GPUs and show resulting one to two orders-of-magnitudes improvements in runtime without sacrificing solution quality compared to traditional sequential algorithms. The resulting algorithm allows to solve very large scale problems arising from scene understanding and biomedical image analysis in a few seconds with small primal-dual gaps.

### Model Selection

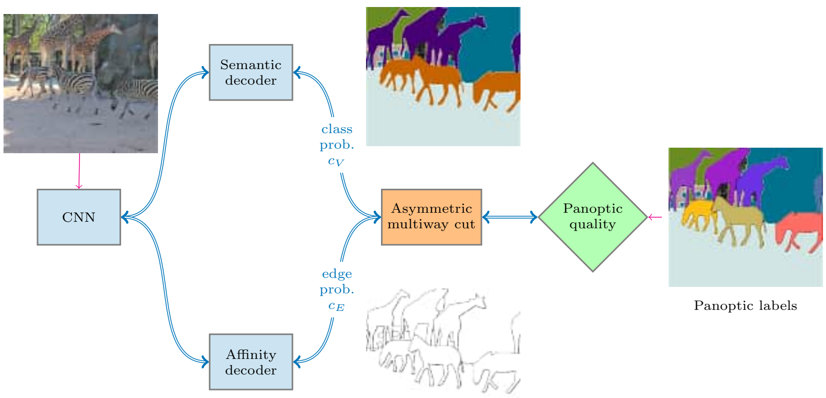



Figure 28.74: End-to-end trainable pipeline for Panoptic segmentation by asymmetric multiway cut (AMWC). The convolutional neural network produces class and affinity scores which are used in the AMWC solver for computing a panoptic labeling. The computation flow marked by arrows indicates forward pass for inference and backward pass for backpropagation.

In [1] we propose an end-to-end trainable architecture for graph clustering. To this end we embed an extension of the multicut problem called asymmetric multiway cut (AMWC) as a layer inside neural networks. At test-time the neural network predicts edge affinities for the optimization problem which produces a clustering. The quality of clustering is evaluated through a non-differentiable metric. We mimic the test-time procedure also during training by enabling backpropagation through the clustering problem and propose a differentiable approximation to the test-time evaluation metric (cf. Figure 28.74). The resulting architecture removes the need for hyperparameter tuning while achieving state-of-the-art results on large scale panoptic segmentation benchmark datasets namely Cityscapes and MSCOCO.

In [3] we address the issue of graph structure design for the multicut problem. We propose a variant of the multicut problem which does not need specification of the graph topology making our approach simpler and potentially better performing. Specifically, we consider the fully connnected graph where the clustering objective is given in a factorized form as inner products of node feature vectors. This allows for an efficient formulation and inference in contrast to multicut clustering, which has at least quadratic representation and computation complexity when working on the complete graph. We propose efficient algorithms for this dense multicut formulation. Empirically our approach produces better clustering than human designed graph structure on large scale instance segmentation benchmark on the Cityscapes dataset. Lastly we achieve up to an order of magnitude speedup over traditional multicut algorithms on the complete graph.

### References


- [1] A. Abbas and P. Swoboda. Combinatorial optimization for panoptic segmentation: A fully differentiable approach. In M. Ranzato, A. Beygelzimer, P. S. Liang, J. W. Vaughan, and Y. Dauphin, eds., *Advances in Neural Information Processing Systems 34 (NeurIPS 2021)*, Virtual, 2021, pp. 15635–15649. Curran Associates, Inc.

- [2] A. Abbas and P. Swoboda. RAMA: A rapid multicut algorithm on GPU. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2022)*, New Orleans, LA, USA, 2022, pp. 8183–8192. IEEE.

- [3] A. Abbas and P. Swoboda. ClusterFuG: Clustering Fully connected Graphs by Multicut. *arXiv preprint arXiv:2301.12159*, 2023.


## 28.9.2 Optimization Algorithms

*Investigators: Ahmed Abbas, Jan-Hendrik Lange and Paul Swoboda*

Structured prediction tasks in machine learning and computer vision are commonly formulated as integer linear programs (ILPs). The most scalable methods to tackle specific problem formulations exploit decompositions of the problem structure in order to solve Lagrangean relaxations. The resulting algorithms are able to outperform ILP solvers that rely on standard linear programming algorithms, such as the simplex and barrier method. However, they are not applicable to instances outside of their dedicated problem class. To this end we study efficient methods for (approximately) solving large scale integer linear programs while still maintaining generality of the approaches.

### Message passing for 0–1 ILPs

In [3] we propose an efficient dual block-coordinate ascent (message passing) method for solving relaxations of 0–1 ILPs. Our approach is based on a representation of each subproblem in the decomposition as a binary decision diagram (BDD). The algorithm then updates dual variables iteratively so as to maximally increase the dual objective in each step. Our method has linear iteration complexity unlike LP solvers, is not restricted to a narrow subclass of problems unlike specialized solvers. We demonstrate the potential of our method on combinatorial problems from MAP inference for Markov Random Fields, quadratic assignment, discrete tomography and cell tracking for developmental biology and show promising performance.

### Parallel Discrete Optimization

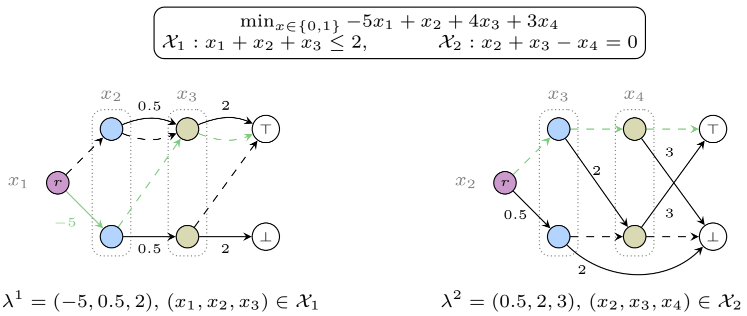



Figure 28.75: Example decomposition of a 0–1 ILP into two subproblems, one for each constraint. Each subproblem is represented by a weighted BDD where solid arcs model the cost $\lambda$ of assigning a 1 to the variable and dashed arcs have 0 cost which model assigning a 0. All $r -$ paths in BDDs encode feasible variable assignments of corresponding subproblems (and $r -$ infeasible). Optimal assignments w.r.t current (non-optimal) $\lambda$ are highlighted in green. Our dual update scheme processes multiple variables in parallel (indicated in same color).

In [2] we present a massively parallel Lagrange decomposition method for solving 0–1 ILPs by extending [3] (cf. Figure 28.75). We propose a new iterative update scheme for solving the Lagrangean dual and a perturbation technique for decoding primal solutions. Our primal and dual algorithms require little synchronization between subproblems and needs only elementary operations without complicated control flow. This allows us to exploit the parallelism offered by GPUs for all components of our method. Our implementation improves upon the runtimes of [3] by up to an order of magnitude. In particular, we come close to or outperform some state-of-the-art specialized heuristics while being problem agnostic.

### Differentiable LP Relaxation

In [1] we further build upon our prior work and propose a fast, scalable, data-driven approach for solving linear relaxations of 0-1 integer linear programs. Specifically, we generalize the algorithm of [2], make it differentiable for end-to-end training, and use a graph neural network to predict its algorithmic parameters. Instead of developing a specialized solver for each problem class, our work allows to train a problem specific solver by making use of example instances without the need for ground-truth collection. Our solver achieves significantly faster performance and better dual objectives than its non-learned version. We also achieve better objective values than specialized approximate solvers for specific problem classes while retaining their efficiency. Our solver is faster by up to an order of magnitude compared to commercial solvers on large scale structured prediction problems.

### References


- [1] A. Abbas and P. Swoboda. DOGE-Train: Discrete Optimization on GPU with End-to-end Training. *arXiv preprint arXiv:2205.11638*, 2022.
- [2] A. Abbas and P. Swoboda. FastDOG: Fast discrete optimization on GPU. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2022)*, New Orleans, LA, USA, 2022, pp. 439–449. IEEE.
- [3] J.-H. Lange and P. Swoboda. Efficient message passing for 0–1 ILPs with binary decision diagrams. In *International Conference on Machine Learning*, 2021, pp. 6000–6010. PMLR.


## 28.9.3 Multiple Object Tracking

*Investigators: Andrea Hornakova and Duy Nguyen and Paul Swoboda*

### Efficient Higher Order Tracking

In [1], we present an efficient approximate message passing solver for the lifted disjoint paths problem (LDP), a natural but NP-hard model for multiple object tracking (MOT).

The solver enables LDP-based tracking for long and crowded sequences of dataset MOT20, which has been out of reach for LDP solvers before due to its size and complexity. Therefore, the work demonstrates that the NP-hard LDP model is applicable for processing massive MOT sequences. The scalability to large scenes is enabled by using approximate solver instead of the globally optimal one on the one hand, and using considerably less computationally demanding features than the tracker in the prior works.

The resulting tracker ApLift (see Figure 28.76 for an illustration) processes also less demanding MOT benchmarks without sacrificing solution quality. It achieved comparable or better results with the state-of-the-art methods (at the time of its publication) on four standard MOT benchmarks (including MOT20).

Being based on on Lagrangean (dual) decomposition, the LDP solver delivers not only an approximate solution but also its gap to the optimum. Therefore, it is a principled approach that is independent of any commercial solver.

### Multi-Camera Tracking

We extend the lifted multi-cut paradigm into the multi-camera setting [2]. This direction is currently drawing attention in the computer vision field due to its superior performance in real-world applications such as video surveillance with crowded scenes or in wide spaces.

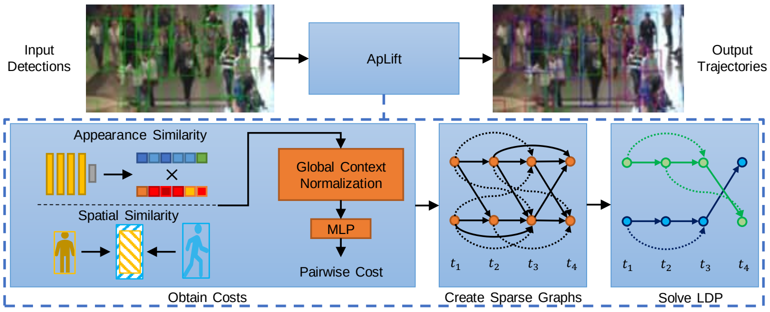



Figure 28.76: Illustration of the tracking pipeline "ApLift" using the lifted disjoint paths problem. First, costs between pairs of detections are computed for guiding the solver's output. There are based on appearance and geometry. Second, a tracking graph is constructed, which is then fed together with the costs to our approximamte LDP solver, which in turn delivers a set of trajectories.

In particular, we propose a mathematically elegant multi-camera multiple object tracking approach based on a spatial-temporal lifted multicut formulation. Our model utilizes state-of-the-art tracklets produced by single-camera trackers as proposals. As these tracklets may contain ID-Switch errors, we refine them through a novel pre-clustering obtained from 3D geometry projections (see Figure 28.77). As a result, we derive a better tracking graph without ID switches and more precise affinity costs for the data association phase. Tracklets are then matched to multi-camera trajectories by solving a globally lifted multicut formulation that incorporates short and long-range temporal interactions on tracklets located in the same camera as well as inter-camera ones. This results in a solution informed by cues from all cameras at once, and the optimal number of trajectories is automatically determined during the optimization. Furthermore, by using nodes in the tracking graph as tracklets, we can accelerate the execution time as the hypothesis space is significantly reduced. Our experimental results on the WildTrack dataset achieve near-perfect performance, outperforming state-of-the-art trackers on Campus while being on par with the PETS-09 dataset.

## References


- [1] A. Horňáková, T. Kaiser, P. Swoboda, M. Rolinek, B. Rosenhahn, and R. Henschel. Making higher order MOT scalable: An efficient approximate solver for lifted disjoint paths. In *ICCV 2021, IEEE/CVF International Conference on Computer Vision*, Virtual Event, 2021, pp. 6310–6320. IEEE.

- [2] D. H. M. Nguyen, R. Henschel, B. Rosenhahn, D. Sonntag, and P. Swoboda. LMGP: Lifted multicut meets geometry projections for multi-camera multi-object tracking. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2022)*, New Orleans, LA, USA, 2022, pp. 8856–8865. IEEE.

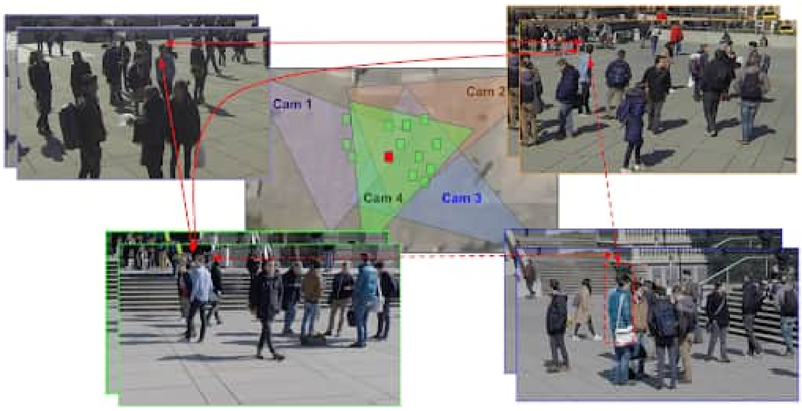



Figure 28.77: Multi-camera tracking with four overlapping cameras. A target object (red rectangle) is occluded at Cam 3 but is still observed at Cam 1, Cam 2, and Cam 4. Taking this correspondence into account (red arrow), we can recover a missing bounding box at Cam 3 (red dashed arrow).

### 28.9.4 Correspondence

*Investigator: Paul Swoboda*

Correspodence problems in computer vision include shape matching, semantic matching and other problems where two or more objects must be brought into alignment to each other. This class of problems has two stages typically: (i) Computing similarity features and (ii) a subsequent matching stage where a one-to-one correspondence between objects is sought. The second stage poses difficulties that can be productively addressed with optimization.

#### Shape Matching

In [2] we propose a scalable combinatorial solver for shape matching, that is matching a triangle mesh so that geometrical consistency is maintained. As input we seek to find matchings between complete shapes but also between sets of partial shapes, which additionally poses the difficulty of estimating which triangles need to be matched. For an illustration of our problem setting see Figure 28.78.

Figure 28.78: Orientation preserving shape matchings on triangle meshed. Matchings between complete shapes (left) and also matches between partial shapes (partial-to-partial shape matching) can be computed. Colors indicate matched triangles.

Our algorithm is based on the primal-dual paradigm, where iteratively better dual indicators of triangle matches are computed and then the most probable matchings are selected, propagation of forced triangle matches are efficiently computed. This procedure is iterated until (i) a final matching is obtained or (ii) conflicts occur and backtracking needs to be done. Our approach has resulted in a fast and scalable solver that could compute in moderate time matchings of triangle meshed that were out of reach for previous solvers.

### Quadratic Assignment Problem/Graph Matching Solvers

The quadratic assignment problem, also called graph matching in the machine learning community, is a classical optimization problem that is notorious for its hardness in the theoretical computer science community. It has found applications most prominently in semantic or keypoint matching, see Figure 28.79. In computer vision a demand for very fast and scalable solvers har arisen that compute good solutions on the type of matching tasks encountered in semantic matching pipelines. This has led to a large number of algorithms of varying quality in the machine learning literature. In order to empirically assess the performance of these algorithms we have undertaken a large comparative study of the most promising algorithms on a wide number of quadratic assignment problems arising mainly from problems in computer vision in our work [1]. This project is designed as an ongoing benchmark for which an accompanying website and new submissions are possible so that future algorithmic developments will be reflected.

### References


- [1] S. Haller, L. Feineis, L. Hutschenreiter, F. Bernard, C. Rother, D. Kainmüller, P. Swoboda, and B. Savchynskyy. A comparative study of graph matching algorithms in computer vision. In S. Avidan, G. Brostow, M. Cissé, G. Farinella, and T. Hassner, eds., *Computer Vision – ECCV 2022*, Tel Aviv, Israel, 2022, LNCS 13683, pp. 636–653. Springer.
- [2] P. Roetzer, P. Swoboda, D. Cremers, and F. Bernard. A scalable combinatorial solver for elastic geometrically consistent 3D shape matching. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2022)*, New Orleans, LA, USA, 2022, pp. 428–438. IEEE.


## 28.9.5 Representation Learning

*Investigators: Paul Swoboda in cooperation with Duy M. H. Nguyen (University of Stuttgart), and Pengtao Xie (University of California San Diego), and Nhat Ho (University of Texas at Austin), and Shadi Albarqouni (University of Bonn), and Daniel Sonntag (DFKI)*

Unsupervised representation learning aims to derive meaningful features from unlabeled data without using any explicit supervision by discovering the underlying structures of data. The resulting pre-trained models can then be utilized for various downstream tasks such as image classification, object detection, or semantic segmentation. There are several approaches have been recently explored, including contrastive learning, generative models, or predictive learning. Our primary focus is on methods that capture objects' internal structures and long-range dependencies, which have important applications in bio-medical domains.

Figure 28.79: Keypoint matching illustration. Given a set of keypoints in images, one-to-one correspondences between those are sought. The underlying combinatorial problem takes into account the geometry of matchings, making the problem computationally demanding.

#### Joint Self-Supervised Image-Volume Representation Learning with Intra-Inter Contrastive Clustering

In [1], we deal with heterogeneous unlabeled data that consists of both 2D images and 3D volumes in the medical field. On this, most prior works have been designed for only one of these data types. As a result, they suffer from major limitations. First, this restricts the capability to fully leverage unlabeled data from numerous sources, which commonly comprise both 2D and 3D data types. Additionally, the use of these pre-trained networks is constrained to downstream tasks with compatible data dimensions. For instance, pretrained 3D-CNN cannot 2D handle object detection and similarly pre-trained 2D-CNN might not be usable for 3D classification tasks (Figure 28.80).

To overcome these challenges, we propose a novel framework for unsupervised joint learning on 2D and 3D data modalities. Given a set of 2D images or 2D slices extracted from 3D volumes, we construct a self-supervised learning (SSL) task through a 2D contrastive clustering problem. The 3D volumes are exploited by computing vectorized embedding at each slice and then assembling a holistic feature through deformable self-attention mechanisms in Transformer, allowing incorporation of the long-range dependencies between slices inside 3D volumes. These holistic features are further utilized to define a novel 3D clustering agreement-based SSL task and masking embedding prediction inspired by pre-trained language models.

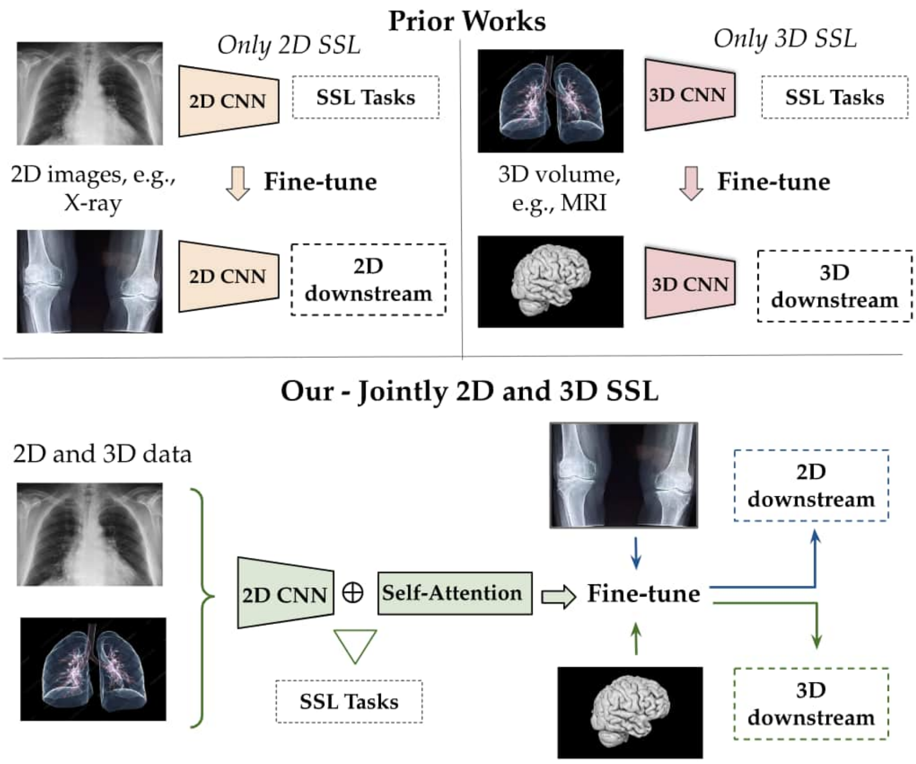



Figure 28.80: The main distinctions between our work and prior studies on 2D and 3D SSL. We can learn representations from diverse data and the pre-trained weights can be trans- ferred for both 2D and 3D downstream tasks.

Experiments on downstream tasks, such as 3D brain segmentation, lung nodule detection, 3D heart structures segmentation, and abnormal chest X-ray detection, demonstrate the effectiveness of our joint 2D and 3D SSL approach. We improve plain 2D Deep-ClusterV2 and SwAV by a significant margin and also surpass various modern 2D and 3D SSL approaches.

### References

• [1] D. M. H. Nguyen, H. Nguyen, M. T. N. Truong, T. Cao, B. T. Nguyen, N. Ho, P. Swoboda, S. Albarqouni, P. Xie, and D. Sonntag. Joint self-supervised image-volume representation learning with intra-inter contrastive clustering. In *Proceedings of the 37th AAAI Conference on Artificial Intelligence*, Washington, DC, USA, 2023. AAAI. Accepted.

## 28.10 Academic Activities

### 28.10.1 Journal Positions

*Bernt Schiele*

- *IEEE Transactions on Pattern Analysis and Machine Intelligence* (Associate Editor in Chief)
- *IEEE Pervasive Computing* (Associate Editor)
- *International Journal of Pervasive Computing and Communications* (Associate Editor)
- *Springer Series Advances in Computer Vision and Pattern Recognition* (Advisory Board Member)

*Margret Keuper*

- *IEEE Transactions on Pattern Analysis and Machine Intelligence* (Associate Editor)

*Dengxin Dai*

- *International Journal of Computer Vision* (Associate Editor)

*Zeynep Akata*

- *IEEE Transactions on Pattern Analysis and Machine Intelligence* (Associate Editor)

*Gerard Pons-Moll*

- *IEEE Transactions on Pattern Analysis and Machine Intelligence* (Associate Editor)

*Jan Eric Lenssen*

- *IEEE Transactions on Pattern Analysis and Machine Intelligence* (Reviewer)

*Shaoshuai Shi*

- *IEEE Transactions on Pattern Analysis and Machine Intelligence* (Reviewer)
- *International Journal of Computer Vision* (Reviewer)
- *IEEE Transactions on Image Processing* (Reviewer)

- *IEEE Transactions on Circuits and Systems for Video Technology* (Reviewer)
- *IEEE Transactions on Robotics* (Reviewer)
- *IEEE Transactions on Multimedia* (Reviewer)
- *IEEE Transactions on Intelligent Transportation Systems* (Reviewer)
- *IEEE Robotics and Automation Letters* (Reviewer)

*Yaoayao Liu*

- *IEEE Transactions on Pattern Analysis and Machine Intelligence* (Reviewer)
- *International Journal of Computer Vision* (Reviewer)
- *IEEE Transactions on Image Processing* (Reviewer)
- *IEEE Transactions on Circuits and Systems for Video Technology* (Reviewer)
- *IEEE Transactions on Multimedia* (Reviewer)
- *IEEE Transactions on Knowledge and Data Engineering* (Reviewer)

*Li Jiang*

- *IEEE Transactions on Pattern Analysis and Machine Intelligence* (Reviewer)
- *International Journal of Computer Vision* (Reviewer)
- *IEEE Transactions on Image Processing* (Reviewer)

*Anna Kukleva*

- *IEEE Transactions on Pattern Analysis and Machine Intelligence* (Reviewer)
- *International Journal of Computer Vision* (Reviewer)
- *IEEE Transactions on Multimedia* (Reviewer)

*Moritz Böhle*

- *IEEE Transactions on Pattern Analysis and Machine Intelligence* (Reviewer)
- *IEEE Transactions on Information Forensics and Security* (Reviewer)

*Nina Shvetsova*

- *IEEE Transactions on Pattern Analysis and Machine Intelligence* (Reviewer)
- *IEEE Transactions on Medical Imaging* (Reviewer)

*Xudong Hong*

- *IEEE Transactions on Pattern Analysis and Machine Intelligence* (Reviewer)
- *Neurocomputing* (Reviewer)

*Haoran Wang*

- *IEEE Transactions on Pattern Analysis and Machine Intelligence* (Reviewer)

*Stephan Alaniz*

- *IEEE Transactions on Pattern Analysis and Machine Intelligence* (Reviewer)

*David Stutz*

- *International Journal of Computer Vision* (Reviewer)
- *IEEE Transactions on Information Forensics and Security* (Reviewer)

*Xianghui Xie*

- *IEEE Transactions on Multimedia*, (Reviewer)

### 28.10.2 Conference and Workshop Positions

#### Membership in program and organization committees

*Bernt Schiele*

- Area Chair of *International Conference on Machine Learning, (ICML)*, 2021
- Area Chair of *AAAI Conference on Artificial Intelligence, (AAAI)*, 2021
- Area Chair of *IEEE/CVF Computer Vision and Pattern Recognition Conference (CVPR)*, 2022
- Area Chair of *European Conference of Computer Vision (ECCV)*, 2022
- Area Chair of *International Conference on Machine Learning, (ICML)*, 2022
- Area Chair of *IEEE/CVF Computer Vision and Pattern Recognition Conference (CVPR)*, 2023
- Area Chair of *IEEE International Conference on Computer Vision (ICCV)* 2023
- Program Committee Member / Reviewer for all major conferences in computer vision and machine learning

*Dengxin Dai*

- Area Chair of the *IEEE/CVF Computer Vision and Pattern Recognition Conference (CVPR)*, 2022
- Area Chair of the *European Conference on Computer Vision (ECCV)*, 2022
- Area Chair of the *IEEE International Conference on Robotics and Automation (ICRA)*, 2022
- Main organizer of the workshop “DeepMTL: Multi-Task Learning in Computer Vision” at ICCV 2021
- Lead organizer of the workshop “Vision for All Seasons” at CVPR 2022
- Lead organizer of the workshop “Vision for All Seasons” at CVPR 2023
- Main organizer of the workshop “Learning with Synthetic Data” at CVPR 2023

*Margret Keuper*

- Reviewer at the *Conference on Neural Information Processing Systems (NeurIPS)*, 2022

- Reviewer at the *International Conference on Machine Learning (ICML)*, 2022
- Reviewer at the *International Joint Conference on Artificial Intelligence and the 25th European Conference on Artificial Intelligence (IJCAI-ECAI)*, 2022
- Reviewer at the *Winter Conference on Applications of Computer Vision (WACV)*, 2022
- Reviewer at the *International Conference on Machine Learning (ICML)*, 2023
- Reviewer at the *International Joint Conferences on Artificial Intelligence (IJCAI)*, 2023
- Senior Program Committee member of the *Association for the Advancement of Artificial Intelligence Conference (AAAI)*, 2022
- Area Chair of the *IEEE/CVF Computer Vision and Pattern Recognition Conference (CVPR)*, 2022
- Area Chair of the *International Conference of Pattern Recognition (ICPR)*, 2022
- Area Chair of the *European Conference on Computer Vision (ECCV)*, 2022
- Area Chair of the *British Machine Vision Conference (BMVC)*, 2022
- Area Chair of the *International Conference on Learning Representations*, 2023
- Area Chair of the *International Conference on Computer Vision (ICCV)*, 2023
- Area Chair of the *IEEE/CVF Computer Vision and Pattern Recognition Conference (CVPR)*, 2023
- Area Chair of the *Conference on Neural Information Processing Systems (NeurIPS)*, 2023

*Zeynep Akata*

- Area Chair of *International Conference on Machine Learning, (ICML)*, 2021
- Area Chair of *International Conference on Machine Learning, (ICML)*, 2022
- Area Chair of *International Conference on Learning Representations, (ICLR)*, 2021
- Area Chair of *International Conference on Learning Representations, (ICLR)*, 2022
- Area Chair of *Neural Information Processing Systems (NeurIPS))*, 2021
- Area Chair of *Neural Information Processing Systems (NeurIPS))*, 2022

*Gerard Pons-Moll*

- Program Chair of the *International Conference on 3D Vision (3DV)*, 2021
- Area Chair of the *IEEE/CVF Computer Vision and Pattern Recognition Conference (CVPR)*, 2022
- Area Chair of the *European Conference on Computer Vision (ECCV)*, 2022
- Area Chair of the *IEEE/CVF Computer Vision and Pattern Recognition Conference (CVPR)*, 2023
- Area Chair of the *International Conference on Computer Vision (ICCV)*, 2023
- Technical Committee of the *SIGGRAPH*, 2022

*Paul Swoboda*

- Area Chair of *International Conference on Machine Learning, (ICML)*, 2022,
- Area Chair of *Neural Information Processing Systems (NeurIPS)*, 2021,
- Area Chair of *Neural Information Processing Systems (NeurIPS)*, 2022,
- Program Committee member of *International Conference on Computer Vision (ICCV)* 2023

*Jan Eric Lenssen*

- Reviewer at *IEEE/CVF Computer Vision and Pattern Recognition Conference (CVPR)*, 2023
- Reviewer at *International Conference on Learning Representations (ICLR)*, 2021, 2023
- Reviewer at *European Conference on Computer Vision (ECCV)*, 2022,
- Reviewer at *Neural Information Processing Systems (NeurIPS)*, 2021, 2022
- Reviewer at *International Conference on 3D Vision (3DV)*, 2022,
- Reviewer at *International Conference on Computer Vision (ICCV)*, 2023

*Paul Swoboda*

- Area Chair of *International Conference on Machine Learning, (ICML)*, 2022,
- Area Chair of *Neural Information Processing Systems (NeurIPS)*, 2021,
- Area Chair of *Neural Information Processing Systems (NeurIPS)*, 2022,
- Program Committee member of *International Conference on Computer Vision (ICCV)* 2023

*Jiangxin Dong*

- Program Committee member of *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 2021, 2022, 2023
- Program Committee member of *International Conference on Computer Vision (ICCV)* 2021, 2023
- Program Committee member of *European Conference of Computer Vision (ECCV)* 2022
- Program Committee member of *Neural Information Processing Systems (NeurIPS)* 2021, 2022, 2023
- Program Committee member of *International Conference on Lerning Representations (ICLR)* 2022

*Li Jiang*

- Reviewer at *International Conference on Computer Vision (ICCV)*, 2021
- Reviewer at *International Conference on Learning Representations (ICLR)*, 2022, 2023
- Reviewer at *IEEE/CVF Computer Vision and Pattern Recognition Conference (CVPR)*, 2022, 2023

- Reviewer at *European Conference on Computer Vision (ECCV)*, 2022
- Reviewer at *Neural Information Processing Systems (NeurIPS)*, 2022
- Reviewer at *ACM SIGGRAPH*, 2023

*Shaoshuai Shi*

- Reviewer at *International Conference on Computer Vision (ICCV)*, 2021, 2023
- Reviewer at *IEEE/CVF Computer Vision and Pattern Recognition Conference (CVPR)*, 2022, 2023
- Reviewer at *European Conference on Computer Vision (ECCV)*, 2022
- Reviewer at *Neural Information Processing Systems (NeurIPS)*, 2022
- Reviewer at *IEEE International Conference on Robotics and Automation (ICRA)*, 2023

*Yaoyao Liu*

- Reviewer at *International Conference on Computer Vision (ICCV)*, 2021
- Reviewer at *IEEE/CVF Computer Vision and Pattern Recognition Conference (CVPR)*, 2022, 2023
- Reviewer at *International Conference on Learning Representations (ICLR)*, 2022, 2023
- Reviewer at *European Conference on Computer Vision (ECCV)*, 2022
- Reviewer at *Neural Information Processing Systems (NeurIPS)*, 2022, 2023
- Reviewer at *International Conference on Machine Learning (ICML)*, 2022, 2023

*David Stutz*

- Reviewer at *International Conference on Computer Vision (ICCV)*, 2021
- Reviewer at *IEEE/CVF Computer Vision and Pattern Recognition Conference (CVPR)*, 2021, 2022
- Reviewer at *International Conference on Learning Representations (ICLR)*, 2022
- Reviewer at *Neural Information Processing Systems (NeurIPS)*, 2021, 2022
- Reviewer at *International Conference on Machine Learning (ICML)*, 2023
- Reviewer at *Association for the Advancement of Artificial Intelligence Conference (AAAI)*, 2021
- Reviewer at *International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2023
- Reviewer at *Conference on Machine Learning and Systems (MLSys)*, 2023

*Bharat Lal Bhatnagar*

- Reviewer at *International Conference on Computer Vision (ICCV)*, 2021, 2023
- Reviewer at *IEEE/CVF Computer Vision and Pattern Recognition Conference (CVPR)*, 2022, 2023
- Reviewer at *European Conference on Computer Vision (ECCV)*, 2022

- Reviewer at *Neural Information Processing Systems (NeurIPS)*, 2021
- Reviewer at *International Conference on 3D Vision (3DV)*, 2021
- Reviewer at *Winter Conference on Applications of Computer Vision (WACV)*, 2021
- Reviewer at *International Conference on Learning Representations (ICLR)*, 2023

*Jovita Lukasik*

- Reviewer at *International Conference on Automated Machine Learning (AutoML)*, 2022, 2023
- Reviewer at *European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD)*, 2022
- Reviewer at *European Conference on Computer Vision (ECCV)*, 2022
- Reviewer at *Winter Conference on Applications of Computer Vision (WACV)*, 2023
- Reviewer at *IEEE/CVF Computer Vision and Pattern Recognition Conference (CVPR)*, 2023
- Reviewer at *International Conference on Computer Vision (ICCV)*, 2023

*Moritz Böhle*

- Reviewer at *IEEE/CVF Computer Vision and Pattern Recognition Conference (CVPR)*, 2023
- Reviewer at *European Conference on Computer Vision (ECCV)*, 2022
- Reviewer at *International Conference on Learning Representations (ICLR)*, 2022, 2023
- Reviewer at *Neural Information Processing Systems (NeurIPS)*, 2022
- Reviewer at *International Conference on Machine Learning (ICML)*, 2023

*Garvita Tiwari*

- Reviewer at *International Conference on Computer Vision (ICCV)*, 2021, 2023
- Reviewer at *ACM SIGGRAPH*, 2021
- Reviewer at *EuroGraphics*, 2021
- Reviewer at *International Conference on 3D Vision (3DV)*, 2021
- Reviewer at *European Conference on Computer Vision (ECCV)*, 2022
- Reviewer at *IEEE/CVF Computer Vision and Pattern Recognition Conference (CVPR)*, 2022, 2023

*Julian Chibane*

- Reviewer at *International Conference on Computer Vision (ICCV)*, 2021, 2023
- Reviewer at *EuroGraphics*, 2021
- Reviewer at *International Conference on 3D Vision (3DV)*, 2022
- Reviewer at *European Conference on Computer Vision (ECCV)*, 2022
- Reviewer at *Neural Information Processing Systems (NeurIPS)*, 2022, 2023

- Reviewer at *IEEE/CVF Computer Vision and Pattern Recognition Conference (CVPR)*, 2022, 2023

*Xudong Hong*

- Reviewer at *CCF International Conference on Natural Language Processing and Chinese Computing (NLPCC)*, 2021
- Reviewer at *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2021
- Reviewer at *Workshop on Computational Approaches to Discourse (CODI@EMNLP)*, 2021
- Reviewer at *Association for Computational Linguistics: ACL Rolling Review (ACL ARR)*, 2022
- Reviewer at *International Conference on Computational Linguistics (COLING)*, 2022
- Reviewer at *Winter Conference on Applications of Computer Vision (WACV)*, 2022
- Reviewer at *CLASP Conference on (Dis)embodiment*, 2022

*Anna Kukleva*

- Reviewer at *International Conference on Computer Vision (ICCV)*, 2021, 2023
- Reviewer at *IEEE/CVF Computer Vision and Pattern Recognition Conference (CVPR)*, 2022, 2023
- Reviewer at *European Conference on Computer Vision (ECCV)*, 2022
- Reviewer at *Winter Conference on Applications of Computer Vision (WACV)*, 2022
- Reviewer at *Neural Information Processing Systems (NeurIPS)*, 2023

*Stephan Alaniz*

- Reviewer at *IEEE/CVF Computer Vision and Pattern Recognition Conference (CVPR)*, 2022, 2023
- Reviewer at *International Conference on Computer Vision (ICCV)*, 2021, 2023
- Reviewer at *European Conference on Computer Vision (ECCV)*, 2022
- Reviewer at *Neural Information Processing Systems (NeurIPS)*, 2023
- Reviewer at *British Machine Vision Conference (BMVC)*, 2022

*Mattia Segu*

- Reviewer at *IEEE International Conference on Robotics and Automation (ICRA)*, 2022,
- Reviewer at *IEEE/CVF Computer Vision and Pattern Recognition Conference (CVPR)*, 2023
- Reviewer at *International Conference on Computer Vision (ICCV)*, 2023
- Reviewer at *International Journal of Computer Vision (IJCV)*, 2023
- Reviewer at *Neural Information Processing Systems (NeurIPS)*, 2023

– Reviewer at *IEEE Access Journal*, 2023

*Steffen Jung*

– Reviewer at *European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD)*, 2022
– Reviewer at *European Conference on Computer Vision (ECCV)*, 2022
– Reviewer at *Neural Information Processing Systems (NeurIPS)*, 2022
– Reviewer at *International Conference on Machine Learning (ICML)*, 2022

*Aymen Mir*

– Reviewer at *International Conference on Computer Vision (ICCV)*, 2021
– Reviewer at *International Conference on 3D Vision (3DV)*, 2021
– Reviewer at *IEEE/CVF Computer Vision and Pattern Recognition Conference (CVPR)*, 2022, 2023
– Reviewer at *European Conference on Computer Vision (ECCV)*, 2022

*Keyang Zhou*

– Reviewer at *International Conference on Computer Vision (ICCV)*, 2021
– Reviewer at *International Conference on 3D Vision (3DV)*, 2021
– Reviewer at *IEEE/CVF Computer Vision and Pattern Recognition Conference (CVPR)*, 2022, 2023
– Reviewer at *European Conference on Computer Vision (ECCV)*, 2022

*Xianghui Xie*

– Reviewer at *IEEE/CVF Computer Vision and Pattern Recognition Conference (CVPR)*, 2022, 2023
– Reviewer at *European Conference on Computer Vision (ECCV)*, 2022
– Reviewer at *International Conference on 3D Vision (3DV)*, 2022
– Reviewer at *International Conference on Computer Vision (ICCV)*, 2023

*Verica Lazova*

– Reviewer at *IEEE/CVF Computer Vision and Pattern Recognition Conference (CVPR)*, 2022
– Reviewer at *International Conference on 3D Vision (3DV)*, 2021, 2022
– Reviewer at *ACM SIGGRAPH*, 2022
– Reviewer at *International Conference on Computer Vision (ICCV)*, 2021, 2023

*Vladimir Guzov*

– Reviewer at *International Conference on Computer Vision (ICCV)*, 2021

- Reviewer at *IEEE/CVF Computer Vision and Pattern Recognition Conference (CVPR)*, 2022, 2023
- Reviewer at *European Conference on Computer Vision (ECCV)*, 2022

*Yue Fan*

- Reviewer at *European Conference on Computer Vision (ECCV)*, 2022
- Reviewer at *Neural Information Processing Systems (NeurIPS)*, 2022
- Reviewer at *IEEE/CVF Computer Vision and Pattern Recognition Conference (CVPR)*, 2023

*Siddhartha Gairola*

- Reviewer at *Neural Information Processing Systems (NeurIPS)*, 2022
- Reviewer at *International Conference on Learning Representations (ICLR)*, 2022
- Reviewer at *International Conference on Machine Learning (ICML)*, 2023

*Xiaohan Zhang*

- Reviewer at *European Conference on Computer Vision (ECCV)*, 2022
- Reviewer at *Neural Information Processing Systems (NeurIPS)*, 2023

*Sukrut Rao*

- Reviewer at *IEEE/CVF Computer Vision and Pattern Recognition Conference (CVPR)*, 2023
- Reviewer at *International Conference on Computer Vision (ICCV)*, 2023

*Nina Shvetsova*

- Reviewer at *IEEE/CVF Computer Vision and Pattern Recognition Conference (CVPR)*, 2022, 2023
- Reviewer at *Neural Information Processing Systems (NeurIPS)*, 2022

*Haoran Wang*

- Reviewer at *IEEE/CVF Computer Vision and Pattern Recognition Conference (CVPR)*, 2022, 2023
- Reviewer at *European Conference on Computer Vision (ECCV)*, 2022

*Xinting Hu*

- Reviewer at *IEEE/CVF Computer Vision and Pattern Recognition Conference (CVPR)*, 2022, 2023
- Reviewer at *European Conference on Computer Vision (ECCV)*, 2022

*Ahmed Abbas*

- Reviewer at *International Conference on Machine Learning (ICML)*, 2022

– Reviewer at *Neural Information Processing Systems (NeurIPS)*, 2023

*Max Losch*

– Reviewer at *IEEE/CVF Computer Vision and Pattern Recognition Conference (CVPR)*, 2022, 2023

#### Membership in steering and other committees

*Bernt Schiele*

– Steering Committee Member of *IEEE International Symposium on Wearable Computing* (since 2017)
– Steering Committee Member of *International Conference on Computer Vision Systems* (since 2001)

### 28.10.3 Invited Talks and Tutorials

*Bernt Schiele*

– Keynote Speech at ACM Multimedia Asia 2021, March 2021 (remote)
– Keynote Speech at Helmholtz AI virtual conference 2021, April 2021 (remote)
– Invited talk at Qualcomm Innovation Fellowship (QIF) Europe 2021 finalists event, May 2021 (remote)
– Invited Talk at TUM AI Lecture Series, June 2021 (remote)
– Invited Talk at Workshop on Safe Artificial Intelligence for Automated Driving (SAIAD) at CVPR 2021, virtual, June 2021 (remote)
– Keynote Speech at Workshop on Long-Tailed Distribution Learning (LTDL) at IJCAI 2021, virtual, August 2021 (remote)
– Keynote Speech at Workshop on Artificial Intelligence with Biased or Scarce Data (AIBSD 2022) at AAAI 2022, virtual, February 2022 (remote)
– Keynote Speech at Kick-off event of the ELLIS unit Stuttgart, Germany, July 2022
– Invited Talk at Explainability in Machine Learning Workshop at the University of Tübingen, March 2023
– Invited Talk at Google Research in Zurich, March 2023

*Dengxin Dai*

– Keynote Speech at the ACM Computer Science in Cars Symposium, December 2022, Ingolstadt
– Keynote Speech at the "3D Perception for Autonomous Driving" workshop at ECCV, October 2022, Aviv, Israel (remote)
– Award Talk at the German Conference on Pattern Recognition, September 2022, Konstanz

- Keynote Speech at the workshop "Beyond supervised learning: addressing data scarcity in intelligent transportation systems" at the Intelligent Vehicles Symposium, July 2022, (remote)
- Invited Talk at the University of Zurich, February 2023

*Zeynep Akata*

- Keynote Talk at ICML Tutorial: Natural XAI, July 2021 (remote)
- Keynote Talk at RWTH AI Colloquium, December 2021 (remote)
- Keynote Talk at SNU Shannon meets Turing Colloquium, December 2021 (remote)
- Keynote Talk at Helmholtz AI Conference, Hamburg, Germany, June 2023
- Invited Talk at ELLIS Unit Cambridge Seminar Series, December 2022 (remote)
- Invited Talk at 30th Year Anniversary of MPI for Informatics, Saarbrücken, November 2023

*Paul Swoboda*

- Invited Talk at Combinatorial Image Analysis Workshop, September 2022
- Invited Talk at GCPR Tutorial, September 2022

*Gerard Pons-Moll*

- Invited Talk at 3DGV Seminar: September 2021 (remote)
- Invited Talk at NVIDIA GTC 2021, November 2021 (remote)
- Invited Talk at Amazon Visual Science for Fashion Forum, January 2021 (remote)
- Invited Talk at Microsoft Research, February 2022 (remote)
- Invited Talk at Disney Research, March 2022 (remote)
- Invited talk at the 4th Workshop on Long-term Human Motion Prediction, 2022 IEEE International Conference on Robotics and Automation (ICRA) Philadelphia (PA), USA, May 23, 2022
- Talk at Machine Learning in Science Conference, Tübingen, July 2022
- Keynote Speaker at ECCV'22 Workshop: 2nd Challenge on Large Scale Point-cloud Analysis for Urban Scenes Understanding (Urban3D), October 2022
- Huawei Workshop on 3D Vision and Neural Rendering, November 2022,
- Max Planck and Amazon Science Hub Inauguration, December 2022

*David Stutz*

- Invited talk at TU Dortmund, May 2021
- Invited talk at MPI for Mathematics in Sciences, October 2021
- Invited talk at University of California, Los Angeles, October 2021
- Invited talk at University of California, Berkeley, November 2021
- Invited talk at University of Cagliari, Sardinia, November 2021

- Invited talk at University College, London, June 2022
- Invited talk at Dataiku, June 2022
- Invited talk at Berliner Hochschule für Technik, November 2022

*Moritz Böhle*

- Spotlight at Explainability in Machine Learning Workshop at the University of Tübingen, March 2023
- Spotlight at XAI4CV Workshop at *IEEE/CVF Computer Vision and Pattern Recognition Conference (CVPR)*, 2022
- Invited talk at Explainable Machine Learning (EML) Group, University of Tübingen, February 2022 (remote)
- Invited talk at Computer Science and Artificial Intelligence Laboratory (CSAIL), Massachusetts Institute of Technology, October 2021

*Bharat Lal Bhatnagar*

- Keynote talk at "Workshop on Visual Perception for Navigation in Human Environments", ECCV, October 2022
- Invited talk at Mercedes Benz, September 2022 (remote)
- Invited talk at Stanford University, July 2022
- Invited talk at BAIR Lab UC Berkeley, July 2022
- Invited talk at Max Planck Institute for Intelligent Systems, September 2021 (remote)
- Invited talk at Zalando, January 2021 (remote)

*Yaoyao Liu*

- Invited talk at Vision and Graphics Seminar, Masachusetts Institute of Technology, April 2023
- Invited talk at External Speaker Series, University of Illinois Urbana-Champaign, April 2023
- Invited talk at VIGR Seminar, Columbia University. March 2023
- Invited talk at Visual Geometry Group (VGG), University of Oxford, November 2022
- Invited talk at Fudan Vision and Learning Laboratory, Fudan University, September 2022
- Invited talk at NExT++ Research Center, National University of Singapore, September 2020

*Julian Chibane*

- Invited talk at Bosch Industry-on-Campus Lab, Bosch Center of AI, December 2022

*Jan Eric Lenssen*

- Invited talk at Google Zürich, Zürich, March 2023

*Garvita Tiwari*

- Invited talk (Virtual) at Google Zürich, Zürich, July 2022

*Xudong Hong*

- Invited talk (Virtual) at Centre for linguistic theory and studies in probability, University of Gothenburg, Sweden, Mar 2021

*Xinting Hu*

- Invited talk at Damo Academy, Alibaba, China, May, 2022

*Stephan Alaniz*

- Spotlight Talk at Explainability in Machine Learning Workshop at the University of Tübingen, March 2023

### 28.10.4 Other Academic Activities

*Bernt Schiele*

- Fellow of the Institute of Electrical and Electronics Engineers IEEE (since 2017)
- Fellow of the International Association of Pattern recognition IAPR (since 2018)
- Fellow of the European Laboratory of Learning and Intelligent Systems ELLIS (since 2019)
- Fellow of the ACM Association for Computing Machinery (since 2022)
- Chairman, DAGM, German Association for Pattern Recognition, Germany (since 2021)
- Member of Preselection Committee Karl Heinz Beckurts Award (since 2016)
- Member of BAR Beratender Ausschuss für EDV-Anlagen in der Max-Planck-Gesellschaft (since 2019)
- Member of the German National Academy of Sciences Leopoldina (since 2021)
- Member of the Honorary Board International Journal of Computer Vision IJCV
- External Reviewer of PhD thesis: Patrick Moritz Follmann, Technical University of Munich, Munich, Germany, 2021
- External Reviewer of PhD thesis: Yazan Jamal Ibrahim Abu Farha, Rheinische Friedrich-Wilhelms-Universität Bonn, Bonn, Germany, 2021
- Opponent for PhD thesis: Joakim Johnander, Linköping University, Sweden, 2022
- External Reviewer of PhD thesis: Ashkan Khakzar, Technical University of Munich, Munich, Germany, 2023
- External Reviewer of PhD thesis: Fabio Cermelli, Polytechnic University of Turin, Turin, Italy, 2023
- External Reviewer of PhD thesis: Matthias De Lange, Katholieke Universiteit Leuven, Leuven, Belgium, 2023

*Dengxin Dai*

- Member of the European Laboratory of Learning and Intelligent Systems ELLIS (since 2021)
- External Reviewer of PhD thesis: Martin Hahner, ETH Zurich, Zurich, Switzerland, 2022
- External Reviewer of PhD thesis: Vaishakh Patil, ETH Zurich, Zurich, Switzerland, 2022
- External Reviewer of PhD thesis: Fabio Pizzati, Inria, Paris, France, 2022

*Zeynep Akata*

- Fellow of the European Laboratory of Learning and Intelligent Systems ELLIS (since 2019)
- External Reviewer of PhD thesis: Tristan Sylvain, University of Montreal, Montreal, Canada, 2021
- External Reviewer of PhD thesis: Thomas Nestmeyer, University of Tübingen, Tübingen, Germany, 2021
- External Reviewer of PhD thesis: Felix Dangel, University of Tübingen, Tübingen, Germany, 2022
- External Reviewer of PhD thesis: Sebastian Blaes, University of Tübingen, Tübingen, Germany, 2022
- External Reviewer of PhD thesis: Judy Borowski, University of Tübingen, Tübingen, Germany, 2022
- External Reviewer of PhD thesis: Guillaume Couairon, University of Grenoble, Grenoble, France 2023
- External Reviewer of PhD thesis: Sebastian Bordt, University of Tübingen, Tübingen, Germany 2023

*Gerard Pons-Moll*

- External Reviewer of PhD thesis: Stephan Alaniz, University of Tübingen, 2022
- External Reviewer of PhD thesis: Alessandro Tonin, University of Tübingen, 2023
- Supervisor and Reviewer of PhD thesis: Bharat Lal Bhatnagar, Saarland University, 2023
- External Reviewer of PhD thesis: Stephan Alaniz, University of Tübingen, 2022
- Second Supervisor and Reviewer of PhD thesis: Hamed Jalali, University of Tübingen, 2023
- External Reviewer of PhD thesis: Michael Niemeyer, University of Tübingen, 2022
- External Reviewer of PhD thesis: Ikhsanul Habibie, Saarland University, 2023
- External Reviewer of PhD thesis: Edgar Schönfeld, Saarland University, 2023
- External Reviewer of PhD thesis: Arne Monsees, University of Tübingen, 2022

- External Reviewer and co-examiner of PhD thesis: Vasileios Choutas, ETH Zurich, 2022
- External Reviewer of PhD thesis: Mohamed Elmustafa Hassan Mohamed Hassan, University of Tübingen, 2023
- External Reviewer of PhD thesis: Mark Benedikt Boss, University of Tübingen, 2023

## 28.11 Teaching Activities

### Summer Semester 2021

- High Level Computer Vision (Bernt Schiele)
- Deep Learning for Vision and Graphics (Seminar), University of Tuebingen (Gerard Pons-Moll)

### Winter Semester 2021/22

- Mathematics for Machine Learning (Course), University of Tuebingen (Gerard Pons-Moll)
- Unsupervised Deep Learning, University of Siegen (Margret Keuper)
- Computer Vision Seminar on Domain Transfer, University of Siegen (Margret Keuper)
- Higher Level Computer Vision, University of Siegen (Margret Keuper)

### Summer Semester 2022

- High Level Computer Vision (Bernt Schiele)
- Convex Analysis and Optimization (Paul Swoboda)
- Deep Learning for Vision and Graphics (Seminar), University of Tuebingen (Gerard Pons-Moll)
- Software Engineering-Team Project, University of Tuebingen (Gerard Pons-Moll)
- Unsupervised Deep Learning, University of Siegen (Margret Keuper)
- Practical Course Machine Learning (Margret Keuper)

### Winter Semester 2022/23

- Virtual Humans (Course), University of Tuebingen (Gerard Pons-Moll)
- Human Motion Synthesis (Seminar), University of Tuebingen (Gerard Pons-Moll)
- Software Engineering-Team Project, University of Tuebingen (Gerard Pons-Moll)
- Computer Vision Seminar on Generative Models, University of Siegen (Margret Keuper)
- Higher Level Computer Vision, University of Siegen (Margret Keuper)

### Master's Theses

- Duy Nguyen, *Lifted Multi-Cut Optimization for Multi-Camera Multi-People Tracking*, Saarland University, 2021
- Anurag Das, *$(SP)^2$Net for Generalized Zero-Label Semantic Segmentation*, Saarland University, 2021
- Enea Duka, *Self-Supervised Representation Learning to Recognize Unintentional Actions*, Saarland University, 2021
- Skender Paturri, *Specialized Head Relational Framework for Spatio-Temporal Action Localization*, Saarland University, 2021
- Praveen Annamalai Nathan, Technical University of Kaiserslautern, 2021
- Xianghui Xie, *Tracking Human Object Interaction from Single RGB Camera*, Saarland University, 2022
- Lotfy Abdel Khaliq, *Multi-Task Learning using Transformers*, Saarland University, 2023
- Dimitrije Antic, Interactive 3D Clothing Segmentation, University of Tubingen, 2023

### Bachelor's Theses

- Navdeeppal Singh, *Adversarial Robustness of Convolutional Dynamic Alignment Networks*, Saarland University, 2021

## 28.12 Dissertations, Habilitations, Awards

### 28.12.1 Dissertations

- Rakshith Shetty, *Adversarial Content Manipulation for Analyzing and Improving Model Robustness*, May 2021 (with distinction)
- Apratim Bhattacharyya, *Long-Term Future Prediction under Uncertainty and Multi-Modality*, September 2021 (with distinction)
- Mohamed Omran, *From Pixels to People: Recovering Location, Shape and Pose of Humans in Images*, September 2021
- Andrea Hornakova, *Lifted Edges as Connectivity Priors for Multicut and Disjoint Paths*, June 2022
- Stephan Alaniz, *Explainability in Deep Learning by Means of Communication*, July 2022
- David Stutz, *Understanding and Improving Robustness and Uncertainty Estimation in Deep Learning*, July 2022 (with distinction)
- Yaoyao Liu, *Learning from Imperfect Data Incremental Learning and Few-shot Learning*, January 2023 (with distinction)
- Bharat Bhatnagar, *Modelling 3D Humans: Pose, Shape, Clothing and Interactions*, March 2023 (with distinction)

### 28.12.2 Offers for Faculty Positions

*Zeynep Akata*

- Institute Director, Helmholtz Munich with Professorship at TU Munich, in negotiation
- Professor at Ellis Institute, in negotiation

*Jiangxin Dong*

- Nanjing University of Science and Technology, Accepted

*Margret Keuper*

- University of Siegen, Accepted
- University of Mannheim, in negotiation

*Li Jiang*

- Chinese University of Hong Kong, in negotiation

*Paul Swoboda*

- University of Mannheim, Accepted
- HHU Düsseldorf, in negotiation

Former members of D2 that received faculty offers during the reporting period include:

*Fabio Galasso*

- Sapienza University of Rome, Accepted

*Seong Joon Oh*

- University of Tübingen, Independent Research Group, Accepted

*Anna Rohrbach*

- TU Darmstadt, Accepted

*Marcus Rohrbach*

- TU Darmstadt, Accepted (as Alexander von Humboldt Professor)

*Xucong Zhang*

- TU Delft, Accepted

### 28.12.3 Awards

- Zeynep Akata: ECVA Young Researcher Award 2022
- Dengxin Dai: DAGM GCPR German Pattern Recognition Award 2022
- Zeynep Akata: DAGM GCPR German Pattern Recognition Award 2021
- Garvita Tiwari, Jan Eric Lenssen, Gerard Pons-Moll (in cooperation with Dimitrije Antic, Nikolaos Sarafianos and Tony Tung): Best Paper Honorable Mention, ECCV 2022, for the paper “Pose-NDF: Modeling Human Pose Manifolds with Neural Distance Fields”
- Zhi Li, Bernt Schiele (in cooperation with Soshi Shimada, Christian Theobalt and Vladislav Golyanik): Best Student Paper Award, 3DV 2022 International Conference on 3D Vision 2022 Prague, for the paper “MoCapDeform: Monocular 3D Human Motion Capture in Deformable Scenes”
- Yongqin Xian, Julian Chibane, Bharat Bhatnagar, Zeynep Akata, Bernt Schiele, Gerard Pons-Moll: Best Paper Honorable Mention, 3DV 2022 International Conference on 3D Vision 2022 Prague, for the paper “Any-Shot GIN: Generalizing Implicit Networks for Reconstructing Novel Classes”
- Vladimir Guzov, Aymen Mir, Gerard Pons-Moll (in cooperation with Torsten Sattler): Best Paper Finalist, CVPR 2021, for the paper “Human POSEitioning System (HPS): 3D Human Pose Estimation and Self-localization in Large Scenes from Body-Mounted Sensors”
- Jiangxin Dong, Bernt Schiele: Best Paper Award of Fraunhofer IGD and the Visual Computing Groups of TU Darmstadt 2021, Honorable Mention in the category “Impact on Science”
- Jan Eric Lenssen: TU Dortmund Dissertation Award 2022, Best Computer Science Dissertation
- Shaoshuai Shi: World Artificial Intelligence Conference Rising Star Award 2021
- Shaoshuai Shi: HKIS Young Scientist Award Honorable Mention 2022
- Julian Chibane: Best Master Degree, University of Würzburg
- Anurag Das: DAGM GCPR Best Master’s Thesis Award 2021
- Anna Kukleva: Grace-Hopper Award, University of Bonn
- Julian Chibane: Meta / Facebook PhD Fellowship Award 2022 – Topic: AR/VR Human Understanding
- Bharat Lal Bhatnagar: Finalist Qualcomm Innovation Fellowship, 2021
- Julian Chibane: CVPR SHARP’21 Workshop Winner
- Shaoshuai Shi, Li Jiang, Dengxin Dai, Bernt Schiele: Champion of Motion Prediction Challenge at WAD, CVPR 2022
- Dengxin Dai: Golden Owl Award from ETH Zurich for exceptional teaching
- Gerard Pons-Moll: Snap Research Award 2021
- Gerard Pons-Moll: Huawei Research Award 2021

- Bernt Schiele: Outstanding Reviewer Award CVPR 2021
- Bharat Lal Bhatnagar: Outstanding Reviewer Award CVPR 2022
- Bharat Lal Bhatnagar: Outstanding Reviewer Award ECCV 2022
- David Stutz: Outstanding Paper Award CVPR 2021 AML-CV Workshop
- David Stutz: Outstanding Reviewer Award CVPR 2021
- David Stutz: Outstanding Reviewer Award CVPR 2022
- Garvita Tiwari: Outstanding Reviewer Award CVPR 2022
- Moritz Böhle: Outstanding Reviewer Award ECCV 2022
- Julian Chibane: Outstanding Reviewer Award CVPR 2022
- Keyang Zhou: Outstanding Reviewer Award ECCV 2022
- Yong Guo: Outstanding Reviewer Award CVPR 2021
- Margret Keuper: Top Reviewer Award NeurIPS 2022
- Mohamed Omran: Outstanding Reviewer Award CVPR 2021
- Mohamed Omran: Outstanding Reviewer Award ICCV 2021
- Mohamed Omran: Outstanding Reviewer Award ICLR 2021

## 28.13 Grants and Cooperations

### Bernt Schiele – Compact Listing of Grants

- Lead Principal Investigator, Industry cooperation with Toyota Europe on "Learning to Categorize Objects in Real-World Scenes"
- Principal Investigator, Google VIA Center
- Principal Investigator, DFG Research Training Group "Neuroexplit Models of Language, Vision and Action"
- Principal Investigator, DAAD Konrad Zuse School ELIZA
- Amazon MPI-Hub sponsorship with Lablets team

### Dengxin Dai – Compact Listing of Grants

- Principal Investigator, Industry cooperation with Toyota Europe on "Multi-task Learning"
- Principal Investigator, one Gift Grant from Meta Research on "Robust Object Detection"
- Co-Investigator, one Horizon Europe Project on "European Lighthouse on Secure and Safe AI".

### Zeynep Akata – Compact Listing of Grants

- Principal Investigator, Else Kröner Fresenius Grant: CLINBrain
- Principal Investigator, Carl Zeiss Stiftung Grant: Certification and Foundations of Safe Machine Learning Systems in Healthcare

– Principal Investigator, Cluster Innovation Grants: Compositionality in Minds and Machines

### Gerard Pons-Moll – Compact Listing of Grants

– Huawei Research Sponsorship
– Amazon MPI-Hub sponsorship with Lablets team
– Amazon MPI-Hub sponsorship with Fashion team

## 28.14 Publications

### Journal articles and book chapters


[1] M. D. Böhle, M. Fritz, and B. Schiele. Optimising for interpretability: Convolutional dynamic alignment networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022.

[2] D. Chen, S. Zhang, J. Yang, and B. Schiele. Norm-aware embedding for efficient person search and tracking. *International Journal of Computer Vision*, 129:3154–3168, 2021.

[3] J. Choe, S. J. Oh, S. Chun, S. Lee, Z. Akata, and H. Shim. Evaluation for weakly supervised object localization: Protocol, metrics, and datasets. *TPAMI*, 2023.

[4] D. Dai, A. B. Vasudevan, J. Matas, and L. Van Gool. Binaural SoundNet: Predicting semantics, depth and motion with binaural sounds. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(1):123–136, 2023.

[5] J. Dong, S. Roth, and B. Schiele. DWDN: Deep Wiener Deconvolution Network for non-blind image deblurring. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(12):9960–9976, 2022.

[6] Y. Fan, A. Kukleva, D. Dai, and B. Schiele. Revisiting consistency regularization for semi-supervised learning. *International Journal of Computer Vision*, 131:626–643, 2023.

[7] R. Gong, W. Li, Y. Chen, D. Dai, and L. Van Gool. DLOW: Domain flow and applications. *International Journal of Computer Vision*, 129:2865–2888, 2021.

[8] J. Grabinski, J. Keuper, and M. Keuper. Aliasing and adversarial robust generalization of CNNs. *Machine Learning*, 111:3925–3951, 2022.

[9] M. Habermann, L. Liu, W. Xu, M. Zollhöfer, G. Pons-Moll, and C. Theobalt. Real-time deep dynamic characters. *ACM Transactions on Graphics (Proc. ACM SIGGRAPH)*, 40(4), Article 94, 2021.

[10] M. Habermann, W. Xu, M. Zollhöfer, G. Pons-Moll, and C. Theobalt. A deeper look into DeepCap. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(4):4009–4002, 2023.

[11] X. Hong, A. Sayeed, K. Mehra, V. Demberg, and B. Schiele. Visual writing prompts: Character-grounded story generation with curated image sequences. *Transactions of the Association for Computational Linguistics*, 11, 2023. E-Text: 2301.08571.

[12] E. Levinkov, A. Kardoost, B. Andres, and M. Keuper. Higher-order multicuts for geometric model fitting and motion segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(1):608–622, 2023.

[13] S. Li, X. Chen, Y. Liu, D. Dai, C. Stachniss, and J. Gall. Multi-scale interaction for real-time LiDAR data segmentation on an embedded platform. *IEEE Robotics and Automation Letters*, 7(2):738–745, 2022.

[14] X. Li, J. Huang, Y. Liu, Q. Zhou, S. Zheng, B. Schiele, and Q. Sun. Learning to teach and learn for semi-supervised few-shot image classification. *Computer Vision and Image Understanding*, 212, Article 103270, 2021.

[15] V. Patil, A. Liniger, D. Dai, and L. Van Gool. Improving depth estimation using map-based depth priors. *IEEE Robotics and Automation Letters*, 7(2):3640–3647, 2022.

[16] S. Shi, L. Jiang, J. Deng, Z. Wang, C. Guo, J. Shi, X. Wang, and H. Li. PV-RCNN++: Point-voxel feature set abstraction with local vector representation for 3D object detection. *International Journal of Computer Vision*, 131:531–551, 2022.

[17] D. Stutz, N. Chandramoorthy, M. Hein, and B. Schiele. Random and adversarial bit error robustness: Energy-efficient and secure DNN accelerators. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(3):3632–3647, 2023.

[18] Q. Sun, Y. Liu, Z. Chen, T.-S. Chua, and B. Schiele. Meta-transfer learning through hard tasks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(3):1443–1456, 2022.

[19] V. Sushko, E. Schönfeld, D. Zhang, J. Gall, B. Schiele, and A. Khoreva. Oasis: Only Adversarial Supervision for Semantic Image Synthesis. *International Journal of Computer Vision*, 130:2903–2923, 2022.

[20] N. Vödisch, O. Unal, K. Li, L. Van Gool, and D. Dai. End-to-end optimization of LiDAR beam configuration for 3D object detection and localization. *IEEE Robotics and Automation Letters*, 7(2):2242–2249, 2022.

[21] Y. Xian, B. Korbar, M. Douze, L. Torresani, B. Schiele, and Z. Akata. Generalized few-shot video classification with video retrieval and feature generation. *TPAMI*, 2022.

[22] W. Xu, Y. Xian, J. Wang, B. Schiele, and Z. Akata. Attribute prototype network for any-shot learning. *IJCV*, 2022.

[23] J.-N. Zaech, D. Dai, A. Liniger, M. Danelljan, and L. Van Gool. Learnable online graph representations for 3D multi-object tracking. *IEEE Robotics and Automation Letters*, 7(2):5103–5110, 2022.

[24] S. Zhang, D. Chen, J. Yang, and B. Schiele. Guided attention in CNNs for occluded pedestrian detection and re-identification. *International Journal of Computer Vision*, 129:1875–1892, 2021.

## Conference articles

[1] A. Abbas and P. Swoboda. Combinatorial optimization for panoptic segmentation: A fully differentiable approach. In M. Ranzato, A. Beygelzimer, P. S. Liang, J. W. Vaughan, and Y. Dauphin, eds., *Advances in Neural Information Processing Systems 34 (NeurIPS 2021)*, Virtual, 2021, pp. 15635–15649. Curran Associates, Inc.

[2] A. Abbas and P. Swoboda. FastDOG: Fast discrete optimization on GPU. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2022)*, New Orleans, LA, USA, 2022, pp. 439–449. IEEE.

[3] A. Abbas and P. Swoboda. RAMA: A rapid multicut algorithm on GPU. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2022)*, New Orleans, LA, USA, 2022, pp. 8183–8192. IEEE.

[4] S. Alaniz, M. Federici, and Z. Akata. Compositional mixture representations for vision and text. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPR 2022)*, New Orleans, LA, USA, 2021, pp. 4201–4210. IEEE.

[5] S. Alaniz, T. Hummel, and Z. Akata. Semantic image synthesis with semantically coupled VQ-model. In *ICLR Workshop on Deep Generative Models for Highly Structured Data (ICLR 2022 DGM4HSD)*, Virtual, 2022. OpenReview.net.

[6] S. Alaniz, M. Mancini, A. Dutta, D. Marcos, and Z. Akata. Abstracting sketches through simple primitives. In S. Avidan, G. Brostow, M. Cissé, G. M. Farinella, and T. Hassner, eds., *Computer Vision – ECCV 2022*, Tel Aviv, Israel, 2022, LNCS 13689, pp. 396–412. Springer.

[7] S. Alaniz, D. Marcos, B. Schiele, and Z. Akata. Learning decision trees recurrently through communication. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2021)*, Nashville, TN, USA (Virtual), 2021, pp. 13518–13527. IEEE.

[8] S. Badirli, Z. Akata, G. O. Mohler, C. Picard, and M. Dundar. Fine-grained zero-shot learning with dna as side information. In *NeurIPS*, 2021.

[9] B. L. Bhatnagar, X. Xie, I. Petrov, C. Sminchisescu, C. Theobalt, and G. Pons-Moll. BEHAVE: Dataset and method for tracking human object interactions. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2022)*, New Orleans, LA, USA, 2022, pp. 15914–15925. IEEE.

[10] A. Bhattacharyya, D. O. Reino, M. Fritz, and B. Schiele. Euro-PVI: Pedestrian vehicle interactions in dense urban centers. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2021)*, Nashville, TN, USA (Virtual), 2021, pp. 6408–6417. IEEE.

[11] M. Böhle, M. Fritz, and B. Schiele. B-cos networks: Alignment is all we need for interpretability. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2022)*, New Orleans, LA, USA, 2022, pp. 10319–10328. IEEE.

[12] M. D. Böhle, M. Fritz, and B. Schiele. Convolutional dynamic alignment networks for interpretable classifications. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2021)*, Nashville, TN, USA (Virtual), 2021, pp. 10029–10038. IEEE.

[13] F. Cermelli, M. Mancini, Y. Xian, Z. Akata, and B. Caputo. Prototype-based incremental few-shot segmentation. In *BMVC*, 2021.

[14] A. Chaudhuri, M. Mancini, Z. Akata, and A. Dutta. Relational proxies: Emergent relationships as fine-grained discriminator. In *Neural Information Processing Systems (NeurIPS)*, 2022.

[15] A. Chaudhuri, M. Mancini, Y. Chen, Z. Akata, and A. Dutta. Cross-modal fusion distillation for fine-grained sketch-based image retrieval. In *33rd British Machine Vision Conference (BMVC 2022)*, London, UK, 2022, Article 499. BMVA Press.

[16] U. Chaudhuri, R. Chavan, B. Banerjee, A. Dutta, and Z. Akata. Bda-sketret: Bi-level domain adaptation for zero-shot sbir. *Neurocomputing*, 2022.

[17] D. Chen, A. Doering, S. Zhang, J. Yang, J. Gall, and B. Schiele. Keypoint message passing for video-based person re-identification. In *Proceedings of the 36th AAAI Conference on Artificial Intelligence*, Virtual Conference, 2022, pp. 239–247. AAAI.

[18] H. Chen, R. Tao, Y. Fan, Y. Wang, M. Savvides, J. Wang, B. Raj, X. Xie, and B. Schiele. SoftMatch: Addressing the quantity-quality tradeoff in semi-supervised learning. In *Eleventh International Conference on Learning Representations (ICLR 2023)*, Kigali, Rwanda, 2023. OpenReview.net. Accepted.

[19] X. Chen, S. Shi, B. Zhu, K. C. Cheung, H. Xu, and H. Li. MPPNet: Multi-frame feature intertwining with proxy points for 3D temporal object detection. In S. Avidan, G. Brostow, M. Cissé, G. M. Farinella, and T. Hassner, eds., *Computer Vision – ECCV 2022*, Tel Aviv, Israel, 2022, LNCS 13668, pp. 680–697. Springer.

[20] Y. Chen, Y. Xian, A. S. Koepke, and Z. Akata. Distilling audio-visual knowledge by compositional contrastive learning. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2021)*, Virtual Conference, 2021, pp. 7012–7021. IEEE.

[21] E. Corona, G. Pons-Moll, G. Alenyà, and F. Moreno-Noguer. Learned vertex descent: A new direction for 3D human model fitting. In S. Avidan, G. Brostow, M. Cissé, G. Farinella, and T. Hassner, eds., *Computer Vision – ECCV 2022*, Tel Aviv, Israel, 2022, LNCS 13662, pp. 145–164. Springer.

[22] A. Das, Y. Xian, D. Dai, and B. Schiele. Weakly-supervised domain adaptive semantic segmentation with prototypical contrastive learning. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2023)*, Vancouver, Canada, 2023. IEEE. Accepted.

[23] A. Das, Y. Xian, Y. He, Z. Akata, and B. Schiele. Urban scene semantic segmentation with low-cost coarse annotation. In *2023 IEEE Winter Conference on Applications of Computer Vision (WACV 2023)*, Waikoloa Village, HI, USA, 2023, pp. 5967–5976. IEEE.

[24] A. Das, Y. Xian, Y. He, B. Schiele, and Z. Akata. $SP^2Net$ for generalized zero-label semantic segmentation. In C. Bauckhage, J. Gall, and A. Schwing, eds., *Pattern Recognition (GCPR 2021)*, Bonn, Germany, 2021, LNCS 13024, pp. 235–249. Springer.

[25] A. Doering, D. Chen, S. Zhang, B. Schiele, and J. Gall. PoseTrack21: A dataset for person search, multi-object tracking and multi-person pose tracking. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2022)*, New Orleans, LA, USA, 2022, pp. 20931–20940. IEEE.

[26] J. Dong, S. Roth, and B. Schiele. Learning spatially-variant MAP models for non-blind image deblurring. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2021)*, Nashville, TN, USA (Virtual), 2021, pp. 4886–4895. IEEE.

[27] A. Dutta, M. Mancini, and Z. Akata. Concurrent discrimination and alignment for self-supervised feature learning. In *IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, 2021.

[28] Q. Fan, M. Segu, Y.-W. Tai, F. Yu, C.-K. Tang, B. Schiele, and D. Dai. Normalization perturbation: A simple domain generalization method for real-world domain shifts. In *International Conference on Learning Representations*, 2023.

[29] Y. Fan, D. Dai, and B. Schiele. CoSSL: Co-learning of representation and classifier for imbalanced semi-supervised learning. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2022)*, New Orleans, LA, USA, 2022, pp. 14554–14564. IEEE.

[30] Y. Fan, A. Kukleva, and B. Schiele. Revisiting consistency regularization for semi-supervised learning. In C. Bauckhage, J. Gall, and A. Schwing, eds., *Pattern Recognition (GCPR 2021)*, Bonn, Germany, 2021, LNCS 13024, pp. 63–78. Springer.

[31] P. Gavrikov, J. Keuper, and M. Keuper. An extended study of human-like behavior under adversarial training. In *The 3rd Workshop of Adversarial Machine Learning on Computer Vision: Art of Robustness, CVPR workshops*, 2023.

[32] J. Geiping, J. Lukasik, M. Keuper, and M. Moeller. DARTS for inverse problems: A study on stability. In *NeurIPS 2021 Workshop on Deep Learning and Inverse Problems (NeurIPS 2021 Deep Inverse Workshop)*, Virtual, 2021. OpenReview.net.

[33] R. Gong, D. Dai, Y. Chen, W. Li, and L. Van Gool. mDALU: Multi-source domain adaptation and label unification with partial datasets. In *ICCV 2021, IEEE/CVF International Conference on Computer Vision*, Virtual Event, 2021, pp. 8856–8865. IEEE.

[34] R. Gong, M. Danelljan, D. Dai, D. P. Paudel, A. Chhatkuli, F. Yu, and L. Van Gool. TACS: Taxonomy adaptive cross-domain semantic segmentation. In S. Avidan, G. Brostow, M. Cissé, G. M. Farinella, and T. Hassner, eds., *Computer Vision – ECCV 2022*, Tel Aviv, Israel, 2022, LNCS 13694, pp. 19–35. Springer.

[35] S. Gong, S. Zhang, J. Yang, D. Dai, and B. Schiele. Bi-level alignment for cross-domain crowd counting. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2022)*, New Orleans, LA, USA, 2022, pp. 7532–7540. IEEE.

[36] S. Gong, S. Zhang, J. Yang, D. Dai, and B. Schiele. Class-agnostic object counting robust to intraclass diversity. In S. Avidan, G. Brostow, M. Cissé, G. M. Farinella, and T. Hassner, eds., *Computer Vision – ECCV 2022*, Tel Aviv, Israel, 2022, LNCS 13693, pp. 388–403. Springer.

[37] L. A. der Goten, Van, T. Hepp, Z. Akata, and K. Smith. Conditional de-identification of 3d magnetic resonance images. In *BMVC*, 2021.

[38] J. Grabinski, P. Gavrikov, J. Keuper, and M. Keuper. Robust models are less over-confident. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, eds., *Advances in Neural Information Processing Systems 35 (NeurIPS 2022)*, New Orleans, LA, USA, 2022, pp. 39059–39075. Currran Assoicate, Inc.

[39] J. Grabinski, S. Jung, J. Keuper, and M. Keuper. FrequencyLowCut pooling – plug & play against catastrophic overfitting. In S. Avidan, G. Brostow, M. Cissé, G. Farinella, and T. Hassner, eds., *Computer Vision – ECCV 2022*, Tel Aviv, Israel, 2022, LNCS 13674, pp. 36–57. Springer.

[40] Y. Guo, D. Stutz, and B. Schiele. Improving robustness by enhancing weak subnets. In S. Avidan, G. Brostow, M. Cissé, G. Farinella, and T. Hassner, eds., *Computer Vision – ECCV 2022*, Tel Aviv, Israel, 2022, LNCS 13684, pp. 320–338. Springer.

[41] Y. Guo, D. Stutz, and B. Schiele. Improving robustness of vision transformers by reducing sensitivity to patch corruptions. In *Conference on Computer Vision and Pattern Recognition 2023*, 2023. https://openreview.net/forum?id=rjM8U3DhlF.

[42] V. Guzov, A. Mir, T. Sattler, and G. Pons-Moll. Human POSEitioning system (HPS): 3D human pose estimation and self-localization in large scenes from body-mounted sensors. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2021)*, Nashville, TN, US (Virtual), 2021, pp. 4318–4329. IEEE.

[43] M. Hahner, C. Sakaridis, M. Bijelic, F. Heide, F. Yu, D. Dai, and L. Van Gool. LiDAR snowfall simulation for robust 3D object detection. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2022)*, New Orleans, LA, USA, 2022, pp. 16343–16353. IEEE.

[44] M. Hahner, C. Sakaridis, D. Dai, and L. Van Gool. Fog simulation on real LiDAR point clouds for 3D object detection in adverse weather. In *ICCV 2021, IEEE/CVF International Conference on Computer Vision*, Virtual Event, 2021, pp. 15263–15272. IEEE.

[45] S. Haller, L. Feineis, L. Hutschenreiter, F. Bernard, C. Rother, D. Kainmüller, P. Swoboda, and B. Savchynskyy. A comparative study of graph matching algorithms in computer vision. In S. Avidan, G. Brostow, M. Cissé, G. Farinella, and T. Hassner, eds., *Computer Vision – ECCV 2022*, Tel Aviv, Israel, 2022, LNCS 13683, pp. 636–653. Springer.

[46] Y. He, N. Yu, M. Keuper, and M. Fritz. Beyond the spectrum: Detecting deepfakes via re-synthesis. In Z.-H. Zhou, ed., *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence (IJCAI 2021)*, Montreal, Canada, 2021, pp. 2534–2541. IJCAI.

[47] A. Horňáková, R. Henschel, B. Rosenhahn, and P. Swoboda. Lifted disjoint paths with application in multiple object tracking. In H. Daumé and A. Singh, eds., *Proceedings of the 37th International Conference on Machine Learning (ICML 2020)*, Virtual Conference, 2020, Proceedings of Machine Learning Research 119, pp. 1539–1548. MLResearchPress.

[48] A. Horňáková, T. Kaiser, P. Swoboda, M. Rolinek, B. Rosenhahn, and R. Henschel. Making higher order MOT scalable: An efficient approximate solver for lifted disjoint paths. In *ICCV 2021, IEEE/CVF International Conference on Computer Vision*, Virtual Event, 2021, pp. 6310–6320. IEEE.

[49] L. Hoyer, D. Dai, and L. Van Gool. DAFormer: Improving network architectures and training strategies for domain-adaptive semantic segmentation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2022)*, New Orleans, LA, USA, 2022, pp. 9914–9925. IEEE.

[50] L. Hoyer, D. Dai, and L. Van Gool. HRDA: Context-aware high-resolution domain-adaptive semantic segmentation. In S. Avidan, G. Brostow, M. Cissé, G. M. Farinella, and T. Hassner, eds., *Computer Vision – ECCV 2022*, Tel Aviv, Israel, 2022, LNCS 13690, pp. 372–391. Springer.

[51] L. Jiang, Z. Yang, S. Shi, V. Golyanik, D. Dai, and B. Schiele. Self-supervised pre-training with masked shape prediction for 3D scene understanding. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2023)*, Vancouver, Canada, 2023. IEEE. Accepted.

[52] S. Jung and M. Keuper. Internalized biases in Fréchet inception distance. In *NeurIPS 2021 Workshop on Distribution Shifts: Connecting Methods and Applications (NeurIPS 2021 Workshop DistShift)*, Virtual, 2021. OpenReview.net.

[53] S. Jung and M. Keuper. Spectral distribution aware image generation. In *Thirty-Fifth AAAI Conference on Artificial Intelligence Technical Tracks 2*, Virtual Conference, 2021, pp. 1734–1742. AAAI.

[54] S. Jung and M. Keuper. Learning to solve minimum cost multicuts efficiently using edge-weighted graph convolutional neural networks. In *Machine Learning and Knowledge Discovery in Databases (ECML PKDD 2022)*, Grenoble, France, 2022, Article 486. ecmlpkdd.org.

[55] S. Jung, J. Lukasik, and M. Keuper. Neural architecture design and robustness: A dataset. In *Eleventh International Conference on Learning Representations (ICLR 2023)*, Kigali, Rwanda, 2023. OpenReview.net. Accepted.

[56] S. Jung, S. Ziegler, A. Kardoost, and M. Keuper. Optimizing edge detection for image segmentation with multicut penalties. In B. Andres, F. Bernard, D. Cremers, S. Frintrop, B. Goldlücke, and I. Ihrke, eds., *Pattern Recognition (DAGM GCPR 2022)*, Konstanz, Germany, 2022, LNCS 13485, pp. 182–197. Springer.

[57] S. Karthik, M. Mancini, and Z. Akata. KG-SP: knowledge guided simple primitives for open world compositional zero-shot learning. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022.

[58] M. Kayser, O.-M. Camburu, L. Salewski, C. Emde, V. Do, Z. Akata, and T. Lukasiewicz. e-ViL: A dataset and benchmark for natural language explanations in vision-language tasks. In *ICCV 2021, IEEE/CVF International Conference on Computer Vision*, Virtual Event, 2021, pp. 1224–1234. IEEE.

[59] Q. Ke, M. Fritz, and B. Schiele. Future moment assessment for action query. In *IEEE Winter Conference on Applications of Computer Vision (WACV 2021)*, Virtual Event, 2021, pp. 3219–3227. IEEE.

[60] J. M. Kim, J. Choe, Z. Akata, and S. J. Oh. Keep CALM and improve visual feature attribution. In *ICCV 2021, IEEE/CVF International Conference on Computer Vision*, Virtual Event, 2021, pp. 8330–8340. IEEE.

[61] Y. Kim, J. M. Kim, Z. Akata, and J. Lee. Large loss matters in weakly supervised multi-label classification. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2022)*, New Orleans, LA, USA, 2022, pp. 14136–14145. IEEE.

[62] M. Kirchhof, K. Roth, Z. Akata, and E. Kasneci. A non-isotropic probabilistic take on proxy-based deep metric learning. In *European Conference on Computer Vision (ECCV)*, 2022.

[63] A. Kukleva, M. Boehle, B. Schiele, H. Kuehne, and C. Rupprecht. Temperature schedules for self-supervised contrastive methods on long-tail data. In *International Conference on Learning Representations*, 2023.

[64] A. Kukleva, H. Kuehne, and B. Schiele. Generalized and incremental few-shot learning by explicit learning and calibration without forgetting. In *ICCV 2021, IEEE/CVF International Conference on Computer Vision*, Virtual Event, 2021, pp. 9000–9009. IEEE.

[65] X. Lai, J. Liu, L. Jiang, L. Wang, H. Zhao, S. Liu, X. Qi, and J. Jia. Stratified transformer for 3D point cloud segmentation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2022)*, New Orleans, LA, USA, 2022, pp. 8490–8499. IEEE.

[66] J.-H. Lange and P. Swoboda. Efficient message passing for 0–1 ILPs with binary decision diagrams. In *International Conference on Machine Learning*, 2021, pp. 6000–6010. PMLR.

[67] K. Li, D. Dai, and L. van Gool. Hyperspectral image super-resolution with RGB image super-resolution as an auxiliary task. In *2022 IEEE Winter Conference on Applications of Computer Vision (WACV 2022)*, Waikoloa Village, HI, USA, 2022, pp. 4039–4048. IEEE.

[68] Y. Li, D. Zhang, M. Keuper, and A. Khoreva. Intra-source style augmentation for improved domain generalization. In *2023 IEEE Winter Conference on Applications of Computer Vision (WACV 2023)*, Waikoloa Village, HI, USA, 2023, pp. 509–519. IEEE.

[69] Z. Li, S. Shi, B. Schiele, and D. Dai. Test-time domain adaptation for monocular depth estimation. In *IEEE International Conference on Robotics and Automation (ICRA 2023)*, 2023. IEEE. Accepted.

[70] Z. Li, S. Shimada, B. Schiele, C. Theobalt, and V. Golyanik. MoCapDeform: Monocular 3D human motion capture in deformable scenes. In *International Conference on 3D Vision*, Hybrid / Prague, Czechia, 2022, pp. 1–11. IEEE.

[71] Z. Liao, J. Yang, J. Saito, G. Pons-Moll, and Y. Zhou. Skeleton-free pose transfer for stylized 3D characters. In S. Avidan, G. Brostow, M. Cissé, G. Farinella, and T. Hassner, eds., *Computer Vision – ECCV 2022*, Tel Aviv, Israel, 2022, LNCS 13662, pp. 640–656. Springer.

[72] W. Lin, A. Kukleva, K. Sun, H. Possegger, H. Kuehne, and H. Bischof. CycDA: Unsupervised cycle domain adaptation to learn from image to video. In S. Avidan, G. Brostow, M. Cissé, G. Farinella, and T. Hassner, eds., *Computer Vision – ECCV 2022*, Tel Aviv, Israel, 2022, LNCS 13663, pp. 698–715. Springer.

[73] Y. Liu, Y. Li, B. Schiele, and Q. Sun. Online hyperparameter optimization for class-incremental learning. In *Proceedings of the 37th AAAI Conference on Artificial Intelligence*, Washington, DC, USA, 2023. AAAI. Accepted.

[74] Y. Liu, B. Schiele, and Q. Sun. Adaptive aggregation networks for class-incremental learning. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2021)*, Nashville, TN, USA (Virtual), 2021, pp. 2544–2553. IEEE.

[75] Y. Liu, B. Schiele, and Q. Sun. RMM: Reinforced memory management for class-incremental learning. In M. Ranzato, A. Beygelzimer, P. S. Liang, J. W. Vaughan, and Y. Dauphin, eds., *Advances in Neural Information Processing Systems 34 (NeurIPS 2021)*, Virtual, 2021, pp. 3478–3490. Curran Associates, Inc.

[76] Y. Liu, B. Schiele, A. Vedaldi, and C. Rupprecht. Continual detection transformer for incremental object detection. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR*, 2023.

[77] J. Lukasik, S. Jung, and M. Keuper. Learning where to look – generative NAS is surprisingly efficient. In S. Avidan, G. Brostow, M. Cissé, G. Farinella, and T. Hassner, eds., *Computer Vision – ECCV 2022*, Tel Aviv, Israel, 2022, LNCS 13683, pp. 257–273. Springer.

[78] Z. Luo, Y. Liu, B. Schiele, and Q. Sun. Class-incremental exemplar compression for class-incremental learning. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR*, 2023.

[79] X. Ma, Z. Wang, Y. Zhan, Y. Zheng, Z. Wang, D. Dai, and C.-W. Lin. Both style and fog matter: Cumulative domain adaptation for semantic foggy scene understanding. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2022)*, New Orleans, LA, USA, 2022, pp. 18900–18909. IEEE.

[80] Y. Ma, Y. Chen, and Z. Akata. Distilling knowledge from self-supervised teacher by embedding graph alignment. In *33rd British Machine Vision Conference (BMVC 2022)*, London, UK, 2022, Article 973. BMVA Press.

[81] O.-B. Mercea, T. Hummel, A. S. Koepke, and Z. Akata. Temporal and cross-modal attention for audio-visual zero-shot learning. In S. Avidan, G. Brostow, M. Cissé, G. M. Farinella, and T. Hassner, eds., *Computer Vision – ECCV 2022*, Tel Aviv, Israel, 2022, LNCS 13680, pp. 488–505. Springer.

[82] O.-B. Mercea, L. Riesch, A. S. Koepke, and Z. Akata. Audio-visual generalised zero-shot learning with cross-modal attention and language. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2022)*, New Orleans, LA, USA, 2022, pp. 10543–10553. IEEE.

[83] P. Müller, A. Braun, and M. Keuper. Impact of realistic properties of the point spread function on classification tasks to reveal a possible distribution shift. In *NeurIPS 2022 Workshop on Distribution Shifts: Connecting Methods and Applications (NeurIPS 2022 Workshop DistShift)*, New Orelans, LA, USA, 2022. OpenReview.net.

[84] M. F. Naeem, Y. Xian, F. Tombari, and Z. Akata. Learning graph embeddings for compositional zero-shot learning. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2021)*, Nashville, TN, US (Virtual), 2021, pp. 953–962. IEEE.

[85] A. Neculai, Y. Chen, and Z. Akata. Probabilistic compositional embeddings for multimodal image retrieval. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPR 2022)*, New Orleans, LA, USA, 2021, pp. 4546–4556. IEEE.

[86] D. H. M. Nguyen, R. Henschel, B. Rosenhahn, D. Sonntag, and P. Swoboda. LMGP: Lifted multicut meets geometry projections for multi-camera multi-object tracking. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2022)*, New Orleans, LA, USA, 2022, pp. 8856–8865. IEEE.

[87] D. M. H. Nguyen, H. Nguyen, M. T. N. Truong, T. Cao, B. T. Nguyen, N. Ho, P. Swoboda, S. Albarqouni, P. Xie, and D. Sonntag. Joint self-supervised image-volume representation learning with intra-inter contrastive clustering. In *Proceedings of the 37th AAAI Conference on Artificial Intelligence*, Washington, DC, USA, 2023. AAAI. Accepted.

[88] A. Oncescu, A. S. Koepke, J. F. Henriques, Z. Akata, and S. Albanie. Audio retrieval with natural language queries. In *Interspeech*, 2021.

[89] G. Pastore, F. Cermelli, Y. Xian, M. Mancini, Z. Akata, and B. Caputo. A closer look at self-training for zero-label semantic segmentation. In *CVPR*, 2021.

[90] G. Pons-Moll, F. Moreno-Noguer, E. Corona, A. Pumarola, and G. Alenyà. SMPLicit: Topology-aware generative model for clothed people. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2021)*, Virtual Conference, 2021, pp. 11870–11880. IEEE.

[91] D. Pu, X. Hong, P.-J. Lin, E. Chang, and V. Demberg. Two-stage movie script summarization: An efficient method for low-resource long document summarization. In *Proceedings of The Workshop on Automatic Summarization for Creative Writing*, Gyeongju, Republic of Korea, 2022, pp. 57–66. Association for Computational Linguistics.

[92] S. Rao, M. Böhle, and B. Schiele. Towards better understanding attribution methods. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2022)*, New Orleans, LA, USA, 2022, pp. 10213–10222. IEEE.

[93] K. Renz, K. Chitta, O.-B. Mercea, A. S. Koepke, Z. Akata, and A. Geiger. PlanT: Explainable planning transformers via object-level representations. In K. Liu, D. Kulic, and J. Ichnowski, eds., *Proceedings of the 6th Annual Conference on Robot Learning (CoRL 2022)*, Auckland, New Zealand, 2022, Proceedings of the Machine Learning Research 205, pp. 459–470. MLResearchPress.

[94] F. Rezaeianaran, R. Shetty, R. Aljundi, D. O. Reino, S. Zhang, and B. Schiele. Seeking similarities over differences: Similarity-based domain alignment for adaptive object detection. In *ICCV 2021, IEEE/CVF International Conference on Computer Vision*, Virtual Event, 2021, pp. 9184–9193. IEEE.

[95] P. Roetzer, P. Swoboda, D. Cremers, and F. Bernard. A scalable combinatorial solver for elastic geometrically consistent 3D shape matching. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2022)*, New Orleans, LA, USA, 2022, pp. 428–438. IEEE.

[96] Y. Rong, W. Xu, Z. Akata, and E. Kasneci. Human attention in fine-grained classification. In *British Machine Vision Conference (BMVC)*, 2021.

[97] K. Roth, M. Ibrahim, Z. Akata, P. Vincent, and D. Bouchacourt. Disentanglement of correlated factors via hausdorff factorized support. In *International Conference on Learning Representations (ICLR)*, 2023.

[98] K. Roth, O. Vinyals, and Z. Akata. Integrating language guidance into vision-based deep metric learning. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022.

[99] K. Roth, O. Vinyals, and Z. Akata. Non-isotropy regularization for proxy-based deep metric learning. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022.

[100] C. Sakaridis, D. Dai, and L. Van Gool. ACDC: The adverse conditions dataset with correspondences for semantic driving scene understanding. In *ICCV 2021, IEEE/CVF International Conference on Computer Vision*, Virtual Event, 2021, pp. 10745–10755. IEEE.

[101] A. Saseendran, K. Skubch, S. Falkner, and M. Keuper. Shape your space: A Gaussian mixture regularization approach to deterministic autoencoders. In M. Ranzato, A. Beygelzimer, P. S. Liang, J. W. Vaughan, and Y. Dauphin, eds., *Advances in Neural Information Processing Systems 34 pre-proceedings (NeurIPS 2021)*, Virtual Event, 2021, pp. 7319–7332. Curran Associates, Inc.

[102] A. Saseendran, K. Skubch, and M. Keuper. Trading off image quality for robustness is not necessary with regularized deterministic autoencoders. In S. Koyejo, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, eds., *Advances in Neural Information Processing Systems 35 (NeurIPS 2022)*, New Orleans, LA, USA, 2022, pp. 26751–26763. Curran Associates, Inc.

[103] E. Schönfeld, V. Sushko, D. Zhang, J. Gall, B. Schiele, and A. Khoreva. You only need adversarial supervision for semantic image synthesis. In *International Conference on Learning Representations (ICLR 2021)*, Vienna, Austria (Virtual), 2021. OpenReview.net.

[104] S. Shi, L. Jiang, D. Dai, and B. Schiele. Motion transformer with global intention localization and local movement refinement. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, eds., *Advances in Neural Information Processing Systems 35 (NeurIPS 2022)*, New Orleans, LA, USA, 2022, pp. 6531–6543. Curran Associates, Inc.

[105] S. Shimada, V. Golyanik, Z. Li, P. Pérez, W. Xu, and C. Theobalt. HULC: 3D HUman motion capture with pose manifold sampLing and dense Contact guidance. In S. Avidan, G. Brostow, M. Cissé, G. M. Farinella, and T. Hassner, eds., *Computer Vision – ECCV 2022*, Tel Aviv, Israel, 2022, LNCS 13682, pp. 516–533. Springer.

[106] S. Sinha, K. Roth, A. Goyal, M. Ghassemi, Z. Akata, H. Larochelle, and A. Garg. Uniform priors for data-efficient learning. In *CVPR*, 2022.

[107] D. Stutz, K. Dvijotham, A. T. Cemgil, and A. Doucet. Learning optimal conformal classifiers. In *The Tenth International Conference on Learning Representations, ICLR 2022, Virtual Event, April 25-29, 2022*, 2022. OpenReview.net.

[108] D. Stutz, M. Hein, and B. Schiele. Relating adversarially robust generalization to flat minima. In *ICCV 2021, IEEE/CVF International Conference on Computer Vision*, Virtual Event, 2021, pp. 7787–7797. IEEE.

[109] G. Sun, T. Probst, D. P. Paudel, N. Popovic, M. Kanakis, J. Patel, D. Dai, and L. Van Gool. Task switching network for multi-task learning. In *ICCV 2021, IEEE/CVF International Conference on Computer Vision*, Virtual Event, 2021, pp. 8271–8280. IEEE.

[110] T. Sun, M. Segù, J. Postels, Y. Wang, L. Van Gool, B. Schiele, F. Tombari, and F. Yu. SHIFT: A synthetic driving dataset for continuous multi-task domain adaptation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2022)*, New Orleans, LA, USA, 2022, pp. 21339–21350. IEEE.

[111] Z. Tian, X. Lai, L. Jiang, S. Liu, M. Shu, H. Zhao, and J. Jia. Generalized few-shot semantic segmentation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2022)*, New Orleans, LA, USA, 2022, pp. 11553–11562. IEEE.

[112] G. Tiwari, D. Antic, J. E. Lenssen, N. Sarafianos, T. Tung, and G. Pons-Moll. Pose-NDF: Modeling human pose manifolds with neural distance fields. In S. Avidan, G. Brostow, M. Cissé, G. Farinella, and T. Hassner, eds., *Computer Vision – ECCV 2022*, Tel Aviv, Israel, 2022, LNCS 13665, pp. 572–589. Springer.

[113] G. Tiwari, N. Sarafianos, T. Tung, and G. Pons-Moll. Neural-GIF: Neural generalized implicit functions for animating people in clothing. In *ICCV 2021, IEEE/CVF International Conference on Computer Vision*, Virtual Event, 2021, pp. 11688–11698. IEEE.

[114] O. Unal, D. Dai, and L. Van Gool. Scribble-supervised LiDAR semantic segmentation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2022)*, New Orleans, LA, USA, 2022, pp. 2687–2697. IEEE.

[115] U. Upadhyay, Y. Chen, and Z. Akata. Robustness via uncertainty-aware cycle consistency. In *Neural Information Processing Systems (NeurIPS)*, 2021.

[116] U. Upadhyay, Y. Chen, T. Hepp, S. Gatidis, and Z. Akata. Uncertainty-guided progressive gans for medical image translation. In *Medical Image Computing and Computer Assisted Intervention (MICCAI)*, 2021.

[117] U. Upadhyay, S. Karthik, Y. Chen, M. Mancini, and Z. Akata. Bayescap: Bayesian identity cap for calibrated uncertainty in frozen neural networks. In *European Conference on Computer Vision (ECCV)*, 2022.

[118] A. B. Vasudevan, D. Dai, and L. Van Gool. Sound and visual representation learning with multiple pretraining tasks. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2022)*, New Orleans, LA, USA, 2022, pp. 14596–14606. IEEE.

[119] H. Wang, L. Ding, S. Dong, S. Shi, A. Li, J. Li, Z. Li, and L. Wang. CAGroup3D: Class-aware grouping for 3D object detection on point clouds. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, eds., *Advances in Neural Information Processing Systems 35 (NeurIPS 2022)*, New Orleans, LA, USA, 2022, pp. 29975–29988. Curran Associates, Inc.

[120] H. Wang, C. Shi, S. Shi, M. Lei, S. Wang, D. He, B. Schiele, and L. Wang. DSVT: Dynamic sparse voxel transformer with rotated sets. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2023)*, Vancouver, Canada, 2023. IEEE. Accepted.

[121] H. Wang, S. Shi, Z. Yang, R. Fang, Q. Qian, H. Li, B. Schiele, and L. Wang. RBGNet: Ray-based grouping for 3D object detection. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2022)*, New Orleans, LA, USA, 2022, pp. 1100–1109. IEEE.

[122] H.-P. Wang, N. Yu, and M. Fritz. Hijack-GAN: Unintended-use of pretrained, black-box GANs. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2021)*, Virtual Conference, 2021, pp. 7868–7877. IEEE.

[123] Q. Wang, D. Dai, L. Hoyer, L. Van Gool, and O. Fink. Domain adaptive semantic segmentation with self-supervised depth estimation. In *ICCV 2021, IEEE/CVF International Conference on Computer Vision*, Virtual Event, 2021, pp. 8495–8505. IEEE.

[124] Q. Wang, O. Fink, L. Van Gool, and D. Dai. Continual test-time domain adaptation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2022)*, New Orleans, LA, USA, 2022, pp. 7191–7201. IEEE.

[125] Y. Wang, H. Chen, Y. Fan, W. Sun, R. Tao, W. Hou, R. Wang, L. Yang, Z. Zhou, L.-Z. Guo, H. Qi, Z. Wu, Y.-F. Li, S. Nakamura, W. Ye, M. Savvides, B. Raj, T. Shinozaki, B. Schiele, J. Wang, X. Xie, and Y. Zhang. USB: A unified semi-supervised learning benchmark for classification. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, eds., *Advances in Neural Information Processing Systems 35 (NeurIPS 2022)*, New Orleans, LA, USA, 2022, pp. 3938–3961. Curran Associates, Inc.

[126] Y. Wang, H. Chen, Q. Heng, W. Hou, Y. Fan, Z. Wu, J. Wang, M. Savvides, T. Shinozaki, B. Raj, B. Schiele, and X. Xie. FreeMatch: Self-adaptive thresholding for semi-supervised learning. In *Eleventh International Conference on Learning Representations (ICLR 2023)*, Kigali, Rwanda, 2023. OpenReview.net. Accepted.

[127] X. Xie, B. L. Bhatnagar, and G. Pons-Moll. CHORE: Contact, human and object reconstruction from a single RGB image. In S. Avidan, G. Brostow, M. Cissé, G. Farinella, and T. Hassner, eds., *Computer Vision – ECCV 2022*, Tel Aviv, Israel, 2022, LNCS 13662, pp. 125–145. Springer.

[128] X. Xie, B. L. Bhatnagar, and G. Pons-Moll. Visibility aware human-object interaction tracking from single rgb camera. In *IEEE Conference on Computer Vision and PatternRecognition (CVPR)*, 2023.

[129] W. Xu, Y. Xian, J. Wang, B. Schiele, and Z. Akata. Vgse: Visually-grounded semantic embeddings for zero-shot learning. In *CVPR*, 2022.

[130] J. Yang, S. Shi, R. Ding, Z. Wang, and X. Qi. Towards efficient 3D object detection with knowledge distillation. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, eds., *Advances in Neural Information Processing Systems 35 (NeurIPS 2022)*, New Orleans, LA, USA, 2022, pp. 21300–21313. Curran Associates, Inc.

[131] Z. Yang, L. Jiang, Y. Sun, B. Schiele, and J. Jia. A unified query-based paradigm for point cloud understanding. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2022)*, New Orleans, LA, USA, 2022, pp. 8531–8541. IEEE.

[132] J.-N. Zaech, A. Liniger, M. Danelljan, D. Dai, and L. Van Gool. Adiabatic quantum computing for multi object tracking. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2022)*, New Orleans, LA, USA, 2022, pp. 8801–8812. IEEE.

[133] X. Zhang, B. L. Bhatnagar, S. Starke, V. Guzov, and G. Pons-Moll. COUCH: Towards controllable human-chair interactions. In S. Avidan, G. Brostow, M. Cissé, G. Farinella, and T. Hassner, eds., *Computer Vision – ECCV 2022*, Tel Aviv, Israel, 2022, LNCS 13665, pp. 518–535. Springer.

[134] K. Zhou, B. L. Bhatnagar, J. E. Lenssen, and G. Pons-Moll. TOCH: Spatio-temporal object correspondence to hand for motion refinement. In S. Avidan, G. Brostow, M. Cissé, G. Farinella, and T. Hassner, eds., *Computer Vision – ECCV 2022*, Tel Aviv, Israel, 2022, LNCS 13663, pp. 1–19. Springer.

[135] Y. Zhou, W. Xiang, C. Li, B. Wang, X. Wei, L. Zhang, M. Keuper, and X. Hua. SP-ViT: Learning 2D spatial priors for vision transformers. In *33rd British Machine Vision Conference (BMVC 2022)*, London, UK, 2022, Article 564. BMVA Press.

## Technical reports and preprints

[1] A. Abbas and P. Swoboda. DOGE-Train: Discrete Optimization on GPU with End-to-end Training. *arXiv preprint arXiv:2205.11638*, 2022.

[2] A. Abbas and P. Swoboda. ClusterFuG: Clustering Fully connected Graphs by Multicut. *arXiv preprint arXiv:2301.12159*, 2023.

[3] S. Agnihotri, J. Gabrinski, and M. Keuper. Context matters: the role of upsampling in pixel-wise prediction tasks. In *Under Submission to the International Conference of Computer Vision (ICCV 2023)*, 2023.

[4] M. D. Böhle, M. Fritz, and B. Schiele. Holistically Explainable Vision Transformers. *arXiv*, 2023. arxiv.org/abs/2301.08669.

[5] E. Duka, A. Kukleva, and B. Schiele. *Leveraging Self-Supervised Training for Unintentional Action Recognition*, 2022. arXiv: 2209.11870.

[6] Y. Fan, A. Kukleva, D. Dai, and B. Schiele. Ssb: Simple but strong baseline for boosting performance of open-set semi-supervised learning. In *Under Submission to the International Conference of Computer Vision (ICCV 2023)*, 2023.

[7] J. Geiping, J. Lukasik, M. Keuper, and M. Moeller. Differentiable architecture search: a one-shot method? In *Under Submission to the AutoML Conference (AutoML23)*, 2023.

[8] X. Hong, V. Demberg, A. Sayeed, Q. Zheng, and B. Schiele. Visual coherence loss for coherent and visually grounded story generation. In *Under Submission*, 2023.

[9] M. M. Losch, D. Stutz, B. Schiele, and M. Fritz. Certified robust models with slack control and large lipschitz constants. In *Under Submission*, 2022.

[10] J. Lukasik, P. Gavrikov, J. Keuper, and M. Keuper. Filter frequency regularization for improved native cnn robustness. In *Under Submission to the International Conference of Computer Vision (ICCV 2023)*, 2023.

[11] P. Müller, A. Braun, and M. Keuper. Classification robustness to common optical aberrations. In *Under Submission to the International Conference of Computer Vision (ICCV 2023)*, 2023.

[12] S. Rao, M. D. Böhle, A. Parchami, and B. Schiele. Using Explanations to Guide Models. *arXiv*, 2023. Under submission at ICCV 2023, IEEE/CVF International Conference on Computer Vision).

[13] S. Rao, M. D. Böhle, and B. Schiele. Better Understanding Differences in Attribution Methods via Systematic Evaluations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023. Under submission. Preprint available at arxiv.org/abs/2303.11884.

[14] M. Segu, B. Schiele, and F. Yu. Darth: Holistic test-time adaptation for multiple object tracking. In *Under Submission to the International Conference of Computer Vision (ICCV 2023)*, 2023.

[15] K. Zhou, B. L. Bhatnagar, B. Schiele, and G. Pons-Moll. *Adjoint Rigid Transform Network: Task-conditioned Alignment of 3D Shapes*, 2021. arXiv: 2102.01161.

## Theses

[1] S. Alaniz. *Explainability in Deep Learning by Means of Communication*. PhD thesis, Universität des Saarlandes, Saarbrücken, 2022.

[2] B. Bhatnagar. *Modelling 3D Humans: Pose, Shape, Clothing and Interactions*. PhD thesis, Universität des Saarlandes, Saarbrücken, 2023.

[3] A. Bhattacharyya. *Long-term future prediction under uncertainty and multi-modality*. PhD thesis, Universität des Saarlandes, Saarbrücken, 2021.

[4] A. Horňáková. *Lifted Edges as Connectivity Priors for Multicut and Disjoint Paths*. PhD thesis, Universität des Saarlandes, Saarbrücken, 2022.

[5] Y. Liu. *Learning from Imperfect Data, Incremental Learning an Few-shot Learning*. PhD thesis, Universität des Saarlandes, Saarbrücken, 2023.

[6] M. Omran. *From Pixels to People*. PhD thesis, Universität des Saarlandes, Saarbrücken, 2021.

[7] R. Shetty. *Adversarial Content Manipulation for Analyzing and Improving Model Robustness*. PhD thesis, Universität des Saarlandes, Saarbrücken, 2021.

[8] D. Stutz. *Understanding and Improving Robustness and Uncertainty Estimation in Deep Learning*. PhD thesis, Universität des Saarlandes, Saarbrücken, 2022.

# 29 D3: Internet Architecture

## 29.1 Personnel

**Head of Group**

Prof. Anja Feldmann, Ph.D.

**Associated Tenure Track Faculty**

Yiting Xia, Ph.D. (from October 2020)

**Researchers**

Dr. Oliver Gasser
Devashish Gosain, Ph.D. (until July 2022, now PostDoc KU Leuven, Belgium)
Savvas Zannettou, Ph.D. (until October 2021, now Assistent Professor TU Delft, Netherlands)
Aleksandr Zavodovski, Ph.D. (May 2022 – Aug. 2022, now researcher Ericsson)
Dr.-Ing. Tobias Fiebig (from April 2022)
Thi Thu Ha Dao (from April 2023)
Mannat Kauer (from April 2023)

**Ph.D. Students**

Seifeddine Fathalli (from April 2019)
Qi Guo (from September 2022)
Pascal Hennen (from January 2023)
Fahad Hilal (from June 2022)
Mohamad Hoseini
Franziska Lichtblau (until April 2022, now SAP SE, Germany)
Friedemann Lipphardt (from April 2021)
Aniss Maghsoudlou (till April 2023, then New Yorker, Berlin)
Cristian Munteanu (from March 2021)
Fariba Osali (from August 2022)
Victor Pădurean (from September 2021)
Mirko Palmer (until June 2022)
Lars Prehn (until April 2023)
Seyedali Rasaii (from October 2021)
Said Jawad Saidi (until August 2022, now Huawei Technologies, Munich Reseach Center, Germany)
Khwaja Zubair Sediqi (from November 2020)
Florian Streibelt (until December 2022, Now SAP SE, Germany)
Florian Steurer (from November 2022)

Emilia Ndilokelwa Weyulu (from Oktober 2018)
Danesh Zeynali (from October 2020)

**Secretary/Technical staff**

Iris Wagner
Rainer May (until October 2022)
Jörg Dorchain

**Associated Researchers/Guest**

Dr. Volker Stocker (Research Group Leader, Weizenbaum Institute, Berlin)
Balakrishnan Chandrasekaran, Ph.D. (Assistant Prof. Vrije Universiteit Amsterdam (since November 2020))
Dr. Christoph Dietzel (Global Head of Products & Research, DE-CIX)
Prof. Georgios Smaragdakis, Ph.D. (Professor TU-Delft)
Daniel Wagner (external Ph.D. candidate DE-CIX, Frankfurt, since December 2020)

## 29.2 Visitors

In 2021, no researchers visited the group in person due to the pandemic situation and the related travel restrictions. Therefore, we offered some researchers the opportunity to do remote interships:

| | | |
|---|---|---|
| Dixi Yao | July 2021–January 2022 | Shanghai Jiao Tong University, China |
| Seyedali Rasaii | April–August 2021 | Sharif University of Technology, Iran |
| Vinicius Dantas de Lima Melo | May–July 2021 | University of Toronto, Canada |
| Ling Chen | May–August 2021 | Boston University, USA |
| Vishal Bindal | May–August 2021 | Indian Institute of Technology, India |
| Felipe Andres Gonzalez Pizarro | June–August 2021 | Universidad Técnica Federico Santa María (UTFSM), Chile |

From February 2022 to February 2023, the following researchers visited our group:

| | | |
|---|---|---|
| Yash Vekaria | June–September 2022 | University of California, Davis, USA (fellowship) |
| Thi Thu Ha Dao | June–August 2022 | University for Advanced Studies, Sokendai, Japan (fellowship) |
| Georgios Smaragdakis | 05–08 June 2022 | TU Delft (research collaboration) |
| Georgios Smaragdakis | 23–25 February 2023 | TU Delft (research collaboration) |

## 29.3 Group Organization

The department has a flat organizational hierarchy and currently consists of the director, two senior researchers, one postdoctoral researcher, 17 Ph.D. students, a secretary, and a technical staff member. In addition, the research groups of Yiting Xia, tenure track faculty member at MPI-INF, and Volker Stocker, Research Group Leader at the Weizenbaum Institute, are associated with the department.

Dr. Savvas Zannettou, Postdoc in the group since December 2019, received a call from TU Delft and joined them in November 2021 as Assistant Professor. Dr. Devashish Gosain, Postdoc in the group from November 2020 to November 2022 moved to KU Leuven in Belgium. Both are still actively collaborating with members of the department. Moreover, the following researchers are closely associated with the department: Daniel Wagner (external Ph.D. candidate, DE-CIX), Dr. Christoph Dietzel (Head of research DE-CIX), as well as Prof. Georgios Smaragdakis (TU Delft).

After two years of more or less stagnant hiring due to the pandemic, we were able to hire a postdoctoral researcher and several new PhD students: Dr.-Ing. Tobias Fiebig joined the group as PostDoc in April 2022, coming from TU Delft, Netherlands. Two students from the Saarland University who did their Masterthesis with INET decided to join the group as Ph.D. students. Another PhD student joined the group via the Computer Science @ Max Planck programm. Three Ph.D. students were selected via IMPRS's open calls, while another Ph.D. student was hired following a direct application. Moreover, one female postdoctoral researcher, who previously visited us, joined us in April this year.

The group meets regularly at least once a week. The meeting serves multiple purposes: (a) discussion of administrative and technical issues, (b) introduction and integration of new members, (c) sharing of research progress, (d) discussion of meta topics, and (e) (especially during the pandemic) boosting the team morale and enhance the feeling of "togetherness". Apart from that, there are often scientific talks held by guest researchers or Bachelor / Master students during or directly after the meeting. Since 2020, the meeting was held only online. Now as the pandemic restrictions have been lifted, we switched the meeting to a hybrid mode, combining the online meeting with an in-person meeting, which has proven to be efficient. We still keep the informal daily 'virtual morning coffee meeting' active to compensate for the lack of random, 'hall-way', social interactions. In 2022, we were finally able to meet in person during a three-days retreat, which was a great opportunity for all to meet and get to know their colleagues, some of which had not met before in person due to the pandemic circumstances. In addition, members of the group meet regularly for social activities such as gaming nights or sportive activities.

The remote collaboration tools introduced during the pandemic, namely, among others, Skype, Zoom, BigBlueButton, Mattermost, Nextcloud, and Git-labs are still being used, as these tools have proven to be an effective way to collaborate within the group and also with external partners, who sometimes live in different countries and time zones. We extended the group's wiki page and continue to use our online calendars where group members can get up-to-date information on infrastructure resources, research data sets and projects, seminars and talks, and conference deadlines. In addition, we introduce a whiteboard where group members can use magnets to indicate if they are in the office, working remotely, or on vacation.

The department currently has **two senior researchers**: Dr. Oliver Gasser and Dr.-Ing. Tobias Fiebig.

**Dr. Oliver Gasser** joined the group as a postdoc in February 2020. He did his Ph.D. on the topic of *Evaluating Network Security Using Internet-wide Measurements* at the Technical University of Munich, which he completed with highest distinction. In July 2021 he was appointed Senior Researcher at the Max Planck Institute for Informatics and is since leading the *Internet Security Measurements* research group. In December 2021 he was also elected Scientific Staff Member Representative. His current research focuses on analyzing the impact of IPv6 deployments in the Internet, investigating security and privacy shortcomings of network protocols, and examining the deployment of different protocols such as MPTCP and various VPN protocols in the Internet. His research areas include Internet-wide measurements of hosts and networks over IPv4 and IPv6, security of network protocols and their deployment, DNS deployment and routing. Dr. Gasser has published multiple papers at conferences such as ACM IMC, ACM CoNEXT, and PAM, and journals such as ACM SIGCOMM CCR. He won best paper paper awards at PAM 2018, 2022 and the community contribution award at ACM IMC 2017. Dr. Gasser served on the technical program committee (TPC) of multiple conferences, including ACM IMC, ACM CoNEXT, PAM, TMA. In addition, he also served as a reviewer for multiple journals, such as ACM SIGCOMM CCR, IEEE/ACM Transactions on Networking, ACM Transactions on Internet Technology, IEEE Communications Letters, IEEE Transactions on Network and Service Management, and Elsevier Computer Networks.

**Dr.-Ing. Tobias Fiebig** joined the group as a postdoc in April 2022. Before that, he held a permanent faculty position as an Assistant Professor at TU Delft, after having completed his PhD on empirically assessing security misconfigurations at Internet scale. In March 2023 he was appointed Senior Researcher at the Max Planck Institute for Informatics and is since leading the *Dependable Digital Infrastructures* research group. His current research focuses on improving our understanding of how we can build and maintain digital infrastructure that is systainable, dependable, and secure, while bridging the gap between technology and human factors. For this, he uses an interdisciplinary mix of methods from a variety of fields, ranging from technical protocol analysis, e.g., analyzing why the email ecosystem got so complex, over network measurements, e.g., measuring how universities migrated to cloud infrastructure of just a few hypergiants, to qualitative methods from the social sciences, e.g., investigating the role of system administrators' gender in their work. Dr.Ing. Fiebig regularly publishes in a variety of venues ranging from security and privacy (IEEE S&P, ACM CCS, USENIX Security, ISOC NDSS, IEEE EuroS&P, PoPETS), networking and systems (USENIX ATC, PAM, IEEE MSR), to Human Factors (ACM CSCW). One of the PAM 2023 papers he co-authored received a best-paper award. Dr.Ing. regularly serves in program committees of top-tier venues, including PAM, IEEE Security & Privacy, and ACM CCS. For his service in the latter in 2022, he received an outstanding reviewer award.

An overview of the research work of Oliver Gasser and Tobias Fiebig can be found in their respective sections.

# 29.4 Internet Traffic Analysis

## 29.4.1 Deep Dive into the IoT Backend Ecosystem

*Investigators: Said Jawad Saidi, Oliver Gasser, and Anja Feldmann in cooperation with Srdjan Matic (IMDEA Software Institute) and Georgios Smaragdakis (TU Delft)*

Internet of Things (IoT) devices are becoming increasingly ubiquitous, e.g., at home, in enterprise environments, and in production lines. To support the advanced functionalities of IoT devices, IoT vendors as well as service and cloud companies operate IoT backends—the focus of this research project. In this project [1], we proposed a methodology to identify and locate them by (a) compiling a list of domains used exclusively by major IoT backend providers and (b) then identifying their server IP addresses. Our developed methodology infers the network and physical location of major IoT backend providers. It relies on a fusion of information from public documentation, passive DNS, and active measurements. For analyzing IoT traffic patterns, we relied on passive network flows from a major European ISP. Our analysis also showed that it is not unusual for IoT protocols, e.g., MQTT, to use non-standard ports or reuse Web ports. The latter makes the identification of IoT backend infrastructure as well as IoT traffic challenging using traditional methods—our proposed methodology resolves this issue. Our analysis focused on the top IoT backends and unveiled diverse operational strategies with regard to deployment, operation, and dependencies: Some IoT backends operate their own infrastructure while others utilize the public cloud. We found that the majority of the top IoT backend providers are located in multiple locations and countries. Still, a handful are located only in one country, which could raise regulatory scrutiny as the client IoT devices are located in other regions. We noticed that a substantial fraction (around 35%) of IoT traffic is exchanged with IoT backend servers outside Europe, which raises both performance and regulatory concerns. We also found that at least six of the top IoT backends rely on other IoT backend providers. Our traffic analysis highlighted that both the IoT population and activity per application differ vastly. While some applications behave more like the typical user-generated traffic, i.e., diurnal patterns, peak evening hours, and downstream-heavy; this is not the case for all IoT applications. We also evaluated if cascading effects among the IoT backend providers are possible in the event of an outage, a misconfiguration, or an attack. Finally, we discussed shortcomings of the current IoT backend ecosystem and assessed the impact of a large-scale outage in one of the major IoT backend providers on IoT connectivity to backend servers.

### References


- [1] S. J. Saidi, S. Matic, O. Gasser, G. Smaragdakis, and A. Feldmann. Deep dive into the IoT backend ecosystem. In C. Barakat and C. Pelsser, eds., *IMC '22, ACM Internet Measurement Conference*, Nice, France, 2022, pp. 488–503. ACM.


## 29.4.2 Analysis of the COVID-19 pandemic

*Investigators: Franziska Lichtblau, Oliver Gasser, Anja Feldmann, Christoph Dietzel (DE-CIX/MPI-INF), Daniel Wagner (DE-CIX/MPI-INF), Georgios Smaragdakis (TU Berlin/MPI-INF) in cooperation with Enric Pujol (BENOCS), Ingmar Poese (BENOCS),*

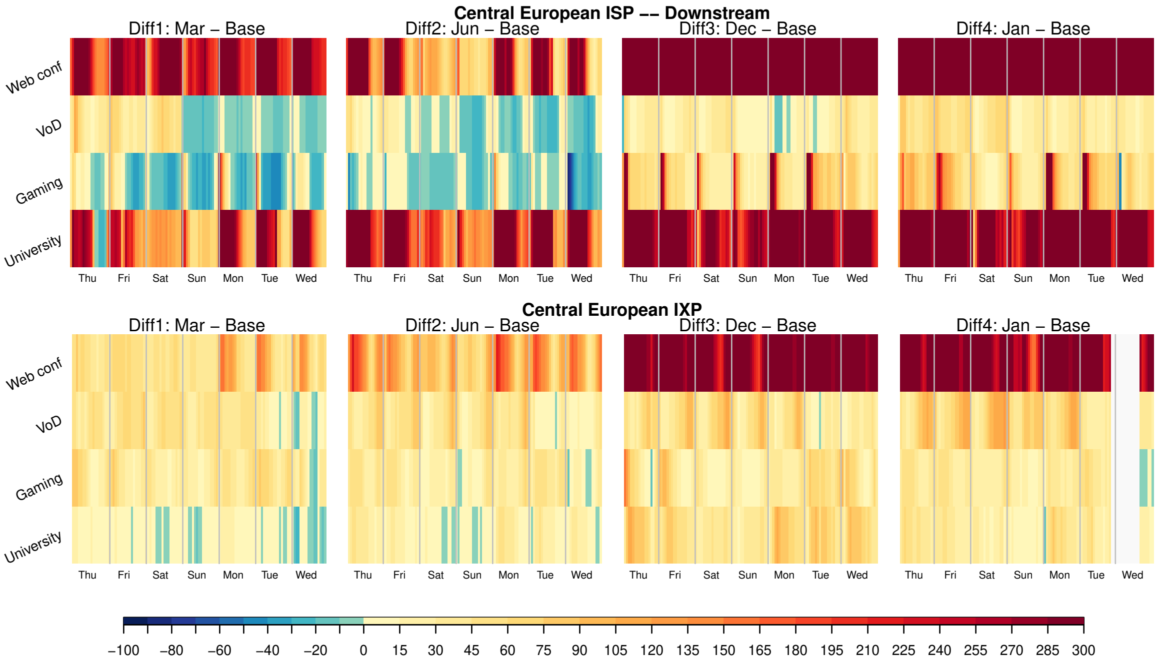



Figure 29.1: ISP (top), IXP-CE (bottom) heatmaps of application classes' traffic at the ISP and IXPs during COVID-19 pandemic: spring and fall waves. Each subplot shows the change in the aggregated traffic volume per hour for the respective class compared to the base week in February 2020. White areas mark missing data.

*Matthias Wichtlhuber (DE-CIX), Juan Tapiador (Universidad Carlos III de Madrid), Narseo Vallina-Rodriguez (IMDEA Networks/ICSI), Oliver Hohlfeld (Brandenburg University of Technology)*

In January 2020 the COVID-19 pandemic, a corona virus variant spreading across the entire globe, reached central Europe. In the beginning, many people underestimated the impact the virus would have on our entire life in the coming months. Once the first lockdown was imposed in March, it quickly became clear that things would drastically change. The Internet was a vital component to maintain as much normality as possible: Many people started to work from home; with shops partially closed, online shopping became even more relevant; artists made first attempts to stream concerts; video streaming at home quickly became one of the few possible leisure activities; and schools relied on remote teaching. With this increased demand the question arose whether the Internet can serve that demand (or if it was near a collapse). To investigate this issue we started the COVID measurement project [1, 2, 3]. The fundamental question of this project is if increased Internet traffic demands of residential users, in particular, for remote working, entertainment, commerce, and education due to the lockdown caused traffic shifts in the Internet core. We rely on a diverse set of vantage points, namely one Internet Service Provider (ISP), three Internet Exchange Point (IXP), and one metropolitan educational network, to shed light on the questions from as many angles as possible. The ISP provides us with a view on residential customers and small enterprises; the IXP facilitates interconnection between different businesses; and the educational network allows us to study the changes on a university campus. We find that the traffic volume increased by 15-20% almost within a week — while overall still modest, this constitutes a large increase within this short time period. However, despite this surge, we observe that the Internet infrastructure can handle the new volume, as most traffic shifts occur outside of traditional peak hours. When looking directly at the traffic sources, it turns out that, while hypergiants (a group of networks responsible for a large fraction of Internet traffic) still contribute a significant fraction of traffic, we see (1) a higher increase in traffic of non-hypergiants, and (2) traffic increases in applications that people use when at home, such as Web conferencing, VPN, services offered by universities and research networks and gaming. Figure 29.1 illustrates this development[1]. We visualize two weeks in the Spring and Fall waves (of the pandemic), namely, the second week in March 2020, June 2020, December 2020, and January 2021, as the difference to a base week before the initial lockdowns began, i.e., February 20–26, 2020. Each column represents one hour of a day.

**Web conferencing:** Web conferencing applications have seen a dramatic surge at both vantage points as the pandemic dramatically changed the way people communicate in their personal and professional lives. Notably, in December and January, the extreme growth also persisted during weekends. This observation indicates that not only has work life moved online but so also have private social activities.

**Video-on-Demand:** Video streaming applications' usage shows high growth. The ISP only sees a moderate growth in the first half of March followed by a reduction of volume below the pre-COVID-19 reference time frame. We attribute this observation to major streaming companies reducing their streaming resolution in Europe by mid-March for 30 days. The IXP sees a similar, but not that much pronounced trend. However, there is a significant increase of traffic in June, December and January, that exceeds 200% (IXP) and 100% (ISP) for some days, especially on weekends, indicating that more people stayed at home, rather than going outside, during leisure time.

**Gaming:** The strong growth of gaming applications is conspicuous at the IXP, especially during the day. While the ISP shows a significant increase during morning hours, it generally exhibits a declining trend in the Spring wave.[2] As the initial download of a game nowadays supersedes the amount of data transferred while playing these high levels may relate to new releases or updates of popular games. Gaming applications, typically used in the evening or at weekends, are now used at any time. The trend starts to flatten in June—this may be due to people going on vacation or spending more time outside. The ISP observed an increase up to 300% in gaming-'related traffic during the fall wave with emphasis on the first half of the day. A similar pattern unfolds at the IXP, but with comparatively smaller increases.

**University networks:** Traffic that originates from such networks at both vantage points behaves similarly, albeit that observed at the ISP shows a pronounced increase. Both vantage points see a high increase in traffic especially during the fall wave with a growth of 100% and more. This growth could be attributed to some European educational networks providing video conferencing solutions, which are now being used by customers of the ISP and IXP.

[1] For details on VPN traffic, refer to [1, 3]

[2] Note, that this effect is mainly caused by unusually high traffic levels in this category during our baseline week in February 2020.

In December 2020 and January 2021 most academic collaborations and teaching activities moved to an online setting. This observation is in line with the comparatively smaller surge of activity at weekends.

While many networks observe increased traffic demands, those providing services to residential users and academic networks, in particular, experience major overall decreases. Yet, in these networks, we can observe substantial increases when considering applications associated with remote working and lecturing. Overall, we indeed find shifts in the traffic usage patterns, especially for residential customers. Nevertheless, from our perspective the Internet coped quite well with the increased demand. At our vantage points quick reaction times, well-organized infrastructures, and flexible capacity planning allowed them to accommodate the quick change and new customer needs. A follow-up article on this work sheds light on the developments during the fall 2020 wave of the pandemic. Here, we highlight that many of the patterns which emerged during spring 2020 did indeed stay and define a "new normal" of Internet usage by eyeball customers.

## References

[1] A. Feldmann, O. Gasser, F. Lichtblau, E. Pujol, I. Poese, C. Dietzel, D. Wagner, M. Wichtlhuber, J. Tapiador, N. Vallina-Rodriguez, O. Hohlfeld, and G. Smaragdakis. The lockdown effect: Implications of the COVID-19 pandemic on internet traffic. In *IMC '20, 20th ACM Internet Measurement Conference*, Virtual Event, USA, 2020, pp. 1–18. ACM.

[2] A. Feldmann, O. Gasser, F. Lichtblau, E. Pujol, I. Poese, C. Dietzel, D. Wagner, M. Wichtlhuber, J. Tapiador, N. Vallina-Rodriguez, O. Hohlfeld, and G. Smaragdakis. A view of internet traffic shifts at ISP and IXPs during the COVID-19 pandemic. In *COVID-19 Network Impacts Workshop*, Virtual Workshop, 2020. IAB.

• [3] A. Feldmann, O. Gasser, F. Lichtblau, E. Pujol, I. Poese, C. Dietzel, D. Wagner, M. Wichtlhuber, J. Tapiador, N. Vallina-Rodriguez, O. Hohlfeld, and G. Smaragdakis. A year in lockdown: How the waves of COVID-19 impact Internet traffic. *Communications of the ACM*, 64(7):101–108, 2021.

### 29.4.3 FlowDNS: Correlating Netflow and DNS Streams at Scale

*Investigators: Aniss Maghsoudlou, Oliver Gasser, and Anja Feldmann in cooperation with Ingmar Poese (BENOCS GmbH)*

Knowing customer's interests, e.g. which Video-On-Demand (VoD) or Social Network services they are using, helps telecommunication companies with better network planning to enhance the performance exactly where the customer's interests lie, and also offer the customers relevant commercial packages. However, with the increasing deployment of CDNs by different services, identification, and attribution of the traffic on network-layer information alone becomes a challenge: If multiple services are using the same CDN provider, they cannot be easily distinguished based on IP prefixes alone. Therefore, it is crucial to go beyond pure network-layer information for traffic attribution. In this work [1], we leverage real-time DNS responses gathered by the clients' default DNS resolvers. Having these DNS responses and correlating them with network-layer headers, we are able to translate CDN-hosted domains to the actual services they belong to. We design a correlation system for this purpose and deploy it at a large European ISP. With our system, we can correlate an average of 81.7% of the traffic with the corresponding services, without any loss on our live data streams. Our correlation results also show that 0.5% of the daily traffic contains malformatted, spamming, or phishing domain names. Moreover, ISPs can correlate the results with their BGP information to find more details about the origin and destination of the traffic.

### References


- [1] A. Maghsoudlou, O. Gasser, I. Poese, and A. Feldmann. FlowDNS: Correlating netflow and DNS streams at scale. In G. Bianchi and A. Mei, eds., *CoNEXT '22, 18th International Conference on Emerging Networking Experiments And Technologies*, Roma, Italy, 2022, pp. 187–195. ACM.


## 29.4.4 Characterizing the VPN Ecosystem in the Wild

*Investigators: Aniss Maghsoudlou, Lukas Vermeulen, and Oliver Gasser in collaboration with Ingmar Poese (BENOCS GmbH)*

With the shift to working remotely after the COVID-19 pandemic, the use of Virtual Private Networks (VPNs) around the world has nearly doubled. Therefore, measuring the traffic and security aspects of the VPN ecosystem is more important now than ever. It is, however, challenging to detect and characterize VPN traffic since some VPN protocols use the same port number as web traffic and port-based traffic classification will not help. VPN users are also concerned about the vulnerabilities of their VPN connections due to privacy issues. In this work [1], we aim at detecting and characterizing VPN servers in the wild, which facilitates detecting the VPN traffic. To this end, we perform Internet-wide active measurements to find VPN servers in the wild, and characterize them based on their vulnerabilities, certificates, locations, and fingerprinting. We find 9.8M VPN servers distributed around the world using OpenVPN, SSTP, PPTP, and IPsec, and analyze their vulnerability. We find SSTP to be the most vulnerable protocol with more than 90% of detected servers being vulnerable to TLS downgrade attacks. Of all the servers that respond to our VPN probes, 2% also respond to HTTP probes and therefore are classified as Web servers. We apply our list of VPN servers to the traffic from a large European ISP and observe that 2.6% of all traffic is related to these VPN servers.

### References


- [1] A. Maghsoudlou, L. Vermeulen, I. Poese, and O. Gasser. Characterizing the VPN ecosystem in the wild. In A. Brunstrom, M. Flores, and M. Fiore, eds., *Passive and Active Measurement (PAM 2023)*, Virtual Event, 2023, LNCS 13882, pp. 18–45. Springer.


## 29.4.5 A Longitudinal View at the Adoption of Multipath TCP

*Investigators: Danesh Zeynali and Oliver Gasser in cooperation with Jorg Ott, Nitinder Mohan (TUM Germany), and Tanya Shreedhar (IIIT-Delhi India)*

The most widely used transport protocol on the Internet, the Transmission Control Protocol (TCP), enables reliable data transmission between any two endpoints on the Internet. The

TCP protocol establishes a connection between two endpoint addresses, a source address, and a destination address. Today, however, devices can have more than one address, for example, if they use both the WiFi interface and the mobile data interface (e.g., 4G) of a cell phone. This is where Multipath TCP (MPTCP) comes in, as it extends traditional TCP to allow the simultaneous use of multiple connection endpoints at the source and destination, resulting in better resource utilization, increased aggregate throughput, and increased resilience to network failures. MPTCP has been under active development since it was standardized by the IETF in the spring of 2013. Because of MPTCP's performance advantages over TCP, several well-known companies have incorporated the protocol into their products and services. Apple uses MPTCP in its iOS devices to improve the user experience related to its system services, such as Siri, Music, Maps, and Wi-Fi Assist. In 2019, Apple provided third-party developers with APIs to use MPTCP in iOS in non-system iOS applications. Korea Telecom is using MPTCP in collaboration with Samsung to provide gigabit speeds over Wi-Fi and LTE to its customers. Moreover, in February 2020, MPTCPv1 was deployed onto Linux and is now available to all users running Linux 5.6 or later.

This project [2] represents the first broad and multi-faceted assessment of MPTCPv0 and MPTCPv1. We examine both the infrastructure in terms of MPTCP-enabled IPv4 and IPv6 addresses and the traffic share at two geographically diverse sites. We identify and remove middleboxes (i.e., systems that interfere with connections on the path) that affect MPTCP in the wild, and investigate whether they also negatively affect MPTCP application traffic. During our 18-month study period, we observed a steady increase in MPTCP-enabled IPs supporting HTTP and HTTPS. In December 2021, we reached 13k+ and 1k for IPv4 and IPv6, respectively. The growth is primarily driven by Apple and Internet service providers around the world that rely on the protocol to improve their services. This project builds upon the findings of the previous MPTCP measurement study [1].

### References


- [1] F. Aschenbrenner, T. Shreedhar, O. Gasser, N. Mohan, and J. Ott. *From Single Lane to Highways: Analyzing the Adoption of Multipath TCP in the Internet*, 2021. arXiv: 2106.07351.
- [2] T. Shreedhar, D. Zeynali, O. Gasser, N. Mohan, and J. Ott. *A Longitudinal View at the Adoption of Multipath TCP*, 2022. arXiv: 2205.12138.


## 29.4.6 Hyper-specific Prefixes: Gotta Enjoy the Little Things in Interdomain Routing

*Investigators: Khwaja Zubair Sediqi, Lars Prehn, and Oliver Gasser*

Autonomous Systems (ASes) exchange reachability information between each other using BGP—the de-facto standard inter-AS routing protocol. While IPv4 (IPv6) routes more specific than /24 (/48) are commonly filtered (and hence not propagated), route collectors still observe many of them. In this work, we take a closer look at those "hyper-specific" prefixes (HSPs). In particular, we analyze their prevalence, use cases, and whether operators use them intentionally or accidentally. While their total number increases over time, most

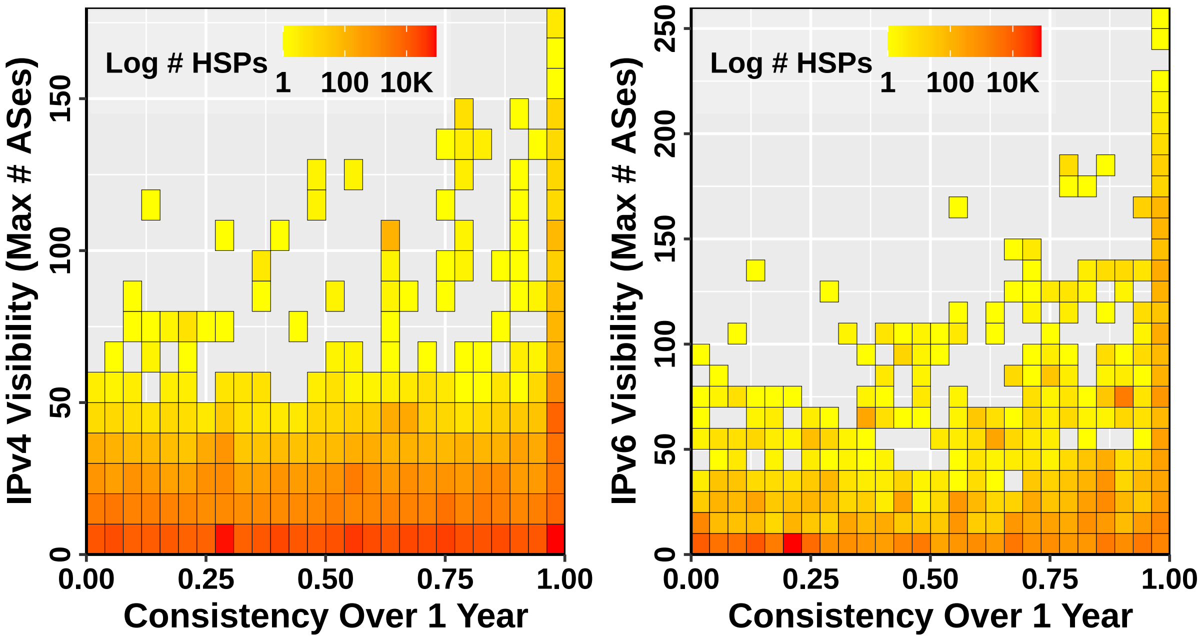



Figure 29.2: Heatmap showing HSP visibility and consistency for IPv4 (left) and IPv6 (right)

HSPs can only be seen by route collector peers. We examine how long HSPs are visible and how far they propagate in terms of the number of ASes (i.e., peer ASes of route collector) on the Internet. We analyzed BGP RIBs and updates for 2020 and plotted them in Figure 29.2 with IPv4 (left) and IPv6 (right) heatmaps.

In Figure 29.2, the y-axis shows the visibility of an HSP (i.e., the maximum number of peer ASes that saw the prefix), and the x-axis represents the consistency (i.e., the duration of time an HSP was visible). Every heatmap cell represents ten feeder ASes on the y-axis and two weeks duration on the x-axis. A cell's color (or heat) represents the number of HSPs within it (using the log10 scale).

We observe that some HSPs have a life span of less than a week, while others are visible throughout the year. Even though most of the HSPs propagate only locally (i.e., to a few ASes), others are globally visible. Most HSPs represent (internal) routes to peering infrastructure or are related to address block relocations or blackholing. While hundreds of operators intentionally add HSPs to well-known routing databases, we observe that many HSPs are possibly accidentally leaked routes [1].

## References


- [1] Z. Sediqi, L. Prehn, and O. Gasser. Hyper-specific prefixes: Gotta enjoy the little things in interdomain routing. *ACM SIGCOMM Computer Communication Review*, 52(2):20–34, 2022.

### 29.4.7 Analyzing the Reachability of large IXPs

*Investigators: Lars Prehn, Franziska Lichtblau, and Anja Feldmann, in cooperation with Christoph Dietzel (Benocs GmBH)*

In this project, we posed the question: How far can peering at large IXPs get us in terms of reachable prefixes and services? To approach this question, we first analyzed and compared Route Server snapshots obtained from eight of the world's largest IXPs. Afterwards, we performed an in-depth analysis of bi-lateral and private peering at a single IXP based on its peering LAN traffic and queries to carefully selected, nearby looking glasses. To assess the relevance of the prefixes available via each peering type, we utilized two orthogonal metrics: the number of domains served from the prefix and the traffic volume that a large eyeball network egress towards it.

Our results show that multi-lateral peering can cover 20% and 40% of the routed IPv4 and IPv6 address space, respectively. We observed that many of those routes lead to out-of-continent locations reachable only via three or more AS hops. Yet, most IXP members only utilize "local" (i.e., single hop) routes. We further infered that IXP members can reach more than half of all routed IPv4 and more than one-third of all routed IPv6 address space via bi-lateral peering. These routes contain almost all of the top 10K egress prefixes of our eyeball network, and hence they would satisfy the reachability requirements of most end users [1].

#### References


- [1] L. Prehn, F. Lichtblau, C. Dietzel, and A. Feldmann. Peering only? Analyzing the reachability benefits of joining large IXPs today. In O. Hohlfeld, G. Moura, and C. Pelsser, eds., *Passive and Active Measurement (PAM 2022)*, Virtual Event, 2022, LNCS 13210, pp. 338–366. Springer.


### 29.4.8 Biases and Sensitivity of Business Relationship Inferences

*Investigators: Lars Prehn and Anja Feldmann*

The business relationships between Autonomous Systems (ASes) can provide fundamental insights into the Internet's routing ecosystem. Throughout the last two decades, many works focused on how to improve the inference of those relationships. Yet, it has proven difficult to assemble extensive ground-truth data sets for validation. Therefore, more recent works rely entirely on relationships extracted from BGP communities to serve as "best-effort" ground-truth.

In this project, we first highlighted the shortcomings of this trend. We showed that the best-effort validation data does not cover relationships between ASes within the Latin American (LACNIC) service region even though 14% of all inferred relationships are from that region. We further showed that the overall precision of 96-98% for peering relationships achieved by three of the most prominent algorithms can drop by 14-25% when considering only peering relationships between Tier-1 and other transit providers [1].

We further systematically analyzed the sensitivity of the most prominent algorithm, ASRank, to short-term routing dynamics by comparing its output for tens of thousands of slightly altered input data sets. Our results showed that short-lived routing dynamics can account for error rate improvements of up to 5.4×, significantly higher than the improvements claimed by recently proposed algorithms. Based on this insight, we suggested a novel evaluation approach that classifies errors as either transient or consistent and facilitates the identification of algorithmic errors.

## References


• [1] L. Prehn and A. Feldmann. How biased is our validation (data) for AS relationships? In *IMC '21, ACM Internet Measurement Conference*, Virtual Event, USA, 2021, pp. 612–620. ACM.


### 29.4.9 Segment Routing IPv6: Is Anybody out there?

*Investigators: Victor Padurean, Oliver Gasser, and Anja Feldmann in cooperation with Randy Bush*

Segment routing is a modern form of source-based routing, i.e., a routing technique where all or part of the routing decision is predetermined by the source or a hop on the path. Since initial standardization efforts in 2013, segment routing seems to have garnered substantial industry and operator support. Especially segment routing over IPv6 (SRv6) is advertised as having several advantages for easy deployment and flexibility in operations in networks. Many people, however, argue that the deployment of segment routing and SRv6 in particular poses a significant security threat if not done with the utmost care. With this project, we aim to measure the SRv6 deployment leakage of segment routing in the wild and, if found, further study its security implications and usages. As of now, no traces of SRv6 were found, yet we submitted a workshop paper at the 7th International Workshop on Traffic Measurements for Cybersecurity (WTMC 2022)[2] describing our methodology. Emulations: Our emulation experiments show that it is possible to detect SRv6 deployment with traceroute, if SRv6 is implemented with inline mode. BGP Data Analysis: We analyze data collected by 10 BGP collectors, looking for attributes that could indicate SRv6 deployment. We find no such attributes. Furthermore, we conduct a thorough investigation of BGP Communities, especially Color Extended Communities, aiming to find a pattern that may indicate SRv6 usage. Again, we do not find any relevant signal. This means that these attributes are either properly filtered by BGP speakers on the path [1] or the egress routers of SRv6 deployments. Tracerouting Campaign: We conduct traceroute measurements to identify SRv6 deployments in the wild, but do not see any traces of SRv6. Either companies do not use SRv6's inline mode as they claim or they filter SRHs from the returned ICMPv6 packets. In this first empirical study on SRv6 we are unable to find traces of SRv6 deployment even for companies that claim to have it deployed in their networks. This lack of leakage might be an indication of good security practices being followed by network operators when deploying SRv6.

## References


[1] T. Krenc, R. Beverly, and G. Smaragdakis. As-level bgp community usage classification. In *Proceedings of the 21st ACM Internet Measurement Conference*, 2021, pp. 577–592.

• [2] V.-A. Pădurean, O. Gasser, R. Bush, and A. Feldmann. SRv6: Is there anybody out there? In *7th IEEE European Symposium on Security and Privacy Workshops (EUROS&P 2022)*, Genoa, Italy, 2022, pp. 252–257. IEEE.

### 29.4.10 Measuring the IPv6-readiness in the Domain Name System

*Investigators: Florian Steurer, Florian Streibelt, Tobias Fiebig, Oliver Gasser, Anja Feldmann*

The exhaustion of available IPv4 address space and the resulting migration to IPv6 networking is one of the major challenges the internet faces today. In an IPv6-only scenario, it is not sufficient for web services or pages to simply expose an IPv6 address under which they can be reached. Additionally, the Domain Name System (DNS) needs to be able to resolve the domain names of pages and all included resources. The aim of this project is to build a more comprehensive picture of the IPv6-readiness of the web with regard to the DNS.

To achieve this, we are developing a scalable measurement pipeline to scan the DNS for potential problems. We will apply this pipeline to assess the current state of IPv6-readiness in the DNS, quantify misconfigurations and derive concrete recommendations for operators and administrators. First findings from passive measurement indicate, that 44.6% of zones found in the Farsight SIE dataset are not IPv6-resolvable [1], highlighting the extent of the problem.

Once we have finished implementation of our measurement tool, we plan to use it on popular domain lists such as the Tranco1M or Alexa1M in order to quantify DNS misconfigurations and to benchmark it against other state-of-the-art tools. Finally, we plan to assess the IPv6-readiness on a per-website basis until the end of 2023.

#### References


- [1] F. Streibelt, P. Sattler, F. Lichtblau, C. H. Gañán, A. Feldmann, O. Gasser, and T. Fiebig. How ready is DNS for an IPv6-only world? In A. Brunstrom, M. Flores, and M. Fiore, eds., *Passive and Active Measurement (PAM 2023)*, Virtual Event, 2023, LNCS 13882, pp. 525–549. Springer.


## 29.5 Innovative Traffic Control to Future-Proof the Internet

### 29.5.1 Collaborative DDos Mitigation

*Investigators: Daniel Wagner, Anja Feldmann, and Christoph Dietzel (DE-CIX/MPI-INF), in cooperation with Daniel Kopp, Matthias Wichtlhuber (DE-CIX), Oliver Hohlfeld (Brandenburg University of Technology), and Georgios Smaragdakis (TU Delft)*

Our daily lives increasingly rely on services in the Internet. With that, the availability of these services has become of unprecedentedly value. At the same time, they evolved into a decent target for attackers to cause harm. A typical Internet attack is the Distributed Denial of Service (DDoS) attack. Here, an attacker is in control of multiple distributed hosts to send coordinated traffic to overwhelm the target's resources and effectively denying the target to serve benign requests. Despite ongoing research on DDoS detection and mitigation paired with improved understanding about adversary strategies, DDoS attacks are still on the rise and at an all time high. To date, attackers incorporate more sophisticated techniques and exploit by far more different mechanisms and protocols to form DDoS attacks at unparalleled threat levels.

In this project [1], we measure the ability of Internet Exchange Points (IXPs) to mitigate amplification DDoS attacks. Located in the heart of the Internet, they are closer to the source of attacks than conventional mitigation facilities that are typically located at the attack's destination, at the egress edge of the Internet. However, the IXPs' location usually lacks a holistic view on the attack traffic, as routes exist towards the target that bypass IXPs. The remaining fraction of the attack that crosses the IXP may not be large enough for local detection mechanisms to detect the traffic as malicious. To cope with this, we propose a collaboration between IXPs to get a more informed view on the attacks and improve the local attack detection. We unify the data of 11 IXPs across Europe and North America and identify 120k amplification DDoS attack events throughout a period of 6 months.

We find that more than 80% of the attack traffic carried by these IXPs is locally undetected as such. With the help of collaboratively exchanging information about traffic activity, a more comprehensive view on globally distributed DDoS attacks can be applied to the local detection. This helps to detect up to 90% of the attack traffic. For exchanging the required information, we further propose a DDoS information exchange platform. This comes with two different trust scenarios from which collaborating parties can choose to exchange their DDoS information.

## References

• [1] D. Wagner, D. Kopp, M. Wichtlhuber, C. Dietzel, O. Hohlfeld, G. Smaragdakis, and A. Feldmann. United we stand: Collaborative detection and mitigation of amplification DDoS attacks at scale. In Y. Kim, J. Kim, G. Vigna, E. Shi, H. Kim, and J. B. Hong, eds., *CCS '21, ACM SIGSAC Conference on Computer and Communications Security*, Virtual Event, Republic of Korea, 2021, pp. 970–987. ACM.

### 29.5.2 Video Streaming with Cross-layer information Sharing

*Investigators: Mirko Palmer, Qi Guo, and Anja Feldmann, in cooperation with Balakrishnan Chandrasekaran (Vrije Universiteit Amsterdam), Ramesh K. Sitaraman and Kevin Spiteri (UMass Amherst, USA), and Malte Appel (Internet Initiative Japan)*

Today, content consumption on the internet is omnipresent. Since the global pandemic and a move towards working from home, the amount of content consumption, specifically, video-streaming has increased substantially. Though, the network conditions are not always ideal to support the high throughput requirements for content consumption. The state-of-the-art solution for overcoming insufficient throughput for video-streaming is to employ some form of adaptive bitrate (ABR) algorithm. An ABR algorithm selects a specific video quality that has a throughput requirement lower than the available throughput. This selection is repeated every few seconds to adjust to account for a change in the available throughput.

These algorithms, however, are not perfect: they can misjudge the network conditions and either download a quality lower than necessary, impacting a users' quality-of-experience (QoE) or select a quality that requires more data than the current network conditions allow, resulting in stalls due to the video not being delivered in time. The latter results in a significant degradation of user's QoE. Virtually all prior work follow a piecemeal approach—either "tweaking" the fully reliable transport layer or making the client "smarter." Departing from prior work, we follow a holistic approach and design a cross-layer video-streaming solution, called VOXEL [1]. We use VOXEL to demonstrate how to combine application-provided "insights" with a partially reliable protocol for optimizing video streaming.

First, we recognize that some video frames are less important than others, i.e., intelligently dropping specific frames does not degrade visual quality, and thus it does not affect end-users' QoEs. We rank the individual frames constituting each video segment in terms of their impact or influence on the overall quality of the video, and use this ranking to determine when (and where) a reliable delivery is required.

To this end, we present a novel ABR algorithm that explicitly trades off losses for improving end-users' video-watching experiences. This synergy of a video streaming tailored transport, a one time in-depth video analysis and a visual quality aware ABR, results in VOXEL reducing the rebuffering, even in challenging network conditions, in the 90th-percentile, by up to 97%, while providing a visual quality that is at least on-par with state-of-the-art streaming solutions. The rebuffering reduction capabilities of VOXEL were evaluated extensively in a full end-to-end system. We conducted several experiments from emulating a diverse set of network conditions in lab to streaming video over the internet from a datacenter in France. Our results from all experiments show that VOXEL, indeed, is at least on par, but in most cases outperforms the state-of-the-art. To evaluate the objective visual impact of dropping frames, we utilized SSIM for its practicability in terms of its robustness compared to its computational complexity.

### References

• [1] M. Palmer, M. Appel, K. Spiteri, B. Chandrasekaran, A. Feldmann, and R. K. Sitaraman. VOXEL: Cross-layer optimization for video streaming with imperfect transmission. In *CoNEXT '21, 17th International Conference on Emerging Networking Experiments And Technologies*, Virtual Event, Germany, 2021, pp. 359–374. ACM.

### 29.5.3 A concept for a P4-programmable IXPs

*Investigators: Daniel Wagner, Anja Feldmann, and Christoph Dietzel (DE-CIX/MPI-INF), in cooperation with Matthias Wichtlhuber (DE-CIX), and Jeremias Blendin (Intel, Barefoot Switch Division)*

Internet Exchange Points (IXPs) are globally distributed Internet infrastructures, that facilitate exchange of local Internet peering traffic. They vary greatly in size and also in their requirements. They are understood to play a critical role in the Internet's ecosystem and harbor the potential to reduce end-to-end latency of any kind of Internet traffic. However, hardware vendors do not focus on special IXP networking requirements due to the small market share of the latter. IXP operators are thus left to invest in expensive hardware with a unsuitable and bloated feature set. Recent endeavors have incorporated the idea of software-programmable networks to efficiently implement solutions tailored for IXPs to build so-called Software Defined Internet Exchanges (SDXes). In this project [1], we propose a concept of leveraging the programming language P4 to implement a P4-based IXP, the P4IX. We detail on a P4 pipeline concept that is split into various stages, each fulfilling certain demands of small, medium, and large IXPs. We account to the major advantages a significantly reduced time to market of solutions highly tailored for IXPs, a tight and automated integration of the business logic into the productive environment and enormously reduced cost. However, challenges come with this proposal. We critically discuss the required shift in the mindset of network operators to software developers and the need for a comprehensive testing environment of the self-implemented features.

### References

- [1] D. Wagner, M. Wichtlhuber, C. Dietzel, J. Blendin, and A. Feldmann. P4IX: A concept for P4 programmable data planes at IXPs. In *FIRA '22, ACM SIGCOMM 2022 Workshop on Future of Internet Routing & Addressing*, Amsterdam, Netherlands, 2022, pp. 72–78. ACM.

## 29.5.4 Advanced Prefix De-aggregation Attacks

*Investigators: Lars Prehn and Oliver Gasser in cooperation with Pawel Foremski (Institute of Theoretical and Applied Informatics, Polish Academy of Sciences)*

The Internet is a critical resource in the day-to-day life of billions of users. To support the growing number of users and their increasing demands, operators have to continuously scale their network footprint—e.g., by joining Internet Exchange Points (IXPs)—and adopt relevant technologies—such as IPv6. IPv6, however, has a vastly larger address space compared to its predecessor, which allows for new kinds of attacks on the Internet routing infrastructure.

In this project, we revisited prefix de-aggregation attacks in the light of these two changes and introduced Kirin—an advanced BGP prefix de-aggregation attack that sources millions of IPv6 routes and distributes them via thousands of sessions across various IXPs to overflow the memory of border routers within thousands of remote ASes. Kirin's highly distributed nature allows it to bypass traditional route-flooding defense mechanisms, such as per-session prefix limits or route flap damping. We analyzed the theoretical feasibility of the attack by formulating it as a Integer Linear Programming problem, tested for practical hurdles by deploying the infrastructure required to perform a small-scale Kirin attack using 4 IXPs, and validated our assumptions via BGP data analysis, real-world measurements, and router testbed experiments. Despite its low deployment cost, we found Kirin capable of injecting lethal amounts of IPv6 routes in the routers of thousands of ASes [1].

### References

- [1] L. Prehn, P. Foremski, and O. Gasser. *Kirin: Hitting the Internet with Millions of Distributed IPv6 Announcements*, 2022. arXiv: 2210.10676.

## 29.5.5 Automatic Detection of Fake Key Attacks in Secure Messaging

*Investigators: Devashish Gosain in cooperation with Tarun Kumar Yadav, Amir Herzberg, Daniel Zappala, and Kent Seamons.*

Popular instant messaging applications such as WhatsApp and Signal provide end-to-end encryption for billions of users. They rely on a centralized, application-specific server to distribute public keys and relay encrypted messages between the users. Therefore, they prevent passive attacks but are vulnerable to some active attacks. A malicious or hacked server can distribute fake keys to users to perform man-in-the-middle or impersonation attacks. While typical secure messaging applications provide a manual method for users to detect these attacks, this burdens users, and studies show it is ineffective in practice.

In our reserach [1], we present a completely automated approach for key verification (KTACA) that is oblivious to users and easy to deploy. We motivate KTACA by designing two approaches to automatic key verification. One approach uses client auditing (KTCA) and the second uses anonymous key monitoring (AKM). Both have relatively inferior security properties, leading to KTACA, which combines these approaches to provide the best of both worlds. We provide a security analysis of each defense, identifying which attacks they can automatically detect. We implement the active attacks to demonstrate they are possible, and we also create a prototype implementation of all the defenses to measure their performance and confirm their feasibility. Finally, we discuss the strengths and weaknesses of each defense, the overhead on clients and service providers, and deployment considerations.

### References

• [1] T. K. Yadav, D. Gosain, A. Herzberg, D. Zappala, and K. Seamons. Automatic detection of fake key attacks in secure messaging. In H. Yin, A. Stavrou, C. Cremers, and E. Shi, eds., *CCS '22, 28th ACM SIGSAC Conference on Computer and Communications Security*, Los Angeles, CA, USA, 2022, pp. 3019–3032. ACM.

## 29.5.6 MiXiM: Mixnet Design Decisions and Empirical Evaluation

*Investigators: Devashish Gosain in cooperation with Ines Ben Guirat and Claudia Diaz (KU Leuven)*

Mixnets are anonymous communication networks that aim to be secure against global adversaries who observe all communications in the underlying network. Real-world deployment of mix networks is challenging and lags behind relatively more recent low-latency systems. Many theoretical results and analyses exist but they do not adequately bridge the gap between theory and practice. One of the main challenges of deployment is deciding on the different mixnet building blocks and if the combination of them is necessarily the best system in terms of anonymity.

The MiXiM framework [1] fills this gap and provides the means to systematically analyze mix network designs from a number of dimensions and supports the mix-network adopter to take practical decisions backed with empirical support. With MiXiM, one can document the metadata (packet sources, destinations, and timings) exposed to mixnet nodes and the underlying network while making an abstraction of data payloads and cryptographic operations. To the best of our knowledge, MiXiM is the first generic mixnet simulation framework that allows the evaluation of anonymity (entropy) for different mixnet designs. In addition, MiXiM captures a number of relevant metrics (latency, bandwidth overhead) that can be used to perform novel analysis. As a result, researchers aiming to build a new mixnet system (or compare existing systems) can use MiXiM and evaluate anonymity and performance trade-offs under different scenarios. This flexible framework allows one to quickly set up experiments to investigate a large combination of mix networks building blocks, such as mixing strategies, network topologies, and the different parameters related to each component. The framework provides a number of metrics covering the anonymity, end-to-end latency, and overheads of mix networks.

## References


- [1] I. Ben Guirat, D. Gosain, and C. Diaz. MiXiM: Mixnet design decisions and empirical evaluation. In G. Livraga and N. Park, eds., *WPES '21, 20th Workshop on Workshop on Privacy in the Electronic Society*, Virtual Event, Republic of Korea, 2021, pp. 33–37. ACM.


### 29.5.7 Deep Packet Inspection in P4

*Investigators: Devashish Gosain in cooperation with Sahil Gupta (Rochester Institute of Technology), Minseok Kwon (Rochester Institute of Technology), and Hrishikesh B Acharya (Rochester Institute of Technology)*

Switches and routers–particularly Software-Defined Network (SDN) switches–have been successfully used to implement network-layer firewalls, flow analysis, and a wide range of other functions. Part of the reason for this remarkable versatility is that a small number of packet headers (source IP, source port, destination IP, destination port, protocol, etc.) are key for a variety of networking tasks. However, more advanced techniques, such as the detection of malicious traffic or malware signatures, require Deep Packet Inspection (DPI), i.e., the inspection of packet payloads and not just packet headers. However, we find that DPI-in-SDN is challenging because, in general, the payload is large and unpredictable in structure compared to packet headers. For example, one payload item–the HTTP application-layer header–has 47 possible fields, and these fields can occur out-of-order, have variable lengths, or can be entirely missing. Switches are designed for high-performance packet forwarding and not for general computation, so even in the case of P4–a language that allows users to freely define headers for their own protocols (which the switch then parses as easily as TCP or IP headers)–the authors of the P4 standard explicitly say that P4 is not intended for DPI. But in this work, we present a system–Deep Packet Inspection in P4 using packet recirculation (DeeP4R)–that performs Deep Packet Inspection in the data plane [1, 2].

DeeP4R is the first firewall to achieve "true Deep Packet Inspection in P4" (which we define as, DPI without realtime help from a controller or external firewall), using only standard P4-compatible switches. When a packet arrives, we use P4 functions to clone it, then apply the recirculate-and-truncate method of pattern matching on the cloned packet. (We loop the packet through the switch, consuming one byte from it with each pass. A Deterministic Finite Automaton keeps track if we have seen the target string.) If the clone is consumed without us seeing the target string (URL), we let the original packet (which has not been altered) pass through; otherwise, we drop it. Our novel method of combining packet cloning with recirculate-and-truncate allows us to perform flexible parsing in P4 and allow non-target traffic to pass through transparently. We do not claim that DeeP4R can handle all Deep Packet Inspection tasks, but it is perfectly capable of application-layer firewall tasks such as URL filtering. To our knowledge, DeeP4R is the first filter able to block URLs directly in the data plane (not taking real-time help from SDN controller, firewall, or special hardware).

In future we will extend our approach to match other strings such as keywords, and to other protocols such as DNS.

We implement, demonstrate, and benchmark the scalability and performance of DeeP4R, which as a dataplane program can process traffic very efficiently on a real switch. For instance, with 5000 domain names to filter and 10000 parallel flows, the latency on DeeP4R on a commodity SDN switch is under 1 milli-second while our firewall server (running standard Linux netfilter firewall) takes over 5 seconds.

## References


- [1] S. Gupta, D. Gosain, G. Grigoryan, M. Kwon, and H. B. Acharya. Demo: Simple deep packet inspection with P4. In *IEEE 29th International Conference on Network Protocols (ICNP 2021)*, Virtual Conference, 2021, pp. 1–2. IEEE.

[2] S. Gupta, D. Gosain, M. Kwon, and H. Acharya. DeeP4R: Deep packet inspection in P4 using packet recirculation. In *International Conference on Computer Communications (INFOCOM)*, 2023. IEEE.


### 29.5.8 On the Anonymity of Peer-To-Peer Network Anonymity Schemes Used by Cryptocurrencies

*Investigators: Devashish Gosain in cooperation with Piyush Kumar Sharma (imec-COSIC, KU Leuven) and Claudia Diaz (imec-COSIC, KU Leuven and Nym Technologies SA)*

Cryptocurrencies are digital currencies that are neither issued nor backed by a centralized banking or financial authority. Instead, they rely on the decentralized verification of cryptographic transactions using blockchain technology, allowing everyone to join and contribute to securing the transaction ledger. The decentralization and scalability aspects of blockchains have received considerable attention and are by now well understood. On the other hand, understanding the privacy properties of these systems presents additional complexity. Transaction anonymity requires protection both on-chain and in the underlying peerto-peer network used to transport the transaction. Ideally, if a transaction is considered private it should not be possible for third parties to identify its source or destination, neither by analyzing blockchain data, nor by analyzing network traffic data available to peers.

But recent work demonstrated that Cryptocurrency systems can be subject to deanonimization attacks by exploiting the network-level communication on their peer-to-peer network. Adversaries who control a set of colluding node(s) within the peer-to-peer network can observe transactions being exchanged and infer the parties involved. Thus, various network anonymity schemes have been proposed to mitigate this problem, with some solutions providing theoretical anonymity guarantees. In this work [1], we model such peer-to-peer network anonymity solutions and evaluate their anonymity guarantees. To do so, we propose a novel framework that uses Bayesian inference to obtain the probability distributions linking transactions to their possible originators. Notably, our anonymity analysis relies on network-level traffic data related to anonymously routing a transaction. Thus, the analysis is identical if originators use the peer-to-peer routing scheme to anonymously broadcast (or to send to selected recipients) a text message, instead of a blockchain transaction. This makes the proposed Bayesian approach applicable to evaluating anonymous peer-to-peer routing schemes in a generic sense. Presently practical deployments of anonymous peer-to-peer schemes relate to blockchain applications, and we focus on these for our evaluation. We demonstrate the generality of our approach by applying it to schemes, Dandelion, Dandelion++ and Lightning network, that rely on fundamentally different concepts for anonymous peerto-peer routing. Dandelion and Dandelion++ implement hop-by-hop probabilistic routing (that ends in broadcast), whereas in Lightning Network transactions are source-routed (all the way to the intended recipient).

We characterize transaction anonymity with those distributions, using entropy as metric of adversarial uncertainty on the originator's identity. We study different configurations and demonstrate that none of them offers acceptable anonymity to their users. For instance, our analysis reveals that in the widely deployed Lightning Network, with 1% strategically chosen colluding nodes the adversary can uniquely determine the originator for about 50% of the total transactions in the network. In Dandelion, an adversary that controls 15% of the nodes has on average uncertainty among only 8 possible originators. Moreover, we observe that due to the way Dandelion and Dandelion++ are designed, increasing the network size does not correspond to an increase in the anonymity set of potential originators. Alarmingly, our longitudinal analysis of Lightning Network reveals rather an inverse trend—with the growth of the network the overall anonymity decreases.

To encourage reproducibility, we make the framework and our analysis public.

### References


- [1] P. Kumar Sharma, D. Gosain, and C. Diaz. On the anonymity of peer-to-peer network anonymity schemes used by cryptocurrencies. In *Network and Distributed System Security Symposium (NDSS 2023)*, San Diego, CA, USA, 2023. Internet Society.


## 29.6 Internet Security Measurements

### 29.6.1 Illuminating Large-Scale IPv6 Scanning in the Internet

*Investigators: Oliver Gasser in cooperation with Philipp Richter and Arthur Berger (Akamai)*

While scans of the IPv4 space are ubiquitous, today still very little is known about scanning activity in the IPv6 Internet. Scans in the IPv4 Internet have been studied for decades, but their IPv6 counterparts remain understudied. Therefore, in this project, we performed a longitudinal and detailed empirical study on large-scale IPv6 scanning behavior in the Internet, based on firewall logs captured at some 230,000 hosts of a major Content Distribution Network (CDN). We presented detailed analyses on large-scale IPv6 scans carried out over the course of 15 months, as seen from a major CDN. We analyzed scan sources, and studied targeted services and addresses. We find that, unlike IPv4 scans, large-scale IPv6 scans are still comparably rare events, and we find them originating only from some 60 ASes. Further, IPv6 scan packets are concentrated on a small number of very active scan sources, with the two most active sources accounting for more than 70% of all logged scan traffic throughout our measurement window (see Figure 29.3). Many large-scale IPv6 scans do not target a single or a small number of specific services, but rather scan large swaths

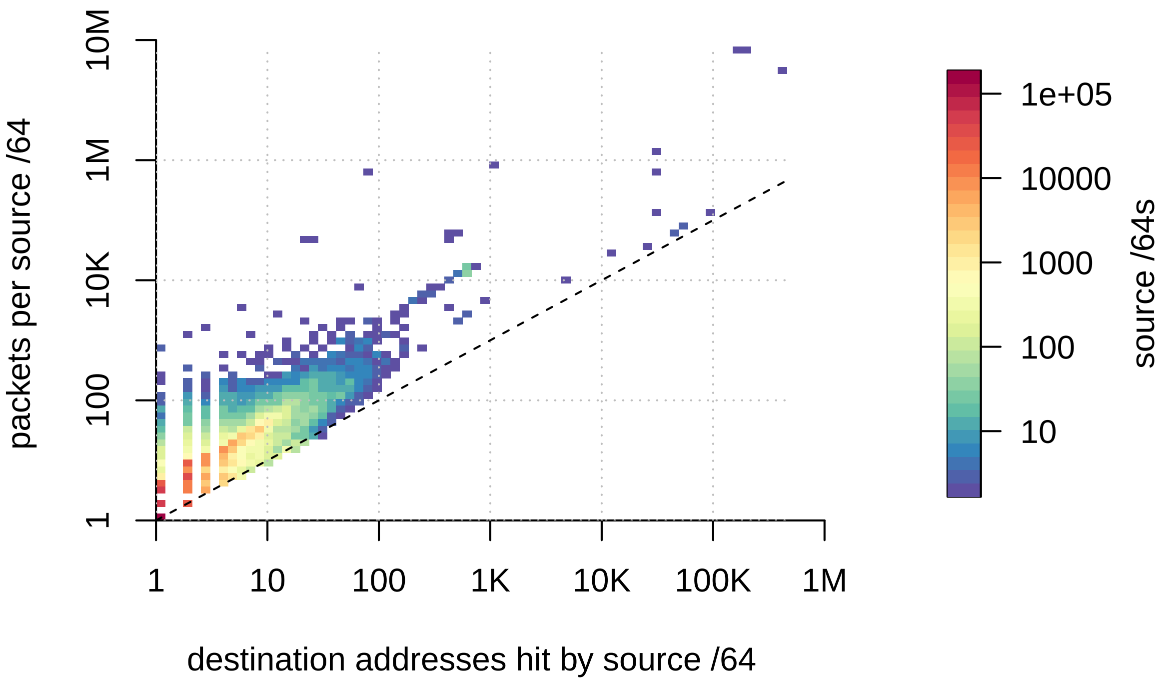



Figure 29.3: IPv6 scanning heatmap.

of port numbers, sometimes exceeding 100 ports targeted per scan. This behavior more closely resembles general and unspecific penetration testing behavior, as opposed to scanning patterns of botnets trying to spread laterally by exploiting individual vulnerabilities. Our findings showed that IPv6 scans in the wild show widely different characteristics from the more well-known IPv4 scans. We contrasted our findings with what can be observed in publicly available data in the MAWI dataset, and we discussed potential reasons for our observations. In this project [1], we developed methods to identify IPv6 scans, assessed current and past levels of IPv6 scanning activity, and studied dominant characteristics of scans, including scanner origins, targeted services, and insights on how scanners find target IPv6 addresses. Where possible, we compared our findings to what can be assessed from publicly available traces. This project identifies and highlights new challenges to detect scanning activity in the IPv6 Internet, and uncovers that today's scans of the IPv6 space show widely different characteristics when compared to the more well-known IPv4 scans.

## References


- [1] P. Richter, O. Gasser, and A. Berger. Illuminating large-scale IPv6 scanning in the internet. In C. Barakat and C. Pelsser, eds., *IMC '22, ACM Internet Measurement Conference*, Nice, France, 2022, pp. 410–418. ACM.

### 29.6.2 Third Time's Not a Charm: Exploiting SNMPv3 for Router Fingerprinting

*Investigators: Oliver Gasser in cooperation with Taha Albakour (TU Berlin), Robert Beverly (Naval Postgraduate School), and Georgios Smaragdakis (TU Delft)*

Remote management functionalities are fundamental to efficient network operation. To address this need, the Simple Network Management Protocol (SNMP) was introduced in the 1980s and has since served as the de facto protocol for fault notification, diagnostics, configuration management, and statistics gathering in IP networks. As a core IP management protocol that is widely implemented, it is unsurprising that SNMP has been both exploited and leveraged as an attack vector—indeed, there are over 400 SNMP-related CVEs. The protocol itself has historically been insecure, with the first standardized versions (SNMPv1 and SNMPv2) including only basic authentication via unencrypted "community strings." Security conscious operators were therefore forced to restrict SNMP access to internal networks. The current SNMPv3 standard, introduced in 2002, is implemented on virtually all modern network equipment. The primary focus of SNMPv3 is to provide a secure version of the protocol by including mechanisms for robust authentication, integrity, and privacy. Of direct relevance to our work is the so-called SNMP "engine ID." During synchronization with a client, the SNMPv3 agent exchanges its engine ID as a unique identifier. As noted in the RFC: the "snmpEngineID is the unique and unambiguous identifier of an SNMP engine. Since there is a one-to-one association between SNMP engines and SNMP entities, it also uniquely and unambiguously identifies the SNMP entity." In this research project [1], we showed that adoption of the SNMPv3 network management protocol standard offers a unique—but likely unintended—opportunity for remotely fingerprinting network infrastructure in the wild. Specifically, by sending unsolicited and unauthenticated SNMPv3 requests, we obtained detailed information about the configuration and status of network devices including vendor, uptime, and the number of restarts. More importantly, the reply contains a persistent and strong identifier that allows for lightweight Internet-scale alias resolution and dual-stack association. By launching active Internet-wide SNMPv3 scan campaigns, we showed that our technique can fingerprint more than 4.6 million devices of which around 350k are network routers. Not only is our technique lightweight and accurate, it is complementary to existing alias resolution, dual-stack inference, and device fingerprinting approaches. Our analysis not only provided fresh insights into the router deployment strategies of network operators worldwide, but also highlighted potential vulnerabilities of SNMPv3 as currently deployed.

### References


- [1] T. Albakour, O. Gasser, R. Beverly, and G. Smaragdakis. Third time's not a charm: Exploiting SNMPv3 for router fingerprinting. In *IMC '21, ACM Internet Measurement Conference*, Virtual Event, USA, 2021, pp. 150–164. ACM.

### 29.6.3 A Multi-Perspective Analysis of Web Cookies

*Investigators: Ali Rasaii, Shivani Singh, Devashish Gosain, and Oliver Gasser*

Web cookies have been the subject of many research studies over the last few years. However, most existing research does not consider multiple crucial perspectives that can influence the cookie landscape, such as the client's location, the impact of cookie banner interaction, and from which operating system a website is being visited. In this project [1], we conduct a comprehensive measurement study to analyze the cookie landscape for Tranco top-10k websites from different geographic locations and analyze multiple different perspectives. One important factor which influences cookies is the use of cookie banners. Most research involving GDPR does not consider interaction with cookie banners (e.g. clicking accept/reject buttons). Thus, we develop the automated tool, BannerClick, to identify and interact with banners with an accuracy of 99% and 96% respectively. We detect banners on about 47% of the Tranco top-10k websites in the EU region, whereas in non-EU regions we find banners on less than 30% of websites. We also investigate the difference in the number of cookies before and after interacting with a cookie banner and find an increase of 5.5x for third-party cookies. Moreover, we analyze the effect of banner interaction on different types of cookies (i.e. first-party, third-party, and tracking). For instance, we observe that websites send, on average, 5.5x more third-party cookies after clicking "accept", underlining that it is critical to interact with banners when performing Web measurements. Additionally, we analyze statistical consistency, evaluate the widespread deployment of consent management platforms, compare landing to inner pages, and assess the impact of visiting a website on a desktop compared to a mobile phone. Our study highlights that all of these factors substantially impact the cookie landscape, and thus a multi-perspective approach should be taken when performing Web measurement studies.

#### References


- [1] A. Rasaii, S. Singh, D. Gosain, and O. Gasser. Exploring the cookieverse: A multi-perspective analysis of web cookies. In A. Brunstrom, M. Flores, and M. Fiore, eds., *Passive and Active Measurement (PAM 2023)*, Virtual Event, 2023, LNCS 13882, pp. 623–651. Springer.


### 29.6.4 Yarrpbox: Detecting Middleboxes at Internet-Scale

*Investigators: Fahad Hilal and Oliver Gasser*

The end-to-end principle is one of the foundations of the original Internet architecture. It states that packets should remain unaltered while in transit between the two endpoints of a communication. This principle is put to the test by middleboxes, i.e., intermediary devices manipulating traffic for purposes other than the standard functions of an IP router. In the current Internet, they fulfill a multitude of tasks such as thwarting of attacks, censoring or monitoring users, address space expansions, or balancing resources. Apart from breaking the end-to-end principle, the deployment of middleboxes comes with additional caveats. They introduce hidden points of failure, thus complicating the debugging of networks. Moreover, they stand in the way of innovations and improvements to protocols and their extensions. The Internet has seen an increased deployment of these middleboxes owing to the value they bring. Therefore, it is important to have a good understanding of the middlebox ecosystem in the Internet.

In this work, we perform a multi-faceted middlebox analysis study. We develop Yarrpbox [1], a tool to efficiently perform middlebox detection measurements on an Internet-scale. Yarrpbox is over 300 times faster than the current state of the art and can conduct large-scale measurements in under 10 hours. With Yarrpbox, we perform IPv4-wide middlebox detection and find that nearly 10% of paths are affected by a total of 5.8k middlebox devices. We perform the first IPv6 study to date, uncovering a lower prevalence of middleboxes in IPv6. Moreover, we show that the location of a vantage point can have an effect on the results, leading to up to 600 more detected middleboxes. Additionally, we characterize middleboxes by mapping them to vendors and resolving aliases.

## References


• [1] F. Hilal and O. Gasser. Yarrpbox: Detecting middleboxes at internet-scale. In *CoNEXT'23, International Conference on Emerging Networking Experiments And Technologies*, 2023. ACM. Accepted for publication.


### 29.6.5 Rusty Clusters? Dusting an IPv6 Research Foundation

*Investigators: Oliver Gasser in cooperation with Johannes Zirngibl, Lion Steger, Patrick Sattler, and Georg Carle (Technical University of Munich)*

The long-running IPv6 Hitlist service is an important foundation for IPv6 measurement studies. It helps to overcome infeasible, complete address space scans by collecting valuable, unbiased IPv6 address candidates and regularly testing their responsiveness. However, the Internet itself is a quickly changing ecosystem that can affect long-running services, potentially inducing biases and obscurities into ongoing data collection means. Frequent analyses but also updates are necessary to enable a valuable service to the community. In this project [1], we showed that the existing IPv6 hitlist is highly impacted by the Great Firewall of China, and we offer a cleaned view on the development of responsive addresses. We evaluated the development of the IPv6 Hitlist over the last four years and new biases introduced by the accumulation of new addresses. Our findings allowed us to filter targets incorrectly tested as responsive. We identified 134 M addresses falsely reported as responsive to UDP/53 by the IPv6 Hitlist since 2018 due to the Great Firewall of China's DNS injection. While the accumulated input showed an increasing bias towards some networks, the cleaned set of responsive addresses is well distributed and showed a steady increase in the number of addresses. Although it is a best practice to remove aliased prefixes from IPv6 hitlists, in our research project we showed that this also removes major content delivery networks. We analyzed aliased prefixes in more detail and investigated whether the initial definition of a single host responsive to a complete prefix remains correct or whether a set of addresses needs to be treated differently. We showed that aliased prefixes host at least 15 M domains including ranked domains from different top lists. More than 98% of all IPv6 addresses announced by Fastly were labeled as aliased and Cloudflare prefixes hosting more than 10 M domains were excluded. In combination with additional findings, we suggest users of the hitlist to include subsets of these prefixes in future research, depending on the hitlist usage, e.g., higher layer protocol scans. Lastly, we evaluated different new address candidate sources, including target generation algorithms to improve the coverage of the current IPv6 Hitlist.

We showed that a combination of different methodologies is able to identify 5.6 M new, responsive addresses. This accounts for an increase by 174 % and combined with the current IPv6 Hitlist, we identified 8.8 M responsive addresses in total. Finally, we updated the IPv6 Hitlist service to allow future research to use our findings within the established service.

#### References


- [1] J. Zirngibl, L. Steger, P. Sattler, O. Gasser, and G. Carle. Rusty clusters?: Dusting an IPv6 research foundation. In C. Barakat and C. Pelsser, eds., *IMC '22, ACM Internet Measurement Conference*, Nice, France, 2022, pp. 395–409. ACM.


### 29.6.6 One Bad Apple Can Spoil Your IPv6 Privacy

*Investigators: Oliver Gasser and Said Jawad Saidi in cooperation with Georgios Smaragdakis (TU Delft)*

IPv6 is being more and more adopted, in part to facilitate the millions of smart devices that have already been installed at home. In this project [1], we found that the privacy of a substantial fraction of end-users is still at risk, despite the efforts by ISPs and electronic vendors to improve end-user security, e.g., by adopting prefix rotation and IPv6 privacy extensions. By analyzing passive data from a large European ISP, we found that around 19% of end-users' privacy can be at risk, i.e., by the lack of IPv6 privacy extensions usage. Privacy extensions are a technique to generate a random IPv6 address suffix, instead of using the persistent and therefore trackable MAC address in the IPv6 address. When we investigated the root causes, we noticed that a single device at home that encodes its MAC address into the IPv6 address can be utilized as a tracking identifier for the entire end-user prefix—even if other devices use IPv6 privacy extensions. Our results showed that IoT devices contribute the most to this privacy leakage and, to a lesser extent, personal computers and mobile devices. To our surprise, some of the most popular IoT manufacturers have not yet adopted privacy extensions that could otherwise mitigate this privacy risk. Finally, we showed that third-party providers, e.g., popular content, application, or service providers, can track up to 17% of subscriber lines in our study due to lack of IPv6 privacy extensions.

#### References


- [1] S. J. Saidi, O. Gasser, and G. Smaragdakis. One bad apple can spoil your IPv6 privacy. *ACM SIGCOMM Computer Communication Review*, 52(2):10–19, 2022.


## 29.7 Online Social Networks

### 29.7.1 Online Hate Speech

*Investigators: Savvas Zannettou in cooperation with Felipe-Gonzalez Pizarro (University of British Columbia, Canada)*

The spread of hate speech and hateful imagery on the Web is a significant problem that needs to be mitigated to improve our Web experience. The problem of hateful content is longstanding on the Web for various reasons. First, there is no scientific consensus on what constitutes hateful content (i.e., no definition of what hate speech is). Second, the problem is complex since hateful content can spread across various modalities (e.g., text, images, videos, etc.), and we still lack automated techniques to detect hateful content with acceptable and generalizable performance. Third, we lack moderation tools to proactively prevent the spread of hateful content on the Web. Our work [1] focuses on assisting the community in addressing the issue of the lack of tools to detect hateful content across multiple modalities. In particular, we contribute to research efforts to detect and understand hateful content on the Web by undertaking a multimodal analysis of Antisemitism and Islamophobia on 4chan's /pol/ using OpenAI's CLIP. This large pre-trained model uses the Contrastive Learning paradigm. We devise a methodology to identify a set of Antisemitic and Islamophobic hateful textual phrases using Google's Perspective API and manual annotations. Then, we use OpenAI's CLIP to identify images that are highly similar to our Antisemitic/Islamophobic textual phrases. By running our methodology on a dataset that includes 66M posts and 5.8M images shared on 4chan's /pol/ for 18 months, we detect 173K posts containing 21K Antisemitic/Islamophobic images and 246K posts that include 420 hateful phrases. Among other things, we find that we can use OpenAI's CLIP model to detect hateful content with an accuracy score of 0.81 (F1 score = 0.54). By comparing CLIP with two baselines proposed by the literature, we find that CLIP outperforms them, in terms of accuracy, precision, and F1 score, in detecting Antisemitic/Islamophobic images. Also, we find that Antisemitic/Islamophobic imagery is shared in a similar number of posts on 4chan's /pol/ compared to Antisemitic/Islamophobic textual phrases, highlighting the need to design more tools for detecting hateful imagery. Finally, we make available (upon request) a dataset of 246K posts containing 420 Antisemitic/Islamophobic phrases and 21K likely Antisemitic/Islamophobic images (automatically detected by CLIP) that can assist researchers in further understanding Antisemitism and Islamophobia.

### References


- [1] F. González-Pizarro and S. Zannettou. *Understanding and Detecting Hateful Content using Contrastive Learning*, 2022. arXiv: 2201.08387.


### 29.7.2 On the Globalization of the QAnon Conspiracy Theory Through Telegram

*Investigators: Mohamad Hoseini, Savvas Zannettou, and Anja Feldmann in cooperation with Philipe Melo (Federal University of Minas Gerais), and Fabrício Benevenuto (Federal University of Minas Gerais)*

QAnon is a far-right conspiracy theory that has implications in the real world, with supporters of the theory participating in real-world violent acts like the US capitol attack in 2021. At the same time, the QAnon theory started evolving into a global phenomenon by attracting followers across the globe and, in particular, in Europe, hence it is imperative to understand how QAnon has become a worldwide phenomenon and how this dissemination has been happening in the online space. This paper performs a large-scale data analysis of QAnon through Telegram by collecting 4.4M messages posted in 161 QAnon groups /channels. Using

Google's Perspective API, we analyze the toxicity of QAnon content across languages and over time. Also, using a BERT-based topic modeling approach, we analyze the QAnon discourse across multiple languages. Among other things, we find that the German language is prevalent in our QAnon dataset, even overshadowing English after 2020. Also, we find that content posted in German and Portuguese tends to be more toxic compared to English. Our topic modeling indicates that QAnon supporters discuss various topics of interest within far-right movements, including world politics, conspiracy theories, COVID-19, and the anti-vaccination movement. Taken all together, we perform the first multilingual study on QAnon through Telegram and paint a nuanced overview of the globalization of QAnon [1, 2].

### References

• [1] M. Hoseini, P. Melo, F. Benevenuto, A. Feldmann, and S. Zannettou. *On the Globalization of the QAnon Conspiracy Theory Through Telegram*, 2021. arXiv: 2105.13020.

[2] M. Hoseini, P. Melo, F. Benevenuto, A. Feldmann, and S. Zannettou. On the globalization of the qanon conspiracy theory through telegram. In *Proceedings of the 15th ACM Web Science Conference 2023*, New York, NY, USA, 2023, WebSci '23, p. 75–85. Association for Computing Machinery.

## 29.7.3 Content Moderation in Social Networks

*Investigators: Savvas Zannettou in cooperation with Shagun Jhaver (Rutgers University, USA), Jeremy Blackburn (Binghamton University, USA), Emiliano De Cristofaro (University College United Kingdom), Gianluca Stringhini (Boston University), Robert West (EPFL, Switzerland), Krishna P. Gummadi (MPI-SWS, Germany)*

Analyzing content moderation online is important for several reasons. First, social media and other online platforms have become major sources of information and communication, shaping public discourse and opinions. The content shared on these platforms can have significant consequences for individuals, communities, and society as a whole. Thus, understanding the challenges and complexities of moderating online content is crucial for ensuring that these platforms are safe and inclusive spaces for all users. Second, content moderation involves a range of technical, social, and ethical issues that require interdisciplinary expertise. Studying content moderation online involves understanding the technical mechanisms used to identify and remove harmful content, as well as the social and cultural contexts in which these mechanisms operate. Moreover, content moderation online is a constantly evolving field, as new technologies and social dynamics emerge. In this line of work, our goal is to analyze and understand multiple aspects of content moderation, including how soft moderation interventions (i.e., warning labels are applied online) and effective they are [4, 2], what happens after online platforms take moderation action on specific online communities (i.e., community bans) [1], and how we can design systems to automatically identify accounts that are state-sponsored trolls and are involved in misinformative campaigns online [3].

### Soft Moderation Interventions

Over the past few years, there is a heated debate and serious public concerns regarding online content moderation, censorship, and the principle of free speech on the Web. To ease these concerns, social media platforms like Twitter, Facebook, and TikTok refined their content moderation systems to support soft moderation interventions. Soft moderation interventions refer to warning labels attached to potentially questionable or harmful content to inform other users about the content and its nature while the content remains accessible, hence alleviating concerns related to censorship and free speech. In our work, we performed one of the first empirical studies on how soft moderation interventions are applied on Twitter [4] and TikTok [2]. In particular, our work for Twitter, uses a mixed-methods approach, to study the users who share tweets with warning labels on Twitter and their political leaning, the engagement that these tweets receive, and how users interact with tweets that have warning labels. Among other things, we find that 72% of the tweets with warning labels are shared by Republicans, while only 11% are shared by Democrats. By analyzing content engagement, we find that tweets with warning labels had more engagement compared to tweets without warning labels. Also, we qualitatively analyze how users interact with content that has warning labels finding that the most popular interactions are related to further debunking false claims, mocking the author or content of the disputed tweet, and further reinforcing or resharing false claims. Finally, we describe concrete examples of inconsistencies, such as warning labels that are incorrectly added or warning labels that are not added to tweets despite sharing questionable and potentially harmful information. Our work on TikTok [2], focuses on the important problem of soft moderation interventions during important health-related events like the COVID-19 pandemic. In particular, we analyze the use of warning labels on TikTok, focusing on COVID-19 videos. First, we construct a set of 26 COVID-19-related hashtags, and then we collect 41K videos that include those hashtags in their description. Second, we perform a quantitative analysis on the entire dataset to understand the use of warning labels on TikTok. Then, we perform an in-depth qualitative study, using thematic analysis, on 222 COVID-19-related videos to assess the content and the connection between the content and the warning labels. Our analysis shows that TikTok broadly applies warning labels on TikTok videos, likely based on hashtags included in the description (e.g., 99% of the videos that contain #coronavirus have warning labels). More worrying is the addition of COVID-19 warning labels on videos where their actual content is not related to COVID-19 (23% of the cases in a sample of 143 English videos that are not related to COVID-19). Finally, our qualitative analysis of a sample of 222 videos shows that 7.7% of the videos share misinformation/harmful content and do not include warning labels, 37.3% share benign information and include warning labels, and 35% of the videos that share misinformation/harmful content (and need a warning label) are made for fun. Our study demonstrates the need to develop more accurate and precise soft moderation systems, especially on a platform like TikTok which is extremely popular among people of younger ages.

### Community Bans

When toxic online communities on mainstream platforms face moderation measures, such as bans, they may migrate to other platforms with laxer policies or set up their own dedicated websites. Previous work suggests that within mainstream platforms, community-level moderation is effective in mitigating the harm caused by the moderated communities. It is, however, unclear whether these results also hold when considering the broader Web ecosystem. Do toxic communities continue to grow in terms of their user base and activity on the new platforms? Do their members become more toxic and ideologically radicalized? In our work [1], we report the results of a large-scale observational study of how problematic online communities progress following community-level moderation measures. We analyze data from r/The_Donald and r/Incels, two communities that were banned from Reddit and subsequently migrated to their own standalone websites. Our results suggest that, in both cases, moderation measures significantly decreased posting activity on the new platform, reducing the number of posts, active users, and newcomers. In spite of that, users in one of the studied communities (r/The_Donald) showed increases in signals associated with toxicity and radicalization, which justifies concerns that the reduction in activity may come at the expense of a more toxic and radical community. Overall, our results paint a nuanced portrait of the consequences of community-level moderation and can inform their design and deployment.

### Detecting State-Sponsored trolls

Growing evidence points to recurring influence campaigns on social media, often sponsored by state actors aiming to manipulate public opinion on sensitive political topics. Typically, campaigns are performed through instrumented accounts, known as troll accounts; despite their prominence, however, little work has been done to detect these accounts in the wild. In our work [3], we present TROLLMAGNIFIER, a detection system for troll accounts. Our key observation, based on analysis of known Russian-sponsored troll accounts identified by Reddit, is that they show loose coordination, often interacting with each other to further specific narratives. Therefore, troll accounts controlled by the same actor often show similarities that can be leveraged for detection. TROLLMAGNIFIER learns the typical behavior of known troll accounts and identifies more that behave similarly. We train TROLLMAGNIFIER on a set of 335 known troll accounts and run it on a large dataset of Reddit accounts. Our system identifies 1,248 potential troll accounts; we then provide a multi-faceted analysis to corroborate the correctness of our classification. In particular, 66% of the detected accounts show signs of being instrumented by malicious actors (e.g., they were created on the same exact day as a known troll, they have since been suspended by Reddit, etc.). They also discuss similar topics as the known troll accounts and exhibit temporal synchronization in their activity. Overall, we show that by using TROLLMAGNIFIER, one can grow the initial knowledge of potential trolls provided by Reddit by over 300%. We argue that our system can be used for identifying and moderating content originating from state-sponsored accounts that aim to perform influence campaigns on social media platforms.

## References

[1] M. Horta Ribeiro, S. Jhaver, S. Zannettou, J. Blackburn, E. De Cristofaro, G. Stringhini, and R. West. *Does Platform Migration Compromise Content Moderation? Evidence from r/The_Donald and r/Incels*, 2020. arXiv: 2010.10397.

• [2] C. Ling, K. Gummadi, and S. Zannettou. *"Learn the Facts About COVID-19": Analyzing the Use of Warning Labels on TikTok Videos*, 2022. arXiv: 2201.07726.

- [3] M. H. Saeed, S. Ali, J. Blackburn, E. De Cristofaro, S. Zannettou, and G. Stringhini. TROLL-MAGNIFIER: Detecting state-sponsored troll accounts on Reddit. In *43rd IEEE Symposium on Security and Privacy (SP 2022)*, San Francisco, CA, USA, 2022, pp. 2161–2175. IEEE.

- [4] S. Zannettou. "i won the election!": An empirical analysis of soft moderation interventions on twitter. In *Proceedings of the Fifteenth International Conference on Web and Social Media (ICWSM 2021)*, Atlanta, GA, USA, 2021, pp. 865–876. AAAI.


## 29.8 Building and Understanding Dependable Infrastructure as a Socio-Technical System

### 29.8.1 Analyzing the Cloudification of Higher Education

*Investigators: Tobias Fiebig and Mannat Kaur in collaboration with Simran Munnot*

The digital transformation of academia is one of the major technical challenges of our time. However, as all challenges, this transformation is not without risks. Following common industry paradigms, universities now commonly look at infrastructure provided just by a few major cloud operators.

In this work, we investigate in how far universities depend on an small set of digital infrastructure providers. We were the first to conduct comprehensive measurements characterizing these developments from 2015 onwards [1], and published further analyses on the organizational implications of these developments [2]. Our work illustrates how the progressing cloudification of academia impacts core-values like academic freedom, and ties differences in cloud adoption between several countries to stark differences in academic cultrue. Furthermore, we provide a clear long-term agenda to presever academic freedom, as well as researchers and students pivacy.

Our work provides though-leadership across multiple independent fields, influencing ongoing work in the legal field, privacy, educational technology, and international governance and policy making. Furthermore, see also the project on centralization, did the novel methods for network measurements we used in this influence the networking field, inspiring further work on centralization [3], also see the associated centralization project.

Besides expanding our technical measurements to a global university-cloud observatory, we are also pursuing more organizationally focused work to establish why academic institutions migrate to cloud infrastructure, which obstacles they face, and why well-known risks are not considered in these migrations.

### References


- [1] T. Fiebig, S. Gürses, C. H. Gañán, E. Kotkamp, F. Kuipers, M. Lindorfer, M. Prisse, and T. Sari. Heads in the clouds? Measuring universities' migration to public clouds: Implications for privacy & academic freedom. *Proceedings on Privacy Enhancing Technologies Symposium (Proc. PETS)*, 2023(2). Accepted 2022.

- [2] T. Fiebig, M. Lindorfer, and S. Gürses. Position paper: Escaping academic cloudification to preserve academic freedom. *Privacy Studies Journal*, 1(1):49–66, 2022.

• [3] F. Streibelt, P. Sattler, F. Lichtblau, C. H. Gañán, A. Feldmann, O. Gasser, and T. Fiebig. How ready is DNS for an IPv6-only world? In A. Brunstrom, M. Flores, and M. Fiore, eds., *Passive and Active Measurement (PAM 2023)*, Virtual Event, 2023, LNCS 13882, pp. 525–549. Springer.

### 29.8.2 Protocol Complexity over the Ages: The Case of Email

*Investigators: Tobias Fiebig and Florian Steurer*

By now, the Internet–despite often feeling 'new'–is an established technology. Many protocols which are instrumental to its functioning (IPv4, DNS) or basic services used by billions (Email) have been around for several decades. Naturally, during that time, these protocols evolved, reacting to a changing thread-landscape and newfound requirements [1].

In this project, we put a spotlight on the developing complexity of these foundational protocols, especially focusing on email. Originally simple, the protocol suite around email has seen an explosion in additions. Our work measuring email deployments' implementation of these additions highlights that the technological developments contribute to a centralization of the ecosystem by making it increasingly harder for smaller operators to support all of these requirements [2].

Our work takes a critical look at the foundations of these protocol suites. While, as we demonstrate, necessary research, similar research lines are often neglected in academia as there is an in-balance between the fundamental research effort required to realize research in this sector in comparison to common KPIs in academic institutions in terms of publications and funding; Hence, our work stands out by addressing fundational challenges, making an important contribution to securing and future profing a critical element of our digital society.

We are currently expanding the measurement platform developped in this project to capture an even more encompasing picture of email deployements. Furthermore, we identified challenges in the available measurement methodologies and best practices when it comes to email measurements, and work on a publication addressing the requirements for conducting ethical email measurements. Finally, we are providing a service to the general public to make our results reproducibly, and to let operators and users assess their email setups under the complex reality of the current state of the protocol at: *https://email-security-scans.org/*

#### References

[1] T. Fiebig, F. Lichtblau, F. Streibelt, T. Krüger, P. Lexis, R. Bush, and A. Feldmann. Learning from the past: designing secure network protocols. In *Cybersecurity Best Practices*, pp. 585–613. Springer, 2018.

• [2] F. Holzbauer, J. Ullrich, M. Lindorfer, and T. Fiebig. Not that simple: Email delivery in the 21st century. In *USENIX ATC '22, USENIX Annual Technical Conference*, Carlsbad, CA, USA, 2022, pp. 295–308. USENIX Association.

### 29.8.3 Humans, Gender, & Security: System Administration Beyond Technology

*Investigators: Tobias Fiebig and Mannat Kaur*

The areas of networking and computer security regularly overlook a core component of secure infrastructure: The people actually building and running these systems [2]. This leads to a situation where computer scientists observe technical effects on the Internet, but are unable to explain the root-causes of these, with a common example being the root-causes of security incidents, which are regularly rooted in simple errors and not complex attacks [1].

This project converges our expertise in system and network operations with methods from the social sciences to investigate how social and cultural effects interact with how technology is shaped. While not purely technical, this work is instrumental for explaining technical measurement results and demonstrating excellence in science by addressing research questiosn end-to-end.

Our recent contributions in this field are an investigation of how COVID-19 lockdowns affected the work of system administrators using an established theoretical model developped at NASA [3]. Our work characterizes how social dynamics inflict on the work of system administrators, highlighting the impact of visibility and coordination issues amplified by the pandemic. Most critically, we find first indications for the significant role of care-work in operating systems. Our subsequent study utilizes a feminist perspective on system administration to investigate the experiences of people with marginalized genders [4].

Our findings shake foundations in this field, demonstrating that non-inclusive and toxic work environments negatively impact the career prospects of system administrators working in respectful and inclusive work environments by limiting their ability to utilize common practices for career development due to the threat of moving to a toxic environment. Additionally, we empirically describe a critical connection between a working environemnt where people can be themselves, and the security and safety of systems. Thereby, we are the first to converge the fields of system/network operations, computer security, safety science, and feminist research/gender studies work.

We are currently pursuing a joint project with TU Delft to further investigate cultural interaction effects between inclusive work environments and the system operations community, conjecturing that cultural priming effects lead to a refusal of practices even though these align with operators' values. Furthermore, We are expanding our investigation into implicit coordination mechanisms in operators' work, and currently prepare a study on expectations and responsibility as perceived by operators and users.

## References

[1] C. Dietrich, K. Krombholz, K. Borgolte, and T. Fiebig. Investigating system operators' perspective on security misconfigurations. In *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security*, 2018, pp. 1272–1289.

[2] M. Kaur, M. van Eeten, M. Janssen, K. Borgolte, and T. Fiebig. Human factors in security research: Lessons learned from 2008-2018. *arXiv preprint arXiv:2103.13287*, 2021.

• [3] M. Kaur, S. Parkin, M. Janssen, and T. Fiebig. "I needed to solve their overwhelmness": How system administration work was affected by COVID-19. *Proceedings of the ACM Human-Computer Interaction (Proc. CSCW)*, 6, CSCW2, Article 390, 2022.

• [4] M. Kaur, H. S. Ramulu, Y. Acar, and T. Fiebig. "Oh yes! over-preparing for meetings is my jam :)": The gendered experiences of system administrators. *Proceedings of the ACM Human-Computer Interaction (Proc. CSCW).* Accepted 2022.

### 29.8.4 Historic Measurements on the Internet's Centralization

*Investigators: Tobias Fiebig, Florian Streibelt, and Florian Scheurer*

As also highlighted in our project on the cloudification of academia, the Internet is centralizing at accelerating speed. Understanding these developments is instrumental for assessing their impact on the Internet, protocol development, and ultimately society–beyond the narow scope of academic cloudification.

In this project, we develop new measurement techniques and infrastructure to aid researchers in investigating centralization, from its historic origins to the current state of the world.

Our contributions so far include public infrastructure–and an accompanying publication introducing the methodology [1]–to perform IP address attribution. Our ongoing work in this domain focuses on further developing these methods and infrastructure, while also combining existing methods to perform large-scale measurement studies to form a comprehensive picture of centralization.

Especially the public infrastructure we provide to the research community at large amplifies the impact of our work, enabling others to advance the field along us.

#### References

[1] F. Streibelt, M. Lindorfer, S. Gürses, C. H. Gañán, and T. Fiebig. Back-to-the-Future Whois: An IP Address Attribution Service for Working with Historic Datasets, 2022.

### 29.8.5 Mental Models of Security and Privacy

*Investigators: Tobias Fiebig and Mannat Kaur*

Security and privacy are essential properties of digital infrastrucutre. Hence, there is a multitude of frameworks, standards, and certifications aimed at ensuring security and privacy in these systems, and experts are tasked with implementing and deploying technology to preserve these values. However, in practice, we regularly find these mechanics failing [2]. The question is: Why?

In this project, we take a closer look at how experts and users from various domains think about security and privacy, i.e., which mental models they hold. We conjecture that assumptions, especially regarding the expertise of domain experts, are often wrong, leading to miscommunication and non-ideal results in the implementation of policies, especially when experts from different domains communicate while holding diverging mental models.

In our recent work on experts' and end-users' mental models concerning corporate VPN infrastructure, we already demonstrated that even security professionals' mental models of VPNs do not necessarily conform with the actual functionality of the technology, and overall do not necessarily diverge significantly from the mental models held by non-experts [1].

We are currently in the process of expanding this work to the general domain of security and privacy. Here, our assumption is that experts in privacy–often with a more legal or policy focused background–have diverging mental models of the technology they govern, while the technology experts ultimate implementing systems conforming to privacy policies hold diverging mental models on privacy topics. This leads to a situation where communication between the parties is significanly inhibited, contributing to concepts like "Privacy-by-Compliance" [3], which lacks actual privacy protections, being unintentionally realized. Hence, with our work we contribute to tangible security and privacy improvements for society.

### References


• [1] V. Binkhorst, T. Fiebig, K. Krombholz, W. Pieters, and K. Labunets. Security at the end of the tunnel: The anatomy of VPN mental models among experts and non-experts in a corporate context. In *31st USENIX Security Symposium*, Boston, MA, USA, 2022, pp. 3433–3450. USENIX.

[2] T. Fiebig. How to stop crashing more than twice: A clean-slate governance approach to IT security. In *2020 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, 2020, pp. 67–74. IEEE.

• [3] T. Fiebig, M. Lindorfer, and S. Gürses. Position paper: Escaping academic cloudification to preserve academic freedom. *Privacy Studies Journal*, 1(1):49–66, 2022.


### 29.8.6 Societal Implications of Digital Infrastructure in a Changing World

*Investigator: Tobias Fiebig*

Our world is changing, and humanity is facing unprecedented challenges. The global climate crisis, collapsing supply chains, the ever increasing threat of pandemics due to the progressing destruction of wildlife habitats. All these developments will ultimately change our world and society.

Given this changing world, we use the synthesis of our research projects to answer the ultimate question of system and network engineering: "What is our role, what is our responsibility, in an ever changing, ever failing future?"

Based on our research, we were able to derive 13 propositions illustrating the state of the Internet, analyzing pressing issues, from the impact of progressing centralization to matters of digital sovereignty [1]. Our work makes scientific results accessible to a wider community, and is especially appreciated by the applyed networking community. Hence, in summary, our work is part of our scientific responsibility to inform and assist society, and help it prepare for the future to come, in a way that is accessible and actionable.

For the future, this project highlights new research directions, and informs our existing research projects' direction, to ultimately conduct research that makes the world a better place.

### References


• [1] T. Fiebig and D. Aschenbrenner. 13 propositions on an Internet for a "Burning World". In G. Sileno, A. Abhishta, and C. Becker, eds., *TAURIN+BGI '22, ACM SIGCOMM 2022 Joint Workshops on Technologies, Applications, and Uses of a Responsible Internet and Building Greener Internet*, Amsterdam, The Netherlands, 2022, pp. 1–5. ACM.

## 29.9 Ongoing Projects

### 29.9.1 Characterizing Information Propagation on Mainstream and Fringe Communities on Telegram

*Investigators: Mohamad Hoseini, and Anja Feldmann in cooperation with Savvas Zannettou (Delft University of Technology), Philipe Melo (Federal University of Minas Gerais), and Fabrício Benevenuto (Federal University of Minas Gerais)*

Online messaging platforms and social media are one of the main modes of communication and bring together millions of people. With well-connected networks capable of reaching thousands of users very quickly, this ecosystem becomes vulnerable to the impacts of misinformation, and fake news campaigns in different aspects. Traditional networks such as Facebook, Twitter, and Instagram concentrate a large part of online content, but alternative platforms have recently emerged and become popular, mainly due to the proposal of free and unmoderated discussion as in mainstream networks. Currently, Telegram is one of the leading messaging platforms hosting fringe communities. Despite the popularity, as a research community, we lack knowledge of how content spreads over this network. Motivated by the importance and impact of messaging platforms on society, we aim to measure the information propagation within the Telegram network, focusing on how public groups and channels exchange messages. We collect and analyze over 130M messages from 9K channels and groups on Telegram. We analyze message forwarding and the lifetime of the messages from 5 different aspects. Among other things, we find inequality in content creation; 6% of the users are responsible for 90% of forwarded messages. We also find that channels mainly play the role of content producers, while the groups act as consumers of forwarded messages. Additionally, we find that 5% of the channels are responsible for 40% of the forwarded messages in the entire dataset. Finally, our lifetime analysis shows that the messages shared in the groups with several active users live longer within the Telegram platform than the messages in the channels.

### 29.9.2 Enabling Multi-hop ISP-Hypergiant Collaboration

*Investigators: Cristian Munteanu, Oliver Gasser, Anja Feldmann in collaboration with Georgios Smaragdakis (TU DELFT, Netherlands) and Ingmar Poese (BENOCS GmbH)*

Today, there is an increasing number of peering agreements between Hypergiants and networks that benefit millions of end-user. However, the majority of Autonomous Systems (more than 50 thousands) do not currently enjoy the benefit of interconnecting directly with Hypergiants to optimally select the path for delivering Hypergiant traffic to their users.

In our research, we develop and evaluate an architecture that can help this long-tail of networks. Our architecture includes multiple scenarios—with and without the cooperation of the transit provider. Indeed, the most basic one does not require any changes to the operation of transit providers, nor any re-negotiation of the relationship between networks and their transit providers or Hypergiants and transit providers.

In our architecture, a network establishes an out-of-band communication channel with Hypergiants that can be two (or more) AS-hops away and, optionally, with the transit provider. This channel enables the exchange of network information to better assign requests of end-users to appropriate Hypergiant servers.

Our analysis using operational data shows that our architecture can optimize, on average, 15% of Hypergiants' traffic and 11% of the overall traffic of networks that are not interconnected with Hypergiants. The gains are even higher during peak hours, where available capacity can be scarce. Our results also show that for some Hypergiants, more than 46% of their traffic delivered to networks via non-direct interconnection can be optimized.

### 29.9.3 Analyzing Internet Measurement Platforms

*Investigators: Lars Prehn in cooperation with Shravan Swaminathan, Pascal Vermeulen (Saarland University), Marcel Flores (Edge.io), Emile Aben (RIPE NCC), Pavlos Sermpezis, Sofia Kostoglou, and Athena Vakali (Aristotle University of Thessaloniki)*

Network operators and researchers frequently use Internet measurement platforms (IMPs), such as RIPE Atlas, RIPE RIS, or RouteViews for, e.g., monitoring network performance, detecting routing events, discovering the AS topology, or optimizing routes. To interpret the results of their measurements, users must understand a platform's limitations and biases.

In this project, we first explored the noise within the data of RIPE RIS and Routeviews. We introduced an classification for typical noise patterns and performed a deep-dive into the noise within the data availability. We found that the archives for publicly available route collector projects persistently suffer from unavailable files and dropped peering sessions. Hereby, we found that patterns occur at the peer, route collector, and region level.

We further introduced a general framework to analyze the multi-dimensional (e.g., across location, topology, network types, etc.) biases of IMPs. Beyond new insights into currently existing IMP biases, our framework provides methods to sub-sample and extend the current set of IMP vantage points. When blind-testing with a common measurement task, we demonstrate that our framework reduces placement bias and that this reduction leads to a more accurate estimation of end-user performance for a large CDN.

Lastly, we conceptualized a system, CorMoRanT, that simplifies the analysis of BGP-based IMP data by generating thousands of pre-computed statistics and partial data views on a daily basis.

### 29.9.4 Detecting Sibling Prefixes in the Wild

*Investigators: Fariba Osali, Zubair Khwaja Sediqi, and Oliver Gasser*

For several decades, IPv4 address has been used on the Internet to identify network device interfaces, effectively allowing the mapping to individual hosts. With IPv4 address exhaustion in recent years, network operators have increasingly implemented IPv6 in their networks. While IPv4 remains the dominant IP version on the Internet, the adoption of IPv6 is increasing. Each IP address consists of a network portion and a host portion, and an IP prefix represents a range of IP addresses with a shared network portion. While many service providers are incorporating both IPv4 and IPv6 in their network, the relationship between IP prefixes of both versions in the wild is unknown.

In this work, we explore the potential relationship between IPv4 and IPv6 prefixes to identify IPv4 and IPv6 sibling prefixes for hosts and services. Both research communities and network operators can benefit from the detection, identification, and classification of sibling prefixes. For example, if network operators want to filter all services on a specific IPv4 prefix, the corresponding IPv6 equivalent class is an excellent candidate.

To achieve this objective, we plan to conduct an investigation into the correlation between IPv4 and IPv6 prefixes in DNS datasets by looking at domains that have both A and AAAA records, as this will allow us to identify cases where the same domain name is associated with both an IPv4 and an IPv6 prefix. Then we plan a longitudinal study on IP prefix changes by tracking the dynamics of IP address and prefix allocation and usage over time.

## 29.10 Weizenbaum Institute: "Work and Cooperation in the Sharing Economy" (until 14 September 2022) / "Digital Economy, Internet Ecosystem, and Internet Policy" (since 15 September 2022)

*Investigators: Anja Feldmann, Stefan Schmid, Volker Stocker, Saba Rebecca Brause, Aaron Kolleck, and Nadine Schawe*

This multidisciplinary research group at the Weizenbaum Institute for the Networked Society examines the efficiency, mechanics and evolution of sharing markets, the impact of the sharing economy on market, industry, and employment structures, the role of data and data access, and the need for reform of existing regulations in the digital economy. The composition of expertise, methods and perspectives within the group enables research which can contribute to a more differentiated understanding of different individual phenomena, as well as to identify and describe possible links between different individual phenomena. In addition, the group examines complementary questions. For example, aspects of competition, regulation, efficiency and evolution of the Internet and the Internet ecosystem, but also of broadband and broadband policy, as well as issues of Internet policy, all of which are critical to understanding digital change and the digital economy.

During the reporting period, the Weizenbaum Institute was successful in securing funding from the BMBF for the next five years (beginning 15th September 2022). The research group played a role in contributing to the proposal and was selected as one of the 16 research groups that will be part of the next funding phase. In this context, the research group underwent reorganization and restructuring, and an additional PI, Prof. Stefan Schmid (TU Berlin, Head of the Internet Architecture and Management Group) was appointed. Starting on September 15th, 2022, the research group was renamed to "Digital Economy, Internet Ecosystem, and Internet Policy", with a modified research focus on topics related to digital infrastructures, broadband and broadband policy, the role of data for platforms in the digital economy, and the evolution of the Internet ecosystem. As of February 2023, the research group is in the process of building its team of PhD researchers and student assistants.

In the reporting period, the research group has successfully developed its PLAMADISO (PLatforms, MArkets, and the DIgital SOciety) into a dynamic and thriving community. Renowned researchers from different disciplines have been invited to present and discuss their cutting-edge research in an international research environment. You can learn more about PLAMADISO by visiting this link: https://plamadiso.weizenbaum-institut.de. Apart from the PLAMADISO Talk Series, where researchers present and discuss their work online, the research group has also organized online symposia and panels, featuring world-class speakers to discuss current topics related to the digital economy across disciplinary boundaries. These formats bring together scholars from different backgrounds such as computer science, management, economics, or the social science and provide platforms for participation by the interested public and other private and public decision-makers. To ensure wider dissemination of the content and make it accessible to a broader audience, the recordings of the events are published on the group's YouTube channel, which you can access here: https://www.youtube.com/@plamadiso. There are already more than 50 video already available.

Besides these outreach and transfer activities, the research group is committed to producing high-quality research and fostering collaborations with leading scholars and institutions worldwide. Members of the research group have been highly productive in producing a series of publications and presenting their research at peer-reviewed international conferences, workshops, and other relevant events. They have collaborated or are collaborating with internationally renowned scholars from prestigious institutions such as the University of Pennsylvania, MIT, TU Delft, and the Universidad Politécnica De Madrid.

In Spring 2022, Volker Stocker spent time in the United States as a visiting researcher to pursue research collaborations with two prominent scholars: Prof. Johannes Bauer, an Economist and Director of the Quello Center at Michigan State University, and Prof. Christopher S. Yoo, a Law scholar and Founding Director of the Center for Technology, Innovation, and Competition at the University of Pennsylvania. For example, one exciting project resulting from this research visit is the joint research project between Volker Stocker and Johannes Bauer of the Quello Center at MSU. They are currently exploring "Innovation Dynamics in the Internet Ecosystem and Digital Economy Policy", which has been accepted for presentation at the upcoming ITS Europe conference in Madrid in June 2023.

In addition to his research work, Volker Stocker holds a leadership role as the acting co-chair of the International Telecommunications Society (ITS) Europe. In this capacity, he has been or will be part of the organizing and program committees for the ITS Europe conferences held in 2021 (online due to COVID-19; https://itseurope.org/2021/), 2022 (in Gothenburg, Sweden; https://itseurope.org/2022/gothenburg/), and 2023 (in Madrid, Spain; https://itseurope.org/2023/). Furthermore, he has been elected to serve on the program committee of the 51st Telecommunications Policy and Research Conference (TPRC) (http://www.tprcweb.com).

The Ph.D. students of the group, Aaron Kolleck and Nadine Schawe, actively worked on their respective Ph.D. research projects during their time as research assistants in the group.

Below, we will provide selected insights into the research activity and publications of the group during the reporting period. For additional information see https://www.weizenbaum-institut.de/en/publications/, https://plamadiso.weizenbaum-institut.de.

### 29.10.1 Evolution of the Internet Ecosystem and Public Policies for Broadband

*Investigators: Volker Stocker and Anja Feldmann in cooperation with Stefan Schmid (TU Berlin) and various collaborators, see publications*

The digital economy is undergoing significant transformational processes that are reshaping industry structures and innovational and competitive dynamics in the Internet ecosystem. With the continual evolution of this ecosystem, modern information and communication technologies, particularly digital infrastructures like broadband (including 5G and beyond) and cloud infrastructures, have become increasingly essential for citizens and businesses – and social and economic participation.

The research group provides insights into the complex interplay between technology, economics, and policy. Their research has examined the evolution of the Internet ecosystem [3] and explored its technological and economic drivers, with a focus on the role of large private networks often owned and operated by hypergiants [4]. The research has identified areas where mismatches exist between existing regulatory frameworks and issues related to the current state of the ecosystem, and has derived policy implications.

In addition to economic, (interdisciplinary) techno-economic, and policy-driven analyses, group members have also investigated new technologies from a computer science perspective [1, 2].

#### References

[1] N. S. Johansen, L. B. Kær, A. L. Madsen, K. Ø. Nielsen, S. Schmid, J. Srba, and R. G. Tollund. Fbr: Dynamic memory-aware fast rerouting. In *2022 IEEE 11th International Conference on Cloud Networking (CloudNet)*, 2022, pp. 55–60. IEEE.

[2] S. Schmid, M. K. Schou, J. Srba, and J. Vanerio. R-mpls: recursive protection for highly dependable mpls networks. In *Proceedings of the 18th International Conference on emerging Networking EXperiments and Technologies*, 2022, pp. 276–292.

[3] V. Stocker and G. Knieps. Digitalizing telecommunications: innovation, complexity and diversity in the internet ecosystem. In *A Modern Guide to the Digitalization of Infrastructure*, pp. 59–91. Edward Elgar Publishing, 2021.

[4] V. Stocker, G. Knieps, and C. Dietzel. The rise and evolution of clouds and private networks–internet interconnection, ecosystem fragmentation. In *TPRC49: The 49th Research Conference on Communication, Information and Internet Policy*, 2021.

### 29.10.2 COVID-19 and the Internet Ecosystem

The COVID-19 pandemic has disrupted the real and the virtual worlds and. highlighted the importance of access to affordable digital infrastructures and services, as well as the necessary skills to makes use of them [3, 4]. The book "Beyond the Pandemic? Exploring the Impact of COVID -19 on Telecommunications and the Internet", edited by Jason Whalley (Northumbria University), Volker Stocker, and William Lehr (MIT) [4] provides broad insights into the pandemic's impact in various contexts. The book features contributions from internationally renowned scholars (from different fields) and industry experts on topics and sectors such as labor, retail, education, cybercrime, misinformation/disinformation, spectrum and spectrum policy, or smart cities. In their chapter [2], Stocker et al. examine the impact of COVID-19 on the Internet, analyzing Internet traffic and exploring the responses of private and public entities to the pandemic crisis. Based on the insights gained, the authors derive lessons for a post-COVID world that can assist better planning for future crises. Stocker et al. [1] provides a more comprehensive high-level analysis of the pandemic's effects based on the various insights from the book chapters. Among other things, the authors discuss how the pandemic accelerated and amplified technology adoption and trends, and how technology helped elleviate the negative consequences of the pandemic for many. However, they explain why not everyone benefitted equally and that (new) questions regarding digital inclusion and divides have elevated important policy debates.

## References

[1] V. Stocker, W. Lehr, and G. Smaragdakis. Beyond the pandemic: Towards a digitally enabled society and economy. In *Beyond the Pandemic? Exploring the Impact of COVID-19 on Telecommunications and the Internet*, ch. 12. Emerald, 2023.

[2] V. Stocker, W. Lehr, and G. Smaragdakis. Covid-19 and the internet: Lessons learned. In *Beyond the Pandemic? Exploring the Impact of COVID-19 on Telecommunications and the Internet*, ch. 2. Emerald, 2023.

[3] V. Stocker and J. Whalley. The internet has coped well with covid-19, but problems remain: Evidence to house of lords committee exploring the impact of covid-19. 15, 2021.

[4] J. Whalley, V. Stocker, and W. Lehr. Beyond the pandemic? exploring the impact of covid-19. 2023.

### 29.10.3 Data Governance Act Proposal

Members of the research group were involved in a multidisciplinary working group at the Weizenbaum Institute focusing on EU Regulations and in particular on the Data Governance Act (DGA) Proposal [1]. Volker Stocker acted as a co-chair of the working group and Nadine Schawe was a valuable member of the group. The working group published a position paper on the DGA Proposal that contains specific recommendations. The group members' participation in the working group underscores their commitment to advancing research in the digital economy and shaping policy debates in this area.

## References

[1] *Data Governance Act Proposal: A position paper by the research groups "Frameworks for Data Markets", "Work and Cooperation in the Sharing Economy", "Trust in Distributed Environments", "Responsibility and the Internet of Things", and "Reorganizing Knowledge Practices" of the Weizenbaum Institute for the Networked Society*, Weizenbaum Series 18. Weizenbaum Institute for the Networked Society - The German Internet Institute, Berlin, 2021.

## 29.11 Academic Activities

### 29.11.1 Conference and Workshop Positions

One point to emphasize is that Anja Feldmann and Lars Prehn contributed to the community by organizing a *shadow TPC* for the ACM Internet Measurement Conference 2022 and acted as Shadow TPC Chairs. The idea of a Shadow TPC (STPC) is to expose students to the academic review process and train the next generation of program committee members. The STPC allows participants to read papers at the submission stage, experience the peer-review process from a reviewer's perspective, and discuss academic work with fellow STPC members. The STPC's peer-review process closely resembled the actual one in function, workload, and timing. STPC reviewers have to abide by the same rules and restrictions applicable to regular TPC members (e.g., anonymity and confidentiality rules). Still, the STPC forwarded the "Best of Shadow TPC" comments – detailed review summaries for two pre-accepted shadow TPC papers with exceptionally high ratings – to the actual TPC and awarded a "Shadow Best Paper". The STPC chairs reviewed all reviews and online discussions as well as lead an onlie TPC meeting to ensure that the learning objectives for the STPC particpants are achieved and that the authors received constructive feedback. The feedback of the students was excellent. They all agreed that they are much better able to understand the process with all its pros and cons.

#### Membership in program and organization committees, reviewing activities

*Prof. Anja Feldmann, Ph.D.:*

- *ACM IMC Conference 2022* Nice, France, October 2022 (Co-TPC-chair Shadow TPC)
- *USENIX Annual Technical* , virtual event, July 2021 (TPC member)
- *ACM Symposium on Operating System Principles (SOSP)*, virtual event, October 2021 (TPC member)
- *ACM SIGCOMM Conference 2021*, virtual event, August 2021 (TPC member)
- *ACM HotNets 2021*, virtual event, November 2021 (TPC member)
- *ACM IMC Conference 2021* virtual event, November 2021 (TPC member)
- *ACM IMC Conference 2022* Nice, France, October 2022 (TPC member)
- *ACM IMC Conference 2023* Montreal, October 2023 (TPC member)
- *ACM CoNEXT 2022* Rome, Italy; December 2022 (TPC member)
- *USENIX NSDI Conference 2022 spring*, (TPC member)
- *USENIX NSDI Conference 2022 fall*, Boston, April 2023 (TPC member)
- *NSDI Conference 2023 spring*, (TPC member)
- *NSDI Conference 2023 fall*, Santa Clara, April 2024 (TPC member)
- *Passive and Active Measurement Conference (PAM) 2022, 2023* virtual event, March 2022, 2023 (TPC member)

*Dr. Oliver Gasser:*

- *IEEE/ACM Transactions on Networking*, Journal of the IEEE Communications Society, the IEEE Computer Society and the ACM Special Interest Group on Data Communications (Reviewer in 2021 and 2022)
- *Computer Communication Review*, the ACM SIGCOMM Newsletter; (Reviewer in 2021)
- *Computer Networks (ComNet)* The International Journal of Computer and Telecommunications Networking (Reviewer in 2021 and 2022)
- *IEEE Transactions on Network and Service Management* Journal (Reviewer in 2021)
- *ACM SIGCOMM 2021 Workshop on Technologies, Applications, and Uses of a Responsible Internet (TAURIN 2021)* virtual event, August 2021 (TPC member)
- *NetSys 2021* virtual event, September 2021 (TPC member)
- *Network Traffic Measurement and Analysis Conference (TMA) 2021* virtual event, September 2021 (TPC member)
- *ACM Internet Measurement Conference 2021* virtual event, November 2021 (TPC member)
- *ACM CoNext Student workshop 2021* virtual event, December 2021 (TPC member)
- *PerFail 2022 – First International Workshop on Negative Results in Pervasive Computing* virtual event, March 2022 (TPC member)
- *Passive and Active Measurement Conference (PAM) 2022* virtual event, March 2022 (TPC member)
- *5th International Workshop on Traffic Measurements (WTMC) 2022* Genova, Italy, June 2022 (TPC member)
- *Network Traffic Measurement and Analysis Conference (TMA) 2022* Enschede, The Netherlands, June 2022 (TPC member)
- *ACM CoNext Student workshop 2022* Rome, Italy, December 2022 (TPC member)

*Dr. Devashish Gosain:*

- *Esorics 2022*, September 2022 (TPC member)
- *Esorics 2023*, September (TPC member)
- *ACM IMC Conference 2022* Nice, France, October 2022 (Shadow TPC member)

*Dr. Volker Stocker:*

- *(Virtual) Talk Series on Platforms, Markets, and the Digital Society (PLAMADISO) at the Weizenbaum Institute for the Networked Society & TU Berlin* virtual event, 2021-ongoing (Organizer)
- *Online symposium "New Perspectives on the Digital Economy: Sharing, Platforms & Regulation"* virtual event, December 2022 (Co-Organizer)
- *ITS Europe: "Realising the digital decade in the European Union – Easier said than done?"* Madrid, Spain, June 2023 (Organizing Committee & TPC Member)

- *TPRC (Telecommunications Policy Research Conference)* Washington DC, USA, September 2023 (TPC member)
- *ITS Europe: "Reining in Digital Platforms? Challenging monopolies, promoting competition and developing regulatory regimes"* Gothenburg, Sweden, June 2022 (Organizing Committee & TPC Member)
- *ITS Europe/ITS Biennial: "Digital societies and industrial transformations: Policies, markets, and technologies in a post-Covid world"* virtual event, June 2021 (Organizing Committee & TPC Member)

*Lars Prehn:*

- *ACM IMC Conference 2022* Nice, France, October 2022 (Co-TPC-chair Shadow TPC)

*Franziska Lichtblau:*

- *Chair of the RIPE Programme committee, 2019-today*, (TPC Chair)

### Membership in steering and other committees

*Prof. Anja Feldmann, Ph.D.:*

- *ACM/IEEE Transactions on Networking (ToN)*, (member of Steering Committee 2014 – 2022)
- *Passive and Active Measurement Conference PAM*, (member of Steering Committee since 2019)
- *ACM IMC Conference* (member of Steering Committee since 2018)
- *ACM CoNEXT* (member of Steering Committee since 2020)
- *MPG Presidential Commission "Governance"*, member since July 2021
- *MPG CPTS Perspectives Commission*, member since 2018
- *MPG Advisory Board Digitisation*, member since 2022
- *MPG IT-Sicherheitskommission*, member since 2019
- *MPG Core Committee "MPI for Security and Privacy"*, member since 2022
- *Uni-Sb: Graduate School Progress Committee*, Member since 2018.
- *IMPRS TRUST*, Spokes Person since 2020.
- *Dioscuri Committee*, member since 2021
- *Supervisory Board of the Karlsruhe Institute of Technology (KIT)*, Member since August 2020.
- *Deutsche Forschungsgemeinschaft e.V.*, Elected Member of the Review Board since November 2019.
- *University Hamburg: Commission on AI and Data Science*, member since November 2021
- *ACM Sigcomm Award Selection committee, member in 2021 and 2022*

- *Expert Committee on Communication and Information of the German Commission for UNESCO e.V.*, Member since 2013.
- *Board of Trustees of the Horst Görtz Institute for IT Security, Ruhr-University, Bochum*, member since 2021
- *ACM ICM Test of Time Selection Committee*, member 2022
- *Scientific Advisory Board of Schloss Dagstuhl – Leibniz Center for Informatics*, member, Reappointment in 2022
- *Scientific Senate of the The German National Research Data Infrastructure (NFDI)*, member, Reappointment in October 2022
- *Eco: Association of the Internet Industry*, Member of the Executive Committee since 2018.
- *Nokia Bell Labs Advisory Council*, since 2022
- *IEEE Mildred Dresselhaus Medal Committee*, since 2023
- *Berlin-Brandenburgische Akademie der Wissenschaften (BBAW)* Member since 2011.
- *Leopoldina*, Member since October 2011. *Academia Europaea*, Member since October 2018.
- *German Academy of Science and Engineering (acatech)*, Member since November 2019.

### 29.11.2 Invited Talks and Tutorials

*Prof. Anja Feldmann, Ph.D.:*

- *"The Internet: Chances and Limits"*, Keynote during IEEE Internet Conference on Communication; virtual event, June, ICC 2021
- *"Internet Traffic Analysis at Scale"* Keynote during VLDB; virtual event, July 2021
- *"Internet Traffic Analysis at Scale"* Keynote during ITC; virtual event, August 2021
- *"P4: Two sides of the same coin"* ACM Sigcomm SPIN Workshop Panelist and N2N Women Workshop Panelist, virtual event, August 2021
- *"Next-Gen-Networks"* Keynote at the Experten-Sounding-Board "Next-Gen-Networks aus Unternehmenssicht: Wald vor lauter Buzzwords kaum zu sehen – Was können und sollen Unternehmen in Sachen ihres eigenen NGN tun?" Frankfurt, Germany, January 2022
- *"Internet Traffic Analysis at Scale"* Guest talk at the Cambridge Systems Research Group seminar series, virtual event, March 2022
- *"Internet Architecture"* Keynote at the Nokia Bell Labs intern programm, virtual event, July 2022
- *"Internet Traffic Analysis at Scale"?* Keynote at the Heinz Nixdorf Symposium "On-The-Fly Computing", Paderborn, September 2022
- *"Anforderungen an die IT-Sicherheit – A Director's Cut"* Datenschutz- und IT Sicherheitstagung of the Max Planck Society, virtual event, November 2022

*Dr.-Ing. Tobias Fiebig:*

- *"Information System Security: Beyond the Sum of its Parts Analyze, Measure, Explain, Improve"* Talk organized by Saarland Informatics Campus Lecture Series; virtual event, July 2022
- *"13 propositions on an Internet for a burning world"* Keynote at the Asia-Pacific Network Information Centre APNIC 54 Conference, Singapore, September 2022
- *"From running systems to getting your head in the clouds"* Keynote at the Meeting of the data protection officers of the MPG, Göttingen, Germany, September 2022
- *13 propositions on an Internet for a burning world* Opening Plenary Talk of the RIPE 85 conference in Nice, France, October 2022

*Dr. Volker Stocker:*

- *Workshop "Shaping the Internet for the Future"* Participant at Max Planck Institute for Innovation and Competition; Munich, Germany, June 2022
- *"Beyond the pandemic? Exploring the impact of Covid on telecommunications and the Internet"* Panel Moderation & Organization at 31st ITS European Conference – Reining in digital platforms? Challenging monopolies, promoting competition and developing regulatory regimes; Gothenburg, Sweden, June 2022
- *"The Rise of Large Private Networks – How Google, Facebook, and other Hypergiants Have Disrupted the Internet Ecosystem"* Invited Talk at Strathclyde Law School – Internet Law Module; virtual event, February 2022
- *"The Rise of Large Private Networks – How Google, Facebook, and other Hypergiants Have Disrupted the Internet Ecosystem"* Invited Talk at Gesellschaft für Netzökonomie, Wettbewerbsökonomie und Verkehrswissenschaft an der Universität Freiburg e.V., University of Freiburg; virtual event, November 2021
- *"The Rise and Evolution of Clouds and Private Networks – Internet Interconnection, Ecosystem Fragmentation"* Invited talk at DigiDem-Kolloquium, Weizenbaum-Institute for the Networked Society; virtual event, November 2021
- *"Digitale Plattformen"* Invited talk/Impulsvortrag Evangelische Bank eG; ChangeHub Berlin, September 2021
- *"What are the potential and challenges for next generation mobile networks?"* Panel Moderation & Organization at 23rd Biennial Conference of the ITS – Digital societies and industrial transformations: Policies, markets, and technologies in a post-Covid world; virtual event, June 2021
- *"Internet interconnection, ecosystem fragmentation — Towards a conceptual foundation of generalized virtual networks"* 23rd Biennial Conference of the ITS – Digital societies and industrial transformations: Policies, markets, and technologies in a post-Covid world; virtual event, June 2021

## 29.12 Teaching Activities

### Summer Semester 2021

– Data Networks Online-Lecture (Anja Feldmann, Oliver Gasser, Devashish Gosain, Savvas Zannettou)

### Winter Semester 2021/2022

– Hot Topics in Data Networks Seminar (Anja Feldmann, Oliver Gasser, Yiting Xia)

### Summer Semester 2022

– Data Networks Online-Lecture (Anja Feldmann, Oliver Gasser, Devashish Gosain, Jialong Li, Yiting Xia)

### Winter Semester 2022/2023

– Hot Topics in Data Networks Seminar (Anja Feldmann, Oliver Gasser, Yiting Xia)

### Master and Bachelor Theses

– Shivani Singh: The Effect of Geographical Location on Web Cookies Bachelor thesis, 2021 (Supervisor: Oliver Gasser)
– Bastian Herra: S(rubbing)C(enter)O(bserver): Measuring ddos-protected prefixes from thousands of vantage points Bachelor thesis, 2021 (Supervisor: Lars Prehn)
– Pascal Vermeulen: Understanding BGP Noise Bachelor thesis, 2021 (Supervisor: Lars Prehn)
– Mika Meyer: An Analysis of MOAS Prefixes Bachelor thesis, 2022 (Supervisor: Zubair Sediqi)
– Lukas Vermeulen: Detection and Analysis of VPN Servers in the Internet Bachelor thesis, 2022 (Supervisor: Aniss Maghsoudlou)
– Dil Paul José: Applying Machine Learning Techniques for Quality Assessment in Live Streaming Master thesis, 2021 (Supervisor: Mirko Palmer)
– Fahad Hilal: Large-scale, Stateless, Load Distributive Middlebox Detection in the Internet Master thesis, 2021 (Supervisor: Oliver Gasser)
– Shavran Swaminathan: Optimizing resource scheduling in parallel systems under tight constraints Master thesis, 2022 (Supervisor: Lars Prehn)
– Redion Xhepa: A large scale analysis of images on 4chan using face recognition Master thesis, 2022 (Supervisor: Savvas Zannettou)
– Gayathri Vijan: This Claim is Disputed - An Empirical Analysis of User Engagement with Twitter Warning Labels Master thesis, 2022 (Supervisor: Savvas Zannettou)
– Peter Jose: Discovering Internet of Things Speaking End points in IPv6 Internet Master thesis, 2022 (Supervisor: Said Jawad Saidi)

- Vinay Shetty: Friend or Foe – Onion routing and cryptocurrencies Master thesis, 2022 (Supervisor: Devashish Gosain)
- Pascal Hennen: Inferring and Emulating Routing Policies Master thesis, 2023 (Supervisor: Cristian Munteanu)

## 29.13 Dissertations, Habilitations, Awards

### 29.13.1 Dissertations

- Niklas Semmler *Data-driven transfer optimizations for big data in the industrial Internet of Things* June 2021
- Franziska Lichtblau *From the Edge to the Core: Towards Informed Vantage Point Selection for Internet Measurement Studies*, March 2022.
- Said Jawad Saidi *Characterizing the IoT Ecosystem at Scale* February 2023
- Mirko Palmer *Towards Enabling Cross-layer Information Sharing to Improve Today's Content Delivery Systems* March 2023
- Tobias Bühler *Improving Internet Path Property Inference* March 2023
- Lars Prehn *Routegazing: Analysing the Evolving Internet Routing Ecosystem* February 2023 (submitted for review)
- Aniss Maghsoudlou *Towards Uncovering Hidden Internet Traffic Characteristics* April 2023 (submitted for review)

### 29.13.2 Offers for Faculty Positions

- Savvas Zannettou: Assistant Professor at TU Delft, Netherlands (accepted)
- Habib Mostafaei: Assistant Professor at TU Eindhoven, Netherlands (tenured)
- Costas Iordanou: Assistant Professor at European University Cyprus

### 29.13.3 Awards

- Daniel Wagner and Anja Feldmann: *IETF Applied Networking Research Prize 2022* for the paper *United We Stand: Collaborative Detection and Mitigation of Amplification DDoS Attacks at Scale* published at ACM Conference on Computer and Communications Security (CCS), 2021
- Florian Streibelt, Franziska Lichtblau, Anja Feldmann, Oliver Gasser, Tobias Fiebig: *Best paper award* at Passive and Active Measurement Conference (PAM) 2023 for the paper *How Ready Is DNS for an IPv6-Only World?*
- Savvas Zannettou: *Best reviewer award* at ACM International Conference on Web and Social Media (ICWSM), 2021
- Tobias Fiebig: *Best reviewer award* at ACM Conference on Computer and Communications Security (CCS), 2022

## 29.14 Grants and Cooperations

### SupraCoNex

SupraCoNex is a joint research project about the developement of new algorithms for sequential reinforcement learning, targetting the optimization of decision-making for resource management in wireless networks.

With the growing number of mobile devices, the need for high-bandwidth connectivity is also increasing. WLAN technologies have emerged as a low-cost and efficient alternative to wireless mobile technologies, which is why not only in the private sector but also increasingly network providers are expanding WLAN-based infrastructure alongside their mobile networks in order to relieve their own networks. Furthermore, especially in the age of the Internet of Things, WLAN technology is the primary wireless communication interface. The ever-increasing demands for fast ubiquitous public and private network access are further exacerbated by mobile users (public hotspot, private homespot) and by the emerging growth of machine-to-machine communication for home and industrial automation (SmartHome, Industry 4.0), area-wide sensing and monitoring (IoT, SmartGrid), connected vehicles (Car-2-X), etc. However, the different markets each result in different quality requirements for wireless network communication. The current resource management in wireless networks with WLAN technologies is only suitable to a very limited extent for meeting the individual application requirements.

The overall research goal of SupraCoNeX is to solve the aforementioned challenges of user and network operators in terms of throughput, latency, robustness and security in WLAN-based wireless networks by dynamically parameterising the PHY, MAC and network layers with efficient multi-armed bandit (MAB) algorithms. SupraCoNeX is the first project that will bridge the gaps between theory and practical applicability of MAB resource allocation algorithms to make low-cost WLAN wireless technology useful for a variety of existing and new applications.

- Starting date: August 2020.
- Duration: 3 years.
- Funding: 2 positions, BMBF
- MPI-INF investigators: Prof. Anja Feldmann, Ph.D. Mirko Palmer
- Partners: Hochschule Nordhausen (project coordination) Technische Universität Braunschweig NewMedia-Net GmbH Fraunhofer-Institut für Nachrichtentechnik Heinrich-Hertz-Institut, Berlin BISDN GmbH, Berlin VDI/VDE Innovation + Technik GmbH, Berlin (project execution)

### 6G Research and Innovation Cluster (6G-RIC)

The 6G Research and Innovation Cluster (6G-RIC) is a research hub designed to lay the scientific and technical foundations for the next generation of mobile communications (6G) across all technology levels, from radio access to core networks and fiber optic transport networks. By pursuing cutting-edge research and international networking, 6G-RIC aims to help establish Germany and Europe as global leaders in the expansion of sustainable 6G technologies. 6G-RIC is focused on researching and developing a secure, flexible and open communications infrastructure as the basis for successful digitization in business and all areas of society.

Our research in this project faces challenges primarily from the radically increased requirements of the targeted 6G applications and the fact that many applications also have significantly increased requirements for latency, reliability, energy efficiency and security. In addition, there are new requirements resulting from the integration of sensor services (network as sensor). These requirements also result in challenges for network management, routing and, last but not least, dynamic fault tolerance. For modern network management, models and finally algorithms are to be developed and tested, with special consideration of demand-driven as well as predictive, fault-tolerance-dependent redundancies, caching possibilities and distribution concepts.

The previously common separate consideration of application and communication can significantly limit the efficiency and resilience of the overall system. We will therefore consider and develop methods that also rely on the concept of semantics-based communication (incl. co-design of control and communication). In this context, semantics-based communication solutions will be developed that deviate from traditional communication paradigms by making the "semantics of information" the basis of the communication process, thus enabling a targeted unification of data generation, information transmission and use.

Clients and agents in the 6G RIC context – not only from the IoT environment, but also due to energy efficiency requirements – will be very differently equipped, but the majority will be rather frugal. It is therefore important to investigate the extent to which it is possible to dynamically roll out intelligent aggregators and evaluation instances that summarise the traffic of the endpoints, streamline it if necessary, cache it and forward it as required.

- Starting date: August 2021.
- Duration: 4 years.
- Funding: 2 positions, BMBF
- MPI-INF investigators: Florian Streibelt Zubair Sediqi
- Partners: Fraunhofer-Institut für Nachrichtentechnik Heinrich-Hertz-Institut, HHI, Berlin Christian-Albrechts-Universität zu Kiel Deutsches Zentrum für Luft- und Raumfahrt e. V. (DLR) Institut für Kommunikation und Navigation, Weßling Friedrich-Alexander-Universität Erlangen-Nürnberg Ferdinand-Braun-Institut gGmbH Leibniz-Institut für Höchstfrequenztechnik, Berlin Fraunhofer-Institut für Offene Kommunikationssysteme FOKUS, Berlin Freie Universität Berlin IHP GmbH Leibniz-Institut für innovative Mikroelektronik, Frankfurt (Oder) Rheinisch-Westfälische Technische Hochschule Aachen Humboldt-Universität zu Berlin Fraunhofer-Institut für Angewandte Festkörperphysik IAF, Freiburg Fraunhofer-Institut für Integrierte Schaltungen IIS, Nürnberg Fraunhofer-Institut für Zuverlässigkeit und Mikrointegration IZM, Berlin Technische Universität Berlin Technische Universität Braunschweig Technische Universität Chemnitz Technische Universität Darmstadt Universität Passau Eberhard Karls Universität Tübingen VDI/VDE Innovation + Technik GmbH, Berlin (project execution)

### Quantum Internet Alliance

The Quantum Internet Alliance is a project funded by the European Comission as a strategic area for investment to build a global quantum internet made in Europe. The Quantum Internet Alliance (QIA), led by QuTech—a collaboration between the TU Delft and TNO—will implement a seven-year program with the aim to develop a full-stack prototype network connecting distant cities. The Max-Planck-Institute for Informatics is part of the sub project named "Quantum Communication Test Track"

- Starting date: 1. October 2022.
- Duration: 3 years.
- Funding: 1 position, EU
- MPI-INF investigators: Prof. Anja Feldmann, Ph.D.
- Partners: TU Munich, Humbold-Universität Berlin, Universität Bonn, TU Delft

### Care, People, Gender, Culture

Digital infrastructure has become a backbone of our society. Over the past decades, we grew to rely on the Internet and related infrastructure for all aspects of modern life. This infrastructure has to be maintained and operated on a variety of levels—from home users' network connections to the data centers hosting cloud infrastructure. This is done by system operators, working for the multi-stakeholder ecosystem of organizations providing such services. Recent research has highlighted ongoing shifts in the nature of system operations and its relation to society, especially in the context of power dynamics, responsibility, equity and care.

This cooperation focuses on a research endeavor to expand existing theoretical knowledge on this interaction of power dynamics, responsibility, and care in the operation of digital systems. Specifically, we will empirically investigate operators' stance towards common mechanics to improve the equity and diversity of a space, especially given the traditionally poor gender diversity in system and network engineering. We conjecture that an interplay of cultural factors observed in previous work leads to resistance to initiatives like the introduction of a Code-of-Conduct, despite operators mostly agreeing with the underlying values of such approaches. At the same time, said resistance is read as toxic refusal by people working towards a more equitable system and network engineering community, especially given the track-record of microaggressions and gender stereotyped exclusion and harassment in system operations. Ultimately, this leads to unnecessary conflict and slowed adoption of best practices, even though the underlying perspectives are actually aligned among community members. As such, our research may contribute significant societal benefit by helping to resolve communication challenges and directing efforts to align more quickly towards adopting best practices for improving equity and diversity.

- Starting date: March 2023.
- Duration: 1 year.
- Funding: 1 position, DFG
- MPI-INF investigators: Dr. Tobias Fiebig
- Partners: Dr. Mannat Kaur, TU Delft, Netherlands

## 29.15 Publications

### Books, proceedings, special issues

[1] *Data Governance Act Proposal: A position paper by the research groups "Frameworks for Data Markets", "Work and Cooperation in the Sharing Economy", "Trust in Distributed Environments", "Responsibility and the Internet of Things", and "Reorganizing Knowledge Practices" of the Weizenbaum Institute for the Networked Society*, Weizenbaum Series 18. Weizenbaum Institute for the Networked Society - The German Internet Institute, Berlin, 2021.

### Journal articles and book chapters

[1] D. Chiba, M. Akiyama, Y. Otsuki, H. Hada, T. Yagi, T. Fiebig, and M. Van Eeten. DomainPrio: Prioritizing domain name investigations to improve SOC efficiency. *IEEE Access*, 10:34352–34368, 2022.

[2] A. Feldmann, O. Gasser, F. Lichtblau, E. Pujol, I. Poese, C. Dietzel, D. Wagner, M. Wichtlhuber, J. Tapiador, N. Vallina-Rodriguez, O. Hohlfeld, and G. Smaragdakis. A year in lockdown: How the waves of COVID-19 impact Internet traffic. *Communications of the ACM*, 64(7):101–108, 2021.

[3] T. Fiebig, S. Gürses, C. H. Gañán, E. Kotkamp, F. Kuipers, M. Lindorfer, M. Prisse, and T. Sari. Heads in the clouds? Measuring universities' migration to public clouds: Implications for privacy & academic freedom. *Proceedings on Privacy Enhancing Technologies Symposium (Proc. PETS)*, 2023(2). Accepted 2022.

[4] T. Fiebig, M. Lindorfer, and S. Gürses. Position paper: Escaping academic cloudification to preserve academic freedom. *Privacy Studies Journal*, 1(1):49–66, 2022.

[5] T. Hoeschele, C. Dietzel, D. Kopp, F. H. P. Fitzek, and M. Reisslein. Importance of Internet Exchange Point (IXP) infrastructure for 5G: Estimating the impact of 5G use cases. *Telecommunications Policy*, 45(3), Article 102091, 2021.

[6] G. Jasser, J. McSwiney, E. Pertwee, and S. Zannettou. 'Welcome to #GabFam': Far-right virtual community on Gab. *New Media & Society*, 2021.

[7] M. Kaur, M. van Eeten, M. Janssen, K. Borgolte, and T. Fiebig. Human factors in security research: Lessons learned from 2008-2018. *arXiv preprint arXiv:2103.13287*, 2021.

[8] M. Kaur, S. Parkin, M. Janssen, and T. Fiebig. "I needed to solve their overwhelmness": How system administration work was affected by COVID-19. *Proceedings of the ACM Human-Computer Interaction (Proc. CSCW)*, 6, CSCW2, Article 390, 2022.

[9] M. Kaur, H. S. Ramulu, Y. Acar, and T. Fiebig. "Oh yes! over-preparing for meetings is my jam :)": The gendered experiences of system administrators. *Proceedings of the ACM Human-Computer Interaction (Proc. CSCW)*. Accepted 2022.

[10] H. Mostafaei, G. Smaragdakis, T. Zinner, and A. Feldmann. Delay-resistant geo-distributed analytics. *IEEE Transactions on Network and Service Management*, 19(4):4734–4749, 2022.

[11] S. J. Saidi, O. Gasser, and G. Smaragdakis. One bad apple can spoil your IPv6 privacy. *ACM SIGCOMM Computer Communication Review*, 52(2):10–19, 2022.

[12] Z. Sediqi, L. Prehn, and O. Gasser. Hyper-specific prefixes: Gotta enjoy the little things in interdomain routing. *ACM SIGCOMM Computer Communication Review*, 52(2):20–34, 2022.

[13] H. Sheikhi, M. Hoseini, and M. Sabaei. k-connected relay node deployment in heterogeneous wireless sensor networks. *Wireless Personal Communications*, 120:3277–3292, 2021.

[14] A. Shukla, K. Hudemann, Z. Vági, L. Hügerich, G. Smaragdakis, A. Hecker, S. Schmid, and A. Feldmann. Runtime verification for programmable switches. *IEEE/ACM Transactions on Networking*, 2023.

[15] V. Stocker and G. Knieps. Digitalizing telecommunications: innovation, complexity and diversity in the internet ecosystem. In *A Modern Guide to the Digitalization of Infrastructure*, pp. 59–91. Edward Elgar Publishing, 2021.

[16] V. Stocker, W. Lehr, and G. Smaragdakis. Beyond the pandemic: Towards a digitally enabled society and economy. In *Beyond the Pandemic? Exploring the Impact of COVID-19 on Telecommunications and the Internet*, ch. 12. Emerald, 2023.

[17] V. Stocker, W. Lehr, and G. Smaragdakis. Covid-19 and the internet: Lessons learned. In *Beyond the Pandemic? Exploring the Impact of COVID-19 on Telecommunications and the Internet*, ch. 2. Emerald, 2023.

[18] V. Stocker and J. Whalley. The internet has coped well with covid-19, but problems remain: Evidence to house of lords committee exploring the impact of covid-19. 15, 2021.

[19] J. Whalley, V. Stocker, and W. Lehr. Beyond the pandemic? exploring the impact of covid-19. 2023.

## Conference and workshop articles

[1] T. Albakour, O. Gasser, R. Beverly, and G. Smaragdakis. Third time's not a charm: Exploiting SNMPv3 for router fingerprinting. In *IMC '21, ACM Internet Measurement Conference*, Virtual Event, USA, 2021, pp. 150–164. ACM.

[2] S. Ali, M. H. Saeed, E. Aldreabi, J. Blackburn, E. De Cristofaro, S. Zannettou, and G. Stringhini. Understanding the effect of deplatforming on social networks. In *WebSci '21, 13th ACM Web Science Conference*, Virtual Event, UK, 2021, pp. 187–195. ACM.

[3] M. Aliapoulios, E. Bevensee, J. Blackburn, B. Bradlyn, E. De Cristofaro, G. Stringhini, and S. Zannettou. A large open dataset from the Parler social network. In *Proceedings of the Fifteenth International Conference on Web and Social Media (ICWSM 2021)*, Atlanta, GA, USA, 2021, pp. 943–951. AAAI.

[4] F. Aschenbrenner, T. Shreedhar, O. Gasser, N. Mohan, and J. Ott. From single lane to highways: Analyzing the adoption of multipath TCP in the internet. In *IFIP Networking 2021, IFIP Networking Conference*, Espoo, Finland, 2021, pp. 1–9. IEEE.

[5] I. Ben Guirat, D. Gosain, and C. Diaz. MiXiM: Mixnet design decisions and empirical evaluation. In G. Livraga and N. Park, eds., *WPES '21, 20th Workshop on Workshop on Privacy in the Electronic Society*, Virtual Event, Republic of Korea, 2021, pp. 33–37. ACM.

[6] V. Binkhorst, T. Fiebig, K. Krombholz, W. Pieters, and K. Labunets. Security at the end of the tunnel: The anatomy of VPN mental models among experts and non-experts in a corporate context. In *31st USENIX Security Symposium*, Boston, MA, USA, 2022, pp. 3433–3450. USENIX.

[7] T. Fiebig and D. Aschenbrenner. 13 propositions on an Internet for a "Burning World". In G. Sileno, A. Abhishta, and C. Becker, eds., *TAURIN+BGI '22, ACM SIGCOMM 2022 Joint Workshops on Technologies, Applications, and Uses of a Responsible Internet and Building Greener Internet*, Amsterdam, The Netherlands, 2022, pp. 1–5. ACM.

[8] D. Gosain, A. Jaiswal, H. B. Acharya, and S. Chakravarty. Telemetron: Measuring network capacity between off-path remote hosts. In L. Khoukhi, S. Oteafy, and E. Bulut, eds., *Proceedings of the IEEE 46th Conference on Local Computer Networks (LCN 2021)*, Edmonton, AB, Canada (Virtual), 2021, pp. 351–354. IEEE.

[9] S. Gupta, D. Gosain, G. Grigoryan, M. Kwon, and H. B. Acharya. Demo: Simple deep packet inspection with P4. In *IEEE 29th International Conference on Network Protocols (ICNP 2021)*, Virtual Conference, 2021, pp. 1–2. IEEE.

[10] S. Gupta, D. Gosain, M. Kwon, and H. Acharya. DeeP4R: Deep packet inspection in P4 using packet recirculation. In *International Conference on Computer Communications (INFOCOM)*, 2023. IEEE.

[11] F. Hilal and O. Gasser. Yarrpbox: Detecting middleboxes at internet-scale. In *CoNEXT'23, International Conference on Emerging Networking Experiments And Technologies*, 2023. ACM. Accepted for publication.

[12] F. Holzbauer, J. Ullrich, M. Lindorfer, and T. Fiebig. Not that simple: Email delivery in the 21st century. In *USENIX ATC '22, USENIX Annual Technical Conference*, Carlsbad, CA, USA, 2022, pp. 295–308. USENIX Association.

[13] M. Hoseini, P. Melo, F. Benevenuto, A. Feldmann, and S. Zannettou. On the globalization of the qanon conspiracy theory through telegram. In *Proceedings of the 15th ACM Web Science Conference 2023*, New York, NY, USA, 2023, WebSci '23, p. 75–85. Association for Computing Machinery.

[14] G. Jungwirth, A. Saha, M. Schröder, T. Fiebig, M. Lindorfer, and J. Cito. Connecting the .dotfiles: Checked-in secret exposure with extra (lateral movement) steps. In *20th International Conference on Mining Software Repositories (MSR 2023)*, Melbourne, Australia, 2023. IEEE. Accepted.

[15] D. Kopp, C. Dietzel, and O. Hohlfeld. DDoS never dies? An IXP perspective on DDoS amplification attacks. In O. Hohlfeld, A. Lutu, and D. Levin, eds., *Passive and Active Measurement (PAM 2021)*, Virtual Event, 2021, LNCS 12671, pp. 284–301. Springer.

[16] T. Krenc, R. Beverly, and G. Smaragdakis. As-level bgp community usage classification. In *Proceedings of the 21st ACM Internet Measurement Conference*, 2021, pp. 577–592.

[17] P. Kumar Sharma, D. Gosain, and C. Diaz. On the anonymity of peer-to-peer network anonymity schemes used by cryptocurrencies. In *Network and Distributed System Security Symposium (NDSS 2023)*, San Diego, CA, USA, 2023. Internet Society.

[18] A. Maghsoudlou, O. Gasser, and A. Feldmann. Zeroing in on port 0 traffic in the wild. In O. Hohlfeld, A. Lutu, and D. Levin, eds., *Passive and Active Measurement (PAM 2021)*, Virtual Event, 2021, LNCS 12671, pp. 547–563. Springer.

[19] A. Maghsoudlou, O. Gasser, I. Poese, and A. Feldmann. FlowDNS: Correlating netflow and DNS streams at scale. In G. Bianchi and A. Mei, eds., *CoNEXT '22, 18th International Conference on Emerging Networking Experiments And Technologies*, Roma, Italy, 2022, pp. 187–195. ACM.

[20] A. Maghsoudlou, L. Vermeulen, I. Poese, and O. Gasser. Characterizing the VPN ecosystem in the wild. In A. Brunstrom, M. Flores, and M. Fiore, eds., *Passive and Active Measurement (PAM 2023)*, Virtual Event, 2023, LNCS 13882, pp. 18–45. Springer.

[21] A. Maghsoudlou, L. Vermeulen, I. Poese, and O. Gasser. Characterizing the VPN ecosystem in the wild. In *International Conference on Passive and Active Network Measurement*, 2023. Springer.

[22] V.-A. Pădurean, O. Gasser, R. Bush, and A. Feldmann. SRv6: Is there anybody out there? In *7th IEEE European Symposium on Security and Privacy Workshops (EUROS&P 2022)*, Genoa, Italy, 2022, pp. 252–257. IEEE.

[23] M. Palmer, M. Appel, K. Spiteri, B. Chandrasekaran, A. Feldmann, and R. K. Sitaraman. VOXEL: Cross-layer optimization for video streaming with imperfect transmission. In *CoNEXT '21, 17th International Conference on Emerging Networking Experiments And Technologies*, Virtual Event, Germany, 2021, pp. 359–374. ACM.

[24] A. Papasavva, J. Blackburn, G. Stringhini, S. Zannettou, and E. De Cristofaro. "Is It a Qoincidence?": An exploratory study of QAnon on Voat. In J. Leskovec, M. Grobelnik, M. Najork, J. Tang, and L. Zia, eds., *The Web Conference 2021 (WWW 2021)*, Ljubljana, Slovenia, 2021, pp. 460–471. ACM.

[25] S. Pletinckx, T.-D. Nguyen, T. Fiebig, C. Kruegel, and G. Vigna. Certifiably vulnerable: Using certificate transparency logs for target reconnaissance. In *8th IEEE European Symposium on Security and Privacy Workshops (EUROS&P 2023)*, Delft, The Netherlands, 2023. IEEE. Accepted.

[26] L. Prehn and A. Feldmann. How biased is our validation (data) for AS relationships? In *IMC '21, ACM Internet Measurement Conference*, Virtual Event, USA, 2021, pp. 612–620. ACM.

[27] L. Prehn, F. Lichtblau, C. Dietzel, and A. Feldmann. Peering only? Analyzing the reachability benefits of joining large IXPs today. In O. Hohlfeld, G. Moura, and C. Pelsser, eds., *Passive and Active Measurement (PAM 2022)*, Virtual Event, 2022, LNCS 13210, pp. 338–366. Springer.

[28] A. Rasaii, S. Singh, D. Gosain, and O. Gasser. Exploring the cookieverse: A multi-perspective analysis of web cookies. In A. Brunstrom, M. Flores, and M. Fiore, eds., *Passive and Active Measurement (PAM 2023)*, Virtual Event, 2023, LNCS 13882, pp. 623–651. Springer.

[29] A. Rasaii, S. Singh, D. Gosain, and O. Gasser. A multi-perspective analysis of Web cookies. In *International Conference on Passive and Active Network Measurement*, 2023. Springer.

[30] M. H. Ribeiro, J. Blackburn, B. Bradlyn, E. De Cristofaro, G. Stringhini, S. Long, S. Greenberg, and S. Zannettou. The evolution of the manosphere across the Web. In *Proceedings of the Fifteenth International Conference on Web and Social Media (ICWSM 2021)*, Atlanta, GA, USA, 2021, pp. 196–207. AAAI.

[31] P. Richter, O. Gasser, and A. Berger. Illuminating large-scale IPv6 scanning in the internet. In C. Barakat and C. Pelsser, eds., *IMC '22, ACM Internet Measurement Conference*, Nice, France, 2022, pp. 410–418. ACM.

[32] M. H. Saeed, S. Ali, J. Blackburn, E. De Cristofaro, S. Zannettou, and G. Stringhini. TROLL-MAGNIFIER: Detecting state-sponsored troll accounts on Reddit. In *43rd IEEE Symposium on Security and Privacy (SP 2022)*, San Francisco, CA, USA, 2022, pp. 2161–2175. IEEE.

[33] S. J. Saidi, A. M. Mandalari, H. Haddadi, D. J. Dubois, D. R. Choffnes, G. Smaragdakis, and A. Feldmann. Detecting consumer IoT devices through the lens of an ISP. In *ANRW '21, Applied Networking Research Workshop*, Virtual Event, USA, 2021, pp. 36–38. ACM.

[34] S. J. Saidi, S. Matic, O. Gasser, G. Smaragdakis, and A. Feldmann. Deep dive into the IoT backend ecosystem. In C. Barakat and C. Pelsser, eds., *IMC '22, ACM Internet Measurement Conference*, Nice, France, 2022, pp. 488–503. ACM.

[35] A. Shukla, K. N. Hudemann, Z. Vági, L. Hügerich, G. Smaragdakis, A. Hecker, S. Schmid, and A. Feldmann. Fix with P6: Verifying programmable switches at runtime. In *IEEE INFOCOM 2021*, Vancouver, Canada, 2021, pp. 1–10. IEEE.

[36] V. Stocker, G. Knieps, and C. Dietzel. The rise and evolution of clouds and private networks–internet interconnection, ecosystem fragmentation. In *TPRC49: The 49th Research Conference on Communication, Information and Internet Policy*, 2021.

[37] F. Streibelt, M. Lindorfer, S. Gürses, C. H. Gañán, and T. Fiebig. Back-to-the-future whois: An IP address attribution service for working with historic datasets. In A. Brunstrom, M. Flores, and M. Fiore, eds., *Passive and Active Measurement (PAM 2023)*, Virtual Event, 2023, LNCS 13882, pp. 209–226. Springer.

[38] F. Streibelt, P. Sattler, F. Lichtblau, C. H. Gañán, A. Feldmann, O. Gasser, and T. Fiebig. How ready is DNS for an IPv6-only world? In A. Brunstrom, M. Flores, and M. Fiore, eds., *Passive and Active Measurement (PAM 2023)*, Virtual Event, 2023, LNCS 13882, pp. 525–549. Springer.

[39] F. Streibelt, P. Sattler, F. Lichtblau, C. H. Gañán, A. Feldmann, O. Gasser, and T. Fiebig. How ready is DNS for an IPv6-only world? In *International Conference on Passive and Active Network Measurement*, 2023. Springer.

[40] F. Tahmasbi, L. Schild, C. Ling, J. Blackburn, G. Stringhini, Y. Zhang, and S. Zannettou. "go eat a bat, Chang!": On the emergence of sinophobic behavior on web communities in the face of COVID-19. In J. Leskovec, M. Grobelnik, M. Najork, J. Tang, and L. Zia, eds., *The Web Conference 2021 (WWW 2021)*, Ljubljana, Slovenia, 2021, pp. 1122–1133. ACM.

[41] D. Wagner, D. Kopp, M. Wichtlhuber, C. Dietzel, O. Hohlfeld, G. Smaragdakis, and A. Feldmann. United we stand: Collaborative detection and mitigation of amplification DDoS attacks at scale. In Y. Kim, J. Kim, G. Vigna, E. Shi, H. Kim, and J. B. Hong, eds., *CCS '21, ACM SIGSAC Conference on Computer and Communications Security*, Virtual Event, Republic of Korea, 2021, pp. 970–987. ACM.

[42] D. Wagner, M. Wichtlhuber, C. Dietzel, J. Blendin, and A. Feldmann. P4IX: A concept for P4 programmable data planes at IXPs. In *FIRA '22, ACM SIGCOMM 2022 Workshop on Future of Internet Routing & Addressing*, Amsterdam, Netherlands, 2022, pp. 72–78. ACM.

[43] Y. Wang, F. Tahmasbi, J. Blackburn, B. Bradlyn, E. De Cristofaro, D. Magerman, S. Zannettou, and G. Stringhini. Understanding the use of fauxtography on social media. In *Proceedings of the Fifteenth International Conference on Web and Social Media (ICWSM 2021)*, Atlanta, GA, USA, 2021, pp. 776–786. AAAI.

[44] Y. Wang, S. Zannettou, J. Blackburn, B. Bradlyn, E. De Cristofaro, and G. Stringhini. A multi-platform analysis of political news discussion and sharing on web communities. In Y. Chen, H. Ludwig, Y. Tu, U. Fayyad, X. Zhu, X. Xu, S. Byna, X. Liu, J. Zyhang, S. Pan, V. Papalexakis, J. Wang, A. Cuzzocrea, and C. Ordonez, eds., *IEEE International Conference on Big Data*, Orlando, FL, USA (Virtual Event), 2021, pp. 1481–1492. IEEE.

[45] T. K. Yadav, D. Gosain, A. Herzberg, D. Zappala, and K. Seamons. Automatic detection of fake key attacks in secure messaging. In H. Yin, A. Stavrou, C. Cremers, and E. Shi, eds., *CCS '22, 28th ACM SIGSAC Conference on Computer and Communications Security*, Los Angeles, CA, USA, 2022, pp. 3019–3032. ACM.

[46] S. Zannettou. "i won the election!": An empirical analysis of soft moderation interventions on twitter. In *Proceedings of the Fifteenth International Conference on Web and Social Media (ICWSM 2021)*, Atlanta, GA, USA, 2021, pp. 865–876. AAAI.

[47] J. Zirngibl, L. Steger, P. Sattler, O. Gasser, and G. Carle. Rusty clusters?: Dusting an IPv6 research foundation. In C. Barakat and C. Pelsser, eds., *IMC '22, ACM Internet Measurement Conference*, Nice, France, 2022, pp. 395–409. ACM.

## Technical reports and preprints

[1] T. Albakour, O. Gasser, R. Beverly, and G. Smaragdakis. *Third Time's Not a Charm: Exploiting SNMPv3 for Router Fingerprinting*, 2021. arXiv: 2109.15095.

[2] M. Aliapoulios, E. Bevensee, J. Blackburn, B. Bradlyn, E. De Cristofaro, G. Stringhini, and S. Zannettou. *An Early Look at the Parler Online Social Network*, 2021. arXiv: 2101.03820.

[3] M. W. M. Alzayat, J. Messias, B. Chandrasekaran, K. Gummadi, and P. Loiseau. *Modeling Coordinated vs. P2P Mining: An Analysis of Inefficiency and Inequality in Proof-of-Work Blockchains*, 2021. arXiv: 2106.02970.

[4] F. Aschenbrenner, T. Shreedhar, O. Gasser, N. Mohan, and J. Ott. *From Single Lane to Highways: Analyzing the Adoption of Multipath TCP in the Internet*, 2021. arXiv: 2106.07351.

[5] F. González-Pizarro and S. Zannettou. *Understanding and Detecting Hateful Content using Contrastive Learning*, 2022. arXiv: 2201.08387.

[6] M. Hoseini, P. Melo, F. Benevenuto, A. Feldmann, and S. Zannettou. *On the Globalization of the QAnon Conspiracy Theory Through Telegram*, 2021. arXiv: 2105.13020.

[7] C. Ling, K. Gummadi, and S. Zannettou. *"Learn the Facts About COVID-19": Analyzing the Use of Warning Labels on TikTok Videos*, 2022. arXiv: 2201.07726.

[8] A. Maghsoudlou, O. Gasser, and A. Feldmann. *Zeroing in on Port 0 Traffic in the Wild*, 2021. arXiv: 2103.13055.

[9] P. Paudel, J. Blackburn, E. De Cristofaro, S. Zannettou, and G. Stringhini. *Soros, Child Sacrifices, and 5G: Understanding the Spread of Conspiracy Theories on Web Communities*, 2021. arXiv: 2111.02187.

[10] L. Prehn, P. Foremski, and O. Gasser. *Kirin: Hitting the Internet with Millions of Distributed IPv6 Announcements*, 2022. arXiv: 2210.10676.

[11] M. H. Ribeiro, S. Zannettou, O. Goga, F. Benevenuto, and R. West. *What Do Fact Checkers Fact-check When?*, 2021. arXiv: 2109.09322.

[12] T. Shreedhar, D. Zeynali, O. Gasser, N. Mohan, and J. Ott. *A Longitudinal View at the Adoption of Multipath TCP*, 2022. arXiv: 2205.12138.

[13] F. Streibelt, M. Lindorfer, S. Gürses, C. H. Gañán, and T. Fiebig. Back-to-the-Future Whois: An IP Address Attribution Service for Working with Historic Datasets, 2022.

# 30 D4: Computer Graphics

## 30.1 Personnel

**Director**

Prof. Dr. Hans-Peter Seidel

**Senior Researchers and Group Leaders**

Dr. Vahid Babaei
Dr. Thomas Leimkühler (from September 2021)
Prof. Dr. Karol Myszkowski
Dr. Gurprit Singh
Dr. Paul Strohmeier (from July 2021)
Dr. Rhaleb Zayer

**Researchers**

Dr. Bin Chen
Dr. Mengyu Chu (until February 2022)
Dr. Courtney Reed (from October 2021 until December 2022)
Dr. Lingyan Ruan (from October 2022)
Dr. Tricard Thibault (from May 2022 until September 2022)
Dr. Quan Zheng (until October 2021)

**PhD Students**

Navid Ansari
Elena Arabadzhiyska (until July 2021)
Mojtaba Bemana
Uğur Çoğalan
Martin Balint (from March 2023)
Jozef Hladký (until August 2021)
Sascha Holl (from April 2023)
Xingchang Huang (from February 2021)
Emiliano Luci (from April 2022)
Ntumba Elie Nsampi (from April 2022)
Nihar Sabnis (from October 2021)
Corentin Salaün
Chao Wang
Krzysztof Wolski

**Research Engineer**

Sebastian Cucerca

**Project Coordination**

Azadeh Asadi (from January 2023)
Bertram Somieski

**Secretaries**

Ellen Fries (until October 2021)
Anne Gesellchen (from January 2023)
Sabine Zimmer (from March 2022)

## 30.2 Visitors

From March 2021 to February 2023, the following researchers visited our group:

| | | |
|---|---|---|
| Halehhossadat Mohammadian Esfahani | 01.08.21–31.03.22 | Sharif University of Technology, Teheran, Iran |
| Felix Willems | 13.02.22 | Forschungszentrum Jülich GmbH |
| Nima Vahidi Ferdowsi | 01.03.22–31.10.22 | Ferdowsi University of Mashhad, Mashhad, Iran |
| Dennis Wittchen | 01.03.22–31.12.22 | University of Technology and Economy, Dresden |
| Prof. Mana Mojadrar | 11.05.22 | HTW Saar, Germany |
| Valentin Martinez-Missir | 01.06.22–30.09.22 | Institut de l'internet et du multimedia, IIM et Ecole Supérieure d'Ingénieurs Léonard de Vinci, ESILV |
| Gabriela Vega | 01.06.22–30.11.22 | Peruvian University of Applied Sciences, UPC, Peru |
| Paul Stumpf | 13.06.22 | Trumpf Laser- und Systemtechnik GmbH |
| Kang Liao | 01.09.22–31.08.22 | Beijing Jiaotong University, China |
| Daniel Jiménez Navarro | 13.02.23–14.02.23 | University of Zaragoza, Spain |

## 30.3 Group Organization

Our research is currently organized into the following five research areas, each having its own small group of coordinators:

- Digital Geometry Processing (R. Zayer)
- Sampling, Image Synthesis and Machine Learning (T. Leimkühler and G. Singh)
- Perception: HDR Imaging, VR, and Material Appearance (K. Myszkowski)
- Computational Design and Fabrication (V. Babaei)
- Sensorimotor Interaction (Paul Strohmeier)

The coordinators coordinate the work in their areas together with Hans-Peter Seidel and they form the D4 steering committee. The steering committee meets on a weekly basis (Tuesday, 11 am) and discusses all group related issues. In particular, it addresses topics such as recruiting, guests and seminars, teaching, project acquisition, mid-term and long-term strategic planning.

The whole group meets thrice a week for the

- D4 lab meeting (Tuesday, 12:30 pm), where organizational issues are discussed and information is distributed by the members of the steering committee,

- D4 graphics colloquium (Tuesday, 1 pm), where visitors present their ongoing work to the group, the computer graphics group at Saarland University and to other interested people, and

- D4/D6 graphics lunch (Thursday, 11:30 am), where people from within D4 and D6 present their work in progress to the groups. The main goal of this meeting is to keep the groups informed on the ongoing projects, collect some feedback and influence further project development at relatively early stages.

Apart from these formal meetings, there are several meetings and discussion groups that also take place frequently, but not on a totally regular basis, such as paper discussion groups that discuss papers of special interest, especially immediately preceding major conference events; technical meetings in special areas that are of particular interest to a specific subset of researchers (often in cooperation with people from the graphics group at Saarland University); internship and practical course meetings where all people involved in internships or FoPras meet and discuss; and last but not least meetings dedicated to single projects.

## 30.4 Digital Geometry Processing

*Coordinator: Rhaleb Zayer*

Digital Geometry Processing is concerned with the representation, analysis, manipulation, and optimization of digital shapes. Thanks to recent advances in 3D acquisition and manufacturing technologies, the usage of geometric data is continuously widening and efficient processing of digital shapes is now central to a variety of applications in areas such as computer graphics, computer-aided design and manufacturing, medical imaging and surgery planning, architecture, entertainment, etc. We develop theoretical models, numerical schemes, and algorithms for challenging problems in digital geometry with focus on performance and on streamlining the processing pipeline. While performance is key towards achieving interactivity, intuitive control of the modeling workflow is equally important. In view of the rapidly growing field of machine learning, particular attention is paid to data-driven modeling, especially in the challenging fields of fluid acquisition and simulation.

### 30.4.1 Discovery of Topological Features

*Investigators: Alexander Weinrauch, Caigui Jiang and Rhaleb Zayer*

Probably one of the most inspiring achievements of topology is the classification theorem for surfaces, which establishes equivalence classes based on the *Euler characteristic* and *orientability.* This development triggered a formidable effort focalized on the so called *Poincaré conjuncture* aiming at extending classification to higher dimensions and in particular to the 3−manifold setting and captured the imagination of generations of mathematicians and the general public alike. Loosely speaking, the classification theorem for surfaces (2-manifolds) tells us that any orientable surface is equivalent to a sphere with a certain number of "handles" sewn onto it. In this respect, the surface can be constructed from a sphere by *topological surgery*, which can be understood as a set of cutting, stitching and deforming operations.

The topic of extracting such topological features is not only relevant as a theoretical pursuit but is of importance to several downstream applications such as mesh parameterization, mesh repair, and feature recognition, and spills beyond to other fields such as biotechnology and bioinformatics. Although a sound theoretical foundation of the subject matter has been laid out in algebraic topology, practical localization of geometrically meaningful handles on surfaces remains a highly challenging topic. The majority of existing solutions avoid operating directly on the surface and resort to simpler graph representation or intermediate models such as skeletons or Reeb graphs thus requiring costly, non-trivial, often cumbersome, pre-processing, optimization and post-processing steps.

To address this challenge, our approach reaches onto first principles in topology embodied by the *loop shrinking* property [2]. Consider a person walking on a given surface while holding a sufficiently long thread from both ends. Had the person been on a sphere, it would be possible to spool it back. On the other hand, if the person is on a torus, it won't be possible to re-spool the thread because it passes through the 2-dimensional hole of the torus. We re-abstracted the shrinking loop property as a continuous growth process steered by diffusion. The diffusion field is modeled using a diffuse interface and the dynamics of the advancing fronts are captured by assuming a multilayered field representation where the creation and merging of layers are respectively steered by front splitting and collision. Our approach is the first to operate directly on the piecewise linear manifold setting. It is thus simple, versatile, and inherits the theoretical guarantees of the shrinking loop. The detection of handle and tunnel loops is robust to feature size, surface irregularities as well as high genus as shown in Figure 30.1. Our natively parallel algorithmic solution takes advantage of graphics hardware and breaks the deadlock in performance.

The method naturally extends to Reeb graph construction as the critical events forming the Reeb graph, namely splitting, merging and vanishing of connected components (fronts), are already detected in our handle and tunnel detection. The embedding of those events are obtained by simply taking the mean of the narrow band at the time those events occurred. Typical results are shown in Figure 30.2

Within the context of topology optimization for load-carrying light-weight structures such as trusses, we presented a method for the design of optimal trusses satisfying functional specifications with minimized material consumption [1]. The improvements upon state of the art are twofolds. First, we formulate an alternating linear programming problem for geometry

Figure 30.1: Typical results of our approach on multiple data sets: From top left, Pegasus, Dancing Children, Chair, Grayloc, Thai Statue, Dragon Phoenix, Dragon Ball, Metal Key, Casting, Botijo Jar, the Napoleon model (with zoomed views), the Drill model (double-sided view) and very high genus test cases, Ball, Metal Table, and a model of the double star system V745 Sco Nova (courtesy of NASA).

Figure 30.2: Geometrically embedded Reeb graphs on a selection of relevant test cases.

optimization. Second, we introduce two sets of complementary topological operations, including a novel subdivision scheme for global topology refinement inspired by Michell's famed theoretical study. Based on these two ideas, we build an efficient computational framework for the design of lightweight trusses. We show that our method achieves trusses with smaller volumes and is faster compared with recent state-of-the-art approaches.

## References


- [1] C. Jiang, C. Tang, H.-P. Seidel, R. Chen, and P. Wonka. Computational design of lightweight trusses. *Computer-Aided Design*, 141, Article 103076, 2021.
- [2] A. Weinrauch, H.-P. Seidel, D. Mlakar, M. Steinberger, and R. Zayer. A variational loop shrinking analogy for handle and tunnel detection and Reeb graph construction on surfaces. *Computer Graphics Forum*, 42(2), 2023.


### 30.4.2 Learning Fluids

*Investigators: Mengyu Chu and Rhaleb Zayer, jointly with D6 (Lingjie Liu, Christian Theobalt)*

The design of art-directable fluid simulations remains notoriously a highly challenging task. Despite growing hardware performance and substantial algorithmic improvements, achieving a desired outcome with a physical simulator often requires a tedious, iterative trial-and-error process. Existing methods rely on established physical models, e.g., the Navier-Stokes equations to drive visible quantities, such as a smoke density, with the motion induced by surrounding fluid. Granted little influence over the underlying physics, it is nigh impossible for users to realize their mental image by direct modifications of the density fields.

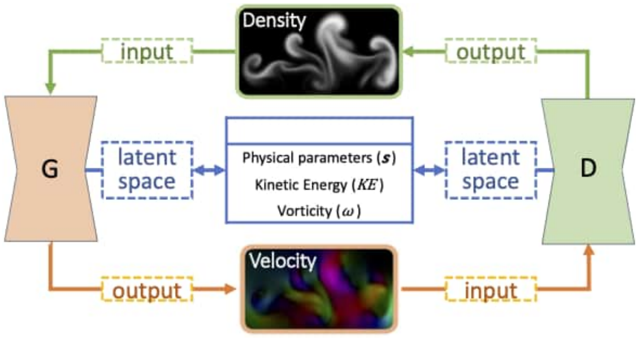



Figure 30.3: In our GAN training, we employ a generator G that translates densities into velocities and a discriminator D checking velocities by inferring densities. Important physical quantities are further encoded in their latent spaces (in blue). Through cyclic adversarial links, sensitive quantitative control is achieved.

Visual effects typically aim to achieve results defined by visible entities that have an obvious semantic meaning for humans, such as the shape of a smoke cloud or the “swirliness” of its motion. Thus, it is advantageous for artists to have a workflow with intuitive knobs and controls to work with these semantic entities, instead of having to tune potentially abstruse physical parameters. This poses an interesting challenge: can we obtain a realistic motion only from a given density configuration? While the problem is ill-posed in the context of traditional physical models, it is an excellent setting for data-driven methods.

We propose a novel deep-learning-based algorithm to encode the physical relationship between a visual quantity, such as the smoke density, and realistic motions as specified by a training data set [1]. Our approach is based on Generative Adversarial Networks (GANs)

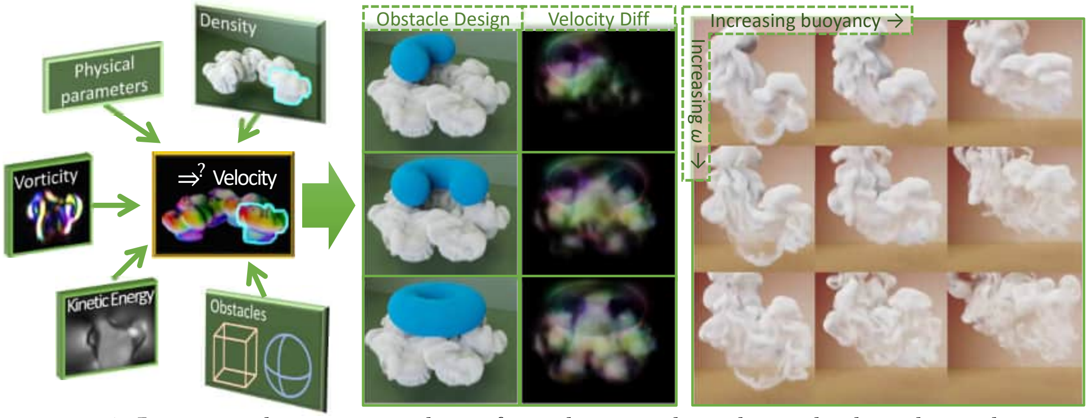



Figure 30.4: Learning the correspondence from density, obstacles and other physical quantities to velocity, our flexible fluid generation method allows for modifications using semantic controls as well as traditional physical parameters. The conditional generation of velocity takes less than 1 second for a 3D volume in resolution of $256^3$. Using our adversarial training with cyclic mapping, we achieve results that can change sensitively with user modifications.

which were shown to be powerful tools to learn learn high-dimensional data distributions, e.g. natural images. However, despite their success, these algorithms have fundamental shortcomings in the presence of multi-modal data, and often *mode-collapse* to a simple subset of the desired outputs. To enable plausible control and mitigate mode collapse, we propose to deeply embed physical quantities in the learned latent-spaces of a GAN, in addition to a cyclic adversarial network design. This prevents the coupled non-linear minimization that needs to be solved during training from converging to undesirable states that largely ignore inputs, as often exhibited by existing conditional GAN algorithms. Our approach, as visualized in Figure 30.3, makes it possible to train networks that learn to synthesize large spaces of flow behavior while at the same time being responsive to changes of a conditioning on semantically meaningful physical quantities, namely physical parameters including buoyancy and boundary conditions, kinetic energy, vorticity, and obstacles. An example of the flexible modifications supported by our networks is shown in Figure 30.4.

### References


- [1] M. Chu, N. Thuerey, H.-P. Seidel, C. Theobalt, and R. Zayer. Learning meaningful controls for fluids. *ACM Transactions on Graphics (Proc. ACM SIGGRAPH)*, 40(4), Article 100, 2021.


### 30.4.3 Dynamic Fluids Reconstruction

*Investigators: Mengyu Chu and Rhaleb Zayer, jointly with D6 (Lingjie Liu, Christian Theobalt)*

High-fidelity reconstruction of dynamic fluids from sparse multiview RGB videos remains a formidable challenge, due to the complexity of the underlying physics as well as the severe occlusion and complex lighting in the captured data. When studying fluid behaviors, one of the core tasks is to estimate the invisible velocity field. In general, velocity can be obtained either by solving physical equations using numerical solvers, or by measuring experimentally, e.g. using particle image velocimetry (PIV), both with different pros and cons. Numerical solvers can achieve high accuracy in solving "forward" fluid problems, but common users cannot easily apply them in an "inverse" manner to handle real life fluid phenomena given the unknown initial and boundary conditions, e.g. steam rising up from a teapot. Experimental techniques, including PIV, allow users to estimate real-life fluid behavior but they remain bound to specialized lab settings and equipment, and the scope is limited to simple scenes where the fluid is the main or sole objective. Furthermore, existing approaches are still susceptible to complex or changing illumination as well as occluding objects.

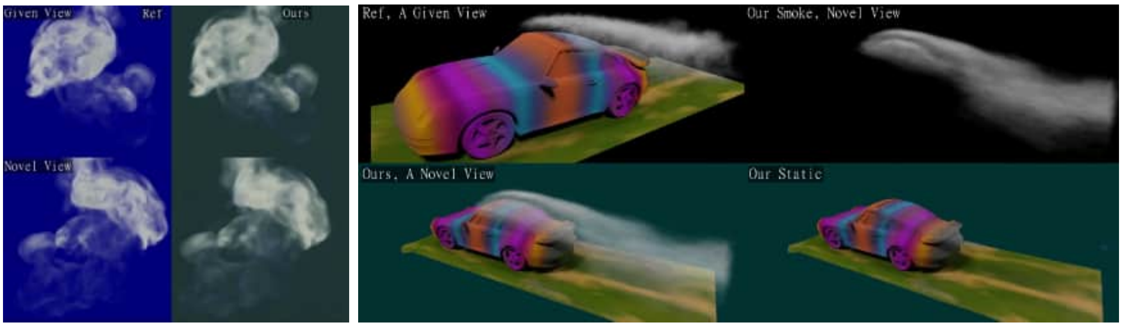



Figure 30.5: Renderings and visualizations of fluid fields reconstructed by our algorithm. Using image sequences from sparse views, we learn continuous radiance and velocity fields to represent of a fluid scene. We can handle synthetic and real fluid scenes as well as hybrid scenes with arbitrary obstacles under unknown lighting conditions.

We aim to relax these constraints, handle scenes with unknown lighting and arbitrary obstacles, and advance the goal of capturing and measuring fluids in natural settings. We propose a novel method for representing dynamic fluid phenomena as continuous space-time functions using image recordings as input [1]. Combining Physics-Informed Neural Networks with Neural Radiance Fields, we aim to leverage physical laws, namely the Navier-Stokes equations, and neural rendering to reveal 4D density and velocity functions from images via end-to-end optimization. Capitalizing on flexibility, we seek a solution that is unrestrained by unknown lighting conditions, geometry information, or boundary conditions, and can be effectively used for analyzing simple buoyant plumes as well as elaborate fluid interactions with static obstacles. The ensuing largely unconstrained task is particularly challenging due the complex nonlinear optimization landscape shaped by physical equations and ambiguity in geometry and lighting, especially with limited camera views. To address these issues, we introduce a regularization term that effectively disentangles the color-opacity ambiguity in the resulting radiance fields. Model-based supervision is proposed to avoid sub-optimal velocity solutions with underestimated vorticity that trivially fulfill the transport equation. With the help of our hybrid scene representation that separates static and dynamic components apart, the reconstruction of hybrid scenes with obstacles is enabled for the first time without additional human labelling. Our method exhibits stability and strong flexibility on a representative set of synthetic and real flow captures as illustrated in Figure 30.5

### References


- [1] M. Chu, L. Liu, Q. Zheng, E. Franz, H.-P. Seidel, C. Theobalt, and R. Zayer. Physics informed neural fields for smoke reconstruction with sparse data. *ACM Transactions on Graphics*, 41(4), Article 119, 2022.


## 30.5 Sampling, Image Synthesis, and Machine Learning

*Coordinators: Thomas Leimkühler and Gurprit Singh*

Synthesizing realistic images typically involves computing high-dimensional light transport integrals. This process is computationally expensive and highly error prone. Traditional rendering algorithms aim towards designing sampling strategies that can ameliorate the error during the estimation of these integrals. Recent advances in neural rendering have opened an exciting line of research where the goal is to synthesize images with the help of deep learning. We develop state-of-the-art algorithms that operate at the intersection between traditional rendering and machine learning-based approaches. At the core, we focus on sampling problems for computer graphics, vision and machine learning, while integrating neural generative priors into the pipeline. We are also interested in interactive rendering and editing of traditional and neural scene representations.

### 30.5.1 Neural Scene Reconstruction with Reflections

*Investigator: Thomas Leimkühler*

View-dependent effects such as reflections pose a substantial challenge for image-based and neural rendering algorithms. Above all, curved reflectors are particularly hard, as they lead to highly non-linear reflection flows as the camera moves. We have introduced a new point-based representation [1] to compute Neural Point Catacaustics allowing novel-view synthesis of scenes with curved reflectors, from a set of casually-captured input photos. At the core of our method is a neural warp field that models catacaustic trajectories of reflections, so complex specular effects can be rendered using efficient point splatting in conjunction with a neural renderer. One of our key contributions is the explicit representation of reflections with a reflection point cloud which is displaced by the neural warp field, and a primary point cloud which is optimized to represent the rest of the scene. Our approach allows interactive high-quality renderings of novel views with accurate reflection flow. Additionally, the explicit representation of reflection flow supports several forms of scene manipulation in captured scenes, such as reflection editing, cloning of specular objects, reflection tracking across views, and comfortable stereo viewing.

## References


- [1] G. Kopanas, T. Leimkühler, G. Rainer, C. Jambon, and G. Drettakis. Neural point catacaustics for novel-view synthesis of reflections. *ACM Transactions on Graphics (Proc. ACM SIGGRAPH Asia)*, 41(6), Article 201, 2022.


### 30.5.2 Editing of Neural Representations

*Investigators: Quan Zheng, Ntumba Elie Nsampi, Adarsh Djeacoumar, Thomas Leimkühler and Gurprit Singh*


Figure 30.6: NeRFshop enables intuitive selection via scribbles and interactive editing of arbitrary NeRF scenes. We show duplicative, affine, and non-affine edits (left to right) from different viewpoints.

Neural fields have recently emerged as a powerful way of representing signals and have witnessed a widespread adoption in particular for visual data. This representation is universal and allows to capture a multitude of modalities. However, to be a true alternative to established specialized representations such as pixel arrays, meshes, point clouds, etc., neural fields are still lacking in a fundamental aspect: They are difficult to edit and hardly amenable to signal processing. We aim to alleviate this limitation by developing both foundational algorithms and practical editing systems.

Neural Radiance Fields (NeRFs) have revolutionized novel-view synthesis for captured scenes, but their implicit representation makes them difficult to edit. Some initial methods have been proposed, but they suffer from limited editing capabilities and/or from a lack of interactivity, and are thus unsuitable for interactive editing of captured scenes. We tackle both limitations and introduce NeRFshop [1], a novel end-to-end method that allows users to interactively select and deform objects through cage-based transformations (Fig. 30.6). NeRFshop provides fine scribble-based user control for the selection of regions or objects to edit, semi-automatic cage creation, and interactive volumetric manipulation of scene content thanks to our GPU-friendly two-level interpolation scheme. Further, we introduce a preliminary approach that reduces potential resulting artifacts of these transformations with a volumetric membrane interpolation technique inspired by Poisson image editing and provide a process that distills the edits into a standalone NeRF representation.

Another limitation of NeRFs is that they are limited to synthesize images under the original fixed lighting condition. Therefore, they are not flexible for the eagerly desired task of relighting. To tackle this problem, several recent methods propose to disentangle reflectance and illumination from the radiance field. These methods can cope with solid objects with opaque surfaces but participating media are neglected. Also, they take into account only direct illumination or at most one-bounce indirect illumination, thus suffer from energy loss due to ignoring the high-order indirect illumination. We propose [2] to learn neural representations for participating media with a complete simulation of global illumination. We estimate direct illumination via ray tracing and compute indirect illumination with spherical harmonics. Our approach avoids computing the lengthy indirect bounces and does not suffer from energy loss. This project is a proof of concept work that demonstrates how to bridge the gap between traditional Monte Carlo rendering and neural radiance fields.

Further, we have developed a general framework to apply a core signal processing technique to neural fields: convolutions. At their core, neural fields only support point samples. In contrast, a convolution requires the continuous integration of values over coordinates weighted by a continuous kernel. In our approach, we consider neural fields to be convolved with piecewise polynomial kernels, which reduce to a sparse set of Dirac deltas after repeated differentiation. Combining this insight with convolution identities on differentiation and integration, our approach requires only a small number of samples from a neural integral field to perform an exact continuous convolution.

### References


- [1] C. Jambon, B. Kerbl, G. Kopanas, S. Diolatzis, T. Leimkühler, and G. Drettakis. NeRFshop: Interactive editing of neural radiance fields. *Proceedings of the ACM on Computer Graphics and Interactive Techniques*, 6(1). Accepted 2023.
- [2] Q. Zheng, G. Singh, and H.-P. Seidel. Neural relightable participating media rendering. In M. Ranzato, A. Beygelzimer, P. S. Liang, J. W. Vaughan, and Y. Dauphin, eds., *Advances in Neural Information Processing Systems 34 (NeurIPS 2021)*, Virtual, 2021, pp. 15203–15215. Curran Associates, Inc.


### 30.5.3 Free-view Rendering with 2D Generative Models

*Investigator: Thomas Leimkühler*

Current 2D Generative Adversarial Networks (GANs) produce photorealistic renderings of portrait images. Embedding real images into the latent space of such models enables high-level image editing. While recent methods provide considerable semantic control over the (re-)generated images, they can only generate a limited set of viewpoints and cannot explicitly control the camera. Such 3D camera control is required for 3D virtual and mixed reality applications. We introduce FreeStyleGAN [1], a new approach that generates an image with StyleGAN defined by a precise 3D camera. This enables faces synthesized with StyleGAN to be used in 3D free-viewpoint rendering, while also allowing all semantic editing provided by GAN methods (Fig. 30.7). We use a few images of a face to perform 3D reconstruction, and we introduce the notion of the GAN camera manifold, the key element allowing us to precisely define the range of images that the GAN can reproduce in a stable manner. We

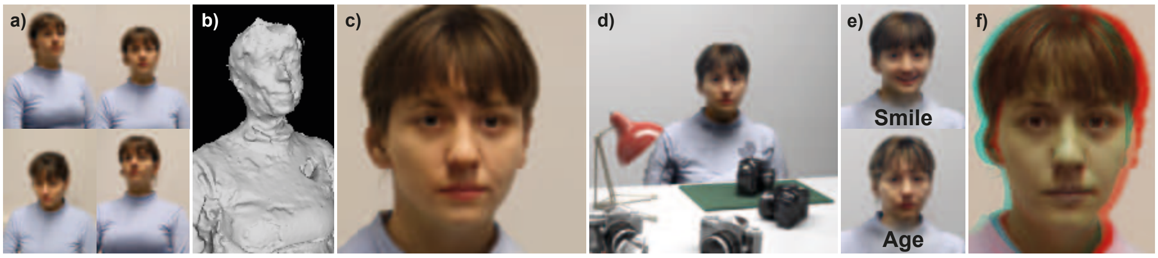



Figure 30.7: FreeStyleGAN takes as input multiple views of a person ($a$), used to reconstruct a coarse 3D mesh ($b$). To render a novel view, we identify the closest camera which corresponds to an image that StyleGAN can generate ($c$). We lift this view to 3D, which allows the integration of our renderings into synthetic 3D scenes ($d$). We inherit the high-quality semantic editing capabilities from StyleGAN (*(e)* smile or aging), and enable stereoscopic rendering ($f$).

train a small face-specific neural implicit representation network to map a captured face to this manifold and complement it with a warping scheme to obtain free-viewpoint novel-view synthesis. We show how our approach – due to its precise camera control – enables the integration of a pre-trained StyleGAN into standard 3D rendering pipelines, allowing e.g., stereo rendering or consistent insertion of faces in synthetic 3D environments. Our solution proposes the first truly free-viewpoint rendering of realistic faces at interactive rates, using only a small number of casual photos as input, while simultaneously allowing semantic editing capabilities, such as facial expression or lighting changes.

## References

• [1] T. Leimkühler and G. Drettakis. FreeStyleGAN: Free-view editable portrait rendering with the camera manifold. *ACM Transactions on Graphics (Proc. ACM SIGGRAPH Asia)*, 40(6), Article 224, 2021.

### 30.5.4 Streaming Rendering

*Investigators: Jozef Hladký and Hans-Peter Seidel*

The landscape of high quality, real-time rendering systems has seen a significant shift over the last years: Real-time rendering is moving towards the cloud and dedicated servers, while display and input happens on inexpensive light-weight devices. However, streaming rendering suffers from increased latency and unreliable connections. High-quality framerate upsampling can hide these issues, especially when capturing shading into an atlas and transmitting it alongside geometric information. The captured shading information must consider triangle footprints and temporal stability to ensure efficient video encoding. Previous approaches only consider either temporal stability or sample distributions, but none focuses on both. With SnakeBinning [1], we have presented an efficient triangle packing approach that adjusts sample distributions and caters for temporal coherence. Using a multi-dimensional binning approach, we enforce tight packing among triangles while creating optimal sample distributions. Our binning is built on top of hardware supported real-time rendering where bins are mapped to

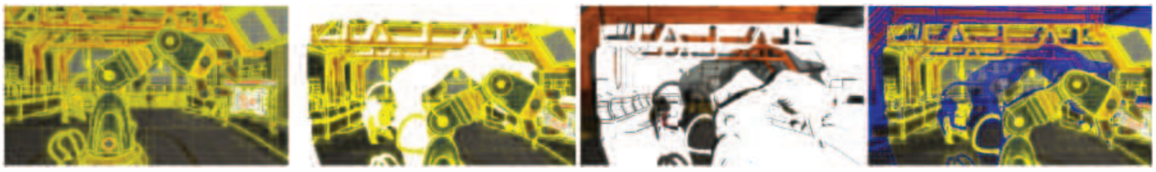


Figure 30.8: The QuadStream architecture enables a server to rasterize and stream a scene to a thin client. The client can then rasterize the scene from a novel viewpoint. From left to right: We decompose a rendered G-Buffer into quad proxies. However, these may not provide enough information for the client to render the scene in the presence of disocclusions at novel views. We therefore augment the QuadStream with quad proxies selectively added from other views within the same view cell, allowing the client to fully reconstruct the scene under motion.

individual pixels in a virtual framebuffer. Fragment shader interlock and atomic operations enforce global ordering of triangles within each bin, and thus temporal coherence according to the primitive order is achieved. Resampling the bin distribution guarantees high occupancy among all bins and a dense atlas packing. Shading samples are directly captured into the atlas using a rasterization pass, adjusting samples for perspective effects and creating a tight packing. A comparison to previous atlas packing approaches shows that our approach is faster than previous work and achieves the best sample distributions while maintaining temporal coherence.

Transmitting a video stream and reprojecting to correct for changing viewpoints fails in the presence of disocclusion events; streaming scene geometry and performing high-quality rendering on the client is not possible on limited-power mobile GPUs. To balance the competing goals of disocclusion robustness and minimal client workload, we have introduced QuadStream [2], a new streaming content representation that reduces motion-to-photon latency by allowing clients to efficiently render novel views without artifacts caused by disocclusion events (Fig. 30.8). Motivated by traditional macroblock approaches to video codec design, we decompose the scene seen into a series of quad proxies. The resulting QuadStream is an approximate geometric representation of the scene that can be reconstructed by a thin client to render both the current view and nearby adjacent views. Our technical contributions are an efficient parallel quad generation, merging, and packing strategy for proxy views covering potential client movement in a scene; a packing and encoding strategy that allows masked quads with depth information to be transmitted as a frame-coherent stream; and an efficient rendering approach for rendering our QuadStream representation into entirely novel views on thin clients. We show that our approach achieves superior quality compared both to video data streaming methods, and to geometry-based streaming.

## References


- [1] J. Hladký, H.-P. Seidel, and M. Steinberger. SnakeBinning: Efficient temporally coherent triangle packing for shading streaming. *Computer Graphics Forum (Proc. EUROGRAPHICS)*, 40(2):475–488, 2021.
- [2] J. Hladký, M. Stengel, N. Vining, B. Kerbl, H.-P. Seidel, and M. Steinberger. Quadstream: A


Figure 30.9: We develop sampling and rendering solutions that are essential for halftoning (left), object placement (middle) and rendering. We take into account the human visual system to model the sampling strategies to obtain perceptually pleasing error distribution during Monte Carlo rendering (right).

quad-based scene streaming architecture for novel viewpoint reconstruction. *ACM Transactions on Graphics (Proc. ACM SIGGRAPH Asia)*, 41(6), Article 233, 2022.

### 30.5.5 Sampling for Stippling, Object Placement and Perceptual Rendering

*Investigators: Corentin Salaun, Karol Myszkowski and Gurprit Singh*

Sampling is at the core of many applications in computer graphics. A lot of research has been dedicated towards developing well distributed sample distributions, e.g. blue noise. Placing objects in a virtual scene or reproducing multi-tones in colored images using multi-class sampling are typical examples where blue noise sampling brings remarkable quality difference. Blue noise sampling has also shown significant impact in improving the convergence properties during Monte Carlo rendering of global illumination effects. Usually, Monte Carlo rendering involves optimising samples within a pixel where numerical integration is performed. Perceptual rendering, however, involves optimizing samples across pixels in such a way that during rendering the noise due to estimation automatically distribute itself in a visually pleasing manner.

Our framework [2] allows sampling for all these tasks—object placement, stippling and getting visually pleasing error distribution on image space during Monte Carlo rendering—in a unified manner. The idea is to represent the problem as a multi-class sampling problem. For example, stippling an image with red (R) and blue (B) colors can result in 3 classes: R, B and RB. We develop a scalable multi-class point optimization formulation based on continuous Wasserstein barycenters. Our formulation is designed to handle hundreds to thousands of optimization objectives and comes with a practical optimization scheme. For perceptual rendering, the samples are optimized irrespective of the scene. The resulting samples can be used for any scene but the resulting quality may not be the best.

We, therefore, develop a perception-oriented framework [1] that can be tailored to a given scene and optimize error during Monte Carlo rendering. We leverage models based on human perception from the halftoning literature. The result is an optimization problem whose solution distributes the error as visually pleasing blue noise in image space. Although the samples are optimized for each scene, the optimization is fast enough to show equal-time improvements compared to the state-of-the-art methods.

### References


- [1] V. Chizhov, I. Georgiev, K. Myszkowski, and G. Singh. Perceptual error optimization for Monte Carlo rendering. *ACM Transactions on Graphics*, 41(3), Article 26, 2022.
- [2] C. Salaün, I. Georgiev, H.-P. Seidel, and G. Singh. Scalable multi-class sampling via filtered sliced optimal transport. *ACM Transactions on Graphics (Proc. ACM SIGGRAPH Asia)*, 41(6), Article 261, 2022.


## 30.5.6 Regression-based Monte Carlo Integration

*Investigators: Corentin Salaun and Gurprit Singh*

Monte Carlo integration has been employed to estimate high-dimensional light transport problems. Monte Carlo estimation involves sampling the integration domain based on some prescribed probability density function, evaluate the function values and then estimate, i.e., compute the mean of these function values. We reformulate Monte Carlo estimation as a regression-based problem where traditional Monte Carlo estimation can be seen as fitting a constant function to the function evaluations. Our idea [1] is to use a polynomial function to fit the function values. This polynomial can be regressed according to the Monte Carlo samples and can be used as a control variate. Control variate estimators are known to reduce the variance in Monte Carlo estimation if the correlation between the integrated function and the control variate is well established. To improve this correlation, we proposed to directly optimize the control variate function, i.e. the polynomial function. By maximizing the correlation, the variance of the estimator is reduced, resulting in a provable improvement over the classical Monte Carlo estimator. This approach requires computing a least square minimization, which is a well-studied problem in mathematics. We tested this method on rendering problems and demonstrated that it can significantly improve the quality of the rendered images.

### References


- [1] C. Salaün, A. Gruson, B.-S. Hua, T. Hachisuka, and G. Singh. Regression-based Monte Carlo integration. *ACM Transactions on Graphics (Proc. ACM SIGGRAPH)*, 41(4), Article 79, 2022.


## 30.5.7 Point Pattern Analysis, Editing and Synthesis

*Investigators: Xingchang Huang and Gurprit Singh*

Depending on how points are distributed over a canvas can change the perception of the underlying density. Any point pattern can be characterized by their density and correlation. Density controls the number of points per unit area and correlation tells us how points are related to its neighbors. We build tools to analyze the local and non-local correlations in point patterns. We build a pipeline that takes an input an exemplar point pattern. We use more traditional Gabor transform-based features. These features when convolved with simple

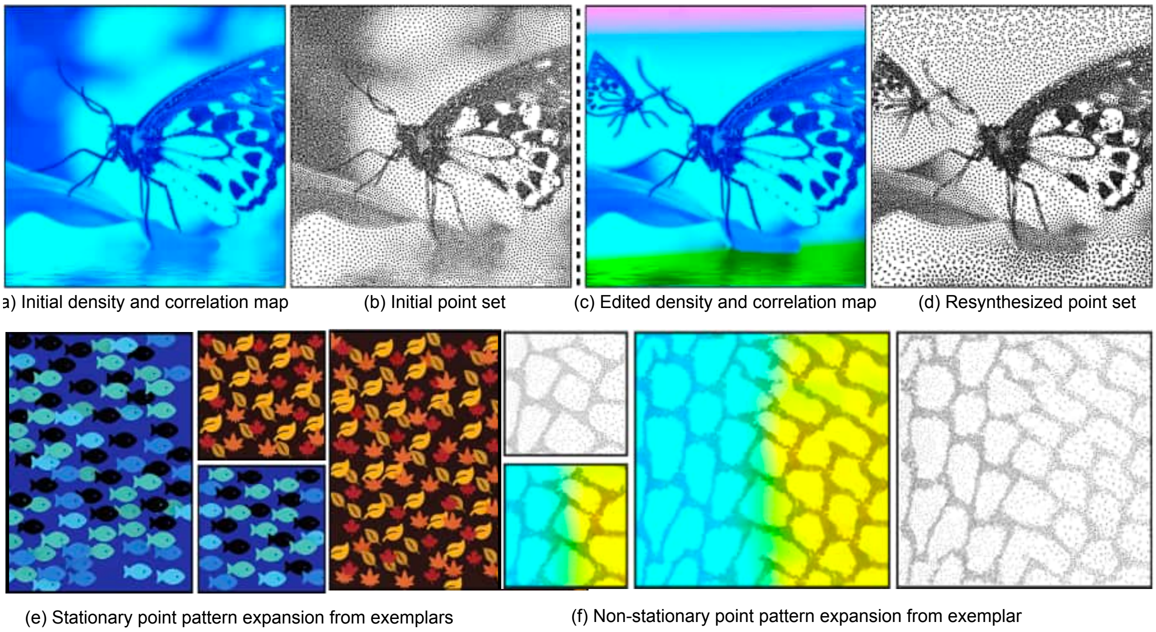



Figure 30.10: We propose a two-dimensional embedding for correlations that allow representing a point pattern (b) as one-channel density and a two-channel correlation map (a). Since these maps are simple three-channel raster images, they can be edited on any image editing software. We perform one such editing in (c) and resynthesize the pointset in (d). Our pipeline also allows non-stationary point pattern expansion (f) without requiring any specialized tools or neural network pipeline. We have also shown how to perform stationary point pattern expansion (e) with data-less optimization and simple random filters.

random filters gives highly expressive feature maps. The resulting framework [1] requires significantly less feature maps compared to existing methods that use VGG-19-based pipeline, without requiring any specific dataset training. Our pipeline extends to multi-class and multi-attribute point patterns.

Analysis and synthesis of point patterns has received a lot of attention in the graphics community. But there are not many tools available to perform editing operations on point patterns. This is primarily due to the lack of a reliable representation for correlations. We are developing a low-dimensional perceptual embedding for point correlations which can be used to encode point correlations through colors. This embedding would allow to embed each point pattern to a common three-channel raster image that can be manipulated with an off-the-shelf image manipulation software. To synthesize points from this three-channel image, we are developing an optimization strategy that carefully handles sharp variations in density and correlation. The resulting framework would allow intuitive and backward-compatible manipulation of point patterns, using feature images. This is an on-going research project.

## References


• [1] X. Huang, P. Memari, H.-P. Seidel, and G. Singh. Point-pattern synthesis using Gabor and random filters. *Computer Graphics Forum (Proc. Eurographics Symposium on Rendering)*, 41(4):169–179, 2022.


## 30.6 Perception: HDR Imaging, VR, and Material Appearance

*Coordinator: Karol Myszkowski*

Understanding complex characteristics of the human visual system (HVS) is important in the context of static and dynamic content depiction. In particular, the quality of rendering systems and display devices might be aligned with HVS limitations, so that visual cues that cannot readily be seen are not computed and displayed [2]. On the other hand, traditional camera sensors and displays feature serious limitations in reproduced contrast and brightness with respect to the real world. High dynamic range imaging (HDRI) [1] aims to compensate for such hardware deficits that ideally should not be easily revealed to the HVS. Along these lines, in Sec. 30.6.1 we investigate HDR video capturing using a dual-exposure sensor and a neural tone mapping solution. In Sec. 30.6.2 we present an efficient Neural Radiance Field (NeRF) solution that enables the rendering of refractive objects based on a limited number of input photographs. In Sec. 30.6.3 we propose perception-based quality metrics specialized to Virtual Reality (VR) rendering and display solutions. Since we consider 3D printing as another aspect of the realistic appearance reproduction of virtual content, we present our solutions for reducing perceptual gaps between displayed and fabricated objects (Sec. 30.6.4).

## References


[1] E. Reinhard, G. Ward, S. Pattanaik, P. Debevec, W. Heidrich, and K. Myszkowski, eds. *High Dynamic Range Imaging: Acquisition, Display, and Image-based Lighting.* Elsevier (Morgan Kaufmann), Burlington, MA, 2. ed., 2010.

[2] M. Weier, M. Stengel, T. Roth, P. Didyk, E. Eisemann, M. Eisemann, S. Grogorick, A. Hinkenjann, E. Krujiff, M. A. Magnor, K. Myszkowski, and P. Slusallek. Perception-driven accelerated rendering. *Computer Graphics Forum*, 36(2):611–643, 2017.


### 30.6.1 High-dynamic Range (HDR) Imaging

*Investigators: Chao Wang, Uğur Çoğalan, Mojtaba Bemana, Bin Chen and Karol Myszkowski*

HDR imaging enables creating images with a wider range of brightness and color by taking and combining multiple exposures captured with different exposure times. While HDR content is becoming more prevalent, tone mapping (TM) operation is still a significant challenge for image visualization due to limited dynamic range on display devices. Neural networks have shown promise in TM, but their performance is limited by training data. In our work [3], we propose a self-supervised tone mapping operator trained at test time for each HDR image using a carefully designed loss function based on contrast perception. We achieve this goal by reformulating classic VGG feature maps into feature contrast maps that normalize local feature differences by their average magnitude in a local neighborhood, allowing our loss to account for contrast masking effects.

HDR video is also gaining popularity as it captures wider dynamic range content than traditional video. HDR video reconstruction is typically done using temporally alternating exposures, but this requires exposure alignment which is difficult for moving content. An attractive alternative are dual-exposure sensors that capture differently exposed and spatially interleaved half-frames in a single shot, which are perfectly aligned by construction. In our work [2], we demonstrate successful compensation for reduced spatial resolution and aliasing in such sensors, improving the quality and dynamic range of reconstructed HDR video compared to single-exposure sensors. We aim to reconstruct three short, medium and long exposures for each frame. The medium exposure is captured for every frame, and serves as a spatial and temporal reference. By using neural networks for denoising, deblurring, and upsampling tasks, we obtain two clean, sharp, and full-resolution exposures for every frame, complemented by warping a missing third exposure.

Lastly, in [1], we explore how dual-exposure sensors can be advantageous in accurately handling HDR scenes with complex motion in the context of video frame interpolation (VFI). These sensors provide sharp short and blurry long exposures that are temporally aligned, allowing for more precise motion sampling within a single camera shot. Our work demonstrates that this facilitates a more complex motion and HDR frame reconstruction in the VFI task. Additionally, we propose a metric for scene motion complexity that provides valuable insights into the performance of VFI methods at test time.

## References


- [1] U. Çoğalan, M. Bemana, K. Myszkowski, H.-P. Seidel, and T. Ritschel. Learning HDR video reconstruction for dual-exposure sensors with temporally-alternating exposures. *Computers and Graphics*, 105:57–72, 2022.
- [2] U. Çoğalan, M. Bemana, H.-P. Seidel, and K. Myszkowski. *Video Frame Interpolation for High Dynamic Range Sequences Captured with Dual-exposure Sensors*, 2022. arXiv: 2206.09485.
- [3] C. Wang, B. Chen, H.-P. Seidel, K. Myszkowski, and A. Serrano. Learning a self-supervised tone mapping operator via feature contrast masking loss. *Computer Graphics Forum (Proc. EUROGRAPHICS)*, 41(2):71–84, 2022.


### 30.6.2 Refractive Novel-view Synthesis

*Investigators: Mojtaba Bemana and Karol Myszkowski*

Traditional photorealistic rendering of real-world scenes proves tedious and challenging due to the need to reconstruct all physical parameters describing the rendered scenes. Recently, implicit scene representations [2, 3] have become a viable alternative to this task, where the entire scene is encoded into the parameters of a neural network. Despite recent advances in scene representations, existing approaches cannot properly reconstruct novel views of transparent objects with complex refraction and require special treatment. Our work presented in [1] tackles this problem by lifting the assumption that light rays are traversing in straight


Figure 30.11: Novel-view synthesis using NeRF [3] (top) as well as our eikonal approach (bottom) for a real scene containing a refractive object.

lines and adapting a physically correct approach to bend the light rays when they intersect with a refractive object in the scene. Specifically, we integrate the physical laws of the eikonal light transport with a state-of-the-art novel view synthesis method (NeRF [3]) and adapt an implicit representation that can model refractive objects with a spatially varying index of refraction, leading to high-quality novel view reconstruction of refractive objects (Figure 30.11) without requiring a dedicated capturing setup or scene configuration.

## References


• [1] M. Bemana, K. Myszkowski, J. R. Frisvad, H.-P. Seidel, and T. Ritschel. Eikonal fields for refractive novel-view synthesis. In M. Nandigjav, N. J. Mitra, and A. Hertzmann, eds., *Proceedings SIGGRAPH 2022 Conference Papers Proceedings (ACM SIGGRAPH 2022)*, Vancouver, Canada, 2022, Article 39. ACM.

[2] M. Bemana, K. Myszkowski, H.-P. Seidel, and T. Ritschel. X-Fields: Implicit neural view-, light- and time-image interpolation. *ACM Transactions on Graphics (Proc. ACM SIGGRAPH Asia)*, 39(6), Article 257, 2020.

[3] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. In *ECCV*, 2020, pp. 405–421.


### 30.6.3 Improving Visual Experience in VR and Games

*Investigators: Krzysztof Wolski and Karol Myszkowski*

Image quality metrics play an important role in computer graphics as they provide objective measurements of the quality and performance of digital images, enabling the automation of multiple applications. While we investigated the distortions that might occur in image space [5], we also noted that the distortions might be introduced much earlier in the graphics pipeline, specifically in the meshes representing 3D objects.

Therefore, in [6], we chose to employ a new perception-based method to measure the distortions in the 3D geometry of faces. This is significant because faces are widely used in gaming, entertainment, and social applications and humans are particularly sensitive to their appearance. In collaboration with Reality Labs (Meta), we generated a dataset composed of 100 high-quality and demographically-balanced face scans. We then subjected these meshes

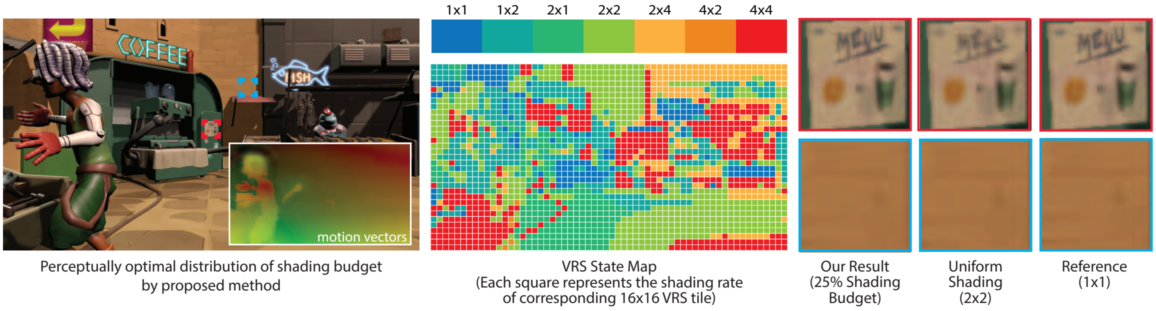



Figure 30.12: Demonstration of the proposed method for adaptive control of local shading and refresh rate. Based on the information contained in the previous frame and motion vectors (left side) method provides the optimal refresh rate and VRS state map (center), which is used to achieve superior visual fidelity compared to the state-of-the-art solutions at the same computational cost (right side).

to distortions that cover relevant use cases in computer graphics and conducted a large-scale perceptual study to subjectively evaluate them. Our dataset consists of over 84,000 quality comparisons, making it the largest-ever psychophysical dataset for geometric distortions. Finally, we demonstrated how our data can be used for applications like state-of-the-art metrics calibration, compression, and level-of-detail rendering.

When the source material is impeccable given the outcomes of the mesh and image quality metrics, it is still important to pay attention to how the given content is rendered and displayed to the end user. Considering cloud gaming or other distributed rendering applications, it is often necessary to limit the rendering budget and hence to find the combination of rendering parameters, such as resolution and refresh rate, that could deliver the best quality. To solve this problem we take advantage of the capability provided by Variable Rate Shading (VRS) and propose a new method for adaptive control of local shading and refresh rate [2]. The method analyzes texture content, on-screen velocities, luminance, and effective resolution and suggests the refresh rate and a VRS state map that maximizes the quality of animated content under a limited budget. The method is based on the new content-adaptive metric of judder, aliasing, and blur, which is derived from the psychophysical models of contrast sensitivity. The proposed metric and adaptive shading method has been implemented as a game engine plugin and further used in our experimental validation that shows a substantial increase in preference for our method over rendering with a fixed resolution and refresh rate (Figure 30.12), and an existing motion-adaptive technique.

Finally, it is crucial to consider the type of display the final image is going to be displayed at. As the concept of Metaverse gets more popular nowadays, we decided to investigate issues that might rise from the use of Virtual Reality (VR) Head-Mounted Displays (HMD). These can span from large physical space required to fully immerse into the virtual world [3], to flicker artifacts that are introduced by low-persistence displays installed in VR headsets [7]. The latter issue can be solved by displaying content at low brightness, allowing to experience smoother motion at lower refresh rates. Moreover, this operation might also prolong the battery life of the standalone HMDs. Unfortunately, such a solution is not flawless, as the binocular depth cues become less reliable at low luminance. In [7], we propose a model of


Figure 30.13: Crossed fusion stereoscopic pairs demonstrating the proposed stereo constancy algorithm. Left: original scene. Right: after stereo constancy processing. The images should be seen on a dimmed display (about $5\,\mathrm{cd/m^2}$ peak brightness) and enlarged to the point that both pairs stretch the width of the screen.

stereo constancy that predicts the precision of binocular depth cues for a given contrast and luminance. We use the model to design a novel contrast enhancement algorithm that compensates for the deteriorated depth perception to deliver good-quality stereoscopic images even for displays of very low brightness (refer to Figure 30.13).

Foveated rendering can greatly improve the computation performance in VR setups [1]. A foveated image can be entirely reconstructed from a sparse set of samples distributed according to the retinal sensitivity of the human visual system, which rapidly decreases with increasing eccentricity. The use of Generative Adversarial Networks has recently been shown to be a promising solution for such a task, as they can successfully hallucinate missing image information. In [4] we consider the problem of efficiently guiding the training of foveated reconstruction techniques such that they are more aware of the capabilities and limitations of the human visual system, and thus can reconstruct visually important image features. Our evaluations revealed significant improvements in the perceived image reconstruction quality compared with the standard GAN-based training approach.

## References


- [1] E. Arabadzhiyska, C. Tursun, H.-P. Seidel, and P. Didyk. Practical saccade prediction for head-mounted displays: Towards a comprehensive model. *ACM Transactions on Applied Perception*, 20(1), Article 2, 2023.
- [2] A. Jindal, K. Wolski, R. K. Mantiuk, and K. Myszkowski. Perceptual model for adaptive local shading and refresh rate. *ACM Transactions on Graphics (Proc. ACM SIGGRAPH Asia)*, 40(6), Article 281, 2021.

[3] A. Serrano, D. Martin, D. Gutierrez, K. Myszkowski, and B. Masia. Imperceptible manipulation of lateral camera motion for improved virtual reality applications. *ACM Transactions on Graphics (Proc. ACM SIGGRAPH Asia)*, 39(6), Article 268, 2020.

- [4] L. Surace, M. Wernikowski, C. Tursun, K. Myszkowski, R. Mantiuk, and P. Didyk. Learning GAN-based foveated reconstruction to recover perceptually important image features. *ACM Transactions on Applied Perception*, 2023.


Figure 30.14: We present a model that reckons with the effects of shape and illumination for predicting material appearance attributes that correlate with human judgments and show that it can be leveraged for several applications. Left: Our predictor can be used to sort material datasets according to desired properties such as perceived glossiness and lightness for target illuminations or shapes, assisting scene design. Right: We demonstrate the capabilities of our predictor for gloss reproduction in 3D printing. While applying the same varnish under different illuminations and shapes would yield different gloss perception, our predictor allows us to find optimal varnish mixtures (insets) for reproducing the desired equivalent gloss under different illuminations and geometries.

[5] K. Wolski, D. Giunchi, N. Ye, P. Didyk, K. Myszkowski, R. Mantiuk, H.-P. Seidel, A. Steed, and R. K. Mantiuk. Dataset and metrics for predicting local visible differences. *ACM Transactions on Graphics*, 37(5), Article 172, 2018.

• [6] K. Wolski, L. Trutoiu, Z. Dong, Z. Shen, K. Mackenzie, and A. Chapiro. Geo-metric: A perceptual dataset of distortions on faces. *ACM Transactions on Graphics (Proc. ACM SIGGRAPH Asia)*, 41(6), Article 215, 2022.

• [7] K. Wolski, F. Zhong, K. Myszkowski, and R. K. Mantiuk. Dark stereo: Improving depth perception under low luminance. *ACM Transactions on Graphics (Proc. ACM SIGGRAPH)*, 41(4), Article 146, 2022.

### 30.6.4 Material Appearance Perception

*Investigators: Bin Chen, Chao Wang and Karol Myszkowski*

Material perception has been investigated in the past, where human rating data of numerous material appearance attributes, such as glossiness and lightness, have been collected for many different BRDFs [4]. However, only a single geometry (sphere) and one illumination scenario were used for stimuli rendering. In our recent work [3], we present a large-scale dataset of human judgments on the appearance attributes of materials under various geometric and illumination conditions. Our dataset contains more than 215,680 responses for 42,120 unique combinations of material properties, shape models, and illumination environments. Based on this massive dataset, firstly, we conduct a comprehensive analysis of how illumination and

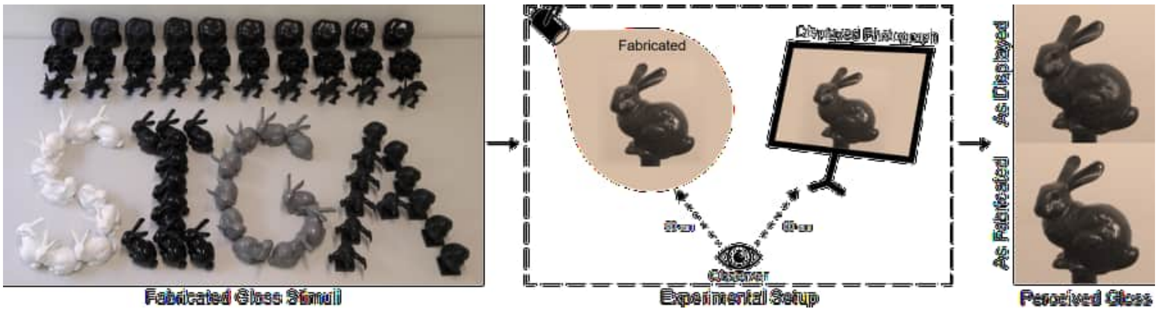



Figure 30.15: We study differences in gloss perception between physical and displayed stimuli. We start by fabricating an extensive set of geometries with different glossy finishes (left). We then use these models in a psychophysical experiment that compares them to their digital counterparts (center). The results demonstrate significant differences in perceived gloss between the two presentation methods (right). We address this gap with a new gloss management system that compensates for the observed differences.

geometry influence the perception of materials across a wide range of diverse appearances; secondly, we train a deep learning architecture that can predict perceptual attributes that are consistent with human judgments. For the first time, we show a reliable predictor that can robustly estimate perceived material attributes from general 2D images (Figure 30.14-left). Unlike previous methods that rely on specific geometry or illumination assumptions, our predictor can operate on the final appearance in an image and compare appearance properties across different geometries and illumination conditions. Finally, we demonstrate several applications that benefit from our predictor such as reproducing appearance using 3D printing technology (Figure 30.14-right), editing BRDF parameters by integrating our predictor in a differentiable renderer framework, designing optimal illumination for enhancing material perception, or recommending suitable materials for scene composition. Moreover, in our follow-up work [2] we further investigate the impact of the interactions between illumination, geometry, and eight different material categories in perceived appearance attributes.

All investigations that we have performed so far focus on reproducing the material appearance on the display devices. In our subsequent work [1], we try to answer a question: could we really reproduce correct material appearance on a desktop display, even if we have accurately rendered images of accurate BRDF using realistic geometry and illumination models? A good match of material appearance between real-world objects and their digital on-screen representations is a crucial requirement for many applications such as fabrication, design, and e-commerce that involve creating or manipulating digital models of physical objects. Unfortunately, most display devices have limited capabilities in terms of viewing angle, reproduced contrast, and luminance range. Moreover, the 2D display lacks important visual cues such as binocular vision, motion parallax, or variable eye accommodation. In this work, we investigate the gap between the perceived glossiness of real-world objects and their digital counterparts on a display device. We conduct psychophysical experiments with a large set of 3D printed samples and their corresponding photographs under different geometric and lighting conditions (refer to Figure 30.15). We analyze how geometry, illumination, and display luminance influence the perception of glossiness and measure the change in perceived glossiness due to the display limitations. Based on our analysis, we propose a method to predict how much the perceived glossiness changes when going from real objects to digital images. We use this prediction to correct the material parameters in a rendering system that generates realistic images of glossy materials. We evaluate our method with additional experiments and show that it improves the match of perceived glossiness between real objects and their visualization on a display device.

### References


- [1] B. Chen, M. Piovarči, C. Wang, H.-P. Seidel, P. Didyk, K. Myszkowski, and A. Serrano. Gloss management for consistent reproduction of real and virtual objects. In S. K. Jung, J. Lee, and A. Bargteil, eds., *Proceedings SIGGRAPH Asia 2022 (ACM SIGGRAPH Asia 2022)*, 2022, Article 35. ACM.
- [2] B. Chen, C. Wang, M. Piovarči, H.-P. Seidel, P. Didyk, K. Myszkowski, and A. Serrano. The effect of geometry and illumination on appearance perception of different material categories. *The Visual Computer*, 37:2975–2987, 2021.
- [3] A. Serrano, B. Chen, C. Wang, M. Piovarči, H.-P. Seidel, P. Didyk, and K. Myszkowski. The effect of shape and illumination on material perception. *ACM Transactions on Graphics (Proc. ACM SIGGRAPH)*, 40(4), Article 125, 2021.

[4] A. Serrano, D. Gutierrez, K. Myszkowski, H.-P. Seidel, and B. Masia. An intuitive control space for material appearance. *ACM Transactions on Graphics (Proc. ACM SIGGRAPH Asia)*, 35(6), Article 186, 2016.


## 30.7 Computational Design and Fabrication

*Coordinator: Vahid Babaei*

In the AI Aided Design and Manufacturing Group, we are deeply interested in developing computational methods that address the digitalization of manufacturing. Inspired by computer graphics methods, in our research on *computational design*, we develop algorithmic tools to *evaluate*, *represent*, and *synthesize* products with improved or completely novel functions. Our research on *intelligent manufacturing* strives to radically change the manufacturing process through bringing the fabrication hardware inside the computation loop, and also empowering this hardware with on-the-fly sensing and computational decision making. The applications of this research span from patient-specific ocular and facial prosthetics with a detailed appearance that matches that of human eyes and skin, to digital manufacturing solutions that enable novel products through deep integration of on-the-fly sensing and computational decision making, and data-driven simulation toolboxes that unify the discovery of novel designs in mechanics, photonics, acoustics, and beyond. In Sec. 30.7.1, we summarize our contributions in inverse design where the forward process is computed via emerging neural surrogate models. In Sec. 30.7.2, we review our work on using reinforcement learning control for challenging scenarios in additive manufacturing. In Sec. 30.7.3, we present our continued efforts to make the case for the use of coordinate based networks as efficient design representation for computational design. Finally, Sec. 30.7.4 summarizes our contribution in the field of inverse design via PDE-constrained topology optimization.

### 30.7.1 Neural Inverse Design

*Investigators: Navid Ansari and Vahid Babaei*

With the proliferation of *neural surrogate models* (NSMs) for design evaluation, we have been encountering a recurring question of how to *invert* these surrogates during design synthesis. This question is therefore being addressed not only in machine learning [3] but also in photonics [5], computer graphics [6] and computational design [4]. What sets our work on *neural inverse design* apart from the rest of the field, is the observation that neural networks, despite their black-box reputation during training, are rather well-behaved mathematical functions once trained. By leveraging the underlying mathematical structure of NSMs, we have introduced several novel methods of design synthesis. In the following, we highlight the two most recent works from our group on this topic.

*Mixed-Integer Neural Inversion.* A general neural inverse problem can be formalized as a search for an input vector $\mathbf{x^0}$, serving our design parameterization, that minimizes a distance, e.g., $\mathcal{L}_1$ norm, between the prediction of a NSM ($F_\theta$), with pretrained weights and biases $\theta$, and a target performance $\mathbf{t}$

$$\underset{\mathbf{x}^0}{\operatorname{argmin}} \left\| F_\theta(\mathbf{x^0}) - \mathbf{t} \right\|_1 . \tag{30.1}$$

This optimization is very challenging as $F_\theta$ is a highly non-linear, non-convex function. In addition, many computational design problems involve discrete variables. That is, $\mathbf{x^0}$ must take integer values. We recently [1] addressed these challenges for omnipresent *piecewise linear* NSMs. A piecewise linear NSM $F_\theta$, is a composition of linear transformations and piecewise linear activation functions, such as the rectified linear unit (ReLU). Our main insight is that we can formulate the neural inverse design as a mixed integer linear programming (MILP) and solve it using powerful mixed-integer solvers. We exploited the piecewise linear structure of neural networks and showed that an equivalent optimization to the above optimization (Eq. 30.1) can be formulated, which involves only linear terms and constraints. In doing so, we eliminated the network's non-linearities at the cost of introducing new binary and continuous variables. We demonstrated that our method significantly facilitates the emerging but challenging combinatorial inverse design tasks.

*Uncertainty-Aware Neural Inversion.* In any neural inverse design framework, the inversion is carried out on a forward neural surrogate model (NSM) that is computed through fitting to the data sampled from a *native forward process* (**NFP**), e.g., a physics simulation. Inversion of a surrogate model, however, is ultimately different than the inversion of the NFP itself. None of the competing methods to date has offered a solution for this crucial gap. Recently [2], we *expected* and *accounted* for any potential mismatch between the data, and consequently the surrogate, on the one hand and the native forward process on the other hand. Our proposed method, *Autoinverse*, realizes this vision by taking into account the predictive uncertainty of the surrogate (both epistemic and aleatoric) and minimizing it during the

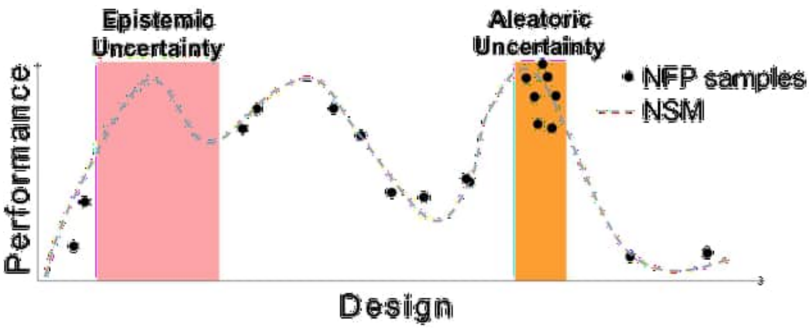



Figure 30.16: A forward Bayesian NSM predicts not only the performance but the associated (epistemic and aleatoric) uncertainty. During the inversion, we penalize uncertainty and thus find solutions where the gap between the surrogate and the NFP is minimal.

inversion (Figure 30.16). Apart from high accuracy, *Autoinverse* automatically enforces the feasibility of solutions, is initialization free, and comes with embedded regularization, thus freeing the inversion approaches from hand-crafted regularizations based on domain knowledge.

### References


- [1] N. Ansari, H.-P. Seidel, and V. Babaei. Mixed integer neural inverse design. *ACM Transactions on Graphics (Proc. ACM SIGGRAPH)*, 41(4), Article 151, 2022.
- [2] N. Ansari, H.-P. Seidel, N. Vahidi Ferdowsi, and V. Babaei. Autoinverse: Uncertainty aware inversion of neural networks. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, eds., *Advances in Neural Information Processing Systems 35 (NeurIPS 2022)*, New Orleans, LA, USA, 2022, pp. 8675–8686. Curran Associates, Inc.

[3] C. Fannjiang and J. Listgarten. Autofocused oracles for model-based design. *Advances in Neural Information Processing Systems*, 33:12945–12956, 2020.

[4] K. Gavriil, R. Guseinov, J. Pérez, D. Pellis, P. Henderson, F. Rist, H. Pottmann, and B. Bickel. Computational design of cold bent glass façades. *ACM Transactions on Graphics (TOG)*, 39(6):1–16, 2020.

[5] J. Jiang, M. Chen, and J. A. Fan. Deep neural networks for the evaluation and design of photonic devices. *Nature Reviews Materials*, pp. 1–22, 2020.

[6] L. Shi, V. Babaei, C. Kim, M. Foshey, Y. Hu, P. Sitthi-Amorn, S. Rusinkiewicz, and W. Matusik. Deep multispectral painting reproduction via multi-layer, custom-ink printing. *ACM Trans. Graph.*, 37(6):271:1–271:15, 2018.


### 30.7.2 Reinforcement Learning for Control of Additive Manufacturing

*Investigators: Thibault Tricard, Kang Liao and Vahid Babaei*

Despite the ongoing advances of computational methods in creating innovative content for manufacturing, the potential of these methods for optimizing or controlling the manufacturing processes in an intelligent way is largely untapped. Our research advocates the integration

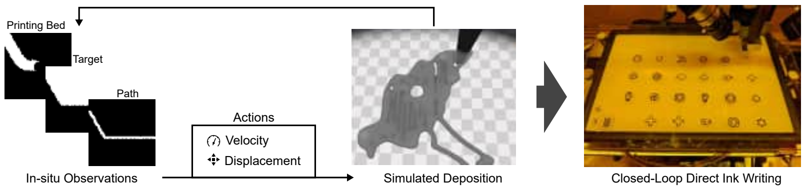



Figure 30.17: We propose a numerical environment suitable for learning close-loop control strategies for additive manufacturing via direct ink writing. Our method observes an in-situ view of the printing process and adjusts the velocity and printing path to achieve the desired deposition. The control policies learned exclusively in simulation can be deployed on real hardware.

of the fabrication hardware into the computational pipeline and not treating the hardware merely as a "consumer" of the created content. In the current reporting period, we have been developing reinforcement learning based controllers for challenging additive manufacturing setups, such as direct ink writing with low-viscosity materials [2], and multi-material fusion [1].

*Reinforcement Learning for Direct Ink Writing.* Enabling additive manufacturing to employ a wide range of novel, functional materials can be a major boost to this technology. However, making such materials printable requires painstaking trial-and-error by an expert operator, as those materials typically tend to exhibit peculiar rheological properties. Even in the case of successfully finding the process parameters, there is no guarantee of the print-to-print consistency due to material differences between batches. These challenges make closed-loop feedback, where the process parameters are adjusted on-the-fly, an attractive option. There are several challenges for designing an efficient controller: the deposition parameters are complex and highly coupled, artifacts occur after long time horizons, simulating the deposition is computationally costly, and learning on hardware is intractable. In a recent work [2], we demonstrate the feasibility of learning a closed-loop control policy for additive manufacturing using reinforcement learning. We show that approximate, but efficient, numerical simulation is sufficient as long as it allows learning the behavioral patterns of deposition that translate to real-world experiences. In combination with reinforcement learning, our model can be used to discover control policies that outperform baseline controllers (Figure 30.17). Furthermore, the recovered policies have a minimal sim-to-real gap. We built a direct ink writing printer and show-cased the power of our RL controller in the real world.

*Reinforcement Learning for Multi-Material Fusion.* 3D printing based on continuous deposition of materials, such as filament-based 3D printing, has seen widespread adoption thanks to its versatility in working with a wide range of materials. An important shortcoming of this type of technology is its limited multi-material capabilities. While there are simple hardware designs that enable multi-material printing in principle, the required software is heavily underdeveloped. A typical hardware design *fuses* together individual materials fed into a single chamber from multiple inlets before they are deposited. This design, however, introduces a time delay between the intended material mixture and its actual deposition. In a very recent work [1], inspired by diverse path planning research in robotics, we show that this mechanical challenge can be addressed via improved printer control. We propose to

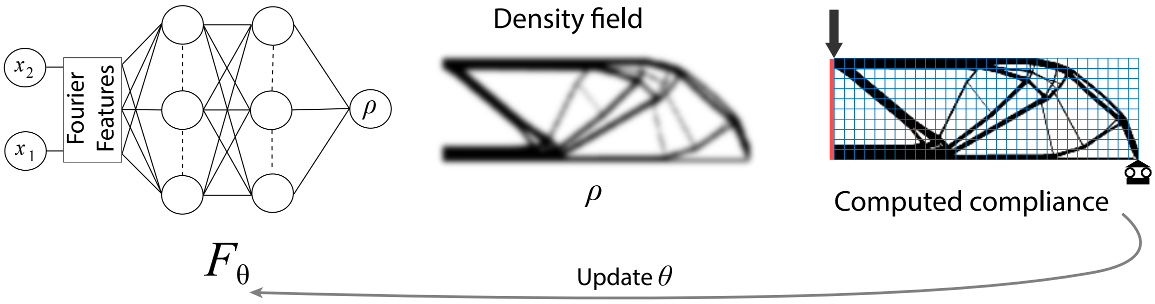



Figure 30.18: A single iteration of our end-to-end differentiable topology optimization via explicit design representation networks.

formulate the search for optimal multi-material printing policies in a reinforcement learning setup. We put forward a simple numerical deposition model that takes into account the non-linear material mixing and delayed material deposition. To validate our system we focus on color fabrication, a problem known for its strict requirements for varying material mixtures at a high spatial frequency. We demonstrate that our learned control policy outperforms state-of-the-art hand-crafted algorithms.

### References


- [1] K. Liao, T. Tricard, M. Piovarči, H.-P. Seidel, and V. Babaei. Learning deposition policies for fused multi-material 3D printing. In *IEEE International Conference on Robotics and Automation (ICRA 2023)*, London, UK, 2023. IEEE. Accepted.
- [2] M. Piovarči, M. Foshey, J. Xu, T. Erps, V. Babaei, P. Didyk, S. Rusinkiewicz, W. Matusik, and B. Bickel. Closed-loop control of direct ink writing via reinforcement learning. *ACM Transactions on Graphics (Proc. ACM SIGGRAPH)*, 41(4), Article 112, 2022.


## 30.7.3 Design Representation Networks

*Investigators: Nikan Doosti and Vahid Babaei*

Advanced manufacturing systems, such as multi-material 3D printers, are continuously growing in build volume and resolution. This rapid growth in hardware technologies is having important consequences for computer aided design algorithms. These algorithms cannot cope with such an exploding design space, resulting in a rapidly widening gap between *what we can compute* and *what we can fabricate*. Consider structural topology optimization (TO) for example – a method of material distribution within a design so as to maximize the structural performance while constraining the design's total weight. The number of elements in a digital design from state-of-the-art structural topology optimization methods [1] is barely **0.1%** of the number of elements (voxels) which a commercially available 3D printer [5] can fit within its build volume.

Inspired by the coordinate-based neural networks used in computer vision for scene representation [4], we recently invented *design representation networks* (DRNs). Similar to an *explicit* analytical function, in a DRN we take any input coordinate, query that coordinate from a special neural network, and output the attribute of the design at that coordinate. One of the main advantages of DRNs is their large capacity in representing high-frequency signals. DRNs, with up to a few million parameters, are extremely lightweight functions considering that they can represent designs with potentially trillion voxels. So far, we have integrated DRNs into two computational design pipelines [3, 6]. In particular, in [3], we solved the structural topology optimization problem where the design was represented via a DRN. We employed the standard Solid Isotropic Material with Penalization (SIMP) approach [2] but instead of working directly with a grid of densities, we optimized for the parameters of a DRN (Figure 30.18). Our computed DRN offers attractive properties, such as mesh-independent results and sub-pixel filtering that leads to appropriate designs for upsampling to higher resolutions. Topology optimization promotes reducing the material waste and making these methods accessible to a wider set of technologies is of great importance to sustainable design.

### References


[1] N. Aage, E. Andreassen, B. S. Lazarov, and O. Sigmund. Giga-voxel computational morphogenesis for structural design. *Nature*, 550(7674):84–86, 2017.

[2] M. P. Bendsoe and O. Sigmund. *Topology optimization: theory, methods, and applications*. Springer Science & Business Media, 2013.

• [3] N. Doosti, J. Panetta, and V. Babaei. Topology optimization via frequency tuning of neural design representations. In E. Whiting, J. Hart, C. Sung, J. McCann, and N. Peek, eds., *Proceedings SCF 2021*, Virtual Event, 2021, Article 1. ACM.

[4] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. In *ECCV*, 2020, pp. 405–421.

[5] Stratasys. Stratasys J8 family, the ultimate multi-material 3D printer. https://www.stratasys.com/3d-printers/j8-series, 2020. [Online; Accessed 15-11-2020].

[6] Q. Zheng, V. Babaei, G. Wetzstein, H.-P. Seidel, M. Zwicker, and G. Singh. Neural light field 3d printing. *ACM Transactions on Graphics (TOG)*, 39(6):1–12, 2020.


### 30.7.4 Shape from Release

*Investigators: Emiliano Luci, Haleh Mohammadian and Vahid Babaei*

The dissolution process is a ubiquitous yet fascinating phenomenon. Objects with the same mass but different shapes can dissolve via different dynamics, resulting in vastly different *release profiles*. This property can be exploited, particularly in pharmaceutics, when designing controlled-release systems for drugs, food, and cosmetics. The surge of 3D printing with its unique capabilities for fabricating complex shapes opens new doors for devising novel controlled-release designs.

While predicting different shapes' dissolution dynamics is an important problem, the more important and challenging question is how to control the dissolution via shape, i. e., designing shapes that lead to a desired release behavior of materials in a solvent over a specific time. We tackled this challenge recently [1] by introducing a computational inverse design pipeline. We first introduced a simple, physically-inspired differentiable forward model of dissolution, which is both efficient and differentiable. We then formulated our inverse

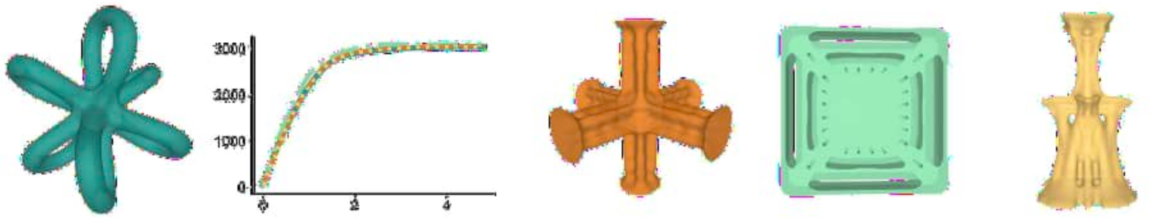



Figure 30.19: Inverse designs evaluated in simulation. From left to right, starting from a source shape, we simulate its release curve (solid green line) and optimize for release equivalent shapes with different fabricability constraints. The resulting release curves are plotted in dashed lines.

design as a PDE-constrained topology optimization that has access to analytical derivatives obtained via sensitivity analysis. Furthermore, we incorporated fabricability terms in the optimization objective that enable physically realizing our designs. Finally, we thoroughly analyzed our approach on a diverse set of examples via both simulation and fabrication (Figure 30.19).

### References

• [1] J. Panetta, H. Mohammadian, E. Luci, and V. Babaei. Shape from release: Inverse design and fabrication of controlled release structures. *ACM Transactions on Graphics (Proc. ACM SIGGRAPH Asia)*, 41(6), Article 274, 2022.

## 30.8 Sensorimotor Interaction

*Coordinator: Paul Strohmeier*

Our experiences of the world emerge from the law-like contingencies between action and perception. The Sensorimotor Interaction Group explores how experiences can be designed by manipulating such sensorimotor contingencies. Specifically, we explore the role of sensorimotor interaction with regard to mediation of information, augmentation of the human body, and vibrotactile rendering of material experiences. The focus of our work has been in building a foundation – methodological, theoretical, and practical – on which to build up future research. Methodologically we have explored new ways of empirically studying subjective experiences, theoretically we have contributed towards the discourse of body-based interfaces and Human-computer INTegration (HInt), and practically we have developed a robust hardware platform for conducting experiments on sensorimotor couplings and rendering material experiences.

### 30.8.1 Mediation & Tacton Design

*Investigators: Nihar Sabnis, Dennis Wittchen, Courtney Reed and Paul Strohmeier*

A core element of human computer interaction is representing and organizing information in ways which are easy for humans to comprehend. This typically happens using icons,

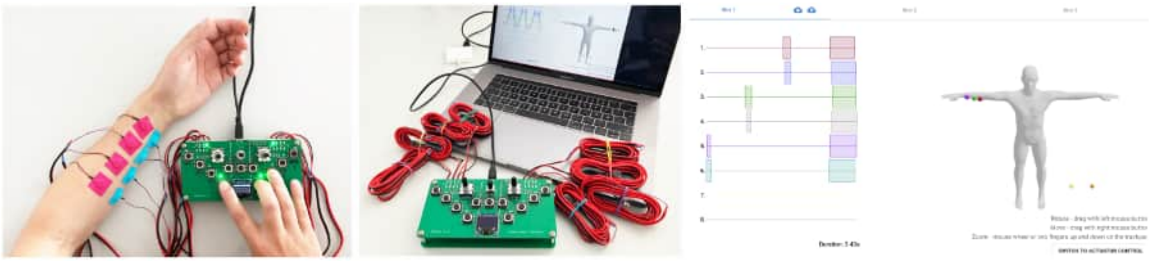


Figure 30.20: TactJam: A software and hardware suit for designing vibrotactile icons on the body.

symbols, or metaphors. We have contributed to this space by designing a prototyping system for designing on-body tactile notifications. This system – TactJam – was deployed in two workshops and has since been used for research purposes at the TU Wien and the Julius-Maximilians-Universität Würzburg. A modified version will be used for a workshop at WorldHaptics 2023 in Delft. The source code and design files are available [3].

In (post)phenomenological accounts of information transfer, a distinction is often made between embodied mediation, and hermeneutic mediation. Embodied mediation is understood in a pre-reflective manner, based on one's lived experience of the world, while hermeneutic mediation is understood based on reflection of the meaning of the symbols within which the information is encoded. In a vibrotactile feedback design study, we highlight that these forms of mediation can be interwoven. We highlight how embodied experiences can be used as building blocks for hermeneutic mediation. We demonstrate that this interweaving broadens the ways in which designers think of symbol encodings, and expand the affective space of tactile symbols [2].

Finally, our works challenge some of the underlying assumptions of Metaphor use. Through an interview study of vocal instructors, we highlight that metaphors which work well have an ambiguity as well as an iterative nature. This ambiguity is why metaphors are a powerful tool, as it allows the individual to apply their own sense-making process, which would not be possible if the metaphor were rigid and clearly specified. These insights can be used to better understand why metaphors in smartphone design are useful, even though they are always in flux, and why the desktop metaphors fail for more naturalistic implementations [1].

## References


- [1] C. N. Reed, P. Strohmeier, and A. McPherson. Negotiating experience and communicating information through abstract metaphor. In *CHI '23, CHI Conference on Human Factors in Computing Systems*, Hamburg, Germany, 2023. ACM. Accepted.
- [2] N. Sabnis, D. Wittchen, G. Vega, C. N. Reed, and P. Strohmeier. Tactile symbols with continuous and motion-coupled vibration: An exploration of using embodied experiences for hermeneutic design. In *CHI '23, CHI Conference on Human Factors in Computing Systems*, Hamburg, Germany, 2023. ACM. Accepted.
- [3] D. Wittchen, K. Spiel, B. Fruchard, D. Degraen, O. Schneider, G. Freitag, and P. Strohmeier. TactJam: An end-to-end prototyping suite for collaborative design of on-body vibrotactile feedback.

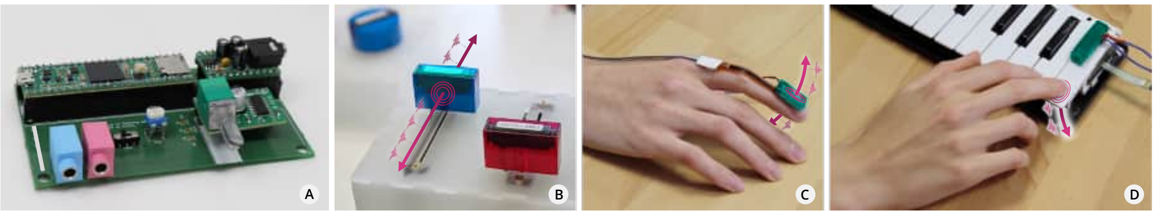



Figure 30.21: Haptic Servos enable rapid rendering of diverse material experiences. (a) The open source Haptic Servo shield, compatible with the Arduino IDE, encapsulates all timing-sensitive elements, to create a rich variety of material experiences. We demonstrate how Haptic Servos can be deployed by example of (b) dynamically rendering the material experience on tangible user interfaces, (c) creating on-body material experiences, and (d) augmenting the experience of everyday objects (here piano key).

In *TEI '22, Sixteenth International Conference on Tangible, Embedded, and Embodied Interaction*, Daejeon, Republic of Korea (Online), 2022, Article 1. ACM.

### 30.8.2 Vibrotactile Rendering

*Investigators: Nihar Sabnis, Dennis Wittchen, Courtney Reed and Paul Strohmeier*

The literature has shown a large range of material experiences that can be rendered using vibration. These include friction, compliance, torsion, bending, elasticity and others. By demonstrating a single prototype capable of creating all the above experiences, we highlight that these experiences are all instances of a larger perceptual structure of vibrotactile signals which change their frequency in correlation with human action. We call such signals motion-coupled. Our work suggests that the specifics of the experience are based on the human action which generate them, much more so than on any one aspect of the algorithm that generates the vibrotactile feedback: qualitatively distinct experiences can be created based on input mapping, even if stimulation parameters and algorithm remain unchanged [1].

In human-computer interaction research, such haptic systems are rarely used. This is partially because they are non-trivial to implement. To make this style of haptic feedback more accessible to non-domain experts, we designed Haptic Servos: self-contained haptic rendering devices which encapsulate all timing-critical elements with a system latency of <5 ms. A workshop demonstrated that users new to Haptic Servos require approximately ten minutes to set up a basic haptic rendering system [1].

We have deployed haptic servos in our own research [2]. We have also modified Haptic Servos to be deployed in interactive shoes. The prototypes are still under development, however a first initial version has been published [3]. In addition to confirming what we already understood about material experience design, anecdotal data collected during initial deployments of the shoe highlight that motion coupled vibration are integrated in the perception of our own actions. Once motion coupled vibration is removed, users report a sense of loss.

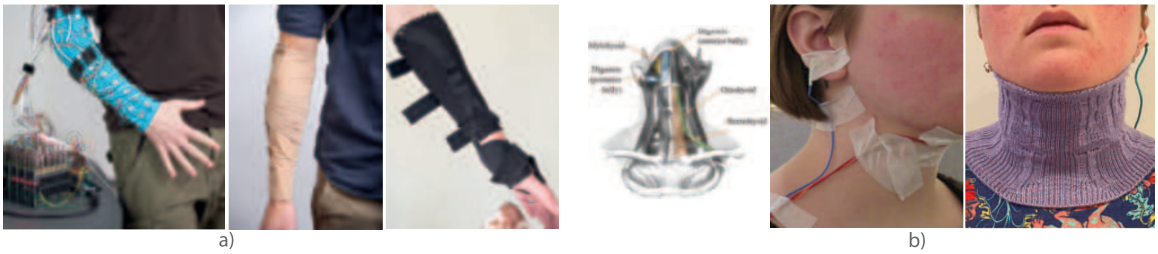



Figure 30.22: Left three panels show SkillSleeves which sense human actions and can actuate humans [2]. Right three panels show Vocal EMS for direct vocal control over vocal augmentation [7].

### References


- [1] N. Sabnis, D. Wittchen, C. N. Reed, N. Pourjafarian, J. Steimle, and P. Strohmeier. Haptic servos: Self-contained vibrotactile rendering system for creating or augmenting material experiences. In *CHI '23, CHI Conference on Human Factors in Computing Systems*, Hamburg, Germany, 2023. ACM. Accepted.
- [2] N. Sabnis, D. Wittchen, G. Vega, C. N. Reed, and P. Strohmeier. Tactile symbols with continuous and motion-coupled vibration: An exploration of using embodied experiences for hermeneutic design. In *CHI '23, CHI Conference on Human Factors in Computing Systems*, Hamburg, Germany, 2023. ACM. Accepted.
- [3] D. Wittchen, V. Marinez-Missir, S. Mavali, N. Sabnis, C. N. Reed, and P. Strohmeier. Designing interactive shoes for tactile augmented reality. In *AHs '23, Augmented Humans International Conference*, Glasgow, UK, 2023, pp. 1–14. ACM.


### 30.8.3 Human Augmentation

*Investigators: Nihar Sabnis, Dennis Wittchen, Courtney Reed and Paul Strohmeier*

Over the last 10 years, there has been a growing interest in the body within HCI research. This is manifested in research areas such as *Human Computer Integration* or more design oriented themes such as *Embodied Interaction*. Yet, even though work in these areas has been conducted for a significant period of time, it is unclear what, if any, progress has been made. This is partially due to it being unclear how progress in these areas might even be defined. To address this, the Sensorimotor Interaction Group has participated in various efforts to better understand this space. This has resulted in an essay on the different ways in which HCI engages with the Human Body – published as an invited Chapter in the Routledge Handbook of Bodily Perception – which highlights four analytical lenses on the body which can be used for understanding how technology engages with it [8]. A *material view*, that is, the body as a material for attaching sensors to, or embedding devices in. A *morphological view*, that is, a structure which can be manipulated by, for example, adding limbs or constraining movement. A *sensorimotor view*, where technology augments the body as it experiences the world through continuous interaction and, finally, an *experiential view*, where technology is used to explore and manipulate the subjective experience of having a body. In a second essay we discuss the design dimensions of Bodily Integration together with a list of challenges for Human Computer Integration [3], here we highlight how augmentation technologies can be better understood by considering their impact on human agency and the experience of body-ownership.

We have also engaged with the Human Augmentation through the design and implementation of novel interfaces. Together with colleagues at Saarland University, we have published a handheld printer which enables collaborative design and implementation of circuits. The printer is aware of its context and user actions and can correct errors introduced by the user – preventing short-circuits, and ensuring high quality traces, even when users hands are unsteady or when materials change [4].

Moving towards systems which integrate more closely with the body, and building on previous work on the design of SkillSleeves [2] – devices which can measure human activity through electromygraphy (EMG) and induce movement through electrical muscle stimulation (EMG) – we implemented SingingKnit, a knitted scarf which measures muscle activity while singing to augment vocal performances [7]. SkillSleeves act as *Morphological* augmentation, as it reconfigures the control-structures of the human body while the SingingKnit goes beyond that, providing *sensorimotor* augmentation. By providing novel feedback for vocalists to explore their behavior and form new action-sound connections with their movement. The new action-sound connections are used to examine vocalists' body perceptions and relationship with their voices (both part of the body and a musical interaction "interface"). The work explores the role of sensorimotor contingencies, and how biofeedback of previously unconscious behaviors can change perception of the body and artistic action [5]. We have also contributed to the space of *experiential* integration, by means of exploring new methods of studying introspective pre-reflective experiences. Here we have explored how micro-phenomenology interview methodology and analyses can be used to explore fine-grained details of experiences when interacting with technologies. This was initially presented in the context of digital musical instruments. Here micro-phenomenology interviews were used to examine experiential modalities and work through the diachronic unfolding of an experience or interaction in time [6]. We have later used these methods to study the experiential dimension of tactile perception, in particular in the context of vibrotactile rendering. We are also actively providing impulses for engaging with the experiential dimension of interaction by hosting a workshop at TEI 2023 [1].

## References


- [1] A. Haynes, C. N. Reed, C. Nordmoen, and S. Skach. Being meaningful: Weaving soma-reflective technological mediations into the fabric of daily life. In *TEI '23, Seventeenth International Conference on Tangible, Embedded, and Embodied Interaction*, Warsaw, Poland, 2023, Article 68. ACM.
- [2] J. Knibbe, R. Freire, M. Koelle, and P. Strohmeier. Skill-sleeves: Designing electrode garments for wearability. In *TEI '21, Fifteenth International Conference on Tangible, Embedded, and Embodied Interaction*, Salzburg, Austria, 2021, Article 33. ACM.
- [3] F. Mueller, N. Semertzidis, J. Andres, M. Weigel, S. Nanayakkara, R. Patibanda, Z. Li, P. Strohmeier, J. Knibbe, S. Greuter, M. Obrist, P. Maes, D. Wang, K. Wolf, L. Gerber, J. Marshall, K. Kunze, J. Grudin, H. Reiterer, and R. Byrne. Human–computer integration: Towards integrating the human body with the computational machine. *Foundations and Trends in Human-Computer Interaction*, 16(1):1–64, 2022.

- [4] N. Pourjafarian, M. Koelle, F. Mjaku, P. Strohmeier, and J. Steimle. Print-A-Sketch: A handheld printer for physical sketching of circuits and sensors on everyday surfaces. In *CHI '22, CHI Conference on Human Factors in Computing Systems*, New Orleans, LA, USA, 2022, Article 270. ACM.
- [5] C. N. Reed and A. P. McPherson. The body as sound: Unpacking vocal embodiment through auditory biofeedback. In *TEI '23, Seventeenth International Conference on Tangible, Embedded, and Embodied Interaction*, Warsaw, Poland, 2023, Article 7. ACM.
- [6] C. N. Reed, C. Nordmoen, A. Martelloni, G. Lepri, N. Robson, E. Zayas-Garin, K. Cotton, L. Mice, and A. McPherson. Exploring experiences with new musical instruments through micro-phenomenology. In *NIME 2022, International Conference on New Interfaces for Musical Expression*, Auckland, New Zealand, 2022. PubPub.
- [7] C. N. Reed, S. Skach, P. Strohmeier, and A. P. McPherson. Singing knit: Soft knit biosensing for augmenting vocal performances. In *AHs '22, Augmented Humans International Conference*, Munich, Germany (Hybrid), 2022, pp. 170–183. ACM.
- [8] P. Strohmeier, A. Mottelson, H. Pohl, J. McIntosh, J. Knibbe, J. Bergström, Y. Jansen, and K. Hornbæk. Body-based user interfaces. In A. J. T. Alsmith and M. R. Longo, eds., *The Routledge Handbook of Bodily Awareness*, pp. 478–502. Routledge, London, 2022.


## 30.9 Software and Datasets

As part of the research process, several libraries, development tools, large corpora of reference data sets, and application frameworks have been developed by members of the group. In this section we describe some of them that evolved to a level where it was appropriate to either distribute them as *open source* projects or let members of other research institutions benefit from software that had been developed in our group. We have also initiated a technology transfer project in order to found a spin-off company.

### 30.9.1 ORACLASE: AI Aided Laser Material Processing

*Investigators: Azadeh Asadi, Sebastian Cucerca, Siddhartha Siddhartha, Balaji Venkatesan and Vahid Babaei*

Lasers are heavily used for *surface activation* where the substrate properties are altered through treatment with laser. A key prerequisite is the process development including finding a set of laser parameters that lead to the desired properties. This is a challenging step which is being done by expert operators but with little efficiency. The fundamental reason is that this step is, intrinsically, a *computationally hard* problem which so far we have been trying to solve manually. In a previous research project [1], we took important steps to address this problem in the context of color laser marking with a computational mindset using artificial intelligence (AI) and other computational algorithms. Our algorithms automatically explore the space of laser marking parameters in search for saturated, diverse, and high-resolution colors. Once we find these colors, we put them together and, using customized image reproduction algorithms, mark arbitrary color images. With the help of an EXIST Transfer of Research grant, in project `Oraclase`, we are now closing the gap between academic research and industrial requirements, wrap our algorithms in a production-ready software, fully test our solution, and make it ready for use by industrial customers. Our software solution will enable a wide range of applications in surface activation and functionalization.

#### References

[1] S. Cucerca, P. Didyk, H.-P. Seidel, and V. Babaei. Computational image marking on metals via laser induced heating. *ACM Transactions on Graphics (Proc. ACM SIGGRAPH)*, 39(4), Article 70, 2020.

### 30.9.2 Automatic Door and Window Profile Classification

*Investigators: Navid Ansari and Vahid Babaei*

This software uses deep learning for classifying different door and window profiles based on the shape of their cross-section. Due to the huge number of classes, classifying the profiles is a very challenging task [1]. In order to be useful in the industry our product should have an accuracy comparable to that of human classification. Moreover, since constantly new profile designs are introduced to the market, our proposed software should be flexible and allow adding of new classes to the classifier. The industry currently relies on humans for profile classification.

To address the large number of classes, we train a single neural network for each class. In this approach we will have a separate 0-1 classifier for each class, where 1 means that the sample belongs to this class and 0 means it is from a different class. A new sample from a new unseen class, can be misclassified by some of our previously trained models. To avoid misclassification, we should feed the new sample to all the models. In case some of them misclassify it as 1, we should add a few instances of the new sample to its 0 class. Once we make sure none of the old models misclassify anymore, we can train a new model for this new sample. Since each model is independent of the others, increasing the number of classes will not affect the accuracy of the software. Moreover, since a single model does not have to see all the data, the size of each dataset remains manageable. Since the calculation time increases linearly with the number of classes, once we use the extra information of cargo and inventory, we can limit the calculation to the classes that we need in a specific time window. The software is accessible via:

https://gitlab.mpi-klsb.mpg.de/nansari/profile-detection/

#### References

[1] Z. Yuan, Z. Guo, X. Yu, X. Wang, and T. Yang. Accelerating deep learning with millions of classes. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXIII 16*, 2020, pp. 711–726. Springer.

### 30.9.3 PFSTOOLS for Processing High Dynamic Range Images and Video

*Investigators: Rafał Mantiuk, Grzegorz Krawczyk, Tunç O. Aydın, and Ivo Ihrke*

The `pfstools` package is a set of command line programs for reading, writing, manipulating and viewing high-dynamic range (HDR) images and video frames. All programs in the package exchange data using a simple generic high dynamic range image format, pfs, and they use unix pipes to pass data between programs and to construct complex image processing operations.

pfstools come with a library for reading and writing pfs files. The library can be used for writing custom applications that can integrate with the existing pfstools programs.

pfstools offer also a good integration with high-level mathematical programming languages, such as MATLAB or GNU Octave. pfstools can be used as the extension of MATLAB or Octave for reading and writing HDR images or simply to store effectively large matrices.

The pfstools package is an attempt to integrate the existing high dynamic range image formats by providing a simple data format that can be used to exchange data between applications.

The pfstools package is accompanied by the pfscalibration and pfstmo packages. The pfscalibration package provides an algorithm for the photometric calibration of cameras and for the recovery of high dynamic range (HDR) images from the set of low dynamic range (LDR) exposures. The pfstmo package contains the implementation of ten state-of-the-art tone mapping operators suitable for convenient processing of both static images and animations.

The pfstools, pfscalibration and pfstmo packages are licensed as an Open Source project under a General Public License (GPL). The project web pages can be found at:

http://www.mpi-inf.mpg.de/resources/pfstools/

The software received wider interest of Open Source community and third party contributors prepared installation packages which are now included in several Linux distributions including Debian, Fedora and Suse. The software was presented on the *Electronic Imaging Conference 2007* and a general introduction to the package was published in the proceedings [1].

### References

[1] R. Mantiuk, G. Krawczyk, R. Mantiuk, and H.-P. Seidel. High dynamic range imaging pipeline: Perception-motivated representation of visual content. In B. E. Rogowitz, T. N. Pappas, and S. J. Daly, eds., *Human Vision and Electronic Imaging XII*, 2007, SPIE, pp. 649212.1–12. SPIE.

### 30.9.4 LocVis – Local Visibility Maps of Artifacts and Distortions in Images

*Investigators: Krzysztof Wolski and Karol Myszkowski*

Numerous applications require a robust metric that can predict whether image differences are visible or not. However, the accuracy of existing visibility metrics, such as HDR-VDP, is often not good enough. In our paper, we argue that the main reason behind this problem is the lack of sufficient image collection with a good coverage of possible distortions [5, 4]. We address this problem by creating a new dataset in collaboration with University of Cambridge.

The *LocVis* dataset[1] consists of 557 images with 170 unique scenes, where the distortions and artifacts were manually labeled by human observers. Many of the images are generated

[1] https://www.repository.cam.ac.uk/handle/1810/274368
http://visibility-metrics.mpi-inf.mpg.de/files/marking_datasets.zip


Figure 30.23: The figure presents examples of stimuli from the dataset. The insets show the closeup of the artifacts.

for up to 3 distortion levels of artifacts, for example, different quality settings of image compression. Each scene is marked by approximately 15-20 observers in a local manner – instead of a single score, i.e. observers were asked to paint the mask and depending on their responses, pixel-wise visibility maps of distortions are recorded by our software tool.

Many of the images used in our dataset are based on the data available in previous studies and they were selected to cover an extensive set of typical artifacts which are frequently observed in computer graphics research and applications (see Fig. 30.23). In order to provide a convenient access for other researchers, we group the images into several subsets depending on the type of artifacts. The subset *mixed* is an extended LOCCG dataset including such distortions such as: high-frequency and structured noise, virtual point light (VPL) clamping, light leaking artifacts, local changes of brightness, aliasing and tone mapping artifacts [2, 3]. The subset *perceptionpatterns* consists of artificial patterns designed to expose well known perceptual phenomena, such as luminance masking, contrast masking and contrast sensitivity. The subsets *aliasing*, *peterpanning*, *shadowacne*, *downsampling* and *zfighting* contain real-time rendering artifacts. The subset *compression* contains distortions due to experimental low-complexity image compression. The subset *deghosting* contains artifacts due to HDR merging, which exposes shortcomings of popular deghosting methods. Images of *ibr* and *cgibr* subsets contain artifacts produced by view-interpolation and image-based rendering methods, and they are mainly adopted from the previous work of Vamsi et. al. [1].

Recently, we added another dataset[2] that contains rendered images using textures of varying resolution and the thresholds at which the artefacts due to the reduction of resolution become visible [4].

[2] https://texture-metrics.mpi-inf.mpg.de/

## References


[1] V. K. Adhikarla, M. Vinkler, D. Sumin, R. Mantiuk, K. Myszkowski, H.-P. Seidel, and P. Didyk. Towards a quality metric for dense light fields. In *30th IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2017)*, Honolulu, HI, USA, 2017, pp. 3720–3729. IEEE Computer Society.

[2] M. Čadík, R. Herzog, R. Mantiuk, R. Mantiuk, K. Myszkowski, and H.-P. Seidel. Learning to predict localized distortions in rendered images. *Computer Graphics Forum (Proc. Pacific Graphics)*, 32(7):401–410, 2013.

[3] M. Čadík, R. Herzog, R. Mantiuk, K. Myszkowski, and H.-P. Seidel. New measurements reveal weaknesses of image quality metrics in evaluating graphics artifacts. *ACM Transactions on Graphics (Proc. ACM SIGGRAPH Asia)*, 31(6), Article 147, 2012.

[4] K. Wolski, D. Giunchi, S. Kinuwaki, P. Didyk, K. Myszkowski, A. Steed, and R. K. Mantiuk. Selecting texture resolution using a task-specific visibility metric. *Computer Graphics Forum (Proc. Pacific Graphics)*, 38(7):685–696, 2019.

[5] K. Wolski, D. Giunchi, N. Ye, P. Didyk, K. Myszkowski, R. Mantiuk, H.-P. Seidel, A. Steed, and R. K. Mantiuk. Dataset and metrics for predicting local visible differences. *ACM Transactions on Graphics*, 37(5), Article 172, 2018.


### 30.9.5 TactJam

*Investigators: Dennis Wittchen and Paul Strohmeier*

TactJam (`https://github.com/TactileVision/TactJam`) is an end-to-end suite for creating and sharing low fidelity prototypes of on-body vibrotactile feedback. TactJam enables users to collaboratively design and share such vibrotactile designs (tactons).

The system consists of a custom hardware controller and a client application (Figure 30.24). The controller supports up to eight vibration motors (ERMs) which can be placed on the body in any spatial configuration. Each actuator is controlled with a dedicated button. This allows users to play and experience spatio-temporal patterns (tactons) live – we call this jamming. Once users have found interesting tactons, they can record them so that the tactons can be shared with others. A maximum of three tactons can be stored on the device at a time. These can be transferred to, or received from, a computer via the client software. Once transferred to the client, the time profiles of all actuators are visualized (Figure 30.24, right). Since users can freely place the actuators on the body, it's important to document these placements to be able to replicate the configuration. This is done by indicating the position of each actuator on a 3D avatar with color-coded spheres. For remote collaboration, the client is connected to an online repository (PostgreSQL database), where all tactons are stored along with additional metadata (e.g., description, tags).

The project's software components are licensed as Open Source projects under the MIT license, whereas the hardware is licensed under Creative Commons (CC-BY4.0). The system was used in a workshop at the *15th Conference on Tangible Embedded and Embodied Interaction (TEI)* in 2021 and the findings of this workshop were published in the proceedings of TEI in 2022 [1].

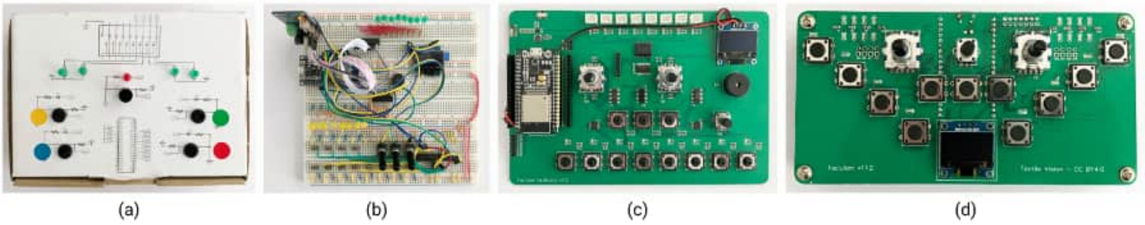



Figure 30.24: (a) Initial prototype of TactJam used for teaching, (b) breadboarded version of TactJam designed for TEI workshop, (c) first PCB version of workshop device and (d) re-designed workshop device with improved ergonomics for bimanual hand-held use.

### References


- [1] D. Wittchen, K. Spiel, B. Fruchard, D. Degraen, O. Schneider, G. Freitag, and P. Strohmeier. TactJam: An end-to-end prototyping suite for collaborative design of on-body vibrotactile feedback. In *TEI '22, Sixteenth International Conference on Tangible, Embedded, and Embodied Interaction*, Daejeon, Republic of Korea (Online), 2022, Article 1. ACM.


### 30.9.6 Haptic Servo

*Investigators: Nihar Sabnis, Dennis Wittchen, Courtney Reed and Paul Strohmeier*

Haptic Servos (https://github.com/sensint/Servo_Haptics) are self-contained vibrotactile rendering devices for creating and augmenting material experiences [1]. With haptic servos, experts and novices alike can create experiences of friction, compliance and torsion. All designs, code and schematics for haptic servos are made open-source to encourage haptic enthusiasts to create their own material experiences. We designed two versions of the haptic servo device: one that can be easily assembled by novices using breakout board components (Figure 30.25, left), and a version with a compact footprint optimized for wearable applications (Figure 30.25, right). To achieve the compact size, we replaced the breakout boards with a custom PCB. All files needed for fabrication are provided with the project's repository. The project's software and hardware components are licensed as Open Source projects under the MIT license. We demonstrate an integrated system to render vibrotactile experiences on tangible user interfaces (TUIs) with an agnostic communication protocol. The vibrotactile experiences were rendered using a Graphical User Interface (https://github.com/sensint/Haptic_Material_Designer). The TUIs with the form factors of knobs and sliders were used as representations of commercially used interactive devices in daily life. These TUIs were used to render the vibrotactile effects, augmenting the material experiences when the user interacts with them.

The firmware consists of a central controller that receives messages from the GUI via a serial interface and forwards them via an I2C-interface to the TUI. We defined a standardized protocol for encoding material properties to enable designers to implement custom GUIs and other applications.

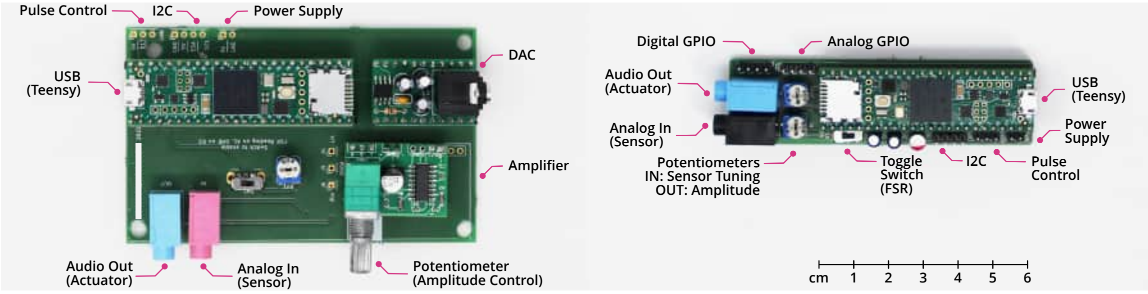



Figure 30.25: Haptic Servo shields using breakout board components for easy assembly (left) and a compact version optimized for wearable applications (right).

## References


- [1] N. Sabnis, D. Wittchen, C. N. Reed, N. Pourjafarian, J. Steimle, and P. Strohmeier. Haptic servos: Self-contained vibrotactile rendering system for creating or augmenting material experiences. In *CHI '23, CHI Conference on Human Factors in Computing Systems*, Hamburg, Germany, 2023. ACM. Accepted.


### 30.9.7 Interactive Shoe

*Investigators: Dennis Wittchen, Nihar Sabnis, Courtney Reed and Paul Strohmeier*

Haptic Shoe [1] is an implementation of an interactive shoe for Tactile Augmented Reality (Figure 30.26). It was designed as a reference platform for augmented shoe design: Even though various approaches and prototypes haven been developed and demonstrated in research, there are almost no open-source implementations available that one can build upon. Also, the design considerations taken into account when building such prototypes are often not highlighted or documented in the corresponding papers or repositories, which makes it hard to re-implement or adapt prior work. With this work, we make implicit decisions explicit, and therefore, described a design space with technical and physiological dimensions, e.g. actuator/sensor selection and integration.

We implemented an example prototype in the form factor of sandals that combines custom designed and 3D printed soles, and straps taken from retail sandals (Figure 30.26, D). Each sole encapsulates four pressure sensors and actuators as well as an IMU (Figure 30.26, A and B). The positions of the sensor-actuator pairs were chosen to sense and augment the dynamics of all gait phases (heel strike to toe lift) as well as balance tasks. We focused on rather simple design and fabrication methods, as well as components and materials that are broadly available to enable easy reproducibility of our system. The sandals are connected to control units (small boxes incl. microcontrollers, DACs, amplifiers, power supply) that are attached to the lower legs. These control units are wirelessly connected (BLE) to a computer, which runs a software to control the vibrotactile experiences. We implemented the augmentation pipeline based on Haptic Servo's, which we then used to create vibrotactile renderings to explore material properties such as compliance in various walking tasks, e.g. walking on different surfaces, climbing stairs, or balancing. Due to the wireless connection, users can freely move in unconstrained environments (indoor and outdoor). The project's software and hardware components are licensed as Open Source projects under the MIT license (https://github.com/sensint/Haptic_Shoe) .


Figure 30.26: The Haptic Shoe consists of four pairs of pressure sensors and vibrotactile actuators and an IMU. The electronics are integrated in two-part 3D printed insole which forms the basis of a sandal.

### References


- [1] D. Wittchen, V. Marinez-Missir, S. Mavali, N. Sabnis, C. N. Reed, and P. Strohmeier. Designing interactive shoes for tactile augmented reality. In *AHs '23, Augmented Humans International Conference*, Glasgow, UK, 2023, pp. 1–14. ACM.


## 30.10 Academic Activities

### 30.10.1 Journal Positions

*Karol Myszkowski is on the editorial board of*

- *Journal of Virtual Reality and Broadcasting* (since 2004),
- *ACM Transactions on Applied Perception* (since 2002),
- *Machine Graphics & Vision* (since 1998).

*Hans-Peter Seidel is on the editorial board of*

- *Visual Informatics (chief editor)* (since 2017),
- *International Journal of Shape Modeling (IJSM)* (since 2001),
- *Computer Aided Geometric Design (CAGD)* (since 1999),
- *Graphical Models (GMOD)* (since 1995).

### 30.10.2 Conference and Workshop Positions

#### Membership in Program Committees

*Vahid Babaei:*

- *EUROGRAPHICS Short Papers*, co-Chair, Saarbrücken, Germany, May 2023.

- *Computational Visual Media*, Shenzhen, China, April 2023.

*Karol Myszkowski:*

- *EUROGRAPHICS Full Technical Papers*, co-Chair, Saarbrücken, Germany, May 2023.
- *EUROGRAPHICS Rendering Symposium*, Prague, Czech Republic, July 2022.
- *EUROGRAPHICS Full Technical Papers*, Reims, France, May 2022.
- *Pacific Graphics*, Wellington, New Zeland, October 2021.
- *EUROGRAPHICS Rendering Symposium*, virtual, June 2021.

*Gurprit Singh:*

- *EUROGRAPHICS Posters*, co-Chair, Saarbrücken, Germany, May 2023.
- *EUROGRAPHICS Diversity & Inclusion*, co-Chair, Saarbrücken, Germany, May 2023.
- *Pacific Graphics*, virtual, October 2021.
- *EUROGRAPHICS Rendering Symposium*, co-Chair, virtual, June 2021.
- *EUROGRAPHICS Short Papers*, virtual, May 2021.

*Paul Strohmeier:*

- *ACM CHI Conference on Human Factors in Computing Systems*, virtual, April 2022.
- *ACM CHI Conference on Human Factors in Computing Systems*, virtual, April 2021.
- *International Augmented Humans Conference*, co-Chair, virtual, February 2021.

*Thomas Leimkühler:*

- *EUROGRAPHICS Short Papers*, Saarbrücken, Germany, May 2023.
- *High Performance Graphics*, virtual, July 2022.
- *High Performance Graphics*, virtual, July 2021.
- *EUROGRAPHICS Short Papers*, Vienna, Austria, May 2021.

### Membership in Organizing Committees

*Courtney Reed:*

- Organizer – Women in Music Information Retrieval (WiMIR) Co-hosted with ISMIR, Online, December 2022.

*Nihar Sabnis:*

- Demo Chair – *International Augmented Humans Conference (AHs)*, Glasgow, March 2023.

*Hans-Peter Seidel:*

- Conference Chair – *EUROGRAPHICS'23*, Saarbrücken, Germany, May 23.
- Conference Steering Committee – *EUROGRAPHICS Annual Conference Series*, European Association for Computer Graphics (EUROGRAPHICS).

- Conference Steering Committee – *Vision, Modeling and Visualization (VMV) Annual Conference Series*.

*Paul Strohmeier:*

- Conference Steering Committee – *International Augmented Humans Conference (AHs)*.
- Conference Steering Committee– *ACM conference on Tangible, Embedded, and Embodied Interaction (TEI)*.

### 30.10.3 Invited Talks and Tutorials

*Vahid Babaei:*

- *Computational design and fabrication*, WSS Scientific Workshop on Development of Active Implants in Fracture Care, University Hospital Saarland, Homburg, October 2022.
- *Focal talk on AI in Manufacturing*, AI Hub @ Karlsruhe Institute of Technology (KIT), Karlsruhe, October 2022.
- *Appearance Management for 3D Printing and Laser Marking* , Optica (former Optical Society of America), virtual, August 2022.

*Karol Myszkowski:*

- *The effect of shape and illumination on material perception for displayed and fabricated content*, Speaker, HiVisComp, Slovakia, February 2023.
- *The effect of geometry and illumination on appearance perception of different material categories*, Speaker, Panel to honor Prof. Kunii, CGI conference, virtual, September 2021.

*Gurprit Singh:*

- *Point correlations for graphics, vision and machine learning*, virtual, Mohamed bin Zayed University of Artificial Intelligence (MBZUAI), December 2022, February 2023.
- *Point correlations for graphics, vision and machine learning*, virtual, Intel Corporation, September 2022.
- *Monte Carlo & quasi-Monte Carlo sampling for image synthesis*, virtual, Weta Digital, New Zealand, June 2022.
- *Point correlations for graphics, vision and machine learning*, virtual, Huawei Technologies Co. London, June 2022.
- *Looking at the world through point correlations*, virtual, University College London (UCL), May 2022.
- *Sampling in Monte Carlo rendering*, virtual, Huawei Technologies Co. Vancouver, March 2022.

*Courtney Reed:*

- *Bodily Understanding and Communication through Sonified Laryngeal sEMG*, Invited Talk at the Joint C4DM-CogSci Body-Centred Perspectives on Human-Human and Human-Machine interaction Seminar and Workshop, Australia, February 2023.

*Nihar Sabnis:*

- *Introduction to vibrotactile Feedback: Designs and Explorations*, Greaves Cotton, India, May 2022.
- *Experiencing and designing for and by the users with vibrotactile feedback*, Jnana Prabodhini, India, April 2022.
- *Principles of vibrotactile feedback design*, Gurumantra Academics, India, April 2022.

*Paul Strohmeier:*

- *Body x Materials*, invited Panelist, CHI, Germany, April 2023.
- *Designing Embodied Mediation with Vibrotactile Feedback*, MPI-IS Stuttgart, Germany, February 2023.
- *Designing Sensorimotor Interaction with Haptic Feedback*, CornellTech NY, USA, December 2022.
- *Cognitive Augmentation*, Impulse Talk at Dagstuhl Seminar, December 2022.
- *Wearables and Haptics*, Masterclass HAW Hamburg, Germany, April 2022.
- *Augmenting Human Expression*, Invited Performance at Augmented Humans 2022 Satallite event, LMU Munich, Germany, March 2022.
- *Body Displays*, MIT Media Lab Virtual, USA, October 2021.
- *Sensorimotor Augmentation*, Keynote at Assistive Augmentation Symposium, Aukland Bioengineering Institute Virtual, New Zealand, September 2021.
- *eTextile Functionalization and HCI*, Salzburg Centre for Smart Materials Virtual, Austria, April 2021.

*Dennis Wittchen:*

- *Haptic Servos: Self-Contained Vibrotactile Rendering System for Creating or Augmenting Material Experiences*, German Pre-CHI 2023, Dresden, Germany, March 2023.

### 30.10.4 Other Academic Activities

*Vahid Babaei:*

- External reviwer for PhD pre-defense: Linkoping University.
- Responsible at MPI Informatics (associate partner) for Erasmus Mundus Joint Master Degree in Photonics for Security, Reliability and Safety: from Instrumentation to Data Science.

*Karol Myszkowski:*

- Reviewer for ERC Advanced Grants.

- External reviewer for PhD theses (only for the period of 2021-2023): Rennes University, DTU Copenhagen, King Abdullah University of Science and Technology (KAUST).

*Hans-Peter Seidel:*

- DFG Senate, German Research Foundation (DFG) (2019-2022).
- ACM SIGGRAPH Academy Committee (2018-2021).
- Senate Evaluation Committee (SAE), Gottfried Wilhelm Leibniz Association (Wissenschaftsgemeinschaft Gottfried Wilhelm Leibniz e.V.) (2014-2022).
- Advisory Board, EUROGRAPHICS Awards Program, European Association for Computer Graphics (since 2014).
- Executive Committee, EUROGRAPHICS Association (since 1992).

*Gurprit Singh:*

- Doctoral Committee Program member, representing MPI as an advance external partner, CYENS Center of excellence, Nicosia, Cyprus (since 2018).
- CIO selection committee member, representing MPI as an advance external partner, CYENS Center of excellence, Nicosia, Cyprus (2020).

*Paul Strohmeier:*

- Reviewer for DFG grants.

*Nihar Sabnis:*

- *Open Science Foundation Award* Saarland Open Science Reporter*innen, 2022.

*Courtney Reed:*

- *Special recognition for Outstanding Reviews* for the 2023 TEI Pictorials Track and the 2023 CHI Papers Track.

## 30.11 Teaching Activities

### Summer Semester 2021

**Lectures:**
Realistic Image Synthesis (P. Slusallek, K. Myszkowski, G. Singh)

### Winter Semester 2021/2022

**Lectures:**
Geometric Modeling (R. Zayer)

### Summer Semester 2022

**Lectures:**
Realistic Image Synthesis (P. Slusallek, K. Myszkowski, G. Singh)

### Winter Semester 2022/2023

**Seminars:**

Computer Vision and Machine Learning for Computer Graphics (C. Theobalt, M. Habermann, T. Leimkühler)

### Master and Bachelor Theses

Suraksha Balachandra: HapTick: a tactile time-telling method, 2022

Nikan Doosti: Design Representation Networks for Topology Optimization, 2021

Varshini Muthukumar: Spectral Analysis for Point Samples on Manifolds, 2021

Sana Nasir: Effects of Augmented Feedback on Performance, 2022

Shishir Reddy: Object Pose Estimation Using Shape Representation, 2021.

## 30.12 Dissertations, Habilitations, Awards

### 30.12.1 Dissertations

**Completed and Defended**

Josef Hladký: Latency Hiding through High Fidelity Novel View Synthesis on Thin Clients using Decoupled Streaming Rendering from Powerful Servers, 16.08.2022.

**Completed and Submitted**

Elena Arabadzhiyska-Koleva: Perceptually Driven Methods for Improved Gaze-Contingent Rendering, 2023.

Mojtaba Bemana, Efficient Image-Based Rendering, 2023.

### 30.12.2 Offers for Faculty Positions

- Mengyu Chu: Assistant Professor, Beijing University, China, (accepted)
- Yulia Gryaditskaya: Lecturer, University of Surrey, UK, 2022
- Petr Kellnhofer: Assistant Professor, TU Delft, Netherlands, 2021
- Ana Serrano: Assistant Professor, University of Zaragoza, Spain, 2021
- Cara Tursun: Assistant Professor, University of Groningen, Netherlands, 2021
- Rhaleb Zayer: Associate Professor, University of East Anglia, UK, (accepted)
- Quan Zheng: Associate Professor, Chinese Academy of Sciences, Institute of Software, Beijing, China 2021.

### 30.12.3 Awards

- Vahid Babaei: *Hermann Neuhaus Award*, Max-Planck Society, 2022
- Hans-Peter Seidel: *SIGGRAPH Academy Award*, ACM SIGGRAPH, 2022

## 30.13 Grants and Cooperations

- Vahid Babaei, EXIST Transfer of Research Grant, German Federal Ministry for Economic Affairs and Climate Action (BMWK), €777,000

- Vahid Babaei, *MAX!mize* Award – the Startup Incubation Program for Max Planck Scientists, Max Planck Foundation, €50,000

- Vahid Babaei with Wojciech Matusik (CSAIL MIT), Szymon Rusinkiewicz (Princeton), Piotr Didyk (USI Lugano), Michal Piovarci and Bernd Bickel (IST Austria): Reinforcement Learning for Direct Ink Writing

- Vahid Babaei with Michal Piovarci (IST Austria): Reinforcement Learning for Multi-Material Fusion

- Vahid Babaei with Julian Panetta (UC Davis): Shape from Release

- Vahid Babaei with Julian Panetta (UC Davis): Design Representation Networks

- Karol Myszkowski with Ana Serrano (Zaragoza University): Material Apperance and Neural Tone Mapping

- Karol Myszkowski with Tobias Ritschel (UCL): Neural Rendering

- Karol Myszkowski with Piotr Didyk (USI Lugano): Foveated Rendering

- Karol Myszkowski with Rafał Mantiuk (Cambridge University): Modern Displays

- Karol Myszkowski with Michal Piovarci (IST Austria): 3D Fabrication and Gloss Perception

- Karol Myszkowski, Horizon 2020, Innovative Training Networks (ITN), RealVision: Hyperrealistic Imaging Experience, grant agreement 765911

- Hans-Peter Seidel, MPI-INF Board Representative, CYENS Centre of Excellence on Interactive Media, Smart systems and Emerging technologies, Nikosia, Cyprus

- Gurprit Singh with Tobias Ritschel (UCL) and Timo Ropinski (Ulm University): Blue Noise for Data Visualization

- Gurprit Singh, Vahid Babaei, Quan Zheng with Gordon Wetzstein (Stanford University) and Matthias Zwicker (University of Maryland): Neural Light Field 3D Printing

- Gurprit Singh, Tianqi Fan with Yifan Xu and Yi Yuan (Netease Fuxi AI Lab): LadyBird

- Gurprit Singh, Quan Zheng with Matthias Zwicker (University of Maryland): Real-time Monte Carlo Denoising
- Gurprit Singh, Gift funding (Euros 167K), Huawei Technologies Co.,Ltd: Fundamental problems in Sampling and Rendering
- Rhaleb Zayer with Markus Steinberger (TU Graz): Fine-grained Performance in Geometry Processing
- Paul Strohmeier with Jürgen Steimle (Saarland University): Tactile Rendering
- Paul Strohmeier and Courtney Reed with Andrew McPherson (Imperial College London): Vocal Augmentation & Metaphors in Vocal Teaching
- Paul Strohmeier with Aske Mottelson (ITU Denmark), Henning Pohl (Aarhus University), Jess McIntosh (University of Copenhagen), Jarrod Knibbe (University of Melbourne), Joanna Bergström (University of Copenhagen), Yvonne Jansen (CNRS), Kasper Hornbæk (University of Copenhagen): Body Based User Interfaces
- Paul Stroheier with Florian Mueller (Monash University), Nathan Semertzidis (Monash University), Josh Andres (Monash University), Martin Weigel (University of Applied Sciences, Giessen), Suranga Nanayakkara (University of Aukland), Rakesh Patibanda (Monash University), Zhuying Li (Monash University), Jarrod Knibbe (University of Melbourne), Stefan Greuter (Deakin University), Marianna Obrist (UCL), Pattie Maes (MIT Media Lab), Dakuo Wang (Northeaster University), Katrin Wolf (Bamberg University), Liz Gerber (Northwestern University), Joe Marshall (University of Nottingham), Kai Kunze (Keio University), Jonathan Grudin (Microsoft), Harald Reiterer (University of Konstanz), Richard Byrne (Monash University): Human Computer Integration
- Paul Strohmeier with Narjes Pourjafarian and Jürgen Steimle (Saarland University): Intelligent Handheld Printer
- Paul Strohmeier and Dennis Wittchen with Oliver Schneider (University of Waterloo), Bruno Fruchard (Inria), Bibhushan Raj Joshi (University of Waterloo), Georg Freitag (HTW Dresden), Donald Degraen (DFKI): Sustainable Haptic Feedback Research & Design

## 30.14 Publications

### Journal articles and book chapters

[1] N. Ansari, H.-P. Seidel, and V. Babaei. Mixed integer neural inverse design. *ACM Transactions on Graphics (Proc. ACM SIGGRAPH)*, 41(4), Article 151, 2022.

[2] E. Arabadzhiyska, C. Tursun, H.-P. Seidel, and P. Didyk. Practical saccade prediction for head-mounted displays: Towards a comprehensive model. *ACM Transactions on Applied Perception*, 20(1), Article 2, 2023.

[3] B. Chen, C. Wang, M. Piovarči, H.-P. Seidel, P. Didyk, K. Myszkowski, and A. Serrano. The effect of geometry and illumination on appearance perception of different material categories. *The Visual Computer*, 37:2975–2987, 2021.

[4] V. Chizhov, I. Georgiev, K. Myszkowski, and G. Singh. Perceptual error optimization for Monte Carlo rendering. *ACM Transactions on Graphics*, 41(3), Article 26, 2022.

[5] M. Chu, L. Liu, Q. Zheng, E. Franz, H.-P. Seidel, C. Theobalt, and R. Zayer. Physics informed neural fields for smoke reconstruction with sparse data. *ACM Transactions on Graphics*, 41(4), Article 119, 2022.

[6] M. Chu, N. Thuerey, H.-P. Seidel, C. Theobalt, and R. Zayer. Learning meaningful controls for fluids. *ACM Transactions on Graphics (Proc. ACM SIGGRAPH)*, 40(4), Article 100, 2021.

[7] U. Çoğalan, M. Bemana, K. Myszkowski, H.-P. Seidel, and T. Ritschel. Learning HDR video reconstruction for dual-exposure sensors with temporally-alternating exposures. *Computers and Graphics*, 105:57–72, 2022.

[8] J. Delanoy, A. Serrano, B. Masia, and D. Gutierrez. Perception of material appearance: A comparison between painted and rendered images. *Journal of Vision*, 21(5), Article 16, 2021.

[9] J. Hladký, H.-P. Seidel, and M. Steinberger. SnakeBinning: Efficient temporally coherent triangle packing for shading streaming. *Computer Graphics Forum (Proc. EUROGRAPHICS)*, 40(2):475–488, 2021.

[10] J. Hladký, M. Stengel, N. Vining, B. Kerbl, H.-P. Seidel, and M. Steinberger. Quadstream: A quad-based scene streaming architecture for novel viewpoint reconstruction. *ACM Transactions on Graphics (Proc. ACM SIGGRAPH Asia)*, 41(6), Article 233, 2022.

[11] X. Huang, P. Memari, H.-P. Seidel, and G. Singh. Point-pattern synthesis using Gabor and random filters. *Computer Graphics Forum (Proc. Eurographics Symposium on Rendering)*, 41(4):169–179, 2022.

[12] C. Jambon, B. Kerbl, G. Kopanas, S. Diolatzis, T. Leimkühler, and G. Drettakis. NeRFshop: Interactive editing of neural radiance fields. *Proceedings of the ACM on Computer Graphics and Interactive Techniques*, 6(1). Accepted 2023.

[13] C. Jiang, C. Tang, H.-P. Seidel, R. Chen, and P. Wonka. Computational design of lightweight trusses. *Computer-Aided Design*, 141, Article 103076, 2021.

[14] A. Jindal, K. Wolski, R. K. Mantiuk, and K. Myszkowski. Perceptual model for adaptive local shading and refresh rate. *ACM Transactions on Graphics (Proc. ACM SIGGRAPH Asia)*, 40(6), Article 281, 2021.

[15] G. Kopanas, T. Leimkühler, G. Rainer, C. Jambon, and G. Drettakis. Neural point catacaustics for novel-view synthesis of reflections. *ACM Transactions on Graphics (Proc. ACM SIGGRAPH Asia)*, 41(6), Article 201, 2022.

[16] T. Leimkühler and G. Drettakis. FreeStyleGAN: Free-view editable portrait rendering with the camera manifold. *ACM Transactions on Graphics (Proc. ACM SIGGRAPH Asia)*, 40(6), Article 224, 2021.

[17] Mallikarjun B R, A. Tewari, A. Dib, T. Weyrich, B. Bickel, H.-P. Seidel, H. Pfister, W. Matusik, L. Chevallier, M. Elgharib, and C. Theobalt. PhotoApp: Photorealistic appearance editing of head portraits. *ACM Transactions on Graphics (Proc. ACM SIGGRAPH)*, 40(4), Article 44, 2021.

[18] F. Mueller, N. Semertzidis, J. Andres, M. Weigel, S. Nanayakkara, R. Patibanda, Z. Li, P. Strohmeier, J. Knibbe, S. Greuter, M. Obrist, P. Maes, D. Wang, K. Wolf, L. Gerber, J. Marshall, K. Kunze, J. Grudin, H. Reiterer, and R. Byrne. Human–computer integration: Towards integrating the human body with the computational machine. *Foundations and Trends in Human-Computer Interaction*, 16(1):1–64, 2022.

[19] J. Panetta, H. Mohammadian, E. Luci, and V. Babaei. Shape from release: Inverse design and fabrication of controlled release structures. *ACM Transactions on Graphics (Proc. ACM SIGGRAPH Asia)*, 41(6), Article 274, 2022.

[20] M. Piovarči, M. Foshey, J. Xu, T. Erps, V. Babaei, P. Didyk, S. Rusinkiewicz, W. Matusik, and B. Bickel. Closed-loop control of direct ink writing via reinforcement learning. *ACM Transactions on Graphics (Proc. ACM SIGGRAPH)*, 41(4), Article 112, 2022.

[21] L. Ruan, B. Chen, J. Li, and M.-L. Lam. AIFNet: All-in-focus image restoration network using a light field-based dataset. *IEEE Transactions on Computational Imaging*, 7:675–688, 2021.

[22] C. Salaün, I. Georgiev, H.-P. Seidel, and G. Singh. Scalable multi-class sampling via filtered sliced optimal transport. *ACM Transactions on Graphics (Proc. ACM SIGGRAPH Asia)*, 41(6), Article 261, 2022.

[23] C. Salaün, A. Gruson, B.-S. Hua, T. Hachisuka, and G. Singh. Regression-based Monte Carlo integration. *ACM Transactions on Graphics (Proc. ACM SIGGRAPH)*, 41(4), Article 79, 2022.

[24] A. Serrano, B. Chen, C. Wang, M. Piovarči, H.-P. Seidel, P. Didyk, and K. Myszkowski. The effect of shape and illumination on material perception. *ACM Transactions on Graphics (Proc. ACM SIGGRAPH)*, 40(4), Article 125, 2021.

[25] P. Strohmeier, A. Mottelson, H. Pohl, J. McIntosh, J. Knibbe, J. Bergström, Y. Jansen, and K. Hornbæk. Body-based user interfaces. In A. J. T. Alsmith and M. R. Longo, eds., *The Routledge Handbook of Bodily Awareness*, pp. 478–502. Routledge, London, 2022.

[26] L. Surace, M. Wernikowski, C. Tursun, K. Myszkowski, R. Mantiuk, and P. Didyk. Learning GAN-based foveated reconstruction to recover perceptually important image features. *ACM Transactions on Applied Perception*, 2023.

[27] C. Wang, B. Chen, H.-P. Seidel, K. Myszkowski, and A. Serrano. Learning a self-supervised tone mapping operator via feature contrast masking loss. *Computer Graphics Forum (Proc. EUROGRAPHICS)*, 41(2):71–84, 2022.

[28] A. Weinrauch, H.-P. Seidel, D. Mlakar, M. Steinberger, and R. Zayer. A variational loop shrinking analogy for handle and tunnel detection and Reeb graph construction on surfaces. *Computer Graphics Forum*, 42(2), 2023.

[29] K. Wolski, L. Trutoiu, Z. Dong, Z. Shen, K. Mackenzie, and A. Chapiro. Geo-metric: A perceptual dataset of distortions on faces. *ACM Transactions on Graphics (Proc. ACM SIGGRAPH Asia)*, 41(6), Article 215, 2022.

[30] K. Wolski, F. Zhong, K. Myszkowski, and R. K. Mantiuk. Dark stereo: Improving depth perception under low luminance. *ACM Transactions on Graphics (Proc. ACM SIGGRAPH)*, 41(4), Article 146, 2022.

## Conference articles

[1] N. Ansari, H.-P. Seidel, N. Vahidi Ferdowsi, and V. Babaei. Autoinverse: Uncertainty aware inversion of neural networks. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, eds., *Advances in Neural Information Processing Systems 35 (NeurIPS 2022)*, New Orleans, LA, USA, 2022, pp. 8675–8686. Curran Associates, Inc.

[2] M. Bemana, K. Myszkowski, J. R. Frisvad, H.-P. Seidel, and T. Ritschel. Eikonal fields for refractive novel-view synthesis. In M. Nandigjav, N. J. Mitra, and A. Hertzmann, eds., *Proceedings SIGGRAPH 2022 Conference Papers Proceedings (ACM SIGGRAPH 2022)*, Vancouver, Canada, 2022, Article 39. ACM.

[3] B. Chen, M. Piovarči, C. Wang, H.-P. Seidel, P. Didyk, K. Myszkowski, and A. Serrano. Gloss management for consistent reproduction of real and virtual objects. In S. K. Jung, J. Lee, and A. Bargteil, eds., *Proceedings SIGGRAPH Asia 2022 (ACM SIGGRAPH Asia 2022)*, 2022, Article 35. ACM.

[4] N. Doosti, J. Panetta, and V. Babaei. Topology optimization via frequency tuning of neural design representations. In E. Whiting, J. Hart, C. Sung, J. McCann, and N. Peek, eds., *Proceedings SCF 2021*, Virtual Event, 2021, Article 1. ACM.

[5] G. Fox, W. Liu, H. Kim, H.-P. Seidel, M. Elgharib, and C. Theobalt. VideoForensicsHQ: Detecting high-quality manipulated face videos. In *IEEE International Conference on Multimedia and Expo (ICME 2021)*, Shenzhen, China (Virtual), 2021, pp. 1–6. IEEE.

[6] I. Habibie, W. Xu, D. Mehta, L. Liu, H.-P. Seidel, G. Pons-Moll, M. Elgharib, and C. Theobalt. Learning speech-driven 3D conversational gestures from video. In *Proceedings of the 21st ACM International Conference on Intelligent Virtual Agents (IVA 2021)*, Virtual Event, Japan, 2021, pp. 101–108. ACM.

[7] A. Haynes, C. N. Reed, C. Nordmoen, and S. Skach. Being meaningful: Weaving soma-reflective technological mediations into the fabric of daily life. In *TEI '23, Seventeenth International Conference on Tangible, Embedded, and Embodied Interaction*, Warsaw, Poland, 2023, Article 68. ACM.

[8] M. Kappel, V. Golyanik, M. Elgharib, J.-O. Henningson, H.-P. Seidel, S. Castillo, C. Theobalt, and M. A. Magnor. High-fidelity neural human motion transfer from monocular video computer vision and pattern recognition. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2021)*, Virtual Conference, 2021, pp. 1541–1550. IEEE.

[9] J. Knibbe, R. Freire, M. Koelle, and P. Strohmeier. Skill-sleeves: Designing electrode garments for wearability. In *TEI '21, Fifteenth International Conference on Tangible, Embedded, and Embodied Interaction*, Salzburg, Austria, 2021, Article 33. ACM.

[10] K. Liao, T. Tricard, M. Piovarči, H.-P. Seidel, and V. Babaei. Learning deposition policies for fused multi-material 3D printing. In *IEEE International Conference on Robotics and Automation (ICRA 2023)*, London, UK, 2023. IEEE. Accepted.

[11] Mallikarjun B R, A. Tewari, T.-H. Oh, T. Weyrich, B. Bickel, H.-P. Seidel, H. Pfister, W. Matusik, M. Elgharib, and C. Theobalt. Monocular reconstruction of neural face reflectance fields. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2021)*, Virtual Conference, 2021, pp. 4789–4798. IEEE.

[12] Mallikarjun B R, A. Tewari, H.-P. Seidel, M. Elgharib, and C. Theobalt. Learning complete 3D morphable face models from images and videos. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2021)*, Virtual Conference, 2021, pp. 3360–3370. IEEE.

[13] J. Nehvi, V. Golyanik, F. Mueller, H.-P. Seidel, M. Elgharib, and C. Theobalt. Differentiable event stream simulator for non-rigid 3D tracking. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPR 2021)*, Virtual Workshop, 2021, pp. 1302–1311. IEEE.

[14] N. Pourjafarian, M. Koelle, F. Mjaku, P. Strohmeier, and J. Steimle. Print-A-Sketch: A handheld printer for physical sketching of circuits and sensors on everyday surfaces. In *CHI '22, CHI Conference on Human Factors in Computing Systems*, New Orleans, LA, USA, 2022, Article 270. ACM.

[15] C. N. Reed. As the luthiers do: Designing with a living, growing, changing body-material. In *CHI' 23 Workshop – Body x Materials*, Hamburg, Germany, 2023. Accepted.

[16] C. N. Reed and A. P. McPherson. The body as sound: Unpacking vocal embodiment through auditory biofeedback. In *TEI '23, Seventeenth International Conference on Tangible, Embedded, and Embodied Interaction*, Warsaw, Poland, 2023, Article 7. ACM.

[17] C. N. Reed, C. Nordmoen, A. Martelloni, G. Lepri, N. Robson, E. Zayas-Garin, K. Cotton, L. Mice, and A. McPherson. Exploring experiences with new musical instruments through micro-phenomenology. In *NIME 2022, International Conference on New Interfaces for Musical Expression*, Auckland, New Zealand, 2022. PubPub.

[18] C. N. Reed, S. Skach, P. Strohmeier, and A. P. McPherson. Singing knit: Soft knit biosensing for augmenting vocal performances. In *AHs '22, Augmented Humans International Conference*, Munich, Germany (Hybrid), 2022, pp. 170–183. ACM.

[19] C. N. Reed, P. Strohmeier, and A. McPherson. Negotiating experience and communicating information through abstract metaphor. In *CHI '23, CHI Conference on Human Factors in Computing Systems*, Hamburg, Germany, 2023. ACM. Accepted.

[20] L. Ruan, B. Chen, J. Li, and M. Lam. Learning to deblur using light field generated and real defocus images. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2022)*, New Orleans, LA, USA, 2022, pp. 16283–16292. IEEE.

[21] N. Sabnis, D. Wittchen, C. N. Reed, N. Pourjafarian, J. Steimle, and P. Strohmeier. Haptic servos: Self-contained vibrotactile rendering system for creating or augmenting material experiences. In *CHI '23, CHI Conference on Human Factors in Computing Systems*, Hamburg, Germany, 2023. ACM. Accepted.

[22] N. Sabnis, D. Wittchen, G. Vega, C. N. Reed, and P. Strohmeier. Tactile symbols with continuous and motion-coupled vibration: An exploration of using embodied experiences for hermeneutic design. In *CHI '23, CHI Conference on Human Factors in Computing Systems*, Hamburg, Germany, 2023. ACM. Accepted.

[23] O. Schneider, B. Fruchard, D. Wittchen, B. R. Joshi, G. Freitag, D. Degraen, and P. Strohmeier. Sustainable haptic design: Improving collaboration, sharing, and reuse in haptic design research. In S. Barbosa, C. Lampe, C. Appert, and D. A. Shamma, eds., *CHI '22, CHI Conference on Human Factors in Computing Systems*, New Orleans, LA, USA, 2022, Article 79. ACM.

[24] D. Wittchen, V. Marinez-Missir, S. Mavali, N. Sabnis, C. N. Reed, and P. Strohmeier. Designing interactive shoes for tactile augmented reality. In *AHs '23, Augmented Humans International Conference*, Glasgow, UK, 2023, pp. 1–14. ACM.

[25] D. Wittchen, K. Spiel, B. Fruchard, D. Degraen, O. Schneider, G. Freitag, and P. Strohmeier. TactJam: An end-to-end prototyping suite for collaborative design of on-body vibrotactile feedback. In *TEI '22, Sixteenth International Conference on Tangible, Embedded, and Embodied Interaction*, Daejeon, Republic of Korea (Online), 2022, Article 1. ACM.

[26] T. Yenamandra, A. Tewari, F. Bernard, H.-P. Seidel, M. Elgharib, D. Cremers, and C. Theobalt. i3DMM: Deep implicit 3D morphable model of human heads. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2021)*, Virtual Conference, 2021, pp. 12798–12808. IEEE.

[27] Q. Zheng, G. Singh, and H.-P. Seidel. Neural relightable participating media rendering. In M. Ranzato, A. Beygelzimer, P. S. Liang, J. W. Vaughan, and Y. Dauphin, eds., *Advances in Neural Information Processing Systems 34 (NeurIPS 2021)*, Virtual, 2021, pp. 15203–15215. Curran Associates, Inc.

## arXiv papers

[1] N. Ansari, H.-P. Seidel, and V. Babaei. *Mixed Integer Neural Inverse Design*, 2021. arXiv: 2109.12888.

[2] E. Arabadzhiyska, C. Tursun, H.-P. Seidel, and P. Didyk. *Practical Saccade Prediction for Head-Mounted Displays: Towards a Comprehensive Model*, 2022. arXiv: 2205.01624.

[3] U. Çoğalan, M. Bemana, H.-P. Seidel, and K. Myszkowski. *Video Frame Interpolation for High Dynamic Range Sequences Captured with Dual-exposure Sensors*, 2022. arXiv: 2206.09485.

[4] I. Habibie, W. Xu, D. Mehta, L. Liu, H.-P. Seidel, G. Pons-Moll, M. Elgharib, and C. Theobalt. *Learning Speech-driven 3D Conversational Gestures from Video*, 2021. arXiv: 2102.06837.

[5] L. Surace, M. Wernikowski, O. T. Tursun, K. Myszkowski, R. Mantiuk, and P. Didyk. *Learning Foveated Reconstruction to Preserve Perceived Image Statistics*, 2021. arXiv: 2108.03499.

[6] C. Wang, B. Chen, H.-P. Seidel, K. Myszkowski, and A. Serrano. *Learning a self-supervised tone mapping operator via feature contrast masking loss*, 2021. arXiv: 2110.09866.

[7] C. Wang, A. Serrano, X. Pan, B. Chen, H.-P. Seidel, C. Theobalt, K. Myszkowski, and T. Leimkühler. *GlowGAN: Unsupervised Learning of HDR Images from LDR Images in the Wild*, 2022. arXiv: 2211.12352.

[8] A. Weinrauch, H.-P. Seidel, D. Mlakar, M. Steinberger, and R. Zayer. *A Variational Loop Shrinking Analogy for Handle and Tunnel Detection and Reeb Graph Construction on Surfaces*, 2021. arXiv: 2105.13168.

# 31 D5: Databases and Information Systems

## 31.1 Personnel

**Director**

Prof. Dr. Gerhard Weikum

**Researchers**

Dr. Klaus Berberich (– June 2021, now HTW Saar)
Dr. Koninika Pal (– August 2021, now IIT Palakkad)
Dr. Paramita Paramita (– September 2022, now Fraunhofer IIS)
Dr. Simon Razniewski (– January 2023, now Bosch AI)
Dr. Julien Roméro (– August 2021, now Télécom SudParis)
Dr. Rishiraj Saha Roy
Dr. Erisa Terolli (– November 2021, now Stevens Institute of Technology, USA)
Dr. Anna Tigunova (February 2023 –)
Dr. Andrew Yates (– January 2022, now University of Amsterdam)

**PhD Students**

Hiba Arnaout
Philipp Christmann (April 2021 –)
Cuong Xuan Chu (– August 2022)
Azin Ghazimatin (– March 2022)
Shrestha Ghosh
Anna Christina Guimarães (– July 2022)
Ghazaleh Haratinezhad Torbati
Vinh Thinh Ho (– July 2022)
Magdalena Kaiser
Preethi Lahoti (– April 2022)
Sreyasi Nag Chowdhury (– September 2021)
Tuan-Phong Nguyen
Sneha Singhania
Anna Tigunova (– April 2022)
Hai Dang Tran

**Secretaries**

Alena Löw (– March 2022)
Larisa Ivanova (May 2022 –)
Petra Schaaf

## 31.2 Visitors

In the time period from April 2021 to April 2023, the following researchers visited our group:

| | | |
|---|---|---|
| Andrew Yates | 19.07.21 | University of Amsterdam |
| Aparna Varde | 03.09.–26.11.21 | Montclair State University, New York |
| Lila Boualili | 21.06.21–31.01.22 | IRIT lab, University of Paul Sabatier, Toulouse |
| Farane Jalali | 29.04.21–30.09.22 | Sharif University of Technology, Iran |
| Gael Varoquaux | 09.03.22 | INRIA Paris |
| Lihu Chen | 01.03.–21.05.22 | Télécom Paris Tech |
| Chen Shani | 21.06.22 | Hebrew University, Jerusalem |
| Jingjing Xu | 06.07.22—30.09.22 | University of Luxembourg |
| Martin Theobald | 28.10.22 | University of Luxembourg |
| Daria Stepanova | 28.10.22 | Bosch, Stuttgart |
| Arijit Khan | 21.11.22 | Aalborg University, Denmark |
| Pauli Miettinen | 05.–06.12.22 | University of Eastern Finland |

## 31.3 Group Organization

Our research has been organized into four technical areas:

- Knowledge Base Construction and Curation (coordinated by Simon Razniewski)
- Information Retrieval and Content Analysis (coordinated by Andrew Yates)
- Question Answering (coordinated by Rishiraj Saha Roy)
- Responsible Data Science (coordinated by Gerhard Weikum)

## 31.4 Knowledge Base Construction and Curation

*Coordinator: Simon Razniewski*

Enhancing computers with "machine knowledge" that can power intelligent applications is a long-standing goal in AI. Major advances on knowledge harvesting - methods for turning noisy Internet content into crisp knowledge structures on entities and relations – have made this formerly elusive vision practically viable today.

Automatically constructed knowledge bases – KBs for short, aka knowledge graphs – are a powerful asset for search, analytics, recommendations and data integration, with intensive use at big industrial stakeholders. Examples are the knowledge graphs for search engines (e.g., Google, Bing, Baidu), question answering (e.g., Apple), social networks (e.g., Facebook, as well as domain-specific KBs (e.g., Bloomberg, Amazon, Alibaba) [1]. These achievements are rooted in academic research and community projects. Our long-term research on the Yago KB has been a trendsetter for this strategic direction. Today, the largest general-purpose KBs with publicly accessible contents are BabelNet (babelnet.org), DBpedia (dbpedia.org), Wikidata (wikidata.org) and Yago (yago-knowledge.org). They contain many millions of entities, organized in hundreds to hundred thousands of semantic classes, and billions of relational facts on entities. Propietary knowledge at large Internet stakeholders even exceed this scale by one or two orders of magnitude.

We have published a comprehensive survey on creation and curation of knowledge bases, dubbed the "machine knowledge book" [4], and we continue to pursue the theme of general-purpose encyclopedic knowledge (see Section 31.4.1). There are new challenges as well, regarding non-standard knowledge aspects, and new opportunities are arising in terms of methodology [3]. With the advent of large Transformer-based language models (aka foundation models), a new approach to knowledge acquisition and representation has made great impact. We believe that this kind of latent knowledge is complementary to structured knowledge bases. While language models tend to have much higher coverage, their outputs often lack explainability and scrutability. In this new field of combining different paradigms for "machine knowledge", we are pursuing a suite of promising directions [2].

**Completeness and Negation** Despite their wealth of facts, none of the major KBs can ever be complete. KBs have been constructed and are maintained with focus on encyclopedic knowledge about prominent and business-relevant entities, and often with strong reliance on Wikipedia. This way of knowledge acquisition misses out on a number of important dimensions, posing open challenges for next-generation KBs. Our research aims to close these gaps in knowledge coverage (see Sections 31.4.2 and 31.4.3).

**Commonsense and Cultural Knowledge** Automatically constructed KBs have mostly focused on harvesting encyclopedic fact knowledge. However, for semantic search and other intelligent applications (e. g., conversational bots in social media), machines need a broader understanding of the world: properties of everyday objects, human activities, plausibility invariants and more. This calls for the goal of distilling commonsense knowledge from Internet sources: properties of objects like size, color, shape, parts or substance of which an object is made of, etc., and knowledge on which objects are used for which activities as well as when and where certain activities typically happen. For example, a rock concert involves musicians, instruments – almost always including drums and guitars, speakers, a microphone for the singer; the typical location is a stage, and so on. Major research endeavors on acquiring commonsense include ConceptNet (`https://conceptnet.io/`), Atomic2020 (`https://allenai.org/data/atomic-2020`) and our own projects *Ascent* and *Candle* (see Section 31.4.4).

**Count Knowledge** Count information such as "number of songs by John Lennon" is ubiquitious, and relevant for many advanced question answering needs. Count information naturally co-exists with instance information ("Let it Be is by John Lennon"), and the two types of information can enhance each other, but also give rise to contradictions. Identifying count information in knowledge bases and text, and giving comprehensive answers to count questions, is an underresearched challenge. Our work has focused on building extractors for count information, and on building question answering systems that consolidate count information from multiple sources, and ground numbers with explanatory instances (see Section 31.4.5).

**Fictional Domain Knowledge** Fiction and fantasy are a core part of human culture, spanning from traditional literature to movies, TV series and video games. Well known fictional domains are, for instance, the Greek mythology, the Mahabharata, Tolkien's Middle earth, the world of Harry Potter, or the Simpsons. These universes contain many hundreds or even thousands of entities and types, and are subject of search-engine queries by fans as well as cultural analysts. For example, fans may query about Muggles who are students of the House of Gryffindor (within the Harry Potter universe). Analysts and researchers in humanities may be interested in understanding character relationships, learning story patterns or investigating gender bias in different cultures. Thus, organizing entities and classes from fictional domains into clean taxonomies is of great value (see Section 31.4.6).

## References

[1] N. F. Noy, Y. Gao, A. Jain, A. Narayanan, A. Patterson, and J. Taylor. Industry-scale knowledge graphs: lessons and challenges. *Commun. ACM*, 62(8):36–43, 2019.

• [2] S. Razniewski, H. Arnaout, S. Ghosh, and F. Suchanek. On the limits of machine knowledge: Completeness, recall and negation in web-scale knowledge bases. *Proceedings of the VLDB Endowment (Proc. VLDB)*, 14(12):3175–3177, 2021.

• [3] G. Weikum. Knowledge graphs 2021: A data odyssey. *Proceedings of the VLDB Endowment (Proc. VLDB)*, 14(12):3233–3238, 2021.

• [4] G. Weikum, L. Dong, S. Razniewski, and F. Suchanek. Machine knowledge: Creation and curation of comprehensive knowledge bases. *Foundations and Trends in Databases*, 10(2-4):108–490, 2021.

### 31.4.1 YAGO Knowledge Base

*Investigators: Gerhard Weikum in cooperation with Thomas Pellisier Tanon (external) and Fabian Suchanek (external)*

Our strategic endeavor on knowledge harvesting has its roots in research on semantic search starting in 2004. Later it became the *Yago-Naga project*, and led to the first release of the *YAGO* knowledge base in February 2007. This work was a trendsetter for the research direction of automatic knowledge base creation at large scale, and a first proof-of-concept and architectural blueprint. This theme was widely adopted by industry in the early 2010s. Major stakeholders like Google, Microsoft, Baidu, Amazon, Alibaba, Bloomberg and others have huge knowledge bases, or "knowledge graphs" as industry calls them, as back-end support for search, recommendation and other mission-critical tasks. Knowledge bases are also frequently used for distant supervision in natural language understanding, for tasks like entity linking, question answering, summarization and more. Lessons from this long-term endeavor as well as new challenges and research opportunities are reflected in the recent keynote paper [7].

The influence and value of Yago has been recognized by the research community through the AIJ Influential Paper Award 2017[1] (for the 2013 Yago2 paper in the Artificial Intelligence Journal [3]), and the W3C Seoul Test of Time Award 2018[2] (for the original WWW 2007 paper on Yago [6]. The W3C Seoul Test of Time Award is given for highly impactful papers

[1] http://aij.ijcai.org/aij-awards-list-of-previous-winners
[2] https://www.iw3c2.org/ToT

that appear in the WWW Conference over its entire history of 30 years. The 2018 award was the fourth paper receiving this paper, the previous three being by Brin and Page on the Google architecture, Sarwar et al. on collaborative filtering for recommender systems, and by Broder et al. on analyzing the web graph. Yago is frequently cited as one of the early blueprints for the automatic construction of large knowledge bases (e.g., [2, 1, 4]).

While earlier versions of Yago mostly built on contents from Wikipedia (and other high-quality web sources), the latest version, *Yago4*, changed the approach by tapping into Wikidata (https://www.wikidata.org/), which is a knowledge base itself, and adding value in the form of an automatically induced, clean and expressive, type system based on schema.org (https://schema.org/). This eliminates inconsistencies and gaps in the Wikidata typing, and enables reasoning with computational-logic tools [5].

Yago is maintained as a joint project of the Max Planck Institute for Informatics and the Télécom ParisTech University. The knowledge base is publicly accessible at http://yago-knowledge.org. Yago is also a key asset for the ambiverseNLU software suite, for recognizing and linking entities and concepts in natural-language texts, developed in our group (https://github.com/ambiverse-nlu).

## References


[1] A. Bosselut, H. Rashkin, M. Sap, C. Malaviya, A. Celikyilmaz, and Y. Choi. COMET: commonsense transformers for automatic knowledge graph construction. In *Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019, Florence, Italy, July 28-August 2, 2019, Volume 1: Long Papers*, 2019, pp. 4762–4779. Association for Computational Linguistics.

[2] X. Dong, E. Gabrilovich, G. Heitz, W. Horn, N. Lao, K. Murphy, T. Strohmann, S. Sun, and W. Zhang. Knowledge vault: a web-scale approach to probabilistic knowledge fusion. In *The 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '14, New York, NY, USA - August 24 - 27, 2014*, 2014, pp. 601–610. ACM.

[3] J. Hoffart, F. M. Suchanek, K. Berberich, and G. Weikum. YAGO2: A spatially and temporally enhanced knowledge base from Wikipedia. *Artificial Intelligence*, 194:28–61, 2013.

[4] I. F. Ilyas, T. Rekatsinas, V. Konda, J. Pound, X. Qi, and M. A. Soliman. Saga: A platform for continuous construction and serving of knowledge at scale. In *SIGMOD '22: International Conference on Management of Data, Philadelphia, PA, USA, June 12 - 17, 2022*, 2022, pp. 2259–2272. ACM.

[5] T. Pellissier Tanon, G. Weikum, and F. Suchanek. YAGO 4: A reason-able knowledge base. In A. Harth, S. Kirrane, A.-C. Ngonga Ngomo, H. Paulheim, A. Rula, A. L. Gentile, P. Haase, and M. Cochez, eds., *The Semantic Web (ESWC 2020)*, Heraklion, Greece, 2020, LNCS 12123, pp. 583 |–596. Springer.

[6] F. M. Suchanek, G. Kasneci, and G. Weikum. Yago: A core of semantic knowledge – unifying WordNet and wikipedia. In C. L. Williamson, M. E. Zurko, P. F. Patel-Schneider, and P. J. Shenoy, eds., *WWW 2007, 16th International World Wide Web Conference*, 2007, pp. 697–706. ACM.

• [7] G. Weikum. Knowledge graphs 2021: A data odyssey. *Proceedings of the VLDB Endowment (Proc. VLDB)*, 14(12):3233–3238, 2021.

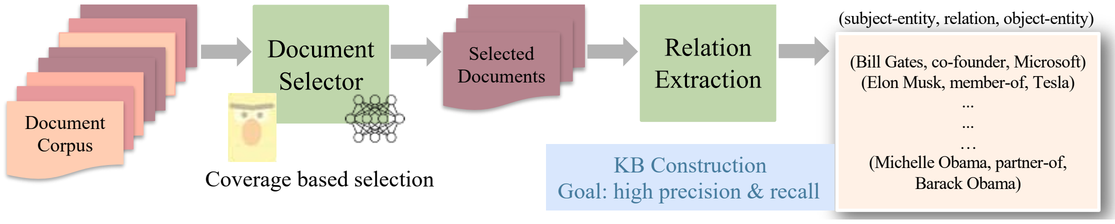



Figure 31.1: RE architecture with document coverage predictor

### 31.4.2 Knowledge Base Coverage

*Investigators: Sneha Singhania, Simon Razniewski, Blerta Veseli, Gerhard Weikum*

Constructing knowledge bases (KBs) with complete information has been a longstanding goal for researchers in natural language processing due to their pivotal role in facilitating a wide range of downstream tasks, such as search, question answering, and translation. However, the high precision and recall requirement to achieve complete information poses a significant challenge in constructing KBs. Most existing techniques primarily focus on achieving high precision and often struggle to strike a balance between precision and recall, resulting in incomplete KBs. In our work [2, 1, 3], we address this issue by proposing novel and efficient methods to estimate the extent of information coverage in various sources, including web documents and pretrained language models. These techniques enable us to effectively materialize information and create KBs with high coverage.

**Estimating Information Coverage in Text Documents** Relation extraction (RE) from text documents producing the structured (subject, relation, object) triples form an essential component in the framework to construct KBs. Intuitively, to achieve high precision and recall, RE methods must process all the available documents to extract relevant information. However, most RE methods are heavy-duty neural networks with computational bottlenecks. In [2], we devise new methods for predicting information coverage in a text document. Specifically, we employ an inexpensive and lightweight neural model for estimating the coverage for a given subject and relation. We curated a diverse dataset and trained our model, which then prioritizes and ranks the most informative documents based on their estimated coverage, allowing us to selectively extract and materialize information. Our approach is not only efficient but also scalable, making it feasible to construct KBs with higher coverage from large-scale sources. Figure 31.1 illustrates this architecture.

**Probing Information from Language Models** Pretrained language models (LMs) have shown to capture world knowledge from large-scale corpora, leading to its widespread usage via prompting to query for structured and factual knowledge. Existing methods focus on probing LMs for a single object using prompts mentioning a subject-relation pair, even though often multiple objects are correct and required for completeness. In [1], we materialize factual knowledge from LMs and explore the viability of KB construction from LMs. We formulate the problem as a *rank-then-select* task as shown in Figure 31.2. For *ranking* candidate objects, we evaluate existing prompting techniques and propose new ones incorporating domain

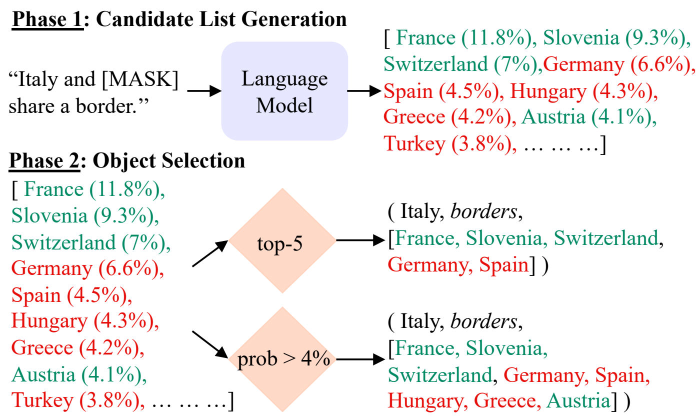



Figure 31.2: Example of LM probing for object ranking and selection

knowledge. For *selecting* objects, we presented various methods that subsets the objects using likelihoods, optimizing for precision and recall. Our results highlight the difficulty of employing LMs for actual knowledge base (KB) construction and pave the way for further research on LMs as KBs.

We hosted our research work as a semantic web challenge, detailed in https://lm-kbc.github.io. The first edition of our challenge gained good participation, and we are hosting the 2nd edition soon. In our ongoing work, we are also devising methods for completing existing KB using LMs and looking into more challenging relations which hold no objects.

## References


- [1] S. Singhania, T.-P. Nguyen, and S. Razniewski, eds. *Knowledge Base Construction from Pre-trained Language Models 2022 (LM-KBC 2022)*, Virtual Event, Hanghzou, China, 2022, CEUR Workshop Proceedings 3274. CEUR-WS.
- [2] S. Singhania, S. Razniewski, and G. Weikum. Predicting document coverage for relation extraction. *Transactions of the Association of Computational Linguistics*, 10:207–223, 2022.
- [3] B. Veseli, S. Singhania, S. Razniewski, and G. Weikum. Evaluating language models for knowledge base completion. In *The Semantic Web (ESWC 2023)*, Hersonissos, Greece, 2023, LNCS. Springer. Accepted.


### 31.4.3 Negation Knowledge

*Investigators: Hiba Arnaout, Simon Razniewski, Gerhard Weikum, Tuan-Phong Nguyen, Jeff Pan (external)*

**Motivation and Problem.** Knowledge bases (KBs) operate under the open-world assumption (OWA), which means statements not contained in the KB should be assumed to have *unknown* truth. The OWA ignores, however, that a significant part of interesting knowledge is *negative*, which cannot be readily expressed in this data model. This work [2, 5, 4, 3, 6, 1] makes the case for explicitly stating *salient* statements that do *not* hold.

| **Statement** | **KB** | **OWA** | **CWA** | **PCWA** |
|---|---|---|---|---|
| (Einstein, award, Nobel Prize in Physics) | yes | true | true | true |
| (Einstein, award, Oscar) | no | unknown | false | unknown |
| (Einstein, member of, Royal Society) | yes | true | true | true |
| (Einstein, member of, Hungarian Academy of Sciences) | no | unknown | false | false |

Table 31.1: Example of comparing OWA, CWA, and PCWA

The popular KBs contain almost only *positive* statements, and this is engraved in the OWA employed on the *Semantic Web*. For instance, it is asserted in Wikidata that *"Stephen Hawking has won the Wolf Prize in Physics"* and thus considered *true*, but the statement that *"Hawking has won the Nobel Prize in Physics"* is absent and thus considered *unknown*. In reality, the reason for the absence of the latter statement is its *falseness*. Being able to distinguish whether a statement is false or unknown would enhance the KBs ability to display more relevant knowledge in *entity summarization* and *question answering*. Currently, KBs answer negative questions by applying the opposite of the OWA, *closed-world assumption* (CWA), which assumes the KB is complete. An intermediate ground is the *partially-closed-world assumption* (PCWA), where *parts* of the data operate under the CWA [8, 7]. The difference between these assumptions is shown in Table 31.1.

**Peer-based Negation Inference.** Let $K^i$ be an (imaginary) ideal KB that perfectly represents reality, i.e., contains exactly those statements that hold in reality. Under the OWA, (practically) available KBs, $K^a$ contain correct statements, but may be incomplete, so the condition $K^a \subseteq K^i$ holds, but not the converse [8]. We distinguish three forms of negative statements:

1. A grounded negative statement $(s, p, o)$ is satisfied if $(s, p, o)$ is *not* $K^i$.
2. A universally negative statement $o\,(s, p, o)$ is satisfied if there is *no* $o$ such that $(s, p, o)$ $K^i$.
3. A conditional negative statement $o\,(s, p, o).(o, p', o')$ is satisfied if there is *no* $o$ such that $(s, p, o)$ and $(o, p', o')$ are $K^i$.

An example of (1.) is *"Tom Cruise never won an Oscar"*, expressed as (Tom Cruise, award, Oscar), an example of (2.) is *"Leonardo DiCaprio has never been married"*, expressed as $x$(Leonardo DiCaprio, spouse, $x$), and an example of (3.) is *"Einstein did not study in any U.S university"*, expressed as $x$(Albert Einstein, educated at, $x$).($x$, location, U.S.).

The goal is to compile a ranked list of *interesting* grounded negative and universally negative statements, given an entity $e$.

The method derives *noteworthy* negative statements by combining information from highly related entities, namely *peers*, with supervised calibration of ranking heuristics, illustrated in Figure 31.3. The intuition behind this method is that similar entities can suggest expectations for relevant statements about a given entity. For instance, many peers of *Stephen Hawking*, namely other physicists, have won the *Nobel in Physics*. We may conclude that the expectation that he did win this prize is feasible, which makes the fact that he did *not* win it an especially useful statement. However, many of his peers were born in *Moscow*, while he was born in

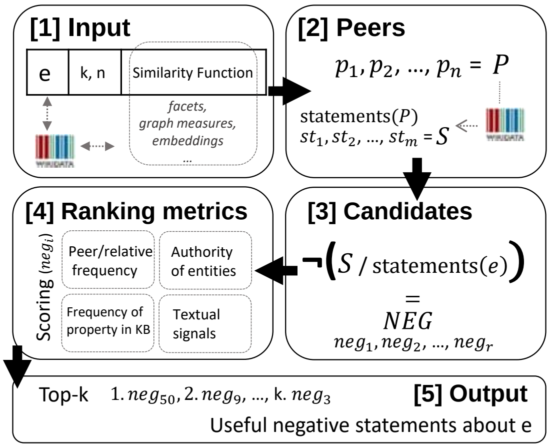



Figure 31.3: Overview of the peer-based negation inference.

*Oxford*. We thus devise supervised ranking methods that take into account various cues such as frequency, importance, unexpectedness, etc.

Table 31.2 shows the top-3 negative statements for *Einstein*. For instance, *Einstein* notably refused to work on the *Manhattan* project, and was suspected of communist sympathies, which makes the assertion that he was *not* a member of the *U.S. Communist Party noteworthy*.

| Random ranking | Property frequency | Ensemble |
|---|---|---|
| $x$(instagram, $x$) | $x$(doctoral student, $x$) | (occupation, astrophysicist) |
| (child, Tarek Sharif) | $x$(candidacy in election, $x$) | (party, Communist Party USA) |
| (award, BAFTA) | $x$(noble title, $x$) | $x$(doctoral student, $x$) |

Table 31.2: Top-3 results for *Einstein* using three different ranking metrics.

**Conditional Negative Statements.** The conditional negative statement is a compromise between two extremes, the grounded negative statements which negates only a *single* assertion, and the universally negative statements, which negates *all* possible assertions for a property. The challenge is, that there is a near-infinite set of *correct* conditional negative statements. And there is a need to identify *noteworthy* ones. The solution in a nutshell: the *peer-based* inference first generates grounded negative statements. Next, subsets of these are *lifted* into more expressive conditional negative statements, using a set of predefined aspects. For instance, it is bulky to list all major universities that *Einstein* did not study at, and it is *not true* that he did not study at any university. However, salient statements are that he *did not study at any U.S. university*, or *at any private university*.

**Wikinegata.** A web portal to showcase the peer-based inference method. Users can explore negative statements about 500K Wikidata entities from various classes, such as organizations, people, and countries. Moreover, users can query the KB using triple-patterns with negated predicates, e.g., *Physicists who did NOT win the Nobel Prize in Physics*. Unlike existing structured search engines, this function returns a ranked list of entities where the negation is useful and often unexpected. Thus, instead of a random list of physicists, the user is shown a set of prominent physicists who did not receive this prize. The system can be accessed at: http://d5demos.mpi-inf.mpg.de/negation.

**Negative Commonsense Knowledge.** Commonsense Knowledge (CSK) are stored in Commonsense Knowledge Bases (CSKB) as (subject, phrase) statements, e.g., (elephant, is a land mammal). Discovering negative statements in CSKB using the peer-based inference method faces two new challenges: 1) inference over well-structured encyclopedic KBs, e.g., Wikidata, does not carry over to verbose and non–canonicalized statements and 2) CSKBs lack a handcrafted taxonomy or a type system with hypernymy relations.

To overcome these challenges, the UnCommonsense method identifies informative negations in CSKBs as follows: given a target concept, e.g., *gorilla*, it computes a set of comparable concepts by employing both an external taxonomy and latent similarity. As in the peer-based inference method, it then postulates the PCWA and considers the positive statements about comparable concepts as candidate negative statements. To eliminate false positives, candidates are scrutinized against related statements in the input CSKB using sentence embeddings, and against a pre-trained language model as an external source of latent knowledge. Finally, the informativeness of negative statements is quantified by statistical scores, and top-ranked negations are generated with provenance showing why they are interesting. For instance, (elephant, can jump) and (gorilla, is territorial), unlike other land mammals. The UnCommonsense knowledge collection can be accessed at: https://uncommonsense.mpi-inf.mpg.de.

## References


[1] H. Arnaout, T.-P. Nguyen, S. Razniewski, and G. Weikum. UnCommonSense in action! Informative negations for commonsense knowledge bases. In T.-S. Chua, H. Lauw, L. Si, E. Terzi, and P. Tsaparas, eds., *WSDM '23, 16th ACM International Conference on Web Search and Data Mining*, Singapore, 2023, pp. 1120–1123. ACM.

[2] H. Arnaout, S. Razniewski, and G. Weikum. Enriching knowledge bases with interesting negative statements. In *Automated Knowledge Base Construction (AKBC 2020)*, Virtual Conference, 2020. OpenReview.

[3] H. Arnaout, S. Razniewski, G. Weikum, and J. Z. Pan. Negative knowledge for open-world Wikidata. In J. Leskovec, M. Grobelnik, M. Najork, J. Tan, and L. Zia, eds., *The Web Conference (WWW 2021)*, Ljubljana, Slovenia, 2021, pp. 544–551. ACM.

[4] H. Arnaout, S. Razniewski, G. Weikum, and J. Z. Pan. Negative statements considered useful. *Journal of Web Semantics*, 71, Article 100661, 2021.

[5] H. Arnaout, S. Razniewski, G. Weikum, and J. Z. Pan. Wikinegata: A knowledge base with interesting negative statements. *Proceedings of the VLDB Endowment (Proc. VLDB)*, 14(12):2807–2810, 2021.

[6] H. Arnaout, S. Razniewski, G. Weikum, and J. Z. Pan. UnCommonSense: Informative negative knowledge about everyday concepts. In M. Al Hasan and L. Xiong, eds., *CIKM '22, 31st ACM International Conference on Information and Knowledge Management*, Atlanta GA USA, 2022, pp. 37–46. ACM.

[7] L. A. Galarraga, C. Teflioudi, K. Hose, and F. Suchanek. Amie: association rule mining under incomplete evidence in ontological knowledge bases. In *WWW*, 2013.

[8] S. Razniewski and W. Nutt. Completeness of queries over incomplete databases. In *VLDB*, 2011.

### 31.4.4 Commonsense and Cultural Knowledge

*Investigators: Tuan-Phong Nguyen, Simon Razniewski, Julien Romero (external), Aparna Varde (external), Gerhard Weikum*

With the continued advances in natural language processing and artificial intelligence, the general public is increasingly coming to expect that systems exhibit what may be considered intelligent behavior. While machine learning allows us to learn models exploiting increasingly subtle patterns in data, our systems still lack explicit and structured representations of commonsense knowledge (CSK). Examples of such knowledge include the fact that fire causes heat, the property of ice being cold, as well as relationships such as that a bicycle is generally slower than a car. Previous work in this area has mostly relied on handcrafted or crowdsourced data, consisting of ambiguous assertions, and lacking multimodal data. The work on ConceptNet (https://conceptnet.io/), for instance, relied on crowdsourcing to obtain an important collection of commonsense data. In the context of this research, we have developed large scale commonsense knowledge base construction methods, Ascent++ [2, 5] and Candle [3], which cover CSK about properties of everyday concepts as well as human traits and behaviors conditioned on socio-cultural contexts respectively. Both methods tap into the large-scale crawl C4 with broad web contents and achieve high performance compared to existing CSK acquisition methods, intrinsically and extrinsically. Our previous projects related to this research include Quasimodo [6], Ascent [4] and Dice [1].

**Refined Commonsense Knowledge from Large-Scale Web Contents** Commonsense knowledge (CSK) about concepts and their properties is helpful for AI applications. Prior works, such as ConceptNet, have compiled large CSK collections. However, they are restricted in their expressiveness to subject-predicate-object (SPO) triples with simple concepts for S and strings for P and O. This paper presents a method called Ascent++ to automatically build a large-scale knowledge base (KB) of CSK assertions, with refined expressiveness and both better precision and recall than prior works. Ascent++ goes beyond SPO triples by capturing composite concepts with subgroups and aspects, and by refining assertions with semantic facets (see Figure 31.4). The latter is essential to express the temporal and spatial validity of assertions and further qualifiers. Furthermore, Ascent++ combines open information extraction (OpenIE) with judicious cleaning and ranking by typicality and saliency scores. For high coverage, our method taps into the large-scale crawl C4 with broad web contents which consists of 365M English web pages. The evaluation with human judgments shows the superior quality of the Ascent++ KB, and an extrinsic evaluation for QA-support tasks underlines the benefits of Ascent++. Notably, Ascent++ outperforms Ascent on all evaluated metrics in our experiments. Ascent is the successor method of Ascent++, which uses handful web search results as knowledge source. We publish the KB and code at https://ascentpp.mpi-inf.mpg.de.

**Extracting Cultural Commonsense Knowledge at Scale** Structured knowledge is important for many AI applications. Commonsense knowledge, which is crucial for robust human-centric

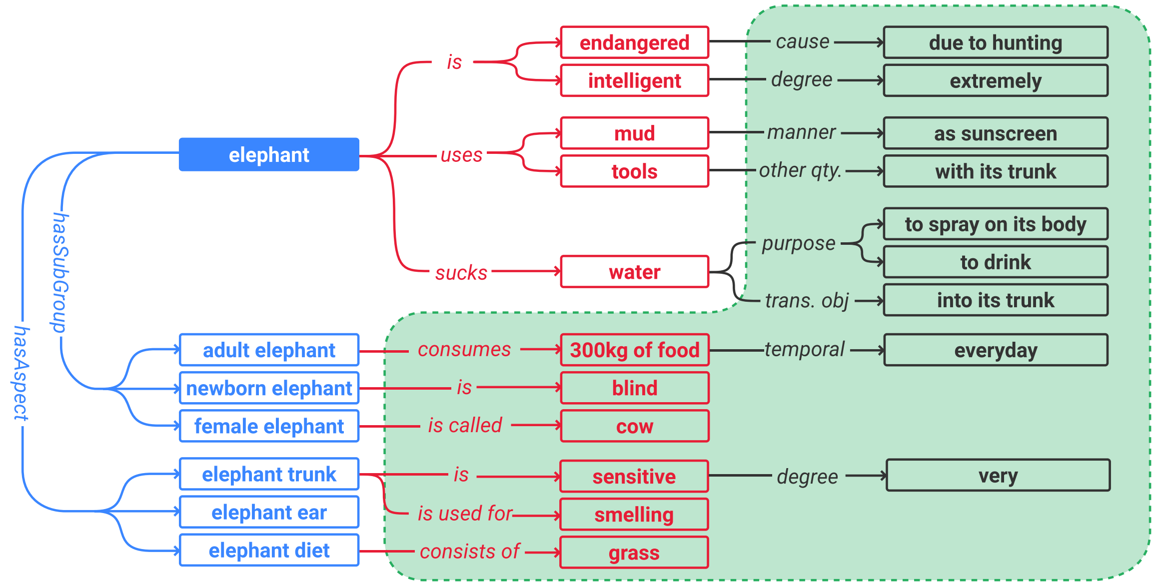



Figure 31.4: Example of ASCENT++knowledge for the concept *elephant*. Prior CSKBs like ConceptNet merely contain assertions outside the green box.

AI, is covered by a small number of structured knowledge projects. However, they lack knowledge about human traits and behaviors conditioned on socio-cultural contexts, which is crucial for situative AI. To fill these gaps, we have developed CANDLE, an end-to-end methodology for extracting high-quality cultural commonsense knowledge (CCSK) at scale [3]. CANDLE extracts CCSK assertions from a huge web corpus (i.e., the C4 crawl) and organizes them into coherent clusters, for 3 domains of subjects (geography, religion, occupation) and several cultural facets (food, drinks, clothing, traditions, rituals, behaviors). CANDLE includes judicious techniques for classification-based filtering and scoring of interestingness. Figure 31.5 shows some example assertions extracted by CANDLE. Experimental evaluations show the superiority of the CANDLE CCSK collection over prior works, and an extrinsic use case demonstrates the benefits of CCSK for the GPT-3 language model. We publish the resource and code at https://candle.mpi-inf.mpg.de.

## References


[1] Y. Chalier, S. Razniewski, and G. Weikum. Joint reasoning for multi-faceted commonsense knowledge. In *Automated Knowledge Base Construction (AKBC 2020)*, Virtual Conference, 2020. OpenReview.

• [2] T.-P. Nguyen, S. Razniewski, J. Romero, and G. Weikum. Refined commonsense knowledge from large-scale web contents. *IEEE Transactions on Knowledge and Data Engineering*, 2022.

• [3] T.-P. Nguyen, S. Razniewski, A. Varde, and G. Weikum. Extracting cultural commonsense knowledge at scale. In *WWW '23, ACM Web Conference*, Austin, TX, USA, 2023. ACM. Accepted.

• [4] T.-P. Nguyen, S. Razniewski, and G. Weikum. Advanced semantics for commonsense knowledge

| | | |
|---|---|---|
| **geography>country** | **Germany** | **drinks** |
| German **beer festivals** in October are a celebration of beer drinking. | | |
| **geography>region** | **East Asia** | **food** |
| **Tofu** is a major ingredient in many East Asian cuisines. | | |
| **geography>region** | **South Asia** | **traditions** |
| In South Asia, **henna** is often used in bridal makeup or to celebrate festivals. | | |
| **occupation** | **lawyer** | **clothing** |
| Lawyers wear **suits** to look professional. | | |
| **occupation** | **firefighter** | **behaviors** |
| Firefighters run into burning buildings to **save lives**. | | |

Figure 31.5: Example assertions of Candle.

extraction. In J. Leskovec, M. Grobelnik, M. Najork, J. Tang, and L. Zia, eds., *The Web Conference 2021 (WWW 2021)*, Ljubljana, Slovenia, 2021, pp. 2636–2647. ACM.

• [5] J. Romero and S. Razniewski. Do children texts hold the key to commonsense knowledge? In Y. Goldberg, Z. Kozareva, and Y. Zhang, eds., *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing (EMNLP 2022)*, Abu Dhabi, United Arab Emirates, 2022, pp. 10954–10959. ACL.

[6] J. Romero, S. Razniewski, K. Pal, J. Z. Pan, A. Sakhadeo, and G. Weikum. Commonsense properties from query logs and question answering forums. In W. Zhu and D. Tao, eds., *CIKM '19, 28th ACM International Conference on Information and Knowledge Management*, Beijing China, 2019, pp. 1411–1420. ACM.

### 31.4.5 Count Knowledge

*Investigators: Shrestha Ghosh, Simon Razniewski, Gerhard Weikum*

**Motivation** Search questions on the web are often related to counts of entities, such as the number of books by an author or the number of songs composed by an artist. Popular question answering datasets contain around 10% of such count questions (cite). While popular question answering systems approach this task in a manner similar to answering factoid questions, it is difficult for a user to comprehend such answers especially since the counts of entities can vary from being very crisply defined to being quite fuzzy. We define count knowledge as information about a class of entities in the form of the cardinality of the entity class as well as enumerations of the individual entities. Figure 31.6 shows an example of count knowledge where the employees of a company are represented as count through the relation *employees* as well as partial entity enumerations through the *employer* relation.

Take for example the number of children of someone, or the number of books by an author or the moons of a planet. These relations are very well-defined and can also be easily enumerated in structured sources such as knowledge bases (KBs). Wikidata for example has the relation *child* that connects a parent entity to all their children entities as well as the relation *numberOfChildren* that connects a parent entity to a literal storing the count of the children. Usually, the enumerations of these well-defined entity classes are complete.

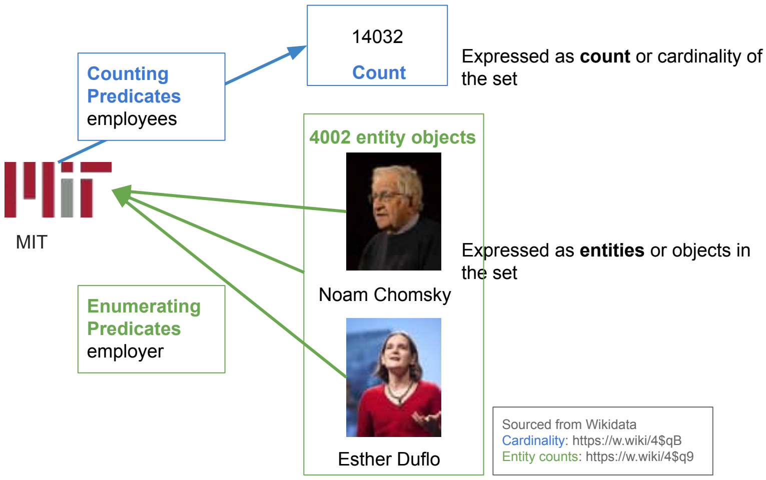



Figure 31.6: Employees of a company as an illustrative example of count knowledge.

When we move to less crisper classes of entities, such as the number of employees of a company, or the number of scientists in the world or the number of exoplanets, it becomes more and ore difficult to find all the entities and count them. It may be so that sources have data on the cardinality of such classes and only a few exemplary instances. Even when sources have counts, there is a chance that the number are estimates.

Another challenge when handling counts is that text sources represent count in multiple formats including numbers, number-words, articles and negation (cite). Counts in text need to be extracted and evaluated as numbers. Some deviation from the ground truth should also be accounted into the evaluation to not heavily penalize systems which are quite close to the correct counts. This is rarely seen in popular question-answering benchmarks which consider exact string match for evaluating all questions.

In context of our description of the count knowledge and its challenges, we work on the following problem statements.

**Extracting Count Information from KBs** We developed the **CounQER** methodology [2] which identifies predicates that store counts of entities (counting predicates) or the entities themselves (enumerating predicates) and aligns semantically related predicates from these two sets. We designed logistic regression classifiers on manually designed features from textual occurrence and KB statistics to identify the predicates. In order to align the predicates, we relied on heuristics that use predicate co-occurrence statistics and linguistic similarity. We demonstrated the usefulness of identifying and aligning count information in KBs for querying [1].

**Answering Entity Counts with Explanations** . We developed the **CoQEx** methodology [3, 4] to infer count answer from a distribution over multiple text snippets and complement it with relevant explanations. The system is illustrated in Figure 31.7. We built a dataset of 5000 questions to train our method on predicting the correct count context and perform evaluations on answer inference, explanations and user comprehension.

CoQEx contextualizes the count by grouping the contexts into three groups: very similar to the inferred answer w.r.t the count and context, contexts denoting semantic subgroups

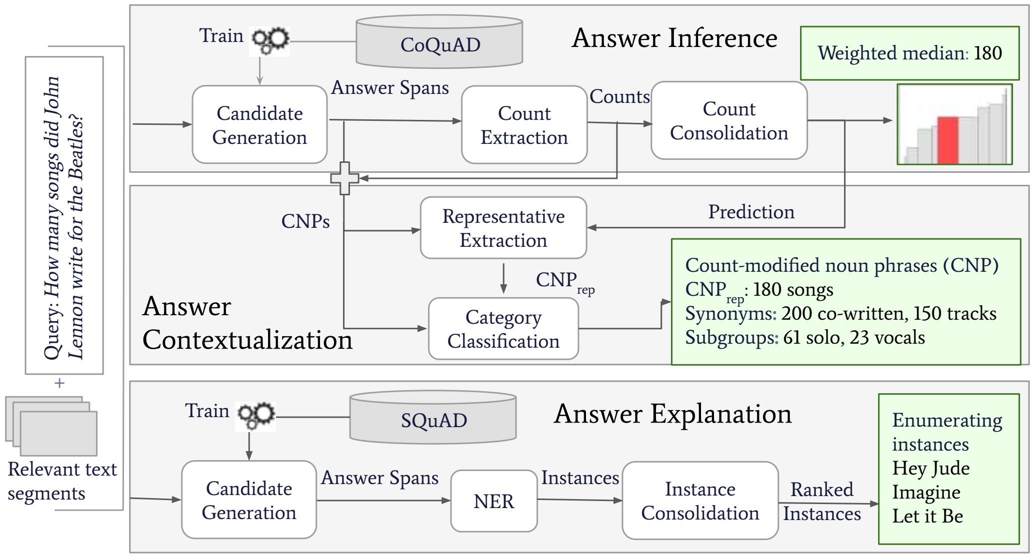



Figure 31.7: Overview of the CoQEx methodology.

and unrelated contexts. Explanations via instances are also provided by extracting instances from text that are compatible with the answer type, which we obtain from the question. We developed a demonstration of our CoQEx system [6] which can be accessed at https://nlcounqer.mpi-inf.mpg.de/.

**Comparing Cardinality of Entity Classes** . Questions on class cardinality comparisons are challenging because of the incompleteness and biases in the digital world of Internet sources. For example, seemingly simple questions like "Are there more lawyers or more programmers? More power plants or more catholic cathedrals?" cannot be reliably answered from Internet sources without major efforts for de-noising and de-biasing. We tackle questions on class cardinality comparison by tapping into three sources – knowledge bases, search engines and language models – for estimates of absolute cardinalities as well as the cardinalities of orthogonal subgroups of the classes. We propose novel techniques for aggregating signals with partial coverage for more reliable estimates and evaluate them on a dataset of 4005 class pairs, achieving an accuracy of 83.7 percent [5].

## References

[1] S. Ghosh, S. Razniewski, and G. Weikum. CounQER: A system for discovering and linking count information in knowledge bases. In A. Harth, V. Presutti, R. Troncy, M. Acosta, A. Polleres, J. D. Fernández, J. Xavier Parreira, O. Hartig, K. Hose, and M. Cochez, eds., *The Semantic Web: ESWC 2020 Satellite Events*, Heraklion, Greece, 2020, LNCS 12124, pp. 84–90. Springer.

[2] S. Ghosh, S. Razniewski, and G. Weikum. Uncovering hidden semantics of set information in knowledge bases. *Journal of Web Semantics*, 64, Article 100588, 2020.

• [3] S. Ghosh, S. Razniewski, and G. Weikum. Answering count queries with explanatory evidence. In E. Amigo, P. Castells, J. Gonzalo, B. Carterett, J. S. Culpepper, and G. Kazai, eds., *SIGIR '22, 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, Madrid, Spain, 2022, pp. 2415–2419. ACM.

- [4] S. Ghosh, S. Razniewski, and G. Weikum. Answering count questions with structured answers from text. *Journal of Web Semantics*, 76, Article 100769, 2023.
- [5] S. Ghosh, S. Razniewski, and G. Weikum. Class cardinality comparison as a fermi problem. In *WWW '23, ACM Web Conference*, Austin, TX, USA, 2023. ACM. Accepted.
- [6] S. Ghosh, S. Razniewski, and G. Weikum. CoQEx: Entity counts explained. In T.-S. Chua, H. Lauw, L. Si, E. Terzi, and P. Tsaparas, eds., *WSDM '23, 16th ACM International Conference on Web Search and Data Mining*, Singapore, 2023, pp. 1168–1171. ACM.


### 31.4.6 Fictional Domain Knowledge

*Investigators: Cuong Xuan Chu, Simon Razniewski, Gerhard Weikum*

**Motivation and Overview** Fictitious contents is richly covered by a variety of Internet sources, like community-based descriptions of movies and TV series (hosted by Wikia, renamed to Fandom), summaries and plots or full-fledged scripts of movies and books (e.g., imdb, librarything, shelfari, ...), or entire book series. The underlying domains often form self-contained universes by themselves, examples being Game of Thrones, the Greek mythology, Buddhism, Tolstoy's novels, Harry Potter or the Simpsons. Knowledge bases on such domains are of interest for two kinds of users: fan communities with entertainment-centric discussions, on one hand, and humanities researchers who aim to analyze historic or modern (sub-) cultures, on the other hand. Extracting salient facts about the entities in historic and fictional domains cannot simply adopt the mainstream state-of-the-art methods. For example, even the inference of an entity's types can be challenging for non-standard types such as wizardry students.

Our research on constructing knowledge bases for such non-standard domains has been directed at three levels of knowledge:

i) organizing entities in a clean type taxonomy that covers both real-world classes and fictional ones (e.g., elves, wizardry students, magic potions) entities and their types [2];
ii) using taxonomic knowledge to infer proper types for mentions in fictional texts, denoting new entities to be incorporated into the knowlege base [3];
iii) extracting relations between entities, as expressed in fictional texts, to populate the knowledge base [4].

Our results along these lines are wrapped up in Cuong Chu's doctoral dissertation [1].

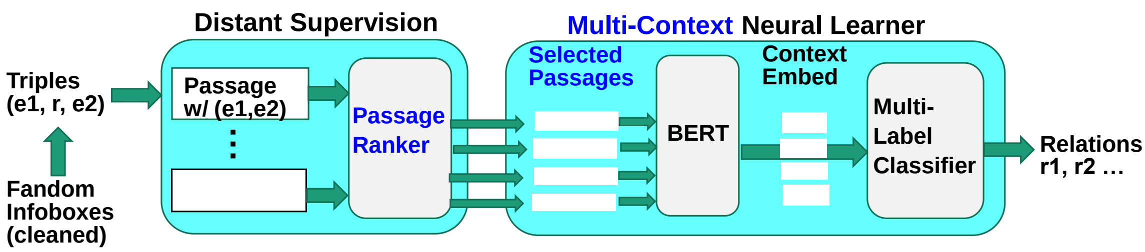



Figure 31.8: KnowFi Architecture for Relation Extraction from Fictional Texts.

**Relation Extraction for Non-Standard Domains** State-of-the-art methods for relation extraction (RE), using deep learning, crucially hinge on spotting training samples in documents and are geared for short input texts. In fictional and cultural domains, the underlying sources include entire novels and other long texts. Standard methods for RE do not work well with these kinds of inputs.

To overcome the limitations, we devised a method that identifies a pool of relevant text passages within long documents, and subsequently ranks the passages in a judicious way. The top-ranked passages for training samples are leveraged for training a Transformer-based neural network, with the salient trait that the network aggregates context from multiple passages. At deployment time, for extracting new facts, the passage ranking is used to counter the sparseness of relevant cues in long texts, greatly improving the output quality of the learned network. Figure 31.8 gives a pictorial overview of this methodology.

Experiments with a variety of inputs (including entire novels) show the viability of the method and its superiority over prior baselines. The method is also applicable to mainstream RE, and our experiments show competitive results also for popular benchmarks like TACRED and DOCRED [6, 5]. Our results have been published in the AKBC Conference [4], a major venue for research on knowledge base construction.

### References


• [1] C. X. Chu. *Knowledge Extraction from Fictional Texts*. PhD thesis, Universität des Saarlandes, Saarbrücken, 2022.

[2] C. X. Chu, S. Razniewski, and G. Weikum. TiFi: Taxonomy induction for fictional domains. In J. McAuley, ed., *Proceedings of The World Wide Web Conference (WWW 2019)*, San Francisco, CA, USA, 2019, pp. 2673–2679. ACM.

[3] C. X. Chu, S. Razniewski, and G. Weikum. ENTYFI: Entity typing in fictional texts. In J. Caverlee and X. B. Hu, eds., *WSDM '20, 13th International Conference on Web Search and Data Mining*, Houston, TX, USA, 2020, pp. 124–132. ACM.

• [4] C. X. Chu, S. Razniewski, and G. Weikum. KnowFi: Knowledge extraction from long fictional texts. In *Automated Knowledge Base Construction (AKBC 2021)*, Virtual Conference, 2021, pp. 1–19. OpenReview.

[5] Y. Yao, D. Ye, P. Li, X. Han, Y. Lin, Z. Liu, Z. Liu, L. Huang, J. Zhou, and M. Sun. Docred: A large-scale document-level relation extraction dataset. In *Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019, Florence, Italy, July 28- August 2, 2019, Volume 1: Long Papers*, 2019, pp. 764–777. Association for Computational Linguistics.

[6] Y. Zhang, V. Zhong, D. Chen, G. Angeli, and C. D. Manning. Position-aware attention and supervised data improve slot filling. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, EMNLP 2017, Copenhagen, Denmark, September 9-11, 2017*, 2017, pp. 35–45. Association for Computational Linguistics.


## 31.5 Information Retrieval and Content Analysis

*Coordinator: Andrew Yates*

Finding relevant documents and extracting information from them are common NLP tasks, with downstream applications such as search, knowledge base construction, and question answering. Consequently, much prior work has considered the core problems of identifying documents relevant to a query, extracting structured information from documents, and analyzing the contents of documents. One might be tempted to assume that these are solved problems given the widespread usage of Information Retrieval through Web search, of Information Extraction through the use of methods that rely on structured data extracted from text (e.g., question answering over knowledge bases), and of Text Analysis through the use of technologies that rely on document-level analysis (e.g., sentiment analysis for monitoring sentiment about a product). These tasks have been extensively studied under specific assumptions and within limited domains, though. This prompts us to develop approaches that relax these assumptions and apply these tasks to more complex domains.

Our work in this space covers a variety of research directions. In the context of Information Retrieval, Sections 31.5.1 and 31.5.2 describe work leveraging improved document understanding to substantially improve retrieval effectiveness and to personalize search in challenging domains, such as for query-based recommendations on refined items like books in the long tail, travel destinations off the beaten path, and more. Section 31.5.3 describes our work applying Information Extraction to the difficult domain of dialogues (e.g., inferring a speaker's hobby from their utterances) in order to support Personal Knowledge Base construction. Finally, in Section 31.5.4 we address the problem of understanding quantities in tables and text with the goal of developing a next generation search engine capable of answering queries containing quantities.

The group has continued to make contributions to core neural information retrieval algorithms, including approaches for performing query expansion and pseudo-relevance feedback with neural models [6, 9], investigations of how to use Transformer models for reranking [8, 7], and investigations of the signals captured by Transformer-based ranking models [1]. Several resources were also released, including a tool for simplifying information retrieval data access and caching in the context of neural information retrieval [4], a tool for visualizing differences between ranking systems [2], and a dataset of queries containing complex information needs [5]. Another notable achievement is the publishing of a comprehensive state-of-the-art survey on Transformer-based neural information retrieval [3].

## References


- [1] L. Boualili and A. Yates. A study of term-topic embeddings for ranking. In J. Kamps, L. Goeuriot, F. Crestani, M. Maistro, H. Joho, B. Davis, C. Gurrin, U. Kruschwitz, and A. Caputo, eds., *Advances in Information Retrieval (ECIR 2023)*, Dublin, Ireland, 2023, LNCS 13981, pp. 359–366. Springer.
- [2] K. M. Jose, T. Nguyen, S. MacAvaney, J. Dalton, and A. Yates. DiffIR: Exploring differences in ranking models' behavior. In F. Diaz, C. Shah, T. Suel, P. Castells, R. Jones, T. Sakai, A. Bellogín, and M. Yushioka, eds., *SIGIR '21, 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, Virtual Event, Canada, 2021, pp. 2595–2599. ACM.
- [3] J. Lin, R. Nogueira, and A. Yates. *Pretrained Transformers for Text Ranking: BERT and Beyond*, Synthesis Lectures on Human Language Technologies 53. Morgan & Claypool Publishers, San Rafael, CA, 2021.
- [4] S. MacAvaney, A. Yates, S. Feldman, D. Downey, A. Cohan, and N. Goharian. Simplified data wrangling with ir_datasets. In F. Diaz, C. Shah, T. Suel, P. Castells, R. Jones, T. Sakai,

A. Bellogín, and M. Yushioka, eds., *SIGIR '21, 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, Virtual Event, Canada, 2021, pp. 2429–2436. ACM.

- [5] I. Mackie, J. Dalton, and A. Yates. How deep is your learning: The DL-HARD annotated deep learning dataset. In F. Diaz, C. Shah, T. Suel, P. Castells, R. Jones, T. Sakai, A. Bellogín, and M. Yushioka, eds., *SIGIR '21, 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, Virtual Event, Canada, 2021, pp. 2335–2341. ACM.
- [6] S. Naseri, J. Dalton, A. Yates, and J. Allan. CEQE: Contextualized embeddings for query expansion. In D. Hiemstra, M.-F. Moens, J. Mothe, R. Perego, M. Potthast, and F. Sebastiani, eds., *Advances in Information Retrieval (ECIR 2021)*, Lucca, Italy (Online Event), 2021, LNCS 12656, pp. 467–482. Springer.
- [7] R. Pradeep, Y. Liu, X. Zhang, Y. Li, A. Yates, and J. Lin. Squeezing water from a stone: A bag of tricks for further improving cross-encoder effectiveness for reranking. In M. Hagen, S. Verbene, C. Macdonald, C. Seifert, K. Balog, K. Nørvåg, and V. Setty, eds., *Advances in Information Retrieval (ECIR 2022)*, Stavanger, Norway, 2022, LNCS 13185, pp. 655–670. Springer.
- [8] X. Zhang, A. Yates, and J. Lin. Comparing score aggregation approaches for document retrieval with pretrained transformers. In D. Hiemstra, M.-F. Moens, J. Mothe, R. Perego, M. Potthast, and F. Sebastiani, eds., *Advances in Information Retrieval (ECIR 2021)*, Lucca, Italy (Online Event), 2021, LNCS 12657, pp. 150–163. Springer.
- [9] Z. Zheng, K. Hui, B. He, X. Han, L. Sun, and A. Yates. Contextualized query expansion via unsupervised chunk selection for text retrieval. *Information Processing & Management*, 58(5), Article 102672, 2021.

### 31.5.1 Neural Information Retrieval with Named Entities

*Investigators: Hai Dang Tran, Andrew Yates*

The direction of combining Neural Information Retrieval (NIR) with knowledge about real-world entities attracts huge interest from research community. The power of NIR is from the context understanding capability of pre-trained language model (PLM). However, the knowledge of PLM about long-tail and recent emerging entities is limited, and this is the place where knowledge source could come to rescue. In other words, knowledge about entities could be a complementary to PLM models, and thus enhance the effectiveness of NIR methods.

**Dense Retrieval with Entity Views.** This work addresses the effectiveness of ad-hoc retrieval with the support of entity knowledge. Specifically, the task is passage ranking: given a query, find the most relevant passages in a large text collection.

Search with dense representations has become an effective alternative to statistical methods that use an inverted index (e.g., BM25) [5]. Ranking with dense representations is efficient with the support of approximate nearest neighbor search [4], which uses a specialized index to quickly identify the document representations (approximately) closest to a query representation. These representations are produced by encoding queries and documents (during an indexing step) as dense vectors using a pre-trained language model (PLM) such as BERT, T5 or their numerous variants. To give a concrete example of their success, dense retrieval approaches have performed well on the TREC Deep Learning collections [1] in terms of both efficiency and effectiveness. For example, a BERT-based model trained with topic-aware sampling (TAS) [3] improves nDCG@10 over a tuned BM25 ranker by a large margin in nDCG@10 on the TREC Deep Learning track while having low query latency.

However, PLM models powering such approaches do not capture full information about real world entities [2]. Especially for uncommon entities, this limitation becomes more clear since PLMs both have difficulty disambiguating entities with similar surface forms [2] and are less likely to have seen information about these entities during pretraining. Even the most powerful PLMs cannot capture information about an entity that was not present in the data or that emerged after the PLM was produced.

These limitations motivate approaches for encoding information about entities outside of the PLM itself. We investigate techniques for enriching the representations of query and document text using such entity representations. In our approach, a textual (query or document) representation from a PLM is combined with an entity representation derived from embeddings of the entities present in the input (query or document) text. We leverage external entity embeddings [7] as a lightweight way to incorporate information about entities. Rather than being tied to a PLM, these embeddings are created and stored independently.

A document may contain many entities; we find that the straightforward approach of creating a single vector representation of all a document's entities does not perform well. Instead, we propose a multiple representation approach in which we create several entity representations with respect to different clusters of entities in a text. This can be understood as creating different views of a document that focus on different groups of entities. These entity views are then combined with a textual representation of the document and indexed for approximate nearest neighbor search. We find that this is an effective and efficient strategy: enriching representations from a strong BERT-based model (TAS) yields significant improvements on the TREC Deep Learning and DL-HARD and MS MARCO benchmarks. Details of this work are in [6], published in CIKM 2022.

Figure 31.9 illustrates the entity views present in a short passage about the scientist Lilli Hornig. The first view is a generic one that captures Lilli Hornig without focusing on any specific aspect. The second and third views capture entities related to Lilli's scientific work at National Science Foundation (NSF), while the remaining two views capture entities related to Lilli's studies at Bryn Mawr College. Our approach captures these different views in order to match different aspects of an entity. For example, the query *Lilli Hornig's work at Virginia* better matches the second and third views related to the NSF, whereas a query about Lilli's personal life or education better matches the last two views.

**Joint Inference for Conversational Search with Emerging Entities.** Conversations often refer to emerging out-of-KB entities or uncommon entities in the long tail. Motivated by this observation, we aim to leverage knowledge about entity mentions into conversational search. Given a current question in a conversation context, we need to find top relevant answers from a collection.

These days, giant PLMs like GPT-3 or ChatGPT show impressive results in chat-style and question-answering conversations. These models follow the philosophy of "bigger is better", leveraging huge amounts of training data and many billions of model paramters. However, due to their enormous size, such models require long training time and expensive compute and energy resources, clearly prohibiting daily re-training. Therefore, PLMs inevitably lack

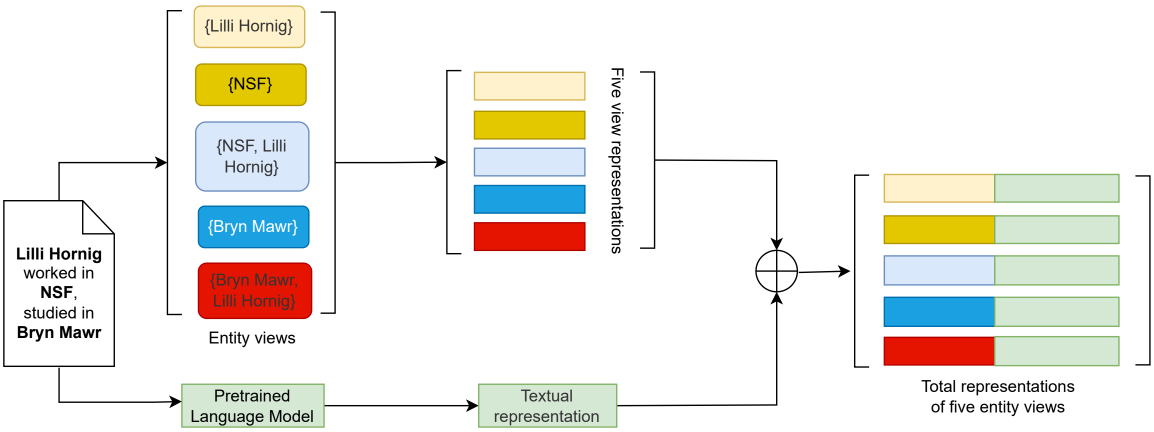



Figure 31.9: Overview of multi-view entity clusters for dense information retrieval

information about recent facts and newly emerging entities, and they also face sparseness and low-confidence issues on topics that involve uncommon entities in the long tail.

This limitation motivates us to investigate how to best handle uncommon and emerging entities in conversational search. We propose to predict types of entity mentions from conversation, and use these as entity knowledge to contextualize and enhance the understanding of query intents. Since not every information in the conversation history is relevant to the current question, judiciously selecting entities and context pieces is crucial. Our approach combines techniques from PLM-based retrieval and KBs for background knowledge with combinatorial optimization methods for selective contextualization.

## References

[1] N. Craswell, B. Mitra, E. Yilmaz, D. Campos, and J. Lin. Overview of the trec 2021 deep learning track. In *Text REtrieval Conference (TREC)*, 2022. TREC.

[2] B. Heinzerling and K. Inui. Language models as knowledge bases: On entity representations, storage capacity, and paraphrased queries. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, 2021, pp. 1772–1791.

[3] S. Hofstätter, S.-C. Lin, J.-H. Yang, J. Lin, and A. Hanbury. Efficiently teaching an effective dense retriever with balanced topic aware sampling. In *Proceedings of the 44th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2021)*, 2021.

[4] J. Johnson, M. Douze, and H. Jégou. Billion-scale similarity search with gpus. *IEEE Transactions on Big Data*, 7(3):535–547, 2019.

• [5] J. Lin, R. Nogueira, and A. Yates. *Pretrained Transformers for Text Ranking: BERT and Beyond*, Synthesis Lectures on Human Language Technologies 53. Morgan & Claypool Publishers, San Rafael, CA, 2021.

• [6] H. D. Tran and A. Yates. Dense retrieval with entity views. In M. Al Hasan and L. Xiong, eds., *CIKM '22, 31st ACM International Conference on Information and Knowledge Management*, Atlanta GA USA, 2022, pp. 1955–1964. ACM.

[7] I. Yamada, A. Asai, H. Shindo, H. Takeda, and Y. Takefuji. Wikipedia2Vec: An optimized tool for learning embeddings of words and entities from wikipedia. *arXiv*, 2018.

### 31.5.2 Personalized Search-based Recommendations

*Investigators: Ghazaleh Haratinezhad Torbati, Andrew Yates, Gerhard Weikum*

The growth of data, ranging from raw information to consumer products and services, increases the demand for personalization. Recommendation systems are at the heart of online shopping and services for music and video streaming, hotels and restaurants, or food recipes [3]. Search-based recommendation or contextual recommendation is a setting where there is a situative context, like a user query or a specific item, that should drive the prediction.

In our work we investigate the extent of user profiling and its effect in performance, we question the default assumptions in the research of recommendation systems, and we study the recommendation in the data-poor regime of long-tail users and long-tail items.

**Search-based Entity Recommendation with Different User Profiles:** Prior work on personalizing recommendation has focused on considering signals from users clicks, likes and ratings capture users' interests. In [5, 6], we investigate tapping into different sources of explicit and implicit signals of interests and tastes: questionnaires and user-to-user chats.

The questionnaire-based user profiles are transparent and scrutable, the users can fully understand and control the information that drives the personalization (e.g., modify or revoke pieces of a profile). However, they require an explicit effort. Most users seem fine with a one-time questionnaire, but few seem ready for periodic updating as their interests and tastes evolve. On the other hand, chat-based user profiles require no effort at all from the user side, and could be easily updated without user intervention. However, the derived models are less transparent to humans and not easily adjustable by users themselves. Also, chat data comes with higher privacy risks.

To compare the performance of different extents of user profiles in search-based entity recommendations, such as books, travel destinations or food recipes, we devise a variety of re-ranking methods: the BM25 family, statistical language models, and neural ranking. Further, we use techniques to enhance the user models by capturing domain-specific vocabularies and by entity-based expansion.

The experiments are based on a collection of online chats and filled out questionnaires from a controlled user study covering three domains: books, travel, food. Figure 31.10 shows excerpts from the questionnaire and the chat collection for an example user. For the query "temples and culture", this user-specific information led to high ranks of travel destinations like Borobudur, Delphi and Ellora – all confirmed as very good recommendations by that user. We released the collected dataset at http://personalization.mpi-inf.mpg.de/.

We evaluate different configurations and compare chat-based user models against concise user profiles from questionnaires. Overall, these two variants perform on par in terms of NCDG@20, but each has advantages on certain domains.

**Search-based Recommendation with Difficult Negative Samples:** Recommender systems have achieved impressive results on benchmark datasets. However, the numbers are often

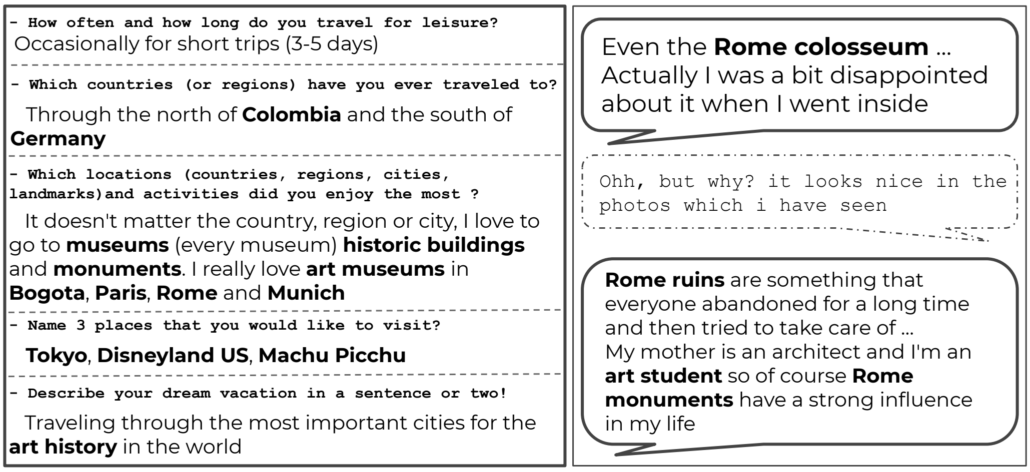



Figure 31.10: Excerpts from user questionnaire and chat on travel domain.

influenced by assumptions made on the data and evaluation mode. We question and revise these assumptions, to study and improve the quality, particularly for the difficult case of search-based recommendations. In this work [4], we focus on the book domain, with generally sparse data, diverse user interests, and a less skewed distribution of item popularity, e.g., compared to movies.

Datasets often contain groups of highly inter-related items, such as books by the same author. Recommending another book from the same author that the users liked is nearly trivial and misses the point of helping users with true discoveries. To this end, we pre-process all data such that training and test sets are disjoint in terms of authors per user (same authors for different users are ok). This way, we measure the true discoveries that a user sees in her recommendations.

It is assumed that recommendations are generated from the user profile alone. In reality, though, there is often a situative context, like a user query or a specific item, that should drive the prediction. Therefore, we focus on search-based recommendations, where the user starts with a tag-based query or an example of a specific liked item, which can be thought of as query-by-example. The user's expectation is to see a ranked list of recommended items that are similar to the query. The system achieves this by (re-)ranking approximate matches to the query item (i.e., similarity search neighbors), and the ranking quality must be measured in experiments. To be concrete, in the evaluation step of recommender systems, instead of drawing the pool of unlabeled items uniformly at random, we propose two new strategies. We either draw them with respect to the user's genre/tag/category distribution, or by running BM25 on the document corpus with the positive test item text as the query.

In many recommendation settings, only positive samples are available for training. In the absence (or extreme scarcity) of explicitly negative labels, recommenders treat all unlabeled items as negative and sample them uniformly for training (and test, as discussed above). Furthermore, adjusting the evaluation setting with respect to the mentioned criteria: disjoint authors per user and search-based mode, would To enable our recommender architecture to cope with search-based situations and to aim for predictions that are truly eye-opening for

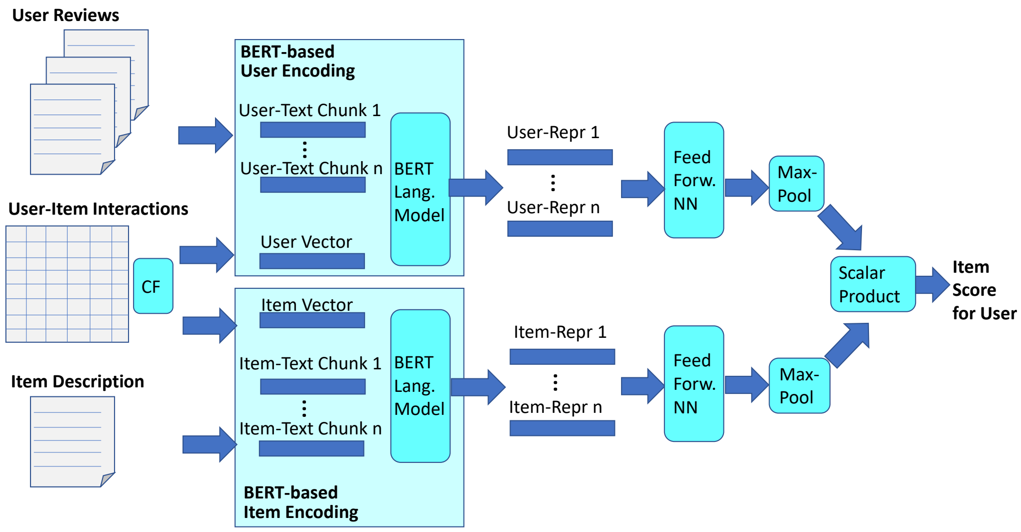



Figure 31.11: Recommender System Architecture

the user, we devise new techniques for generating negative training samples. The prevalent approach then is to randomly draw negative samples uniformly from unlabeled data points. We propose two new techniques for sampling training points, both emulating a search-based recommendation scenario. One approach is to obtain negative training points by drawing them from the unlabeled points that have a category/genre tag in common with one of the points in the user's training item. This reflects the user's general taste without knowing anything about the actual user queries at test-time. The expected effect is to obtain negative training samples that are closer to the user interest and thus more difficult to discriminate. Another approach is to clone each sampled unlabeled item: a positive item weighted by its relatedness to the user, and a negative item with the complement weight. The weight is calculated by averaging the item-item similarity between the sampled item and user's positive training items. The intuition for this technique, inspired by [1], is to improve learning by having more informatively labeled negative samples even if these carry only partial weight by their lower confidence in their labels.

We devise a general framework for recommender systems, which allows us to run a variety of methods by specific configurations. We focus on a Transformer architecture that couples BERT-based encoders for items and users with a (pre-computed) latent matrix for interaction signals (if available). On top, we place feed-forward layers with max-pooling for supervised learning. The prediction scores for user-item pairs yield the per-user ranking of items. This architecture can take as input a spectrum of user and item information including user reviews and item descriptions and also different kinds of meta-data. Figure 31.11 depicts this architectural framework.

Our experiment show the need for more research on such underexplored cases of difficult evaluation. They also indicate that our smarter sampling techniques for training-time negatives are a promising direction [4].

**Recommendations for Long Tail Users and Items:** The success of the recommender systems is rooted in the ample amount of user- item interaction data. In this work, we study the underexplored case of long-tail users and long-tail items, using two book communities as experimental data.

We extend the framework explained above (Figure 31.11, by additional techniques to cope with expressive text from user reviews to select the most informative cues. User text, especially the reviews are an interleaved mix of descriptive elements (e.g., "the unusual murder weapon"), sentiment expressions (e.g., "it was fun to read") and irrelevant filler material (e.g., "I read only on weekends"), and only the first aspect is helpful for content-based profiling (as the sentiment is already in the interaction data). Second, the total text length often exceeds the bound for a chunk that can be fed into large language models such as BERT. Aggregation via max-pooling is all but a perfect solution, as it cannot preserve all relevant signals. To overcome these issues, we devised three techniques for selecting the most informative parts of user text (as appose to randomly selecting sentences) to fill one or more bounded-length chunks: *Weighted Phrases*: sorting n-grams by their tf-idf weights, *Weighted Sentences*: sorting sentences by their normalized accumulated idf weights, and *Similar Sentences*: sorting sentences with respect to their similarity with the item description using Sentence-BERT [2].

We are interested in reviewing the recommendation performance on data-poor scenarios and for long-tail users with sparse data. To this end we sample the datasets into different flavours is terms of data sparsity by controlling the item's degree in the sampled user-item graph. And for the evaluation of the system, we group the users with respect to their data scarcity. We observe that in data-poor regime, the collaborative-signal is not sufficient and incorporating item and user text is crucial.

## References

[1] J. Bekker and J. Davis. Learning from positive and unlabeled data: a survey. *Mach. Learn.*, 109(4):719–760, 2020.

[2] N. Reimers and I. Gurevych. Sentence-bert: Sentence embeddings using siamese bert-networks. In *2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP, Nov 3-7, 2019*, 2019, pp. 3980–3990. Association for Computational Linguistics.

[3] F. Ricci, L. Rokach, and B. Shapira, eds. *Recommender Systems Handbook*. Springer US, 2022.

• [4] G. H. Torbati, G. Weikum, and A. Yates. Search-based recommendation: The case for difficult predictions. In Y. Ding, J. Tang, J. Sequeda, L. Aroyo, C. Castillo, and G.-J. Houben, eds., *The ACM Web Conference 2023 (WWW 2023)*, Austin, TX, USA, 2023, pp. 318–321. ACM.

[5] G. H. Torbati, A. Yates, and G. Weikum. Personalized entity search by sparse and scrutable user profiles. In H. O'Brain and L. Freund, eds., *CHIIR '20, Fifth ACM SIGIR Conference on Human Information Interaction and Retrieval*, Vancouver, BC, Canada, 2020, pp. 427–431. ACM.

• [6] G. H. Torbati, A. Yates, and G. Weikum. You get what you chat: Using conversations to personalize search-based recommendations. In D. Hiemstra, M.-F. Moens, J. Mothe, R. Perego, M. Potthast, and F. Sebastiani, eds., *Advances in Information Retrieval (ECIR 2021)*, Lucca, Italy (Online Event), 2021, LNCS 12656, pp. 207–223. Springer.

### 31.5.3 Information Extraction from Conversations

*Investigators: Anna Tigunova, Andrew Yates, Paramita Mirza, Gerhard Weikum*

**Motivation and problem statement.** Intelligent assistants have become an essential part of life, being able to deal with a wide range of routine tasks and provide factual information. Still, there is a growing demand for creating user-oriented chatbot agents, which are able to hold personalized conversations. To avoid the necessity for the user to provide their information, the agent should be able to learn the personal facts directly from their utterances. Thus, we are interested in automated methods to learn personal facts from dialogues.

Personal facts span from sociological data (age or profession), interests, skills, relationships with other people to sentiments and personality. Inferring personal facts from conversations can not be directly addressed by existing Information Extraction (IE) methods. Prior work on IE generally focuses on formal texts, such as news articles and Wikipedia pages. The colloquial nature of dialogues provides an additional challenge: the utterances are short, noisy and colloquial.

We investigate the prediction of values for personal attributes (e.g., predicting value *swimming* for the subject's attribute *hobby*). We devised three neural models for personal attribute prediction: HAM [5], CHARM [6] and PRIDE [4]. We also created novel datasets supporting our research [2]. This line of research is comprehensively covered in Anna Tigunova's doctoral disseration [1].

**Hidden Attribute Models (HAMs).** Conversational data contains a lot of noise, many words and whole utterances providing no valuable information. Moreover utterances considered outside of their context are often useless.

Following these observations we propose *Hidden Attribute Models* (HAMs) [5]. Given a set of input utterances $U$, each one consisting of terms $t_1, \ldots, t_M$, HAMs work in three steps:

1. Create a representation of the utterance, considering the importance of terms: $R_i^{utt} = f_{utt}(t_1, t_2, \ldots, t_M)$
2. Create a representation of the speaker from all their utterances, considering their importance: $R^{sp} = f_{sp}(R_1^{utt}, \ldots, R_N^{utt})$
3. Given the speaker's representation, score the attribute values: $a_k = f_{obj}(R^{sp})$

$R_i^{utt}$ and $R^{sp}$ are represented by attention mechanisms, which assign importance to the components of both speaker and utterance representations. As prediction HAM returns the value with the highest score $a_k$ from the fully-connected layer acting as $f_{obj}$.

We conducted extensive experiments on predicting such personal attributes as *profession*, *hobby*, *age* and *family status*, which showed that HAM outperforms the baseline models for user profiling.

However, HAM's predictions are restricted to a short list of values, which do not cover all possibilities for long-tailed personal attributes. To be able to handle unlimited value lists we introduce a new retrieval-based model, CHARM.

**Conversational Hidden Attribute Models (CHARMs)** [6] make a step towards making predictions for personal attributes with very large number of values (like hobbies). To aid the prediction of values not represented in the training data, CHARM links input utterances to ane external document collection.

CHARM's pipeline consists of two steps: *keyword extraction* and *document ranking.*

1. **Keyword extraction.** In the first stage, CHARM uses a *term scoring model* to evaluate terms in input utterances. The score of a term indicates how useful the term is for making a prediction. The top scoring $K$ terms are selected to form the query for the document ranking stage. The selected terms should be descriptive of the predicted attribute in general, instead of its particular values, to allow *zero-shot* learning of the attribute values (unseen during training).

2. **Document ranking.** In the second stage, CHARM uses a web document collection where each document is automatically associated with several attribute values (for example, the document collection can consist of all leaf pages from Wikipedia's list of hobbies). The *retrieval model* assigns relevance scores to each document based on the query obtained from the keyword extraction step.

   As each attribute value can have several retrieved documents associated with it, the document relevance scores are aggregated to form the final scores for attribute values. Finally, the top scoring attribute values are returned as predictions.

Getting the predictions for the attribute values by utilizing an external document collection makes CHARM flexible for the large and open-ended lists of attribute values. Having trained CHARM once, we only need to update the document collection to enable CHARM to predict new values. Our experiments show that CHARM outperforms all baselines in the zero-shot setup (sets of training and test attribute values are disjoint).

Additionally, we created a demonstration platform, which allows users to explore CHARM's predictions on their own conversational data [3].

**Predicting Relationships in Conversations (PRIDE).** In this setting, we aim to automatically extract interpersonal relationships of conversation interlocutors, so as to enrich personal knowledge bases towards enhancing personalized search, recommenders and chatbots. To infer speakers' relationships from dialogues we propose PRIDE, a neural multi-label classifier, based on BERT and Transformer for creating a conversation representation [4]. PRIDE utilizes dialogue structure and augments it with external knowledge about speaker features and conversation style. Unlike prior works, we address multi-label prediction of fine-grained relationships (e.g., co-worker, classmate, teacher/student, doctor/patient etc.).

We release large-scale datasets, based on screenplays of movies and TV shows, with directed relationships of conversation participants. Extensive experiments on both datasets show superior performance of PRIDE compared to the state-of-the-art baselines.

## References

• [1] A. Tigunova. *Extracting Personal Information from Conversations.* PhD thesis, Universität des Saarlandes, Saarbrücken, 2022.

[2] A. Tigunova, P. Mirza, A. Yates, and G. Weikum. RedDust: a large reusable dataset of Reddit user traits. In N. Calzolari, F. Béchet, P. Blache, K. Choukri, C. Cieri, T. Declerck, S. Goggi, J. Mariani, H. Mazo, A. Moreno, J. Odiik, and S. Piperidis, eds., *Twelfth Language Resources and Evaluation Conference (LREC 2020)*, Marseille, France, 2020, pp. 6118–6126. ELRA.

• [3] A. Tigunova, P. Mirza, A. Yates, and G. Weikum. Exploring personal knowledge extraction from conversations with CHARM. In L. Lewin-Eytan, D. Carmel, E. Yom-Tov, E. Agichtein, and E. Gabrilovich, eds., *WSDM '21, 14th International Conference on Web Search and Data Mining*, Virtual Event, Israel, 2021, pp. 1077–1080. ACM.

• [4] A. Tigunova, P. Mirza, A. Yates, and G. Weikum. PRIDE: Predicting relationships in conversations. In M.-F. Moens, X. Huang, L. Specia, and S. W.-t. Yih, eds., *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP 2021)*, Punta Cana, Dominican Republic, 2021, pp. 4636–4650. ACL.

[5] A. Tigunova, A. Yates, P. Mirza, and G. Weikum. Listening between the lines: Learning personal attributes from conversations. In J. McAuley, ed., *Proceedings of The World Wide Web Conference (WWW 2019)*, San Francisco, CA, USA, 2019, pp. 1818–1828. ACM.

[6] A. Tigunova, A. Yates, P. Mirza, and G. Weikum. CHARM: Inferring personal attributes from conversations. In B. Webber, T. Cohn, Y. He, and Y. Liu, eds., *The 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP 2020)*, Online, 2020, pp. 5391–5404. ACM.

### 31.5.4 Quantity Search

*Investigators: Vinh Thinh Ho, Koninika Pal, Yusra Ibrahim, Simon Razniewski, Klaus Berberich, Gerhard Weikum, Dragan Milchevski (external), Daria Stepanova (external), Jannik Stroetgen (external)*

**Introduction.** Quantities are more than numeric values. They denote physical, technical, financial and other measures for entities, expressed in numbers with associated units. Search queries often include quantities, such as "athletes who ran 100m under 9.9 seconds" or "Internet companies with quarterly revenue above $2 Billion". Such queries with filter conditions on quantities are an important building block for downstream analytics, and pose challenges when the content of interest is spread across a huge number of ad-hoc data sources on the internet.

Processing such queries requires understanding the quantities, where capturing the surrounding context is an essential part of it. Although modern search engines or QA systems handle entity-centric queries well, they consider numbers and units as simple keywords, and therefore fail to understand the condition (less than, above, etc.), the unit of interest (seconds, dollar, etc.), and the context of the quantity (100m race, quarterly revenue, etc.). As a result, search engines often fail for quantity queries with filter conditions.

The following examples illustrate this point.

| Query | Expected Results(s) |
|---|---|
| 1: richest business woman in Europe | Francoise Bettencourt (L'Oreal) |
| 2: net worth of Francoise Bettencourt | 76.1 Bio USD |
| 3: European business women worth more than 500 Mio Euro | Susanne Klatten (BMW), Charlene de Carvalho (Heineken) ... |
| 4: sprinters who ran 100 m under 9.9 s | Usain Bolt, Tyson Gay, Yohan Blake, Justin Gatlin ... |

Queries 1 and 2 are answered properly by major search engines and QA assistants, but they are just simple lookup queries without filter conditions for quantities. For query 3, search engines return “ten blue links” to web pages of mostly poor quality: results include pages such as “EUR 500m EIB backing for youth and female focused”, “Opening Statement Ursula von der Leyen European Parliament”, “List of Europeans by net worth – Wikipedia”, “The German Mittelstand: Facts and figures about ... – BMWi”. Likewise, query 4 lacks proper interpretation and query evaluation; results include pages with titles like “10-second barrier - Wikipedia”, “100 metres – Wikipedia”, “Ultimate 100-Meter Time: 9.27 Seconds? | Runner’s World”. Some of these contain proper answers, but the user would have to browse through them rather than receiving entity lists – not exactly what we expect from a modern search engine with a huge knowledge graph as a back-end asset. Open knowledge graphs like Wikidata do not help much either: they are very rich in terms of entities, but very sparsely populated as far as quantitative properties are concerned.

In this project, we develop methods for extracting and contextualizing entity-quantity facts from text and tables in web pages, in order to support quantity querying and extracting entity-quantity facts to augment knowledge bases [2]. Ultimately, we aim to support high-stake information needs by advanced users, such as analysts, journalists, scientists and other knowledge workers [10]. Ideally, an analyst would run her entire data analysis over web contents as easily as posing a keyword query or single-sentence question:

- Which runners have complete 10 or more marathons under 2:10:00?
- Which is the top-10 most energy-efficient hybrid car models?
- How does the carbon footprint of Japanese cars compare to US-made cars?
- Which vaccinations have which coverage in the 50 population-wise largest countries?

**Qsearch: Answering Quantity Queries from Text.** To tap into text sources, we developed Qsearch [3, 4], based on three main components. First, entity-quantity pairs and essential context cues are extracted from text corpora, such as Wikipedia articles or web crawls, by training an LSTM neural network combined with a CRF. The output of this sequence tagging are triples $(e, q, X)$ – *(entity, quantity, context)*, called *Qfacts*. For instance, from the text

*“BMW i8 has a price of 138k Euros in Germany and its range on battery is 50 to 60 km”*

we extract two Qfacts:

(*BMW i8*; *€138,000*; *price, Germany* ) and (*BMW i8*; *50-60 km*; *range, battery* ).

Extracted Qfacts are organized and indexed in a data repository, where entities are linked to their semantic types from a knowledge base. Second, a query is decomposed into *entity type*, *quantity condition* and *context* cue words, which are also cast into a triple $(t^*, q^*, X^*)$, called *Qquery*. For example, the user input

*“hybrid cars with electric range above 60 km”*

is modeled as a Qquery:

(*hybrid car*; *>60 km*; *electric, range* ).

The query processing matches the Qquery against the extracted Qfacts. Third, the query results are judiciously ranked using language models, with embeddings for the relatedness of cue words (e.g., matching “electric” in the query against “battery” in a Qfact).

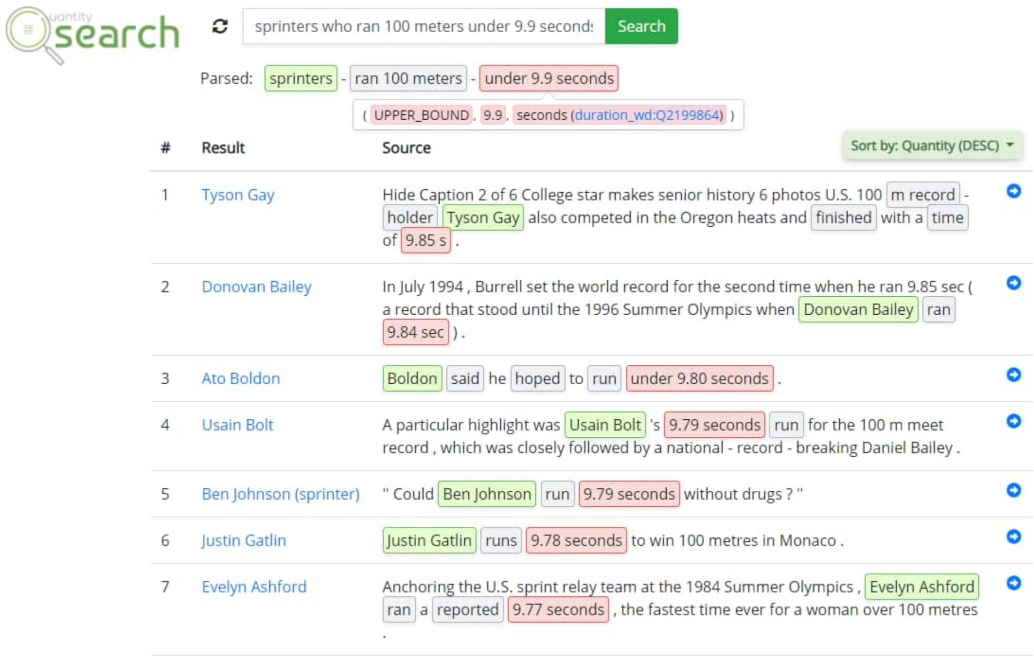



Figure 31.12: Illustration of Qsearch

Qsearch has been published as a full paper at ISWC 2019 [3], and has been demonstrated at WSDM 2020 [4]. The system is accessible at `https://qsearch.mpi-inf.mpg.de/`. Figure 31.12 shows a screenshot with top results for the query "sprinters who ran 100 meters under 9.9 seconds". Note that these contain some spurious answers, which underlines the difficulty of the task.

**QuTE: Answering Quantity Queries from Web Tables.** Many relevant Qfacts about entities are not expressed in textual web contents or are too difficult to extract. Therefore, we extend the scope of quantity extraction and search to a different kind of input modality: ad-hoc *web tables*, embedded in HTML pages or in the form of spreasheets and open data (in RDF, JSON etc.). Starting with the seminal work of [8, 9], extracting knowledge from web tables and bringing this content into the realm of search has been studied intensively [1, 11]. Nevertheless, queries with quantity filters are largely underexplored.

To take advantage of the semi-structured characteristics of web tables, we developed the QuTE system [5, 6]. The two main stages of QuTE are the extraction of Qfacts from web tables, carried out as an offline computation, and the online query processing. The first stage comprises components for entity linking, quantity detection and normalization, column alignment, and Qfact contextualization. The second stage involves matching and ranking of Qfacts against user queries, largely following Qsearch but with enhancements for the corroboration of candidate results.

The most important step in the QuTE architecture is *column alignment*, which maps each quantity column its its proper entity column, in order to extract Qfacts from the right column pairs. The problem and its difficulty are illustrated by Table 31.3. As the example data

| Team | Stadium | Capacity | Coach | Value (in Bio) |
|---|---|---|---|---|
| Bayern | Allianz Arena | ca. 75000 | Hansi Flick | 2.549 Euro |
| Real | Bernabéu | 81,044 | Zidane | 3.649 Euro |
| Man City | unknown | n/a | Pep Guardiola | 2.055 GBP |
| Chelsea | Stamford Bridge | 40,834 | Frank Lampard | 1.958 GBP |
| Liverpool | Anfield | 53,394 | Jürgen Klopp | ca. 1.7 GBP |

Table 31.3: Web table example on football teams.

contains three different entity columns, it is a-priori unclear to which column the quantity columns Capacity and Value refer. Simple heuristics like leftmost or nearest-left fail here: leftmost would incorrectly align Capacity with Team, and nearest-left would erroneously align Value with Coach. Prior work with supervised classifiers or information-theoretic inference are not robustly working either for these kinds of complex tables.

Column alignment plays a decisive role extracting Qfacts. The key novelty of our method is a smart way of assessing candidate alignments by inspecting the Qfacts that would result from a given alignment. To this end, we compute Qfact confidence scores via evidence from large text collections, with type-based inference to overcome sparseness problems [5].

Extracted Qfacts are contextualized using cues from the table itself and from surrounding elements: table caption, table column headers, same-row values, page title, DOM-tree headings and text passages preceding and following the table. The query processing component takes advantage of these contextualizations for better matching and more accurate answers. Moreover, we developed a new method based on consistency learning for corroborating candidate facts at query time, re-ranking them, and pruning false positives.

QuTE has been published as a full paper at WWW 2021 [5], and a system demonstration was given at SIGMOD 2021 [6]. QuTE is accessible at https://qsearch.mpi-inf.mpg.de/table/. Figure 31.13 shows a screenshot with top results for the query "sprinters who ran 100 meters under 9.9 seconds".

**Enhancing Knowledge Bases with Quantity Facts.** Knowledge bases (KBs) about the world's entities should include quantity properties, such as heights of buildings, running times of athletes, energy efficiency of car models, energy production of power plants, and more. State-of-the-art KBs, such as Wikidata, cover many relevant entities but often miss the corresponding quantities. Prior work on extracting quantity facts from web contents focused on high precision for top-ranked outputs, but did not tackle the KB coverage issue. This work devises a recall-oriented approach to narrow this gap in KB coverage [7].

Our method is based on iterative learning for extracting quantity facts, with two novel contributions to boost recall for KB augmentation without sacrificing the quality standards of the knowledge base. The first contribution is a query expansion technique to capture a larger pool of fact candidates. The second contribution is a novel technique for harnessing observations on value distributions for self-consistency. Experiments with extractions from more than 13 million web documents demonstrate the benefits of our method

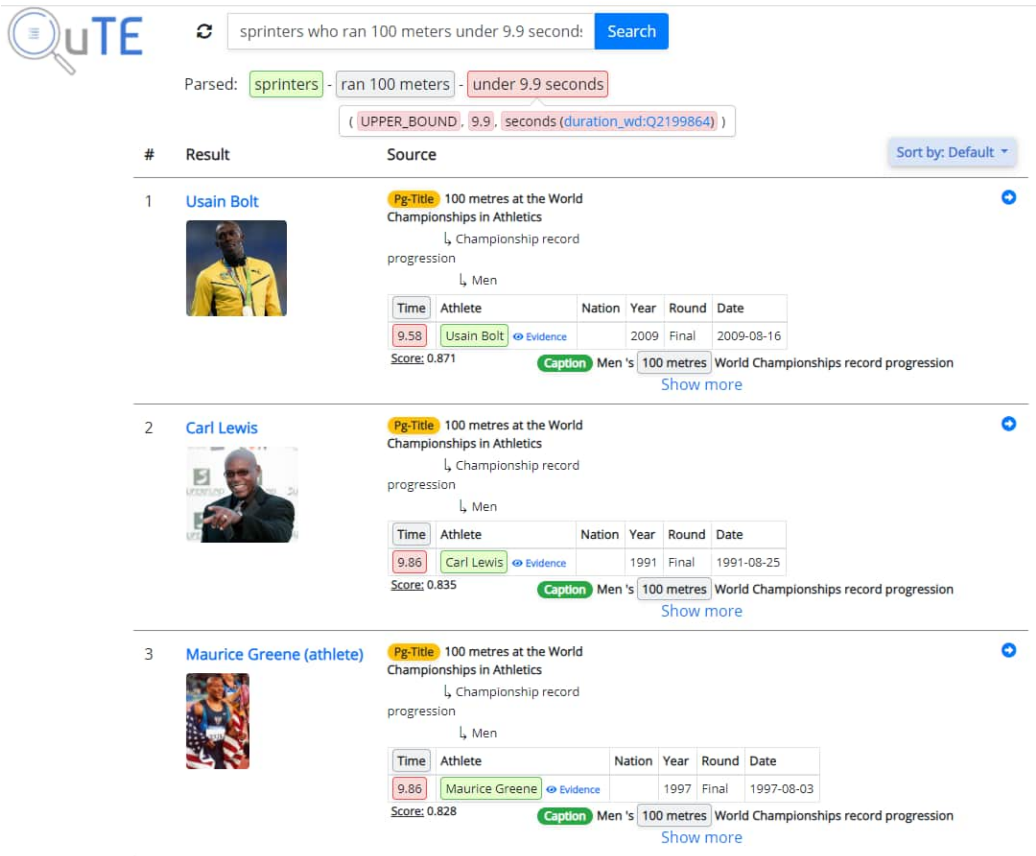



Figure 31.13: Illustration of QuTE

## References


[1] M. J. Cafarella, A. Y. Halevy, H. Lee, J. Madhavan, C. Yu, D. Z. Wang, and E. Wu. Ten years of webtables. *Proc. VLDB Endow.*, 11(12):2140–2149, 2018.

• [2] V. T. Ho. *Entities with quantities: extraction, search, and ranking.* PhD thesis, Saarland University, Saarbrücken, Germany, 2022.

[3] V. T. Ho, Y. Ibrahim, K. Pal, K. Berberich, and G. Weikum. Qsearch: Answering quantity queries from text. In C. Ghidini, O. Hartig, M. Maleshkova, V. Svátek, I. Cruz, A. Hogan, J. Song, and M. Lefrançois, eds., *The Semantic Web – ISWC 2019*, Auckland, New Zealand, 2019, LNCS 11778, pp. 237–257. Springer.

[4] V. T. Ho, K. Pal, N. Kleer, K. Berberich, and G. Weikum. Entities with quantities: Extraction, search, and ranking. In J. Caverlee and X. B. Hu, eds., *WSDM '20, 13th International Conference on Web Search and Data Mining*, Houston, TX, USA, 2020, pp. 833–836. ACM.

• [5] V. T. Ho, K. Pal, S. Razniewski, K. Berberich, and G. Weikum. Extracting contextualized quantity facts from web tables. In J. Leskovec, M. Grobelnik, M. Najork, J. Tang, and L. Zia, eds., *The Web Conference 2021 (WWW 2021)*, Ljubljana, Slovenia, 2021, pp. 4033–4042. ACM.

• [6] V. T. Ho, K. Pal, and G. Weikum. QuTE: Answering quantity queries from web tables. In G. Li, Z. Li, S. Idreos, and D. Srivastava, eds., *SIGMOD '21, International Conference on Management of Data*, Xi'an, Shaanxi, China, 2021, pp. 2740–2744. ACM.

• [7] V. T. Ho, D. Stepanova, D. Milchevski, J. Strötgen, and G. Weikum. Enhancing knowledge bases with quantity facts. In F. Laforest, R. Troncy, E. Simperl, D. Agarwal, A. Gionis, I. Herman, and L. Médini, eds., *WWW '22, ACM Web Conference*, Virtual Event, Lyon, France, 2022, pp. 893–901. ACM.

[8] G. Limaye, S. Sarawagi, and S. Chakrabarti. Annotating and searching web tables using entities, types and relationships. *Proc. VLDB Endow.*, 3(1):1338–1347, 2010.

[9] S. Sarawagi and S. Chakrabarti. Open-domain quantity queries on web tables: annotation, response, and consensus models. In S. A. Macskassy, C. Perlich, J. Leskovec, W. Wang, and R. Ghani, eds., *The 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '14, New York, NY, USA - August 24–27, 2014*, 2014, pp. 711–720. ACM.

[10] G. Weikum. Entities with quantities. *Bulletin of the Technical Committee on Data Engineering*, 43(1):4–8, 2020.

[11] S. Zhang and K. Balog. Web table extraction, retrieval, and augmentation: A survey. *ACM Trans. Intell. Syst. Technol.*, 11(2):13:1–13:35, 2020.


# 31.6 Question Answering

*Coordinator: Rishiraj Saha Roy*

Question answering (QA) is one of the most active topics across the fields of artificial intelligence (AI), natural language processing (NLP), and information retrieval (IR) today. In the context of MPI-INF Department D5, QA is among the most prominent applications of large and curated knowledge bases (KBs) like YAGO, DBpedia, Wikidata and proprietary ones in enterprises. Question answering, in full generality, covers a very broad range of tasks. So let us set the right scope of our own research with the following example of a multi-turn conversation with one question and one answer in each turn ($q_i$'s and $a_i$'s represent questions and answers, respectively):

$q_1$: *When was the first Marvel movie released in Germany?*
$a_1$: Iron Man
$q_2$: *No, the film's year of release..?*
$a_2$: 2008
$q_3$: *Next in the Marvel Series?*
$a_3$: The Incredible Hulk
$q_4$: *Where in US is the Hulk storyline set?*
$a_4$: Virginia
$q_5$: *I meant where in Virginia?*
$a_5$: Culver University

To summarize, we deal with questions that have an objective and fact-based answer (typically a small set of entities or strings, usually of set size one), often referred to as factoid or factual questions. The above-mentioned dialogue-like QA setting has become an expected mode of information access with most digital assistants today, like Siri, Alexa, Cortana or the Google Assistant. The research challenges that emerge from such a scenario are manifold: (i) recognizing and disambiguating entities in the question (`Marvel Cinematic Universe` and `Germany` in $q_1$); (ii) dealing with complex information needs involving multiple entities, relations, and temporal conditions (two entities in $q_1$ and $q_4$; temporal conditions in $q1 - q3$); (iii) tackling a conversational setup with implicit context ($q_2 - q_5$ have incomplete specification of user intent); (iv) leveraging question reformulations ($q_2$ and $q_5$ are reformulations of $q_1$ and $q_4$, respectively); and (v) harnessing heterogeneous information sources ($q_1, q_2$ are best answerable by a KB; $q_3$ by a Web table or Wikipedia infobox; and $q_4, q_5$ by text from the Web). These themes precisely correspond to the problems addressed by our research group in the last two years, and are presented in the following sections. The developed methods are designed to be *explainable* to end-users, *efficient* in time and memory consumption, and *robust* with respect to variations in benchmarks and domains.

### 31.6.1 Named Entity Recognition and Disambiguation

*Investigators: Johannes Hoffart (external), Dragan Milchevski (external), Luciano del Corro (external), Ghazaleh Torbati, Gerhard Weikum*

Named entity recognition and disambiguation, NER and NED or NERD for short, is a key pillar for many tasks in natural language understanding, including information extraction, fact checking and question answering. It involves recognizing entity mentions in textual contents and disambiguating them by linking mentions to entities in a knowledge base.

AmbiverseNLU is a toolkit comprising a suite of methods from our group's research over the past decade, including named entity recognition, entity disambiguation (also referred to as entity linking), concept linking, open information extraction, relational fact extraction and ranking. Its original core is the AIDA software for entity linking, developed by [1]. This publication is among the highest cited papers on named entity disambiguation (with more than 1,200 citations according to Google Scholar, as of February 2023).

AmbiverseNLU has been made available as open source[3]. The software is widely used, for our own ongoing research as well as by other research groups worldwide. AmbiverseNLU can be used as a standalone natural language processing toolkit, as a Web Service, or directly for end users as a Web-based platform. See Figure 31.14 for an example run of AmbiverseNLU's NERD output on a natural language sentence.

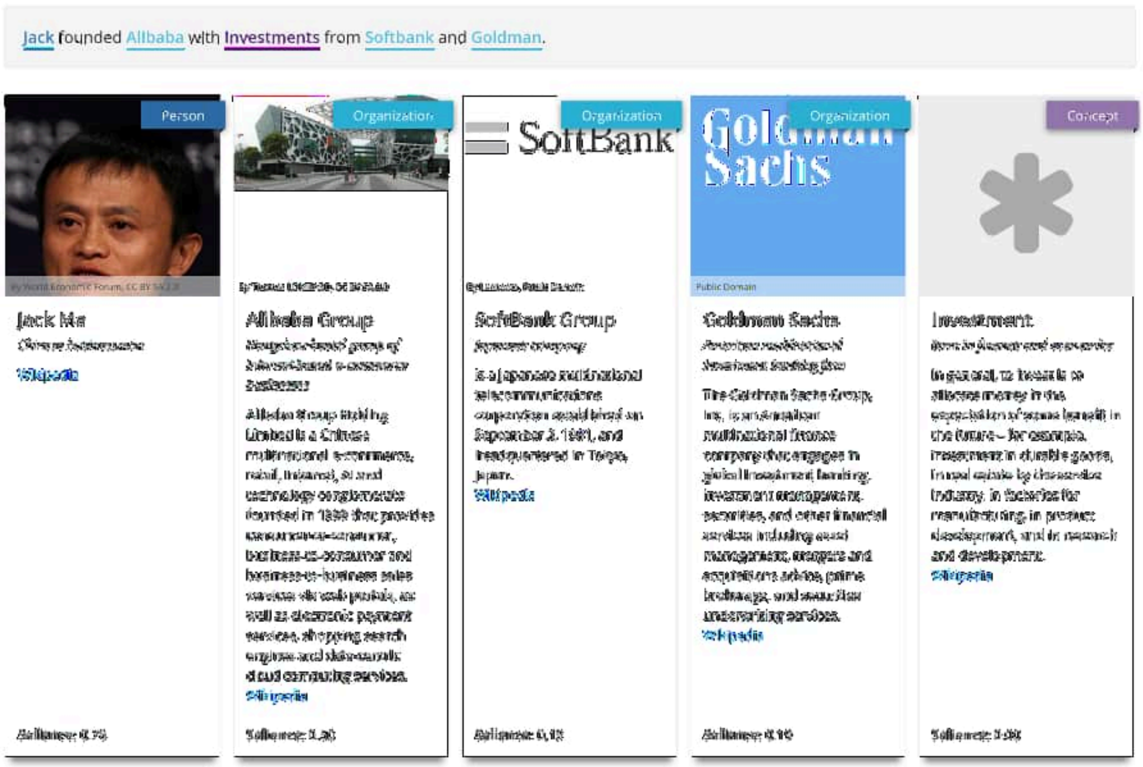



Figure 31.14: Web-based platform providing AmbiverseNLU functionality.

## References

[1] J. Hoffart, M. A. Yosef, I. Bordino, H. Fürstenau, M. Pinkal, M. Spaniol, B. Taneva, S. Thater, and G. Weikum. Robust disambiguation of named entities in text. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, Edinburgh, Scotland, UK, 2011, pp. 793–803. The Association for Computational Linguistics.

[3]https://ambiversenlu.mpi-inf.mpg.de/

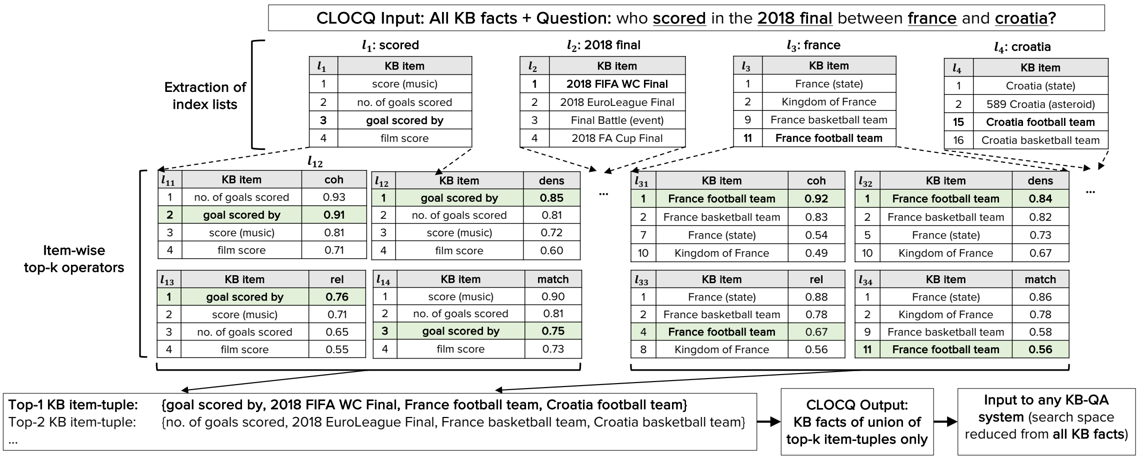



Figure 31.15: An overview of the CLOCQ algorithm.

## 31.6.2 Complex Question Answering

*Investigators: Philipp Christmann, Rishiraj Saha Roy, Gerhard Weikum, Soumajit Pramanik (external), Zhen Jia (external)*

### Search space reduction for complex question answering

One of the first tasks in complex question answering is to reduce the volume of the space to search an answer in. This volume may be measured by the number of KB (equivalently, KG) facts, Web documents, tables, and so on. In question answering over knowledge bases or knowledge graphs (KB-QA/KG-QA), systems often prune the search space for candidate answers using NERD methods like AmbiverseNLU, which was presented in Section 31.6.1. Such NERD methods map mentions in questions (single words or short phrases) to KB entities, and the QA system subsequently uses only the KB facts containing these entities as its *search space* for locating the answer(s). However, general-purpose NERD tools have major limitations in this context: (i) they are not tailored for downstream use by KB-QA systems; (ii) they usually disambiguate only named entities and disregard words and phrases that denote general concepts, types or predicates; and (iii) they typically output merely the top-1 entity per mention, missing out on further candidates that can serve as relevant cues. Even methods designed for short input texts, like TAGME [4] and ELQ [7], have such limitations.

To address these concerns, we propose CLOCQ [2, 1] (Contracting answer spaces with Lists and top-$k$ Operators for Complex QA, pronounced "Clock"), a time- and memory-efficient method that operates over *all KB items* to produce top-$k$ candidates for entities, types, concepts, and predicates. Consider the following complex question on the FIFA Football World Cup 2018: *Who scored in the 2018 final between France and Croatia?*

Most systems for complex KB-QA tackle the answering process in two phases. First, they disambiguate question tokens to KB entities. These entities *establish a reduced search space* for the QA system, that can either be an *explicit* set of facts containing these KB entities or involve *implicit* grounding to a small zone in the KB via structured queries containing these entities. Second, depending upon the approach in the first phase, KB-QA systems either search for the answer in the retrieved facts, or build a complex query in SPARQL-like syntax that would return the answer when executed [10]. CLOCQ tries to improve the effectiveness and the efficiency of the first phase above. The output of CLOCQ is thus a small set of disambiguated KB items and facts containing these items, and this is fed into the answering phase. Answer presence in the KB subspace inherently sets an upper bound on the performance of the downstream KB-QA system, making fast and effective search space reduction a vital step in the QA pipeline.

CLOCQ first builds inverted lists of KB items per question word with *term matching scores* based on TF-IDF. Top-ranked items from these lists, up to a certain depth, are then scored and ranked by a combination of *global signals*, like semantic coherence between items and connectivity in the KB graph, and *local signals* like relatedness to the question and term-matching score. These scoring signals are computed at question time: this is made feasible with CLOCQ's novel KB representation and storage model, that substantially speeds up lookups with respect to existing solutions.

The threshold algorithm (TA) [3] is used for extracting the top-$k$ candidates for each question term separately. Since it may not always be obvious how to choose $k$ for every term, we propose an entropy-based mechanism for making this choice automatically. The union of the per-term top-$k$ items forms a pool of relevant KB items, and their KB facts is the output of CLOCQ that would be passed on to the answering phase of a KB-QA system. Figure 31.15 gives an overview of the CLOCQ algorithm. An API for CLOCQ is available at `https://clocq.mpi-inf.mpg.de`. It has been accessed already more than 20 million times by external users within a year of making it public.

### Answering complex questions over heterogeneous sources

While KGs capture a large part of the world's encyclopedic knowledge, they are inherently incomplete. This is because they cannot stay up-to-date with the latest information, so that emerging and ephemeral facts (e.g., teams losing in semi-finals of sports leagues, or celebrities dating each other) are not included. Also, user interests go way beyond the predicates that are modeled in KGs like Wikidata. As a result, answering over text from the open Web, like websites of newspapers or magazines, is an absolute necessity.

The fragmented landscape of QA research over KGs (KG-QA) and text (Text-QA) has led to a state where methods from one paradigm are completely incompatible with the other. Systems operating over knowledge graphs are not equipped at all to compute answers from text passages. The reason behind this is two-fold. First, QA over KGs assumes one of two major paradigms: (i) an explicit structured query (in a SPARQL-like language) is constructed that, when executed over the KG, retrieves the answer. Such structured queries cannot be executed over a textual source; and (ii) they perform approximate graph search without an explicit query, using graph algorithms, graph embeddings, or graph convolutions. The basic idea here is to learn to traverse the graph in symbolic or latent space, and these cannot work over text inputs as there is no natural underlying graph representation.

On the other hand, powerful deep learning methods for text-QA are not geared at all for tapping into KGs and other RDF data. This is because these algorithms are basically classifiers trained to predict the start and end position of the most likely answer span in a given piece of text, a sequence-based setup that does not agree with RDF triples.

Works on *heterogeneous QA* incorporate text (typically articles in Wikipedia or search results) as a supplement to RDF triples, or the other way around. However, usually KG and text are combined merely to improve the rankings of candidate answers through additional evidence, but the answers themselves come either from the KG alone or from text alone. Moreover, none of these methods is geared to handle the extreme, but realistic situations, where the input is either only a KG or only a text corpus. State-of-the-art methods in heterogeneous QA, GRAFT-Net [12] and PullNet [11], suffer from shortcomings that involve an indispensable reliance on the KG source. Specifically, they require trained KG entity embeddings, KG entity-linked text, shortest paths in KGs for creating training data, and can only provide such KG entities as answers. Thus, they cannot operate over arbitrary text corpora with an open vocabulary where ad-hoc phrases could be answers. A very recent method, UniK-QA [8], takes an approach of unifying sources via verbalization of every knowledge evidence, but still cannot deal with more complex information needs.

Problems are particularly exacerbated for complex questions with multiple entities and predicates (setting for this work), like: *director of the western for which Leo won an Oscar?* Here, SPARQL queries become particularly brittle for KG-QA systems, graph search becomes more complicated, or text evidence for Text-QA systems has to be joined from more than one document. For example, the correct query for our running example would be something like: `SELECT ?x WHERE {?y director ?x . ?y castMember LeonardoDiCaprio . ?y genre western . LeonardoDiCaprio awarded AcademyAward [forWork ?y] .}`

Hardly any KG-QA system would get this mapping onto KG predicates and entities perfectly right. Even though graph-search methods do not formulate explicit queries, they still need to get the mapping onto KB entities and predicates right. For Text-QA, computing answers requires stitching together information from several texts, as it is not easy to find a single Web page that contains all relevant cues in the same zone.

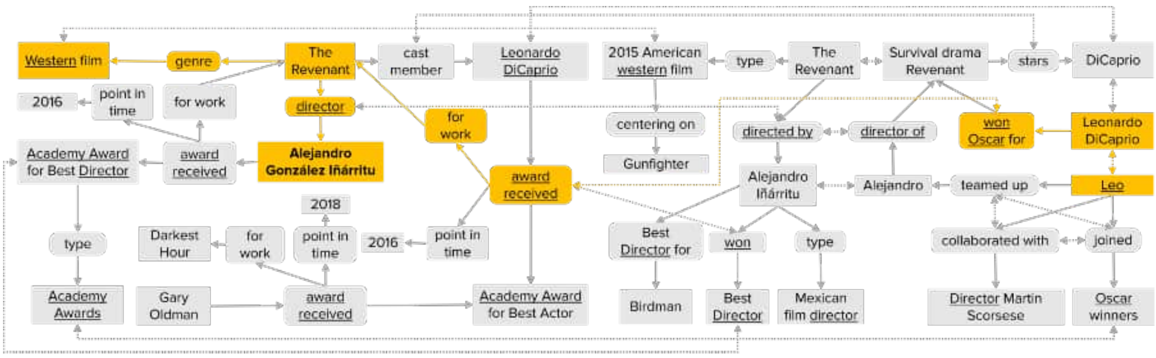



Figure 31.16: Context graph (XG) built by UNIQORN in the heterogeneous answering setup for the question $q$ = *director of the western for which Leo won an Oscar?* Anchors are nodes with (partly) underlined labels; answers are in bold. Orange subgraphs are Group Steiner Trees.

To overcome these limitations, we propose UNIQORN, a Unified framework for question answering over RDF knowledge graphs and Natural language text, that addresses these limitations [9]. Our proposal hinges on two key ideas:

- Instead of attempting to compute perfect translations of questions into crisp SPARQL queries (which seems elusive for complex questions), we relax the search strategy over KGs by devising graph algorithms for Group Steiner Trees, this way connecting question-relevant cues from which candidate answers can be extracted and ranked.
- Handling KG-QA, Text-QA and heterogeneous setups with the same unified method, by building a noisy KG-like context graph from KG or text inputs on-the-fly for each question, using named entity recognition and disambiguation (NERD), open information extraction (Open IE) and BERT fine-tuning.

In a nutshell, UNIQORN works as follows. Given an input question, we first perform NERD on the question using the AmbiverseNLU software and then retrieve question-relevant *evidences* from one or more knowledge sources via CLOCQ [2], using a fine-tuned BERT model. From these evidences, that are either KG facts or text fragments, UNIQORN constructs a *context graph (XG)* that contains question-specific entities, predicates, types, and candidate answers. Depending upon the input source, this XG thus either consists of: (i) KG facts defining the neighborhood of the question entities, or, (ii) a quasi-KG dynamically built by joining Open IE triples extracted from text snippets, or, (iii) both, in case of retrieval from heterogeneous sources. Triples in the XG originate from evidences that are deemed question-relevant by a fine-tuned BERT model. We identify *anchor nodes* in the XG that match phrases in the question. Treating the anchors as terminals, Group Steiner Trees (GST) are computed that contain candidate answers. These GSTs establish a joint context for disambiguating mentions of entities, predicates, and types in the question. Candidate answers are ranked by simple statistical measures rewarding redundancy within the top-$k$ GSTs. Figure 31.16 illustrates this unified approach for heterogeneous QA. UNIQORN thus belongs to the family of methods that locate an answer using graph search and traversal, and does not build an explicit structured query [10].

### Complex question answering with temporal intent

While complex KG-QA in general has been a major topic, little attention has been paid to the special case of *temporal questions*. Such questions involve explicit or implicit notions of constraining answers by associated timestamps in the KG. This spans a spectrum, starting from simpler cases such as *when was obama born?*, *where did obama live in 2001?*, and *where did obama live during 9/11?*, to more complex temporal questions like: *where did obama's children study when he became president?*

Complex questions must consider multi-hop constraints (`Barack Obama` `child` `Malia Obama, Sasha Obama` `educated at` `Sidwell Friends School`), and reason on the overlap of the intersection of time points and intervals (the start of the presidency in 2009 with the study period at the school, 2009 – 2016). A simplified excerpt of the relevant zone in the Wikidata KG necessary for answering the question, is shown in Figure 31.17. This research addresses these challenges that arise for complex temporal questions.

Early works on temporal QA over unstructured text sources involve various forms of question and document parsing, but do not carry over to KGs with structured facts comprised of entities and predicates. The few works specifically geared for time-aware QA over KGs (like [5]) uses a small set of hand-crafted rules for question decomposition and temporal

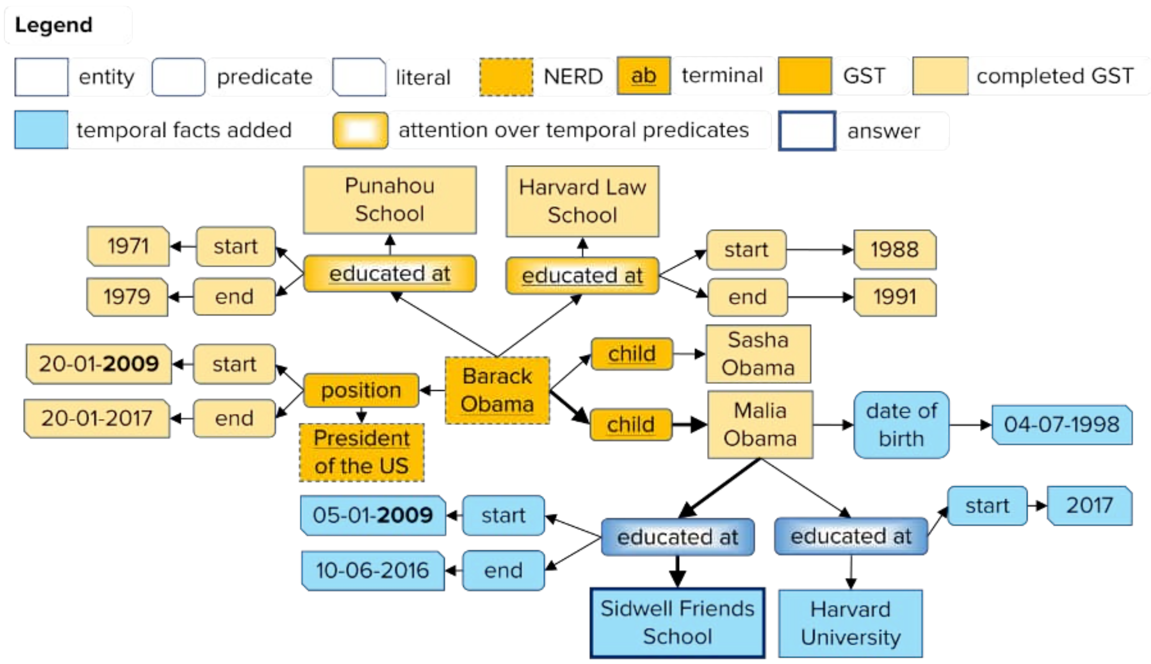



Figure 31.17: Wikidata excerpt showing the relevant KG zone for *where did obama's children study when he became president?* with answer `Sidwell Friends School`.

reasoning. This approach needs human experts for the rules and does not cope with complex questions. This research presents EXAQT: EXplainable Answering of complex Questions with Temporal intent, a system that does not rely on manual rules for question understanding and reasoning [6]. EXAQT answers complex temporal questions in two steps:

– Identifying a compact, tractable *answer graph* that contains all cues required for answering the question, based on dense-subgraph algorithms and fine-tuned BERT models (building on our own work, the UNIQORN model [9] desribed earlier);

– A *relational graph convolutional network (R-GCN)* [12] to infer the answer in the graph, augmented with signals about time.

The two stages work as follows (partly illustrated in Figure 31.17). EXAQT first fetches all KG facts of entities mentioned in the question (`Barack Obama, President of the United States`: dashed outline boxes) using our in-house toolkits AmbiverseNLU and CLOCQ. The resulting noisy set of facts is distilled into a tractable set by means of a fine-tuned BERT model (admitting information about the children Malia and Sasha, but not Michelle Obama). To construct a KG subgraph of all question-relevant KG items and their interconnections from this set, Group Steiner Trees (GST) are *computed* (dark orange nodes, terminals or keyword matches underlined: "obama", "president", "child", "educated at" and *completed* (light orange nodes). The last and decisive step at this point augments this candidate answer graph with pertinent *temporal facts*, to bring in cues (potentially multiple hops away from the question entities) about relevant dates, events and time-related predicates. To this end, we use an analogous BERT model for identifying question-relevant temporal facts (blue nodes:

educational affiliations of Malia and Sasha and their dates). The resulting *answer graph* is the input of the second stage.

In the second stage, inspired by the popular GRAFT-Net model [12] and related work, we construct an R-GCN that learns entity embeddings over the answer graph and casts answer prediction into a node classification task. However, R-GCNs as used in prior works are ignorant of temporal constraints. To overcome this obstacle, we augment the R-GCN with time-aware entity embeddings, attention over temporal relations, and encodings of timestamps, temporal signals, and temporal question categories. In our running example, temporal attention helps EXAQT focus on `educated at` as a question-relevant relation (partly shaded nodes). The time-enhanced representation of `Barack Obama` flows through the R-GCN (thick edges) and boosts the likelihood of `Sidwell Friends School` as the answer (node with thick borders), which contains 2009 (in bold) among its temporal facts. By producing such concise KG snippets for each question (as colored in Figure 31.17), EXAQT yields explainable evidence for its answers.

## References


• [1] P. Christmann, R. S. Roy, and G. Weikum. CLOCQ: A toolkit for fast and easy access to knowledge bases. In *Datenbanksysteme für Business, Technologie und Web (BTW 2023), 20. Fachtagung des GI-Fachbereichs „Datenbanken und Informationssysteme“ (DBIS), 06.-10. März 2023, Dresden, Germany, Proceedings*, 2023, LNI P-331, pp. 579–591. Gesellschaft für Informatik e.V.

• [2] P. Christmann, R. Saha Roy, and G. Weikum. Beyond NED: Fast and effective search space reduction for complex question answering over knowledge bases. In *WSDM '22, Fifteenth ACM International Conference on Web Search and Data Mining*, Tempe, AZ, USA (Virutal Event), 2022, pp. 172–180. ACM.

[3] R. Fagin, A. Lotem, and M. Naor. Optimal aggregation algorithms for middleware. *Journal of computer and system sciences*, 66(4), 2003.

[4] P. Ferragina and U. Scaiella. TAGME: On-the-fly annotation of short text fragments (by Wikipedia entities). In *CIKM*, 2010.

[5] Z. Jia, A. Abujabal, R. Saha Roy, J. Strötgen, and G. Weikum. TEQUILA: Temporal question answering over knowledge bases. In A. Cuzzocrea, J. Allan, N. Paton, D. Srivastava, R. Agrawal, A. Broder, M. Zaki, S. Candan, A. Labrinidis, A. Schuster, and H. Wang, eds., *CIKM'18, 27th ACM International Conference on Information and Knowledge Management*, Torino, Italy, 2018, pp. 1807–1810. ACM.

• [6] Z. Jia, S. Pramanik, R. Saha Roy, and G. Weikum. Complex temporal question answering on knowledge graphs. In G. Demartini, G. Zuccon, J. S. Culpepper, Z. Huang, and H. Tong, eds., *CIKM '21, 30th ACM International Conference on Information & Knowledge Management*, Virtual Event, Australia, 2021, pp. 792–802. ACM.

[7] B. Z. Li, S. Min, S. Iyer, Y. Mehdad, and W.-t. Yih. Efficient one-pass end-to-end entity linking for questions. In *EMNLP*, 2020.

[8] B. Oguz, X. Chen, V. Karpukhin, S. Peshterliev, D. Okhonko, M. Schlichtkrull, S. Gupta, Y. Mehdad, and S. Yih. UniK-QA: Unified representations of structured and unstructured knowledge for open-domain question answering. In *NAACL*, 2022.

• [9] S. Pramanik, J. Alabi, R. Saha Roy, and G. Weikum. *UNIQORN: Unified Question Answering over RDF Knowledge Graphs and Natural Language Text*, 2021. arXiv: 2108.08614.

• [10] R. Saha Roy and A. Anand. *Question Answering for the Curated Web: Tasks and Methods in QA over Knowledge Bases and Text Collections.* Synthesis Lectures on Information Concepts, Retrieval, and Services. Morgan & Claypool, San Rafael, CA, 2021.

[11] H. Sun, T. Bedrax-Weiss, and W. Cohen. PullNet: Open domain question answering with iterative retrieval on knowledge bases and text. In *EMNLP*, 2019.

[12] H. Sun, B. Dhingra, M. Zaheer, K. Mazaitis, R. Salakhutdinov, and W. Cohen. Open domain question answering using early fusion of knowledge bases and text. In *EMNLP*, 2018.


### 31.6.3 Conversational Question Answering

*Investigators: Magdalena Kaiser, Philipp Christmann, Rishiraj Saha Roy, Gerhard Weikum*

#### Learning from question reformulations

Our first contribution on learning from reformulations for conversational question answering (ConvQA) over KGs is CONQUER (Conversational Question Answering with Reformulations) [6]. CONQUER applies reinforcement learning (RL) to learn from question reformulations which serve as implicit feedback signals that the previous system response was unsatisfying. In this way, it can learn directly from a stream of user interactions without having access to gold answers, that are not readily available in reality. This is a key step towards learning in a more natural setting, and is a major deviation from current benchmark-driven static setups.

Due to the incompleteness arising from the ad hoc conversational style, questions like *What was the next from Marvel?* (question $q_{21}$ in the example below) are hard to answer in one shot. However, users can often recognize wrong system responses based on the expected answer type (e.g. a person instead of a movie), or their domain knowledge. In this case, they can provide light-weight implicit feedback by reformulating their question (a different phrasing of the same intent). Below is a snippet from a conversation consisting of such a sequence of new intents, reformulations and system responses:

$q_{21}$: *What was the next from Marvel?* (**New intent**)
$a_{21}$: `Stan Lee` (**Wrong answer**)
$q_{22}$: *What came next in the series?* (**Reformulation**)
$a_{22}$: `Marvel Cinematic Universe` (**Wrong answer**)
$q_{23}$: *The following movie in the Marvel series?* (**Reformulation**)
$a_{23}$: `Spider-Man: Far from Home` (**Correct answer**)
$q_{31}$: *Released on?* (**New intent**)

Figure 31.18 shows an excerpt of the KG relevant to answer $q_{21}$. Entities (as well as literals and types) are depicted in red, and predicates are in blue. Given the current (say $q_{21}$) and the previous utterances (say, *When was Avengers Endgame released in Germany?*), CONQUER creates and maintains a set of *context entities* from the KG that are the most relevant to the conversation so far (*Avengers Endgame*, *Marvel Cinematic Universe*, *Captain Marvel*). These are starting points for RL *agents*, and are shown in thick-bordered boxes in the figure.

Inspired by works on KG reasoning, like [5], an agent now walks over the KG to entities in its neighborhood. The destination entities of this walk are candidate answers for the current question, that are aggregated to produce the final answer (*Spider-Man: Far from Home*, shaded in the figure).

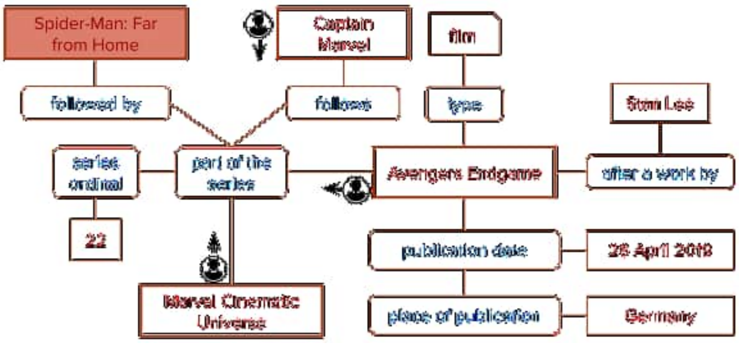



Figure 31.18: KG excerpt with RL agents and their chosen walking directions.

The goal is to learn a similarity function between questions and KG facts so that given an input question, the agent can select the KG edge that most likely leads to an answer. Since it is intractable to learn transition probabilities for all outgoing edges for each context entity and question due to an infinite number of possibilities, the agent learns a *parameterized policy* where parameters are estimated using a neural network called the *policy network*. The input to the network consists of the encodings of the current (and previous) utterances and the KG facts (*actions*) involving the context entity. The output is a probability distribution over the available actions from which the agent *samples* an action (during training) or takes the *top action* (at answering time). We apply the policy-gradient algorithm REINFORCE [12] to update the weights of the policy network.

These updates depend on the received reward: if the next user utterance has a new intent, the agent has most likely selected the correct action and returned the correct answer. Therefore, it receives a positive reward (+1). If a reformulation is observed as next utterance (indicating a wrong system response), it receives a negative reward (-1) instead. We have fine-tuned a BERT model which predicts whether two questions have the same intent or not. Note that this reward is noisy since the predictor can make mistakes, and the user might not behave as expected (e.g. asking a different question with a new intent regardless of whether the previous answer was correct). Yet, CONQUER is very robust to noise and is able to significantly outperform the state-of-the-art.

Since there is no public dataset available that contains question reformulations, we conducted a user study to create such a resource, termed ConvRef. More precisely, ConvRef was created by enhancing our ConvQuestions benchmark [2] by collecting up to four reformulations for each question. These reformulations were generated by real users while interacting with a baseline QA model created by randomly initializing CONQUER's policy network. ConvRef contains about 11k conversations with more than 200k turns in total, out of which 205k are reformulations.

In CONQUER, we proposed learning from users' reformulations since reformulations by average users are easier to collect than expert-labeled gold answers [6, 9]. However, access to

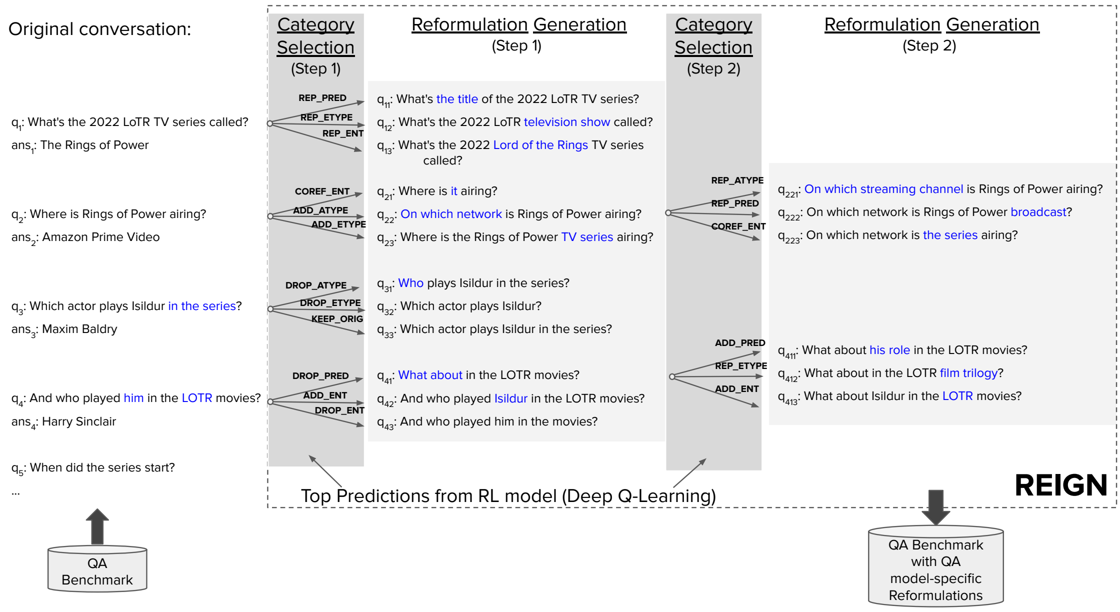



Figure 31.19: Learning from generated reformulations with REIGN.

large numbers of real users is not easily feasible (outside of big search-engine companies). Moreover, a typical user is probably not too keen on spending time for many reformulations of a given intent. Another concern is that humans, when issuing reformulations, do not know upfront which variations would guide the QA model towards better training. These issues make training ConvQA models from reformulations at scale very challenging.

Thus, in our second contribution that makes use of reformulations for training, we create large volumes of system-generated reformulations by means of a novel taxonomy. We use a reinforcement learning model to judiciously select a set of *helpful* reformulations to enable more robust ConvQA training.

Automatic generation in absence of large volumes of training data requires an understanding of the basic underlying categories in the input. Along these lines, our first proposal is a functional taxonomy of question reformulations, given two constraints: (i) the utterances are potentially incomplete w.r.t. an information need, and (ii) the questions and answers need to be mapped to and derived from a large KG. Inspired by string edit distances, we posit that such transformations belong to four major operations: (i) insertion, (ii) substitution, (iii) deletion, and (iv) coreferencing (specific to ConvQA). These operations can be applied to four constituents of the user utterance (operands): (i) entities, (ii) relations, (iii) question entity types, and (iv) answer entity types. Removing unrealistic $\langle operator, operand \rangle$ pairs from consideration (coreferencing for relations and types) and adding an identity operation that keeps the original question, we end up with 14 transformation rules in this framework.

These rules are either directly applied to questions from a ConvQA benchmark, or used to fine-tune a seq2seq transformer model, to generate a large number of previously unseen reformulations. While the goal is to improve the training of a specific ConvQA model, it is unlikely that all generated variations would lead to an improved performance. Therefore, given a QA model, we select the more helpful variants for each training instance using an RL approach, inspired by [1]. The RL model (Deep Q-Networks, or DQN [7]) is guided using intrinsic and extrinsic rewards corresponding to answering performance. We refer to this principle as *guided reformulation generation*. Our model receives the encoded conversational question (state representation) as input and outputs the so-called Q-values $Q(s, a)$ for each state-action pair to quantify the usefulness of taking action $a$ in state $s$ under a policy $\pi$. We propose two state representations: (i) encoding the questions using BERT, and (ii) creating a one-hot encoding based on the presence/absence of key elements in the question (i.e. entity, predicate, and type information). The actions consists of 14 reformulation categories corresponding to our 14 transformation rules. To make the most of the available training data, our method includes both single-step and multi-step transformations of the original questions. Finally, the original data is augmented with the reformulations most helpful for the given QA model (predictions from the DQN). This new data is then used to train the ConvQA system. This model-aware data augmentation framework is referred to as REIGN (=REInforced GeNeration). An overview of our REIGN framework is depicted in Figure 31.19. A sample conversation is shown to exemplify the question transformations (based on our taxonomy) that are considered in REIGN.

### Answering over heterogeneous sources

State-of-the-art works on ConvQA make use of single information sources: either curated knowledge bases (KB), or a text corpus, or a set of Web tables, but only one of these. This can inherently limit the answer coverage of ConvQA systems. Consider the following conversation as an example:

$q_1$: *Who wrote the book Angels and Demons?*
$a_1$: Dan Brown
$q_2$: *the main character in his books?*
$a_2$: Robert Langdon
$q_3$: *who plays him in the films?*
$a_3$: Tom Hanks
$q_4$: *to which headquarters was robert flown in the book?*
$a_4$: CERN
$q_5$: *and the city was?*
$a_5$: Meyrin
$q_6$: *how long is the novel?*
$a_6$: 768 pages

Some of these questions can be answered using a KB ($q_1$, $q_3$, $q_5$), or Web tables ($q_1$, $q_3$, $q_6$) as they ask about salient information about entities, and some via text sources ($q_2$, $q_3$, $q_4$) as they are more likely to be contained in content and discussion. However, none of these sources represents the whole information required to answer *all questions* of this conversation. A smart ConvQA system should, therefore, be able to tap into more than one kind of knowledge repository, to improve answer recall and to boost answer confidence by leveraging multiple kinds of evidence across sources.

Existing research on ConvQA has considered solely one kind of information source for deriving answers. Further, when specializing on a given source, methods often adopt source-specific design choices that do not generalize well. For example, representations of the conversational context, like KB subgraphs or text passages, are often specifically modeled for the knowledge repository at hand, which is incompatible with heterogeneous sources. Methods for question rewriting and question resolution convert short user utterances into full-fledged questions making the question intent *explicit* [10, 11]. However, this adds major complexity and may lose valuable cues from the conversation flow. Further, these methods face evidence retrieval problems arising from long and potentially verbose questions. There has been work on single-question QA over heterogeneous sources trying to deal with KBs, text, and tables [8]. Their approach is designed for simple questions, though, and cannot easily be extended to the challenging ConvQA setting. Finally, none of the prior works on ConvQA produce human-interpretable structures that could assist end users in case of erroneous system responses.

To overcome these limitations, we propose CONVINSE [3] (CONVQA with INtermediate Representations on Heterogeneous Sources for Explainability), an end-to-end framework for conversational QA on a mixture of sources. CONVINSE consists of three main stages:

- *Question Understanding (QU)*,
- *Evidence Retrieval (ER)*, and
- *Heterogeneous Answering (HA)*.

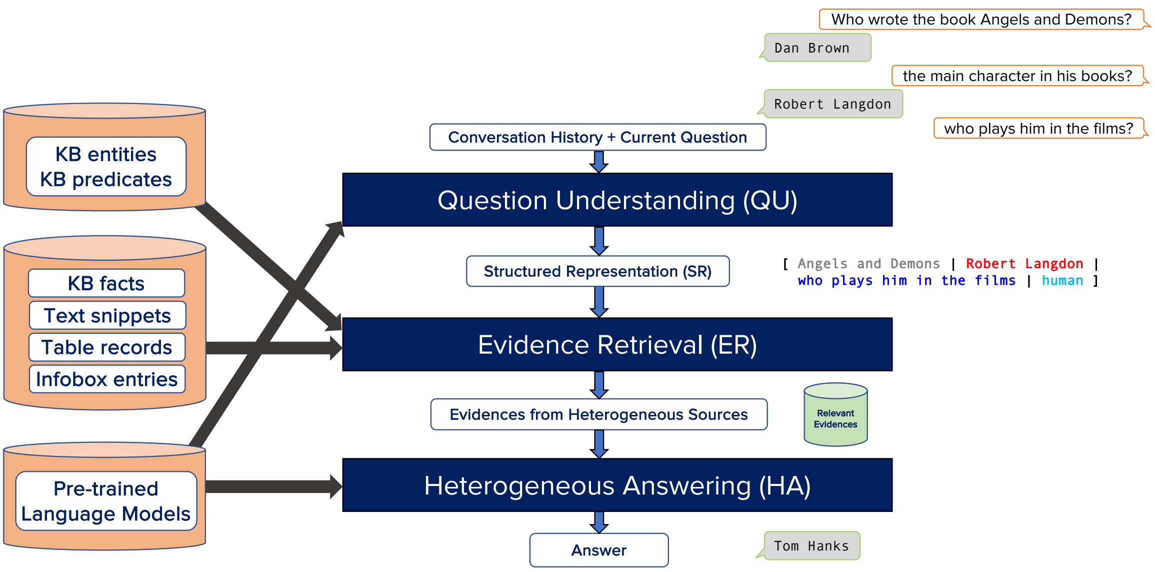



Figure 31.20: An overview of the CONVINSE pipeline, illustrating the three main stages of the approach, the input sources to these stages (on the left), and the respective results (on the right). The output is the answer and corresponding explanation, consisting of derived intermediate results and explaining evidences.

Figure 31.20 gives an overview of the pipeline. The first stage, QU, addresses the challenges of incomplete user utterances introduced by the conversational setting. We derive an *intent-explicit structured representation (SR)* that captures the complete information need. SRs are frame-like structures for a question that contain designated slots for open-vocabulary lexical representations of entities in the conversational context (marked gray) and the current question (red), relational predicates (blue), and expected answer types (cyan). An example SR for $q_3$ of the running example is: `[Angels and Demons|Robert Langdon|who plays him in the films|human]`. SRs can be viewed as concise gists of user intents, intended to be in a form independent of any specific answering source. They are self-contained interpretable representations of the user's information need, and are inferred using fine-tuned transformer models trained on data generated by distant supervision from plain sequences of QA pairs.

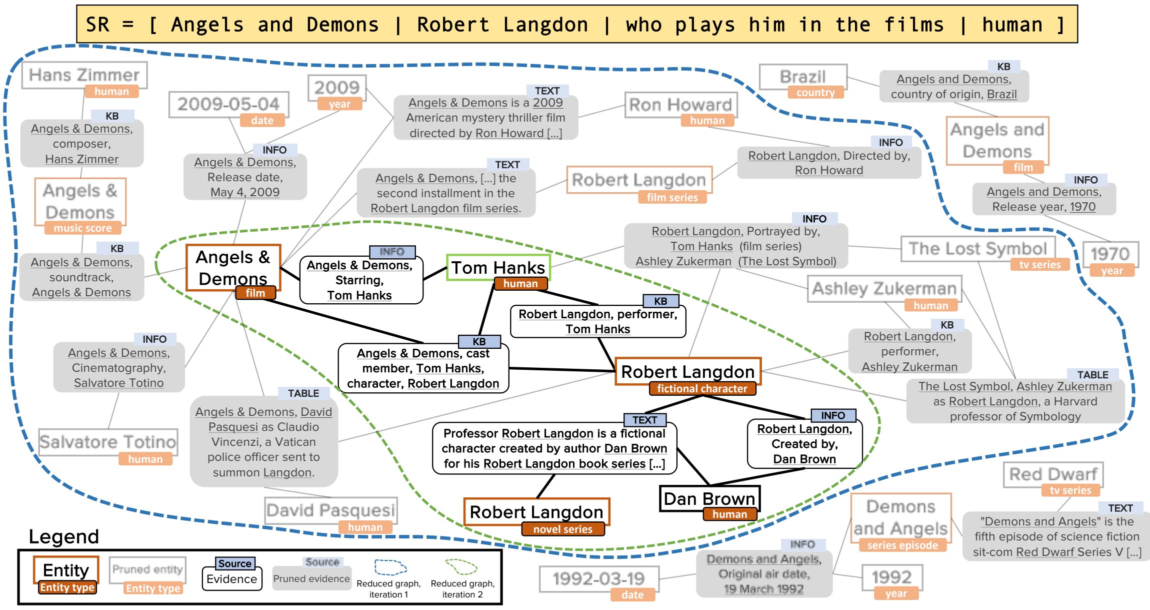



Figure 31.21: An excerpt of the heterogeneous graph for answering $q_3$. The graph is iteratively reduced, by GNN inference, to compute the answer and key evidences. The subgraph surrounded by the blue dotted line is the result of the first iteration, the green dotted line indicates the graph after the second iteration. From this smaller subgraph, the final answer (`Tom Hanks`) is inferred.

The second stage, ER, exploits the CLOCQ-algorithm outlined in Section 31.6.2 to judiciously retrieve question-relevant evidences (KB facts, text sentences, table records, or infobox entries) from each information source. CLOCQ identifies and disambiguates KB entities, and retrieves relevant KB facts for a given SR. For the disambiguated KB entities, the corresponding Wikipedia page is looked up, to extract text sentences, table records and infobox entries. These heterogeneous evidences are verbalized on-the-fly, to obtain a homogeneous set of textual evidences. The top-$k$ pieces of evidence are passed on to the answering stage.

The third and final stage, HA, aims to obtain the answer from these evidences, leveraging their relationships and connections. We first construct a heterogeneous graph from the retrieval outputs, consisting of nodes of two different types: evidences and KB entities. Within this graph, an edge runs between an evidence and a KB entity, if the KB entity is mentioned within the evidence. An excerpt of such a graph can be seen in Figure 31.21. We then iteratively apply dedicated graph neural networks (GNNs) on the graph for computing the best answers and supporting evidences in a small number of steps. A key novelty is that each iteration reduces the graph in size, and only the final iteration yields the answer and a small user-comprehensible set of explanatory evidences [4].

To evaluate the pipeline, we construct ConvMix, the first benchmark for ConvQA over heterogeneous sources. ConvMix is a crowdsourced dataset that contains questions with answers emanating from the Wikidata KB, the full text of Wikipedia articles, and the collection of Wikipedia tables and infoboxes. ConvMix contains 2800 conversations with five turns (14k utterances), and 200 conversations with ten turns (2k utterances), their gold answers and respective knowledge sources for answering. Conversations are accompanied by metadata like entity annotations, completed questions, and paraphrases.

## References

[1] C. Buck, J. Bulian, M. Ciaramita, W. Gajewski, A. Gesmundo, N. Houlsby, and W. Wang. Ask the right questions: Active question reformulation with reinforcement learning. In *ICLR*, 2018.

[2] P. Christmann, R. Saha Roy, A. Abujabal, J. Singh, and G. Weikum. Look before you hop: Conversational question answering over knowledge graphs using judicious context expansion. In W. Zhu and D. Tao, eds., *CIKM '19, 28th ACM International Conference on Information and Knowledge Management*, Beijing China, 2019, pp. 729–738. ACM.

• [3] P. Christmann, R. Saha Roy, and G. Weikum. Conversational question answering on heterogeneous sources. In E. Amigo, P. Castells, J. Gonzalo, B. Carterett, J. S. Culpepper, and G. Kazai, eds., *SIGIR '22, 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, Madrid, Spain, 2022, pp. 144–154. ACM.

• [4] P. Christmann, R. Saha Roy, and G. Weikum. Explainable conversational question answering over heterogeneous sources via iterative graph neural networks. In *SIGIR '23, 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*, Taipei, Taiwan, 2023. ACM. Accepted.

[5] R. Das, S. Dhuliawala, M. Zaheer, L. Vilnis, I. Durugkar, A. Krishnamurthy, A. Smola, and A. McCallum. Go for a walk and arrive at the answer: Reasoning over paths in knowledge bases using reinforcement learning. In *ICLR*, 2018.

• [6] M. Kaiser, R. Saha Roy, and G. Weikum. Reinforcement learning from reformulations in conversational question answering over knowledge graphs. In F. Diaz, C. Shah, T. Suel, P. Castells, R. Jones, T. Sakai, A. Bellogín, and M. Yushioka, eds., *SIGIR '21, 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, Virtual Event, Canada, 2021, pp. 459–469. ACM.

[7] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, et al. Human-level control through deep reinforcement learning. *Nature*, 518(7540), 2015.

[8] B. Oguz, X. Chen, V. Karpukhin, S. Peshterliev, D. Okhonko, M. Schlichtkrull, S. Gupta, Y. Mehdad, and S. Yih. UniK-QA: Unified representations of structured and unstructured knowledge for open-domain question answering. In *NAACL*, 2022.

[9] P. Ponnusamy, A. R. Ghias, C. Guo, and R. Sarikaya. Feedback-based self-learning in large-scale conversational AI agents. In *IAAI (AAAI Workshop)*, 2020.

[10] S. Vakulenko, S. Longpre, Z. Tu, and R. Anantha. Question rewriting for conversational question answering. In *WSDM*, 2021.

[11] N. Voskarides, D. Li, P. Ren, E. Kanoulas, and M. de Rijke. Query resolution for conversational search with limited supervision. In *SIGIR*, 2020.

[12] R. J. Williams. Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine learning*, 8(3-4), 1992.

## 31.7 Responsible Data Science

*Coordinator: Gerhard Weikum*

Humans leave more and more digital traces in online systems, both laterally across a variety of platforms and longitudinally over several years. This rich online information is leveraged by platform providers to learn user profiles and enhance algorithmic services from recommenders, filters, classifiers and rankers all the way to largely automating decision-making processes such as loan requests, visa applications or invitations for job interviews. As a consequence, users are faced with increasing risks of losing privacy, receiving unwanted visibility and being treated in an unfair and potentially discriminating way in online systems. Another frequently arising concern is that users are puzzled with recommended items, lacking user-comprehensible and faithful explanations. In simple cases, this creates confusion about predictions for potentially liked products, but more serious issues arise with recommendations of politically polarizing news or social-media discussions.

Starting in 2015, we have addressed a suite of sub-topics in this broad space of societally responsible data science. Initially, we focused on trustworthiness of online information [9, 10, 8], user-controllable privacy and its tensions with trust and utility [5, 3, 4], and fairness in algorithmic predictions [2, 4, 7, 6].

Our recent work along these lines have mostly focused on transparency and explainability.

Section 31.7.1 addresses the controllability and scrutability of machine-learning models, coping with deployment-time risks such as data shifts. Section 31.7.2 presents our work on user-comprehensible explainability, with focus on recommender systems. Section 31.7.3 investigates properties and patterns of online discussion forums, as a major domain where bias and polarization can be problematic. Section 31.7.4 connects methods from data mining and machine learning, to enhance the interpretability of descriptive and predictive models.

Much of this work is carried out in the context of the ERC Synergy Grant imPACT (2015-2022) on Privacy, Accountability, Compliance and Trust. Our results contribute to trust, transparency and explainability of models and methods for the paradigm of *responsible data science* [1, 11].

### References

[1] S. Abiteboul, G. Miklau, J. Stoyanovich, and G. Weikum, eds. *Data, Responsibly (Dagstuhl Seminar 16291)*, Wadern, Germany, 2016, Dagstuhl Reports 6. Schloss Dagstuhl.

[2] A. J. Biega, K. P. Gummadi, and G. Weikum. Equity of attention: Amortizing individual fairness in rankings. In *SIGIR'18, 41st International ACM SIGIR Conference on Research and Development in Information Retrieval*, Ann Arbor, MI, USA, 2018, pp. 405–414. ACM.

[3] A. J. Biega, R. Saha Roy, and G. Weikum. Privacy through solidarity: A user-utility-preserving framework to counter profiling. In *SIGIR'17, 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*, Shinjuku, Tokyo, Japan, 2017, pp. 675–684. ACM.

[4] J. A. Biega. *Enhancing Privacy and Fairness in Search Systems.* PhD thesis, Universität des Saarlandes, Saarbrücken, 2019.

[5] J. A. Biega, K. P. Gummadi, I. Mele, D. Milchevski, C. Tryfonopoulos, and G. Weikum. R-susceptibility: An IR-centric approach to assessing privacy risks for users in online communities. In *SIGIR'16, 39th International ACM SIGIR Conference on Research and Development in Information Retrieval*, Pisa, Italy, 2016, pp. 365–374. ACM.

• [6] P. Lahoti. *Operationalizing Fairness for Responsible Machine Learning.* PhD thesis, Universität des Saarlandes, Saarbrücken, 2022.

[7] P. Lahoti, K. Gummadi, and G. Weikum. Operationalizing individual fairness with pairwise fair representations. *Proceedings of the VLDB Endowment (Proc. VLDB)*, 13(4):506–518, 2019.

[8] K. Popat. *Credibility Analysis of Textual Claimswith Explainable Evidence.* PhD thesis, Universität des Saarlandes, Saarbrücken, 2019.

[9] K. Popat, S. Mukherjee, J. Strötgen, and G. Weikum. Where the truth lies: Explaining the credibility of emerging claims on the Web and social media. In *WWW'17 Companion*, Perth, Australia, 2017, pp. 1003–1012. ACM.

[10] K. Popat, S. Mukherjee, A. Yates, and G. Weikum. DeClarE: Debunking fake news and false claims using evidence-aware deep learning. In E. Riloff, D. Chiang, J. Hockenmaier, and T. Jun'ichi, eds., *The Conference on Empirical Methods in Natural Language Processing (EMNLP 2018)*, Brussels, Belgium, 2018, pp. 22–32. ACL.

[11] J. Stoyanovich, B. Howe, and H. V. Jagadish. Responsible data management. *Proc. VLDB Endow.*, 13(12):3474–3488, 2020.

### 31.7.1 Fair and Responsible Machine Learning

*Investigators: Preethi Lahoti, Gerhard Weikum, Krishna Gummadi (external)*

Over the past years, we have witnessed growing concerns about the unintended societal impact ofmachine learning ( ML) systems. This research focuses on developing models and techniques for responsible development and deployment of ML systems. A crucial component of building responsible ML systems is that the system's predictions are *fair*, and do not adversely affect individuals based on user's demographic or otherwise sensitive attributes (such as gender, ethnic group etc.). That is, they can be safely applied to make predictions for individuals from all demographic groups, without undue risk of discriminating certain groups.

**Achieving Group Fairness Without Access to Protected Attributes** Most of the prior works on fairness assume that protected/sensitive attributes such as race and gender are specified upfront and that the model has access to these attributes to mitigate unfairness, typically by enforcing group fairness constraints in the ML optimization objective [2, 11]. In practice, however, factors like privacy and regulation often prohibit ML models from collecting or using protected attributes at all. This raises the question how we can train a ML model to with group.fairness awareness when we do not have access to protected data attributes, neither at training nor inference time? To this end, we have developed Adversarially Reweighted Learning (ARL) [6, 5], an optimization method that leverages the notion of computationally-identifiable errors. In particular, we hypothesize that non-protected features and ML output labels are valuable for computationally-identifying systematic errors due to unfairness, and can be used to co-train an adversarial re-weighting technique for improving the ML performance for the worst-case unobserved groups. Our results show that ARL improves Rawlsian Max-Min fairness, with notable AUC improvements for worst-case protected groups in multiple datasets, outperforming state-of-the-art alternatives.

**Learning Fair Representations by Incorporating Expert Knowledge on Fairness** In this work we advance the original notion of individual fairness proposed by Dwork et al [1]. A key limitation of the prior work in operationalizing individual fairness is the assumption that a *fairness-aware distance metric* is specified beforehand. While standard metrics like Euclidean distance over numerical feature spaces are straightforward to plug in, they often fail to capture to sensitive aspects in the data and the crucial issue of equally deserving individuals. We address this limitation by proposing a practically viable operationalization of the individual fairness paradigm that does not rely on human specification of a distance metric. Instead, we propose easier and more intuitive forms of eliciting expert fairness knowledge by relative comparisons of individuals, leading to a formal notion of *pairwise fairness graphs*. The problem of fair representation learning is then cast into a *graph embedding problem*. We propose *PFR* (pairwise fair representations) [7, 5], an approach to learn a pairwise fair representation of the data, which captures both data-driven similarity between individuals and pairwise fairness judgements in fairness graphs. Comprehensive experiments with synthetic and real-life data on recidivism prediction (COMPAS) and violent neighbourhood prediction (Crime and Communities) demonstrate the practical viability of our model and its advantages over state-of-the-art baselines.

**Responsible Deployment and Risk Mitigation** So far, we have focused on training fair ML models. However, their success crucially hinges on the amount and quality of training data, and also on the assumption that the data distribution for the deployed system stays the same and is well covered by the training samples. However, this cannot be taken for granted. Saria et al. [10] categorize limitations and failures of ML systems into several regimes, including data shifts (between training-time and deployment-time distributions), high data variability (such as overlapping class labels near decision boundaries) and model limitations (such as log-linear decision boundaries vs. neural ML). Trustworthy ML needs models and tools for predicting potential failure risks and analyzing the underlying sources of uncertainty. Unfortunately, systems often fail silently without any warning, often while showing high confidence in their predictions.

This work addresses the challenge of predicting, analyzing and mitigating failure risks for classifier systems. The goal is to provide the system with *uncertainty scores* for its predictions, so as to (a) reliably predict test-time inputs for which the system is likely to fail, and (b) detect the kind of uncertainty that induces the risk, so that (c) appropriate mitigation actions can be pursued. Equipped with such *uncertainty scores*, a deployed system could take meaningful actions when faced with previously unseen data points. For instance, if the system's output is likely to be erroneous, it is better to abstain or hand the data over to a human expert, rather than making a classification error that could adversely affect human lives. When many production data points are out-of-distribution, collecting additional training samples and rebuilding the system would be a remedy.

We propose the *Risk Advisor* [8, 9, 5], a post-hoc *meta-learning* approach to estimate uncertainties of a *fully trained* classifier, and to give guidance on the underlying sources of uncertainty. *Risk Advisor* is *model-agnostic*, and can be applied to any ML system, given only black-box access to the classifier, its training data, and its classification outputs. A meta-learner is then trained to predict the classification errors, as a first-cut insight into the trustworthiness of the learned model in different regions of the data space.

The meta-learner yields a single metric only, which does not reveal anything about the nature of the errors and their underlying causes. Therefore, the *Risk Advisor* refines the analysis by leveraging the notions of *aleatoric* and *epistemic* uncertainties [3] to distinguish between risks caused by distribution shifts between training data and deployment data, inherent data variability, and model limitations. These information-theoretic measures allow us to quantify different risks, and enable us to recommend specific mitigation steps towards trustworthy applications. The latter include judicious abstention in classifiers and obtaining additional training data for critical regions. Figure 31.22 illustrates the working of the Risk Advisor, by a simplified synthetic-data example.

Our approach for detecting failure risks is constructive by offering guidance on potential mitigation actions, like abstentions for criticial deployment-data points or requesting more training samples for re-building the production system. Comprehensive experiments with a variety of synthetic and real-world datasets show that *Risk Advisor* effectively predicts deployment-time failure risks and the source of uncertainty, outperforming state-of-the-art baselines [4].

## References


[1] C. Dwork, M. Hardt, T. Pitassi, O. Reingold, and R. S. Zemel. Fairness through awareness. In S. Goldwasser, ed., *Innovations in Theoretical Computer Science 2012, Cambridge, MA, USA, January 8-10, 2012*, 2012, pp. 214–226. ACM.

[2] M. Hardt, E. Price, and N. Srebro. Equality of opportunity in supervised learning. In *Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems 2016, December 5-10, 2016, Barcelona, Spain*, 2016, pp. 3315–3323. https://proceedings.neurips.cc/paper/2016/hash/9d2682367c3935defcb1f9e247a97c0d-Abstract.html.

[3] E. Hüllermeier and W. Waegeman. Aleatoric and epistemic uncertainty in machine learning: An introduction to concepts and methods. *Machine Learning*, pp. 1–50, 2021.

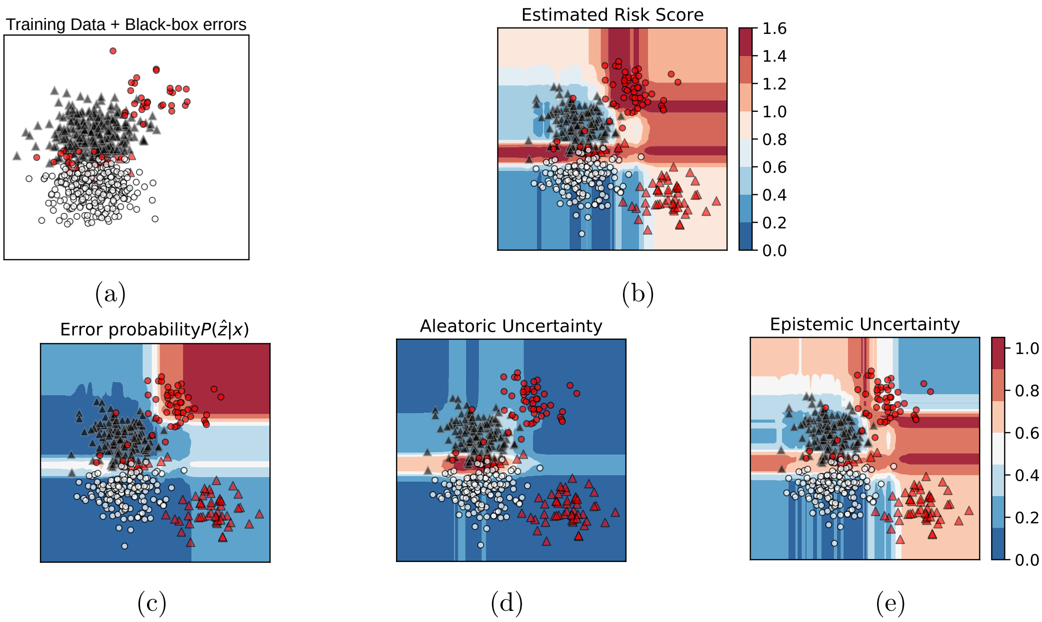



Figure 31.22: Risk Advisor: (a) Training input to *meta-learner*; (b) *meta-learner's* estimated overall *risk score*; (c, d, e) decomposition of the overall *risk score* into its various constituting components, i.e., *model*, *aleatoric* and *epistemic* uncertainties, that capture errors due to (c) model limitations, (d) data variability and noise, and (e) data shift, respectively.

[4] H. Jiang, B. Kim, M. Y. Guan, and M. R. Gupta. To trust or not to trust A classifier. In *Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2018, December 3-8, 2018, Montréal, Canada*, 2018, pp. 5546–5557. https://proceedings.neurips.cc/paper/2018/hash/7180cffd6a8e829dacfc2a31b3f72ece-Abstract.html.

• [5] P. Lahoti. *Operationalizing Fairness for Responsible Machine Learning*. PhD thesis, Universität des Saarlandes, Saarbrücken, 2022.

[6] P. Lahoti, A. Beutel, J. Chen, K. Lee, F. Prost, N. Thain, X. Wang, and E. Chi. Fairness without demographics through adversarially reweighted learning. In H. Larochelle, M. A. Ranzato, R. Hadsell, M.-F. Balcan, and H.-T. Lin, eds., *Advances in Neural Information Processing Systems 33 (NeurIPS 2020)*, Virtual Event, 2020. Curran Associates, Inc.

[7] P. Lahoti, K. Gummadi, and G. Weikum. Operationalizing individual fairness with pairwise fair representations. *Proceedings of the VLDB Endowment (Proc. VLDB)*, 13(4):506–518, 2019.

• [8] P. Lahoti, K. Gummadi, and G. Weikum. Detecting and mitigating test-time failure risks via model-agnostic uncertainty learning. In J. Bailey, P. Miettinen, Y. S. Koh, D. Tao, and X. Wu, eds., *21st IEEE International Conference on Data Mining (ICDM 2021)*, Auckland, New Zealand (Virtual Conference), 2021, pp. 1174–1179. IEEE.

• [9] P. Lahoti, K. Gummadi, and G. Weikum. Responsible model deployment via model-agnostic uncertainty learning. *Machine Learning*, 112:939–970, 2022.

[10] S. Saria and A. Subbaswamy. Tutorial: Safe and reliable machine learning. *CoRR*, abs/1904.07204, 2019. E-Text: 1904.07204.

[11] M. B. Zafar, I. Valera, M. Gomez-Rodriguez, and K. P. Gummadi. Fairness constraints: Mechanisms for fair classification. In *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, AISTATS 2017, 20-22 April 2017, Fort Lauderdale, FL, USA*, 2017, Proceedings of Machine Learning Research 54, pp. 962–970. PMLR.

### 31.7.2 Explainable Recommendations

*Investigators: Azin Ghazimatin, Rishiraj Saha Roy, Gerhard Weikum*

**Overview.** The society's increasing reliance on algorithms and machine learning to make decisions calls for models and methods for interpretable, scrutable and trustworthy AI. On the theme of transparency and explainability of recommender systems, we have developed new ways of generating user-comprehensible explanations for recommended items, using only information about the user herself – that is, not disclosing any cues about other users, for privacy preservation. Our method operates in a counterfactual way, and is applicable to both random-walk-based methods and neural-network methods [3, 6]. On top of this work, we devised a method to leverage user feedback on both a newly recommended item and the item that is shown as explanatory support [4]. A user may, for example, dislike the recommendation but express liking the explanation item. Our method, called ELIXIR, learns latent aspects that the user pays attention to when comparing different items. It uses pair-wise feedback to enhance a matrix-factorization model for capturing item-item similarities as well as user profiles, with latent aspects. Experimental evaluation demonstrates the viability of ELIXIR and its advantages over prior works. Our method with feedback on pairs of recommended and explanation items is compared, among other baselines, to feedback on item pairs that are sampled from the space of all possible pairs guided by active learning. The experimental results show that our method clearly outperforms all baselines.

This line of research is wrapped up in the doctoral dissertation of Azin Ghazimatin [1], which won the GI DBIS Dissertation Award of the German Computer Society [2].


Figure 31.23: ELIXIR.

**ELIXIR: Learning from User Feedback on Explanations to Improve Recommender Models.** Explanations are often used to improve user acceptance. However, the presence of explanations alone is not sufficient for scrutability: enabling the user to control and influence the future behavior of the recommender. This has motivated us to turn the role of explanations around and investigate how they can contribute to enhancing the quality of the generated recommendations themselves. We devised ELIXIR, a human-in-the-loop framework for learning from user feedback on explanations [4]. ELIXIR (*E*fficient *L*earning from Item-level e*X*planations *I*n *R*ecommenders) enables the recommenders to leverage user feedback on pairs of recommendation and explanation items. The collected feedback is then used to learn a user-specific preference vector.

ELIXIR extends the explainable recommenders with components for feedback collection and feedback incorporation. To collect feedback, ELIXIR assumes presence of item-level explanations, where items in user's history are used as explanations. For instance, the statement "*because you watched The Prestige*" is an example of an item-level explanation for the movie recommendation *Fight Club*. Pairing the recommendation item with each of its top explanation items, ELIXIR allows users to give feedback on their similarity. Figure 31.23 illustrates some instances of pair-level user feedback. User $u$ is recommended with the movie *Fight Club* which she does not like, and thus gives it a low rating. This single rating, however, is not sufficient for the system to learn the disliked aspects of the movie. When paired with explanation items, user can provide more fine-grained feedback. For instance, by liking the pairs (*Fight Club*, *The Prestige*) and (*Fight Club*, *7 Years in Tibet*), user rules out *surprise ending* and *Brad Pitt* as disliked aspects of the movie *Fight Club*.

To incorporate user pair-level feedback into the recommender model, ELIXIR imposes a soft constraint for learning a user-specific preference vector $\vec{w}_u$. This vector models user's bias with respect to each latent feature and is combined with item feature vectors to produce user-specific item representations. Vector $\vec{w}_u$ is learned such that the new item representations reflect user's pair-level feedback, i.e., the liked (disliked) pairs of items get closer to (farther from) each other in the vector space. We instantiated ELIXIR using generalized graph recommendation via Random Walk with Restart [5]. Insightful experiments with a real user study show significant improvements in movie and book recommendations over item-level feedback.

## References

• [1] A. Ghazimatin. *Enhancing Explainability and Scrutability of Recommender Systems.* PhD thesis, Universität des Saarlandes, Saarbrücken, 2021.

• [2] A. Ghazimatin. Enhancing explainability and scrutability of recommender systems. In B. König-Ries, S. Scherzinger, W. Lehner, and G. Vossen, eds., *BTW 2023*, Dresden, Germany, 2023, LNI P-331, pp. 633–640. GI.

[3] A. Ghazimatin, O. Balalau, R. Saha Roy, and G. Weikum. PRINCE: Provider-side interpretability with counterfactual explanations in recommender systemsxts. In J. Caverlee and X. B. Hu, eds., *WSDM '20, 13th International Conference on Web Search and Data Mining*, Houston, TX, USA, 2020, pp. 196–204. ACM.

• [4] A. Ghazimatin, S. Pramanik, R. Saha Roy, and G. Weikum. ELIXIR: Learning from user feedback on explanations to improve recommender models. In J. Leskovec, M. Grobelnik, M. Najork, J. Tang, and L. Zia, eds., *The Web Conference 2021 (WWW 2021)*, Ljubljana, Slovenia, 2021, pp. 3850–3860. ACM.

[5] A. N. Nikolakopoulos and G. Karypis. Recwalk: Nearly uncoupled random walks for top-n recommendation. In *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining, WSDM 2019, Melbourne, VIC, Australia, February 11-15, 2019*, 2019, pp. 150–158. ACM.

• [6] K. H. Tran, A. Ghazimatin, and R. Saha Roy. Counterfactual explanations for neural recommenders. In F. Diaz, C. Shah, T. Suel, P. Castells, R. Jones, T. Sakai, A. Bellogin, and M. Yushioka, eds., *SIGIR '21, 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, Virtual Event, Canada, 2021, pp. 1627–1631. ACM.

### 31.7.3 Analyzing Online Discussion Forums

*Investigators: Anna Guimarães, Erisa Terolli, Gerhard Weikum*

**Overview.** Analyzing discussions in social media has been a big topic in data science for more than a decade. Most works focus on specific dimensions and phenomena: structure of user interactions (friendships, likes etc.), dynamics of content dissemination (re-tweets, virality etc.), sentiments in posts, all the way to extreme polarization like hate speech and trolling. There is much less work on jointly investigating the combined signals from multiple dimensions. In particular, the interplay of content topics, post sentiments and user feedback in online discussions is underexplored.

Our research in this space aims to narrow this gap. It focuses on analyzing and characterizing different kinds of controversies in content-rich platforms with extended multi-user discussions, like Reddit and health forums. The goal is to understand the combined effects of salient topics in a discussion thread, sentiments expressed by different users, and the dynamic feedback (likes, up- and downvotes) among the participating users. This involves obtaining insights on the following research questions:

- What are the key factors in heavily polarized discussions between two (political) parties?
- Beyond such binary cases, what general kinds of controversial discussions are observed and how can they be categorized?
- What are the factors that indicate the onset of a controversy? Which signals allow us to predict subsequent issues with down-votes or even bans of posts?
- Shifting attention to the specific case of health forums, what are the key characteristics of the discussions in this thematic setting?

This line of research is wrapped up in the doctoral dissertation of Anna Guimaraes [2].

**X-Posts Explained: Analyzing and Predicting Controversial Contributions in Thematically Diverse Reddit Forums** Reddit hosts a number of communities dedicated to local or global politics. Users can contribute to discussions in these communities by submitting relevant news articles and by posting comments in response to these articles or in response to other users' posts. Unlike other social media platforms, posts on Reddit can receive positive and negative votes from users, which are then used to internally curate a discussion. We leverage this explicit community feedback mechanic to define two types of posts: **normal posts**, which receive mostly positive votes, and **X-posts**, which receive a significant amount of negative votes. In this work, we identify elements of Reddit discussions that lead to the occurrence of X-posts and how their presence impacts the development of future discussions. We conduct computational studies with three popular sub-reddit forums: US politics, world-wide politics, soccer, and personal relationships. The rationale for investigating this diverse scope is that "controversiality" is not a universal notion, but rather dependent on its context and on the culture of individual communities.

We devise a set of community-specific logistic regression classifiers to predict future occurrences of X-posts based on the initial posts of a discussion. Our model uses features relating to the textual content of initial posts, the topical cohesiveness of posts in the discussion, and the sentiments (positive, negative, and neutral) they express. Our classifiers achieve F1-scores between 65 and 75 percent, which is in line with the performance of other predictive models for social media data in general [1].

In addition to enabling the prediction of X-posts, our models yield interesting insights about the specific indicators associated with the presence of X-posts in each community. Mentions about specific personalities, such a famous politicians or athletes, for instance, are more strongly associated with controversiality in political and sports-related communities, whereas as the presence of negative sentiments is a bigger element of controversiality in communities dedicated to the discussion of world news. Additionally, we find that topical cohesiveness is strongly tied to the absence of X-posts in every community, which indicates that discussions are more resilient to disagreements and strong opinions than to abrupt changes in topic. This work has been published in the 2021 AAAI Conference on Web and Social Media (ICWSM) [4].

**Analyzing Health Discussion Forums** As an exemplary case of health discussion forums, we devise methods to analyzes the contents of three major platforms – healthboards.com, patient.info and relevant sub-reddits – on three important medical conditions: high blood pressure, depression and diabetes. By leveraging our expertise on entity detection, details of prevalent drugs and their dosages are extracted when present. The systematic analysis of this large-scale data sheds light into the intensity of discussions, the salient topics and the level of detail, with comparisons across the three platforms and the three medical themes. An interesting finding is that anti-depression medications are frequently mentioned in US-centric discussions (healthboards) whereas the UK-based forum (patient) does not exhibit this drug focus and rather emphasizes mutual support among affected users. These and other results appear in our publication at the 2021 ACM Conference on Computer-Supported Cooperative Work and Social Computing (CSCW) [3].

## References

[1] J. Cheng, M. S. Bernstein, C. Danescu-Niculescu-Mizil, and J. Leskovec. Anyone can become a troll: Causes of trolling behavior in online discussions. In *Proc. of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing, CSCW*, 2017, pp. 1217–1230. ACM.

• [2] A. Guimarães. *Data Science Methods for the Analysis of Controversial Social Media Discussions*. PhD thesis, Universität des Saarlandes, Saarbrücken, 2022.

- [3] A. Guimarães, E. Terolli, and G. Weikum. Comparing health forums: User engagement, salient entities, medical detail. In S. Ding, S. Fussell, A. Monroy-Hernández, S. Munson, I. Shklovski, and M. Naaman, eds., *CSCW '21 Companion*, Virtual Event, USA, 2021, pp. 57–61. ACM.
- [4] A. Guimarães and G. Weikum. X-posts explained: Analyzing and predicting controversial contributions in thematically diverse Reddit forums. In *Proceedings of the Fifteenth International Conference on Web and Social Media (ICWSM 2021)*, Atlanta, GA, USA, 2021, pp. 163–172. AAAI.


### 31.7.4 Explainable Machine Learning

*Investigators: Jonas Fischer, Jilles Vreeken*

The field of *data mining (DM)* provides methods for the *descriptive analysis* of multivariate datasets, aiming to reveal informative patterns that capture regularities in the data. Patterns can take simple forms like frequently co-occurring attribute values, or can be specified by more expressive pattern languages. Pattern mining is typically an unsupervised methodology, and builds on discrete math. The field of *machine learning (ML)*, on the other hand, provides methods for *predictive modeling* of multivariate data, with the goal of classifying data points or generating values of unknown attributes. In doing this, the methods aim to generalize the learned model, so as to prepare for new data that is not seen in the original input. Machine learning methods are typically supervised, relying on training data, and they are based on differentiable objective functions over real-valued vectors.

The DM and ML fields have many techniques in common, but apply them under different paradigms, as sketched above. As a consequence, the two research communities, on data mining and on machine learning, have largely disregarded each other. Our work addresses this very gap between the two fields, by investigating different aspects of how data mining can support machine learning and vice versa. A full suite of results is described in the dissertation of Jonas Fischer [1].

On the theme of leveraging DM for ML, we devised methods that make neural networks more interpretable. The approach adapts pattern mining techniques to activations of neurons when the network makes predictions. The method identifies compact subsets of neurons from two adjacent layers that contribute most to the network's output. Obviously, this faces a huge search space, and a large part of the solution goes into making the computation efficient. The resulting method is called ExplaiNN, and experimentally evaluated with CNNs for image data. These results have been published in the ICML 2021 conference [2]. Another way of using pattern mining to enhance the interpretability of neural network predictions is the PREMISE method [4], targeted at the case when the trained network produces errors and with focus on NLP use-cases. Our solution devises an algorithm that identifies succinct subsets of features that contribute most to particular labels of a classifier and to the erroneous outputs where the predicted label misses the ground truth. The method is experimentally studied for the NLP task of named entity recognition and for visual question answering on images. These results have been published in the ICML 2022 conference [4].

The reverse direction leverages neural machine learning for the data mining task of describing a dataset in terms of few succinct patterns. Towards this goal, we devised a neural auto-encoder with differentiable reconstruction loss. The network is equipped with a hidden layer of capturing patterns. Rounding these learned values to binary outputs yields the desired patterns; hence the name BINAPS (Binary Pattern Networks) for this approach. Our method resembles earlier works on matrix factorization with rounding to integers or Boolean values, but addresses the unexplored case of neural learning. The BINAPS method scales to large datasets with a large number of features. These results have been published in the KDD 2021 conference [3].

### References


- [1] J. Fischer. *More than the sum of its parts.* PhD thesis, Universität des Saarlandes, Saarbrücken, 2022.
- [2] J. Fischer, A. Oláh, and J. Vreeken. What's in the box? Exploring the inner life of neural networks with robust rules. In M. Meila and T. Zhang, eds., *Proceedings of the 38th International Conference on Machine Learning (ICML 2021)*, Virtual Event, 2021, Proceedings of the Machine Learning 139, Article 26. MLR Press.
- [3] J. Fischer and J. Vreeken. Differentiable pattern set mining. In F. Zhu, B. C. Ooi, C. Miao, G. Cong, J. Tang, and T. Derr, eds., *KDD '21, 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, Virtual Event, Singapore, 2021, pp. 383–392. ACM.
- [4] M. A. Hedderich, J. Fischer, D. Klakow, and J. Vreeken. Label-descriptive patterns and their application to characterizing classification errors. In K. Chaudhuri, S. Jegelka, S. Le, S. Csaba, N. Gang, and S. Sabato, eds., *Proceedings of the 39th International Conference on Machine Learning (ICML 2022)*, Baltimore, MA, USA, 2022, Proceedings of the Machine Learning Research 162, pp. 8691–8707. https://proceedings.mlr.press/v162/hedderich22a.html.


## 31.8 Academic Activities

### 31.8.1 Journal Positions

*Simon Razniewski:*

- Special Semantic Web Journal (SWJ) Issue on Wikidata (Guest editor 2023)
- Special Semantic Web (SWJ) Issue on Commonsense Knowledge and Reasoning (Guest editor 2021)

*Gerhard Weikum:*

- *Transactions of the Association for Computational Linguistics (TACL)* (standing reviewer team since 2021)
- *Proceedings of the VLDB (PVLDB)* (associate editor 2021/2022)

*Andrew Yates:*

- *Information Processing and Management* (member of the editorial board since 2019)

### 31.8.2 Book Positions

*Gerhard Weikum:*

- *Member of the Editorial Board of the ACM Books Series* (2014-2022)

### 31.8.3 Conference and Workshop Positions

#### Membership in program and organization committees

*Hiba Arnaout:*

- *2022 Conference on Empirical Methods in Natural Language Processing (EMNLP 2022)*, Abu Dhabi, United Arab Emirates, December 2022 (PC member)
- *2021 Conference on Empirical Methods in Natural Language Processing (EMNLP 2021)*, Punta Cana, Dominican Republic, November 2021 (PC member)
- *2023 International Joint Conference on Artificial Intelligence (IJCAI 2023)*, Cape Town, South Africa, August 2023 (PC member)
- *2022 International Joint Conference on Artificial Intelligence (IJCAI 2022)*, Vienna, Austria, July 2022 (PC member)
- *2021 International Joint Conference on Artificial Intelligence (IJCAI 2021)*, Montreal, Canada, August 2021 (PC member)
- *2021 International Conference on Information and Knowledge Management (CIKM 2021)*, Queensland, Australia, November 2021 (PC member)
- *2021 International Semantic Web Conference (ISWC 2021)*, Virtual conference, October 2021 (PC member)
- *2023 Annual Meeting of the Association for Computational Linguistics (ACL 2023)*, Toronto, Canada, July 2023 (PC member)

*Klaus Berberich:*

- *Conference on Human Information Interaction and Retrieval, CHIIR 2021*, Online, March 2021 (PC member)
- *43rd European Conference on IR Research, ECIR 2021*, Lucca, Italy, March 2021 (PC member)
- *The 14th International Conference on Web Search and Data Mining, WSDM 2021*, Jerusalem, Israel, March 2021 (PC member)
- *The 16th Conference of the European Chapter of the Association of Computational Linguistics, EACL 2021*, Kyiv, Ukraine, April 2021 (PC member)
- *37th International Conference on Data Engineering, ICDE 2021*, Chania, Greece, April 2021 (PC member for demos)
- *The Web Conference 2021, WWW 2021*, Ljubljana, Slovenia, April 2021 (PC member)
- *The Semantic Web – 18th International Conference, ESWC 2021*, Hersonissos, Greece, June 2021 (Area Chair for NLP and Information Retrieval)
- *44th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2021*, Montreal, Canada, July 2021 (Senior PC member)
- *The Joint Conference of the 59th Annual Meeting of the Association of Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, ACL-IJCNLP 2021*, Bangkok, Thailand, August 2021 (PC member)

*Philipp Christmann:*

- *29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, KDD 2023*, Long Beach, California, USA, August 2023 (PC member)
- *46th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2023*, Taipei, Taiwan, July 2023 (PC member)
- *31st ACM International Conference on Information and Knowledge Management, CIKM 2022*, Atlanta, Georgia, USA, October 2022 (PC member)
- *28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, KDD 2022*, Washington, DC, USA, August 2022 (PC member)
- *7th China Conference on Knowledge Graph and Semantic Computing, CCKS 2022*, Qinhuangdao, China, August 2022 (PC member)
- *15th ACM International WSDM Conference, WSDM 2022*, Online, February 2022 (PC member)
- *30th ACM International Conference on Information and Knowledge Management, CIKM 2021*, Online, November 2021 (PC member)

*Jonas Fischer:*

- *SIAM International Conference on Data Mining, SDM21*, Virtual Conference, April 2021 (PC member)

*Shrestha Ghosh:*

- *2023 European Semantic Web Conference (ESWC 2023)*, Hersonissos, Greece, May 2023 (Research Track PC member)
- *2022 Conference on Frontiers in Computing and Systems (COMSYS-2022)*, Punjab, India, December 2022 (PC member)
- *2022 Conference on Information and Knowledge Management (CIKM 2022)*, Georgia, USA, October 2022 (Short Paper Track PC member)
- *2022 International Semantic Web Conference (ISWC 2022)*, Online event hosted by Hangzhou, China, October 2022 (Research Track PC member)
- *2021 Conference on Information and Knowledge Management (CIKM 2021)*, Online event, November 2021 (PC member)
- *2021 International Semantic Web Conference (ISWC 2021)*, Online event, October 2021 (Research Track PC member)
- *2021 Wikidata Workshop at ISWC 2021*, Online event, October 2021 (PC member)

*Magdalena Kaiser:*

- *15th ACM International WSDM Conference (WSDM 2022)*,online, February 2022 (PC member)
- *36th AAAI Conference on Artificial Intelligence (AAAI 2022)*, online, February 2022 (PC member)

- *45th International ACM SIGIR Conference on Researchand Development in Information Retrieval (SIGIR 2022)*, Madrid,July 2022 (PC member, Demo Track)
- *46th International ACM SIGIR Conference on Researchand Development in Information Retrieval (SIGIR 2023)*, Taipei,July 2023 (PC member, Demo Track)
- *61st Annual Meeting of the Association forComputational Linguistics (ACL 2023)*, Toronto, July 2023 (PCmember)

*Preethi Lahoti:*

- *Workshop on Responsible AI, ICLR 2021*, Vienna, Austria, May 2021 (Area Chair)
- *2021 European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML/PKDD 2021)*, Bilbao, Spain, September 2021 (PC member)

*Alexander Marx:*

- *SIAM International Conference on Data Mining (SDM21)*, virtual, 2021 (PC member)

*Paramita Mirza:*

- *The Web Conference 2021*, April 2021 (PC member)
- *The 16th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2021)*, online, April 2021 (PC member)
- *The 2021 Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL 2021)*, Mexico City, Mexico, June 2021 (PC member)
- *The 59th Annual Meeting of the Association for Computational Linguistics (ACL 2021)*, Bangkok, Thailand, August 2021 (PC member)
- *Workshop on Personal Knowledge Graphs, Co-located with the 3rd Automatic Knowledge Base Construction Conference (AKBC'21)*, online, October 2021 (Co-organizer)
- *The 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP 2021)*, Punta Cana, Dominican Republic, November 2021 (PC member)
- *The 17th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2023)*, Dubrovnik, Croatia, May 2023 (PC member)
- *The 61st Annual Meeting of the Association for Computational Linguistics (ACL 2023)*, Toronto, Canada, July 2023 (PC member)

*Tuan-Phong Nguyen:*

- *The 21st International Semantic Web Conference (ISWC 2022)*, Virtual Event, October 2022 (PC member)
- *Knowledge Base Construction from Pre-trained Language Models (LM-KBC) Challenge @ 21st International Semantic Web Conference (ISWC 2022)*, Virtual Event, October 2022 (Organizer)
- *2023 Extended Semantic Web Conference (ESWC 2023)*, Hersonissos, Greece, May 2023 (PC member)

- *The 61st Annual Meeting of the Association for Computational Linguistics (ACL 2023) – System Demonstrations Track*, Toronto, Canada, July 2023 (PC member)

*Koninika Pal:*

- *11th IEEE International Conference on Knowledge and System Engineering, KSE 2019*, Da Nang, Vietnam, October 2019 (PC member)
- *29th ACM International Conference on Information and Knowledge Management, CIKM 2020*, Online, October 2020 (PC member)

*Simon Razniewski:*

- *The Web Conference 2021, WWW 2021*, Ljubljana, Slovenia, April 2021 (PC member)
- *The Semantic Web – 18th International Conference, ESWC 2021*, Hersonissos, Greece, June 2021 (PC Member)
- *2021 Annual Conference of the North American Chapter of the Association for Computational Linguistics, NAACL 2021*, Minneapolis, June 2021 (PC member)
- *59th annual meeting of the Association for Computational Linguistics, ACL 2021*, Virtual Event, July 2021 (PC member)
- *2021 International Joint Conference on Artificial Intelligence, IJCAI 2021*, Virtual Event, August 2021 (Area Chair)
- *2021 Conference on Empirical Methods in Natural Language Processing (EMNLP 2021)*, Punta Cana, Dominican Republic, November 2021 (PC member)
- *2021 International Conference on Information and Knowledge Management (CIKM 2021)*, Queensland, Australia, November 2021 (senior PC member)
- *The 21st International Semantic Web Conference (ISWC 2022)*, October 2022, Hangzhou, China (Workshop Co-organizer; Wikidata)
- *The 21st International Semantic Web Conference (ISWC 2021)*, October 2021, Virtual event (Workshop Co-organizer; Wikidata)
- *The 31st International World Wide Web Conference (WWW 2022)*, April 2022, Vitual Conference (Senior PC member)
- *The 17th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2023)*, May 2023, Dubrovnik, Croatia (Area Chair)
- *The 20th European Semantic Web Conference (ESWC 2023)*, May 2023, Hersonissos, Greece (senior PC member)
- *The 49th International Conference on Very Large Data Bases (VLDB 2023)*, August 2023, Vancouver, Canada (PC member)
- *The 32nd International Joint Conference on Artificial Intelligence (IJCAI 2023)*, August 2023, Cape Town, South Africa (PC member)
- *The 4th Conference on Language, Data and Knowledge (LDK 2023)*, September 2023, Vienna, Austria (PC member)
- *The 31st ACM International Conference on Information and Knowledge Management (CIKM 2022)*, October 2022, Georgia, USA (Senior PC member)

- *The 15th ACM International WSDM Conference (WSDM 2022)*, February 2022, Virtual Conference (PC member)
- *The 19th European Semantic Web Conference (ESWC 2022)*, May-June 2022, Hersonissos, Greece (PC member)
- *Conference on Empirical Methods in Natural Language Processing (EMNLP 2022)*, December 2022, Abu-Dhabi, UAE (PC member)
- *The 48th International Conference on Very Large Data Bases (VLDB 2022)*, September 2022, Sydney, Australia (PC member)
- *The 37th Annual Conference on Neural Information Processing (NeurIPS 2022)*, November-December 2022, Conference (reviewer)

*Rishiraj Saha Roy:*

- *2021 International Joint Conference on Artificial Intelligence, IJCAI 2021*, Virtual Event, August 2021 (PC member)
- *44th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2021*, Virtual Event, July 2021 (PC member)
- *59th annual meeting of the Association for Computational Linguistics, ACL 2021*, Virtual Event, July 2021 (PC member)
- *2021 Annual Conference of the North American Chapter of the Association for Computational Linguistics, NAACL 2021*, Minneapolis, June 2021 (PC member)

*Sneha Singhania:*

- *2020 Conference on Information and Knowledge Management (CIKM)*, Online, (PC Member)
- *2021 Conference on Information and Knowledge Management (CIKM)*, Online, (PC Member)
- *2022 Conference on Information and Knowledge Management (CIKM)*, Atlanta, USA, (PC Member)
- *2022 International Semantic Web Conference (ISWC)*, Hangzhou, China, (PC Member)
- *2023 European Semantic Web Conference (ESWC)*, Hersonissos, Greece, (PC Member)
- *2023 Association of Computational Linguistics (ACL)*, Toronto, Canada, (PC Member)

*Anna Tigunova:*

- *Workshop on Personal Knowledge Graphs (AKBC 2021)*, October 2021 (PC member)

*Gerhard Weikum:*

- ACL Conference on Empirical Methods for Natural Language Processing (EMNLP) 2021
- European Conference on Information Retrieval (ECIR) 2023
- ACM Conference on Research and Development in Information Retrieval (SIGIR) 2021 (Senior PC), 2022, 2023

- International Conference on Truth and Trust Online (TTO) 2021, 2022
- International Conference on Very Large Data Bases (VLDB) 2022 (Senior PC / Editorial Board)
- ACM Conference on Web Search and Data Mining (WSDM) 2021 (Senior PC), 2022 (Senior PC)

*Andrew Yates:*

- *16th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2021)* , April 2021 (Area Chair)
- *43rd European Conference on Information Retrieval (ECIR 2021)* , April 2021 (PC Member)
- *44th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2021)*, July 2021 (PC Member)
- *Conference on Empirical Methods in Natural Language Processing (EMNLP 2021)* , November 2021 (Senior Area Chair)
- *30th ACM International Conference on Information and Knowledge Management (CIKM 2021)* , November 2021 (PC Member)
- *15th ACM International Conference on Web Search and Data Mining (WSDM 2022)* , February 2022 (PC Member)
- *44th European Conference on Information Retrieval (ECIR 2022)* , April 2022 (PC Member)
- *45th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2022)* , July 2022 (PC Member)
- *31st ACM International Conference on Information and Knowledge Management (CIKM 2022)* , November 2022 (PC Member)

### Membership in steering and other committees

*Gerhard Weikum:*

- *Steering Committee of ACM Conference on Web Search and Data Mining (WSDM) (since 2019)*
- *Steering Committee of Conference on Design of Experimental Search and Information Retrieval Systems (DESIRES) (since 2018)*
- *Award Committee for the Best Paper Award of the ACM WSDM Conference (2021)*
- *Award Committee for the Test-of-Time Award of the ACM CIKM Conference (2021)*
- *Award Committee for the ACM India Early Career Award (2020-2022)*
- *Award Committee for the GI-DBIS Dissertation Award (2023)*
- *Award Committee for the IEEE John von Neumann Medal (since 2021)*

### 31.8.4 Invited Talks and Tutorials

*Hiba Arnaout:*

- *Negative Statements Considered Useful*, invited talk at the Institute for Language, Cognition and Computation (ILCC), The University of Edinburgh, Edinburgh, United Kingdom, November 2022
- *Completeness, Recall, and Negation in Open-World Knowledge Bases*, tutorial at KR 2021, Virtual Conference, November 2021, at ISWC 2021, Virtual Conference, October 2021, and at WWW 2022, Virtual Conference, April 2022
- *On the Limits of Machine Knowledge*, tutorial at VLDB 2021, Copenhagen, Denmark, August 2021

*Philipp Christmann:*

- *Conversational Question Answering on Heterogeneous Sources*, invited talk at Search-oriented Conversational AI (SCAI) 2022, Madrid, July 2022
- *Conversational Question Answering on Heterogeneous Sources*, invited talk at Amazon Alexa AI Berlin, remote, June 2022

*Shrestha Ghosh:*

- *Completeness, Recall, and Negation in Open-World Knowledge Bases*, tutorial at The Web Conference (WWW 2022), Online event hosted by Lyon, France, April 2022
- *Completeness, Recall, and Negation in Open-World Knowledge Bases*, tutorial at 2021 International Conference on Principles of Knowledge Representation and Reasoning (KR 2021), Online event, November 2021
- *Completeness, Recall, and Negation in Open-World Knowledge Bases*, tutorial at 2021 International Semantic Web Conference (ISWC 2021), Online event, October 2021
- *On the Limits of Machine Knowledge: Completeness, Recall and Negation in Web-scale Knowledge Bases*, tutorial at 2021 Conference on Very Large Data Bases (VLDB 2021), Copenhagen, Denmark, August 2021
- *CountInformation in Knowledge Bases and Text*, invited talk atthe Data, Intelligence & Graphs (DIG) team, Télécom Paris, Palaiseau,France, July 2021.

*Magdalena Kaiser:*

- *Reinforcement Learning from Reformulations for ConvQA overKGs*, invited talk at the Glasgow Information Retrieval Group at University of Glasgow, online, October 2021

*Paramita Mirza:*

- *Building Personal Knowledge Graphs from User-generated Content and Conversations*, invited talk at MLT-DFKI, 2021
- *Building Personal Knowledge Graphs from User-generated Content and Conversations*, invited talk at Fraunhofer IIS, 2022

- *Building Personal Knowledge Graphs from User-generated Content and Conversations*, invited talk at GESIS, 2022

*Simon Razniewski:*

- *Completeness, Recall, and Negation in Open-World Knowledge Bases*, tutorial at International Semantic Web Conference (ISWC 2021), October 2021
- *Completeness, Recall, and Negation in Open-World Knowledge Bases*, tutorial at 47th International Conference on Very Large Data Bases (VLDB 2021), August 2021
- *Completeness, Recall, and Negation in Open-World Knowledge Bases*, tutorial at 18th International Conference on Principles of Knowledge Representation and Reasoning (KR 2021), November 2021
- *Completeness, Recall, and Negation in Open-World Knowledge Bases*, tutorial at 31st International World Wide Web Conference (WWW 2022), April 2022

*Rishiraj Saha Roy:*

- *Learning from User Feedback on Explanations to Improve Recommender Models*, invited talk at Flipkart Private Limited, Virtual Event, April 2021
- *Reinforcement Learning from Reformulations in Conversational Question Answering*, invited talk at Lowe's Companies, Inc., Virtual Event, July 2021
- *Conversational Question Answering over Knowledge Graphs*, invited talk at IIT Bombay, Virtual Event, April 2022
- *Learning from Reformulations in Conversational Question Answering over Knowledge Graphs*, invited talk at the Knowledge Graph Conference (KGC 2022), Virtual Event, May 2022
- *Conversational Question Answering over Knowledge Graphs*, invited talk at Defense Research and Development Organization (DRDO) India, Virtual Event, January 2023
- *Question Answering over Knowledge Graphs*, invited talk at Stevens Institute of Technology, Virtual Event, February 2023
- *New directions in Conversational Question Answering*, invited talk at the SouthWest Jiaotong University (SWJTU) China, Virtual Event, March 2023

*Anna Tigunova:*

- *Extracting personal information from conversations,*, invited talk at Workshop on Personal Knowledge Graphs at AKBC 2021

*Gerhard Weikum:*

- *15 Years of Knowledge Graphs: Achievements, Challenges, Opportunities*, Invited Talk at Web Conference, Future of Web Track, Virtual Edition, April 2021
- *Content Moderation in Online Communities*, Panel Statement, Symposium on Digitization and Democracy, Joint Symposium of National Science Academies of Germany, Israel and USA, June 2021

- *Knowledge Graphs 2021: a Data Odyssey*, Keynote at VLDB Conference, Copenhagen, Denmark, August 2021
- *Knowledge Graphs 2021: Achievements, Challenges, Opportunities*, Keynote at AIML Conference, Bangalore, India, October 2021
- *Knowledge Graphs 2022: Achievements, Challenges, Opportunities*, Distinguished Lecture at Simon Fraser University, Vancouver, Canada, May 2022
- *What Computers Know and What They Should Know*, Keynote on the Occasion of the 50th Anniversary of Computer Science at TU Darmstadt, Germany, May 2022

*Andrew Yates:*

- *Transformers for Text Ranking*, invited talk at Goldman Sachs R&D, 2021
- *Pretrained Transformers for Text Ranking: BERT and Beyond*, tutorial at WSDM 2021
- *Pretrained Transformers for Text Ranking: BERT and Beyond*, tutorial at SIGIR 2021

### 31.8.5 Other Academic Activities

*Klaus Berberich:*

- Steering Committee of Section on Information Retrieval, German Computer Society, since 2018

*Shrestha Ghosh:*

- *IMPRS-TRUST Admission Committee* (member as PhD representative), July 2021 – July 2022

*Gerhard Weikum:*

- *Scientific Advisory Board, L3S Research Center, Leibniz University Hannover (-2022)*
- *Scientific Advisory Board, Department of Mathematics and Informatics, University Basel (since 2022)*
- *Award Committee for the Best Paper Award of the ACM WSDM Conference (2021)*
- *Award Committee for the Test-of-Time Award of the ACM CIKM Conference (2021)*
- *Award Committee for the ACM India Early Career Award (2020-2022)*
- *Award Committee for the GI-DBIS Dissertation Award (2023)*
- *Award Committee for the IEEE John von Neumann Medal (since 2021)*
- *Panel Member ERC Advanced Grants (2022)*

## 31.9 Teaching Activities

### Summer Semester 2022

- Automated Knowledge Base Construction (S. Razniewski)

### Master and Bachelor Theses

- Philipp Christmann: CLOCQ: Contextualization using Index Lists and Top-k Operators for Complex Question Answering, Master's thesis, 2021 (Supervisor: R. Saha Roy and G. Weikum)
- Thac-Thong Nguyen: Grounding Depression Detection in Clinical Questionnaires, Master's thesis, 2021 (Supervisor: A. Yates)
- Ahmed Sohail Anwari: Learning Filters to Improve Social Media Search, Master's thesis, 2021 (Supervisor: A. Yates)
- Suhas Shrinivasan: Knowledge Base Stability, Master's thesis, 2021 (Supervisor: S. Razniewski)
- Kevin Martin Jose: Improving Efficiency of Dense Retrieval Methods with Query Expansion, Master's thesis, 2021 (Supervisor: A. Yates)
- Abdallah Bashir: Leveraging Self Supervised Learning in Domain-specific Language Models, Master's thesis, 2021 (Supervisor: G. Weikum, K. Pal)
- Yongqing Wang: Coreference resolution for extracting quantity-facts from multiple sentencess, Master's thesis, 2022 (Supervisor: G. Weikum, K. Pal)

## 31.10 Dissertations, Habilitations, Awards

### 31.10.1 Dissertations

- Mohamed Gad-Elrab: *Explainable Methods for Knowledge Graph Refinement and Exploration via Symbolic Reasoning*, 2021.
- Azin Ghazimatin: *Enhancing Explainability and Scrutability of Recommender Systems*, 2021.
- Sreyasi Nag Chowdhury: *Text-Image Synergy for Multimodal Retrieval and Annotation*, 2021.
- Cuong Xu Chu: *Knowledge Extraction from Fictional Texts*, 2021.
- Panagiotis Mandros: *Discovering robust dependencies from data*, 2021.
- Alexander Marx: *Information-Theoretic Causal Discovery*, 2021.
- Xiaoyu Shen: *Deep latent-variable models for neural text generation*, 2021.
- Jonas Fischer: *More than the sum of its parts – pattern mining, neural networks, and how they complement each other*, 2022.
- Anna Guimarães: *Data Science Methods for the Analysis of Controversial Social Media Discussions*, 2022.
- Preethi Lahoti: *Operationalizing Fairness for Responsible Machine Learning*, 2022.
- Anna Tigunova *Extracting Personal Information from Conversations*, 2022.
- Vinh Thinh Ho: *Entities with Quantities: Extraction, Search and Ranking*, 2022.

### 31.10.2 Awards

- Asia Biega: *GI DBIS Dissertation Award* 2021 for *Enhancing Privacy and Fairness in Search Systems*
- Philipp Christmann: *Outstanding PC Award, ACM WSDM Conference* 2022
- Azin Ghazimatin: *GI DBIS Dissertation Award* 2023 for *Enhancing Explainability and Scrutability of Recommender Systems*
- Rishiraj Saha Roy: *Outstanding Senior PC Award, AAAI Conference* 2021
- Gerhard Weikum: *Konrad Zuse Medal* 2021
- Gerhard Weikum: *Outstanding Senior PC Award, ACM WSDM Conference* 2022

## 31.11 Grants and Cooperations

#### ERC Synergy Grant 610150: imPACT, 2015-2022

The imPACT project is funded by the European Research Council (ERC), with a total of ca. 10 Mio. Euros over 6 years (2015-2020, extended until June 2022), by awarding an ERC Synergy Grant to the four principal investigators Michael Backes (Saarland University), Peter Druschel (MPI for Software Systems), Rupak Majumdar (MPI for Software Systems), and Gerhard Weikum (MPI for Informatics). imPACT stands for “Privacy, Accountability, Compliance, and Trust for the Internet of Tomorrow”. It addresses the challenge of providing these PACT properties in the Internet. A key goal is to understand and master the different roles, interactions and relationships of users and their joint effect on the four PACT properties and the potential tensions between them. The focus is on a user-centric perspective, taking into consideration modern user behaviors in a wide spectrum of online communities and commercial services.

#### BigMax – MaxNet on Big-Data-Driven Material Discovery (2017-2020)

BigMax is the short-hand name for the MaxNet on Big-Data Driven Material Discovery, funded by the Max Planck Society. The network comprises partners from 9 different Max Planck Institutes, mostly in the domain of material science. Our research in this context aims to investigate theory and methods for using data mining and machine learning techniques for discovering insights from material science data.

#### DFG Grant on Negative Knowledge at Web Scale (2021-2023)

The NeKnoWS project is funded by the German Research Foundation (DFG) with a total of 312,000 Euro over 3 years (2021-2023), via an individual engagement grant awarded to Simon Razniewski. NeKnoWS stands for “Negative Knowledge at Web Scale”. It addresses the challenge of compiling salient knowledge about assertions that do not hold in the real world. It combines statistical inference methods with text extraction and novel ranking methods, in order to identify, among the vast space of possible negative assertions, those that are both likely correct negations, and also salient.

## 31.12 Publications

### Books and proceedings


[1] J. Lin, R. Nogueira, and A. Yates. *Pretrained Transformers for Text Ranking: BERT and Beyond*, Synthesis Lectures on Human Language Technologies 53. Morgan & Claypool Publishers, San Rafael, CA, 2021.

[2] R. Saha Roy and A. Anand. *Question Answering for the Curated Web: Tasks and Methods in QA over Knowledge Bases and Text Collections.* Synthesis Lectures on Information Concepts, Retrieval, and Services. Morgan & Claypool, San Rafael, CA, 2021.


### Journal articles and book chapters


[1] H. Arnaout, S. Razniewski, G. Weikum, and J. Z. Pan. Negative statements considered useful. *Journal of Web Semantics*, 71, Article 100661, 2021.

[2] L. De Stefani, E. Terolli, and E. Upfal. Tiered sampling: An efficient method for counting sparse motifs in massive graph streams. *ACM Transactions on Knowledge Discovery from Data*, 15(5), Article 79, 2021.

[3] S. Ghosh, S. Razniewski, and G. Weikum. Answering count questions with structured answers from text. *Journal of Web Semantics*, 76, Article 100769, 2023.

[4] P. Lahoti, K. Gummadi, and G. Weikum. Responsible model deployment via model-agnostic uncertainty learning. *Machine Learning*, 112:939–970, 2022.

[5] T.-P. Nguyen, S. Razniewski, J. Romero, and G. Weikum. Refined commonsense knowledge from large-scale web contents. *IEEE Transactions on Knowledge and Data Engineering*, 2022.

[6] S. Singhania, S. Razniewski, and G. Weikum. Predicting document coverage for relation extraction. *Transactions of the Association of Computational Linguistics*, 10:207–223, 2022.

[7] G. Weikum, L. Dong, S. Razniewski, and F. Suchanek. Machine knowledge: Creation and curation of comprehensive knowledge bases. *Foundations and Trends in Databases*, 10(2-4):108–490, 2021.


### Conference and workshop articles


[1] J. Ali, P. Lahoti, and K. P. Gummadi. Accounting for model uncertainty in algorithmic discrimination. In M. Fourcade, B. Kuipers, S. Lazar, and D. Mulligan, eds., *AIES '21, Fourth AAAI/ACM Conference on Artificial Intelligence, Ethics and Society*, Virtual Conference, 2021, pp. 336–345. ACM.

[2] H. Arnaout, T.-P. Nguyen, S. Razniewski, and G. Weikum. UnCommonSense in action! Informative negations for commonsense knowledge bases. In T.-S. Chua, H. Lauw, L. Si, E. Terzi, and P. Tsaparas, eds., *WSDM '23, 16th ACM International Conference on Web Search and Data Mining*, Singapore, 2023, pp. 1120–1123. ACM.

[3] H. Arnaout, S. Razniewski, G. Weikum, and J. Z. Pan. Negative knowledge for open-world Wikidata. In J. Leskovec, M. Grobelnik, M. Najork, J. Tan, and L. Zia, eds., *The Web Conference (WWW 2021)*, Ljubljana, Slovenia, 2021, pp. 544–551. ACM.

[4] H. Arnaout, S. Razniewski, G. Weikum, and J. Z. Pan. Wikinegata: A knowledge base with interesting negative statements. *Proceedings of the VLDB Endowment (Proc. VLDB)*, 14(12):2807–2810, 2021.

[5] H. Arnaout, S. Razniewski, G. Weikum, and J. Z. Pan. UnCommonSense: Informative negative knowledge about everyday concepts. In M. Al Hasan and L. Xiong, eds., *CIKM '22, 31st ACM International Conference on Information and Knowledge Management*, Atlanta GA USA, 2022, pp. 37–46. ACM.

[6] H. Arnaout, T.-K. Tran, D. Stepanova, M. H. Gad-Elrab, S. Razniewski, and G. Weikum. Utilizing language model probes for knowledge graph repair. In *Wiki Workshop 2022*, Virtual Event, 2022. https://wikiworkshop.org/2022/.

[7] A. B. Biswas, H. Arnaout, and S. Razniewski. Neguess: Wikidata-entity guessing game with negative clues. In O. Seneviratne, C. Pesquita, J. Sequeda, and L. Etcheverry, eds., *Proceedings of the ISWC 2021 Posters, Demos and Industry Tracks (ISWC-Posters-Demos-Industry 2021)*, Virtual Conference, 2021, CEUR Workshop Proceedings 2980, Article 350. CEUR-WS.org.

[8] K. Budhathoki, M. Boley, and J. Vreeken. Discovering reliable causal rules. In C. Demeniconi and I. Davidson, eds., *Proceedings of the SIAM International Conference on Data Mining (SDM 2021)*, Virtual Conference, 2021, pp. 1–9. SIAM.

[9] E. Chang, X. Shen, D. Zhu, V. Demberg, and H. Su. Neural data-to-text generation with LM-based text augmentation. In P. Merlo, ed., *The 16th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2021)*, Online, 2021, pp. 758–768. ACL.

[10] I. Chernyavsky, A. S. Varde, and S. Razniewski. CSK-Detector: Commonsense in object detection. In *IEEE International Conference on Big Data*, Osaka, Japan, 2022, pp. 6609–6612. IEEE.

[11] P. Christmann, R. Saha Roy, and G. Weikum. Beyond NED: Fast and effective search space reduction for complex question answering over knowledge bases. In *WSDM '22, Fifteenth ACM International Conference on Web Search and Data Mining*, Tempe, AZ, USA (Virutal Event), 2022, pp. 172–180. ACM.

[12] P. Christmann, R. Saha Roy, and G. Weikum. Conversational question answering on heterogeneous sources. In E. Amigo, P. Castells, J. Gonzalo, B. Carterett, J. S. Culpepper, and G. Kazai, eds., *SIGIR '22, 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, Madrid, Spain, 2022, pp. 144–154. ACM.

[13] P. Christmann, R. Saha Roy, and G. Weikum. Question entity and relation linking to knowledge bases via CLOCQ. In G. Singh, R. Mutharaju, P. Kapanipathi, N. Mihindukulasooriya, M. Dubey, R. Usbeck, and D. Banerjee, eds., *Joint Proceedings of SemREC 2022 and SMART 2022 co-located with 21st International Semantic Web Conference (ISWC 2022)*, Hybrid Event, Hanghzou, China, 2022, CEUR Workshop Proceedings 3337, pp. 33–47. CEUR-WS.org.

[14] P. Christmann, R. Saha Roy, and G. Weikum. CLOCQ: A toolkit for fast and easy access to knowledge bases. In B. König-Ries, S. Scherzinger, W. Lehner, and G. Vossen, eds., *BTW 2023*, Dresden, Germany, 2023, LNI P-331, pp. 579–591. GI.

[15] J. Fischer, F. B. Ardakani, K. Kattler, J. Walter, and M. H. Schulz. CpG content-dependent associations between transcription factors and histone modifications. *PLoS One*, 16(4), Article 0249985, 2021.

[16] J. Fischer, A. Oláh, and J. Vreeken. What's in the box? Exploring the inner life of neural networks with robust rules. In M. Meila and T. Zhang, eds., *Proceedings of the 38th International Conference on Machine Learning (ICML 2021)*, Virtual Event, 2021, Proceedings of the Machine Learning 139, Article 26. MLR Press.

[17] J. Fischer and J. Vreeken. Differentiable pattern set mining. In F. Zhu, B. C. Ooi, C. Miao, G. Cong, J. Tang, and T. Derr, eds., *KDD '21, 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, Virtual Event, Singapore, 2021, pp. 383–392. ACM.

[18] A. Ghazimatin, S. Pramanik, R. Saha Roy, and G. Weikum. ELIXIR: Learning from user feedback on explanations to improve recommender models. In J. Leskovec, M. Grobelnik, M. Najork, J. Tang, and L. Zia, eds., *The Web Conference 2021 (WWW 2021)*, Ljubljana, Slovenia, 2021, pp. 3850–3860. ACM.

[19] S. Ghosh, S. Razniewski, and G. Weikum. Answering count queries with explanatory evidence. In E. Amigo, P. Castells, J. Gonzalo, B. Carterett, J. S. Culpepper, and G. Kazai, eds., *SIGIR '22, 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, Madrid, Spain, 2022, pp. 2415–2419. ACM.

[20] S. Ghosh, S. Razniewski, and G. Weikum. CoQEx: Entity counts explained. In T.-S. Chua, H. Lauw, L. Si, E. Terzi, and P. Tsaparas, eds., *WSDM '23, 16th ACM International Conference on Web Search and Data Mining*, Singapore, 2023, pp. 1168–1171. ACM.

[21] A. Guimarães, E. Terolli, and G. Weikum. Comparing health forums: User engagement, salient entities, medical detail. In S. Ding, S. Fussell, A. Monroy-Hernández, S. Munson, I. Shklovski, and M. Naaman, eds., *CSCW '21 Companion*, Virtual Event, USA, 2021, pp. 57–61. ACM.

[22] A. Guimarães and G. Weikum. X-posts explained: Analyzing and predicting controversial contributions in thematically diverse Reddit forums. In *Proceedings of the Fifteenth International Conference on Web and Social Media (ICWSM 2021)*, Atlanta, GA, USA, 2021, pp. 163–172. AAAI.

[23] M. A. Hedderich, J. Fischer, D. Klakow, and J. Vreeken. Label-descriptive patterns and their application to characterizing classification errors. In K. Chaudhuri, S. Jegelka, S. Le, S. Csaba, N. Gang, and S. Sabato, eds., *Proceedings of the 39th International Conference on Machine Learning (ICML 2022)*, Baltimore, MA, USA, 2022, Proceedings of the Machine Learning Research 162, pp. 8691–8707. https://proceedings.mlr.press/v162/hedderich22a.html.

[24] V. T. Ho, K. Pal, S. Razniewski, K. Berberich, and G. Weikum. Extracting contextualized quantity facts from web tables. In J. Leskovec, M. Grobelnik, M. Najork, J. Tang, and L. Zia, eds., *The Web Conference 2021 (WWW 2021)*, Ljubljana, Slovenia, 2021, pp. 4033–4042. ACM.

[25] V. T. Ho, K. Pal, and G. Weikum. QuTE: Answering quantity queries from web tables. In G. Li, Z. Li, S. Idreos, and D. Srivastava, eds., *SIGMOD '21, International Conference on Management of Data*, Xi'an, Shaanxi, China, 2021, pp. 2740–2744. ACM.

[26] V. T. Ho, D. Stepanova, D. Milchevski, J. Strötgen, and G. Weikum. Enhancing knowledge bases with quantity facts. In F. Laforest, R. Troncy, E. Simperl, D. Agarwal, A. Gionis, I. Herman, and L. Médini, eds., *WWW '22, ACM Web Conference*, Virtual Event, Lyon, France, 2022, pp. 893–901. ACM.

[27] Z. Jia, S. Pramanik, R. Saha Roy, and G. Weikum. Complex temporal question answering on knowledge graphs. In G. Demartini, G. Zuccon, J. S. Culpepper, Z. Huang, and H. Tong, eds., *CIKM '21, 30th ACM International Conference on Information & Knowledge Management*, Virtual Event, Australia, 2021, pp. 792–802. ACM.

[28] K. M. Jose, T. Nguyen, S. MacAvaney, J. Dalton, and A. Yates. DiffIR: Exploring differences in ranking models' behavior. In F. Diaz, C. Shah, T. Suel, P. Castells, R. Jones, T. Sakai, A. Bellogín, and M. Yushioka, eds., *SIGIR '21, 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, Virtual Event, Canada, 2021, pp. 2595–2599. ACM.

[29] L.-A. Kaffee, S. Razniewski, G. Amaral, and K. S. Alghamdi, eds. *Wikidata Workshop 2022*, Virtual Event, Hangzhou, China, 2022, CEUR Workshop Proceedings 3262. CEUR-WS.

[30] M. Kaiser, R. Saha Roy, and G. Weikum. Reinforcement learning from reformulations in conversational question answering over knowledge graphs. In F. Diaz, C. Shah, T. Suel, P. Castells, R. Jones, T. Sakai, A. Bellogín, and M. Yushioka, eds., *SIGIR '21, 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, Virtual Event, Canada, 2021, pp. 459–469. ACM.

[31] P. Lahoti, K. Gummadi, and G. Weikum. Detecting and mitigating test-time failure risks via model-agnostic uncertainty learning. In J. Bailey, P. Miettinen, Y. S. Koh, D. Tao, and X. Wu, eds., *21st IEEE International Conference on Data Mining (ICDM 2021)*, Auckland, New Zealand (Virtual Conference), 2021, pp. 1174–1179. IEEE.

[32] S. MacAvaney, A. Yates, S. Feldman, D. Downey, A. Cohan, and N. Goharian. Simplified data wrangling with ir_datasets. In F. Diaz, C. Shah, T. Suel, P. Castells, R. Jones, T. Sakai, A. Bellogín, and M. Yushioka, eds., *SIGIR '21, 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, Virtual Event, Canada, 2021, pp. 2429–2436. ACM.

[33] I. Mackie, J. Dalton, and A. Yates. How deep is your learning: The DL-HARD annotated deep learning dataset. In F. Diaz, C. Shah, T. Suel, P. Castells, R. Jones, T. Sakai, A. Bellogín, and M. Yushioka, eds., *SIGIR '21, 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, Virtual Event, Canada, 2021, pp. 2335–2341. ACM.

[34] A. Marx and J. Fischer. Estimating Mutual Information via Geodesic $k$NN. In *Proceedings of the SIAM International Conference on Data Mining (SDM 2022)*, Alexandria, VA, USA, 2022, pp. 415–423. SIAM.

[35] A. Marx, L. Yang, and M. van Leeuwen. Estimating conditional mutual information for discrete-continuous mixtures using multidimensional adaptive histograms. In *Proceedings of the SIAM International Conference on Data Mining (SDM 2021)*, Virtual Conference, 2021, pp. 387–395. SIAM.

[36] O. A. Mian, A. Marx, and J. Vreeken. Discovering fully oriented causal networks. In *Thirty-Fifth AAAI Conference on Artificial Intelligence*, Vancouver, Canada, 2021, pp. 8975–8982. AAAI.

[37] P. Mirza, M. Abouhamra, and G. Weikum. AligNarr: Aligning narratives on movies. In F. Xia, W. Li, and R. Navigli, eds., *The 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (ACL-IJCNLP 2019)*, Virtual, 2021, pp. 427–433. ACL.

[38] S. Nag Chowdhury, R. Bhowmik, H. Ravi, G. de Melo, S. Razniewski, and G. Weikum. Exploiting image-text synergy for contextual image captioning. In M. Mosbach, M. A. Hedderich, S. Pezzelle, A. Mogadala, D. Klakow, M.-F. Moens, and Z. Akata, eds., *Proceedings of the Third Workshop on Beyond Vision and LANguage: inTEgrating Real-world kNowledge (LANTERN)*, Kyiv, Ukraine (Online), 2021, pp. 30–37. ACL.

[39] S. Nag Chowdhury, S. Razniewski, and G. Weikum. SANDI: Story-and-images alignment. In P. Merlo, J. Tiedemann, and R. Tsarfaty, eds., *The 16th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2021)*, Online, 2021, pp. 989–999. ACL.

[40] S. Nag Chowdhury, R. Wickramarachchi, M. H. Gad-Elrab, D. Stepanova, and C. Henson. Towards leveraging commonsense knowledge for autonomous driving. In O. Seneviratne, C. Pesquita, J. Sequeda, and L. Etcheverry, eds., *International Semantic Web Conference*

*(ISWC) 2021: Posters, Demos, and Industry Tracks*, Virtual Conference, 2021, CEUR Workshop Proceedings 2980, Article 396. CEUR-WS.org.

[41] S. Naseri, J. Dalton, A. Yates, and J. Allan. CEQE: Contextualized embeddings for query expansion. In D. Hiemstra, M.-F. Moens, J. Mothe, R. Perego, M. Potthast, and F. Sebastiani, eds., *Advances in Information Retrieval (ECIR 2021)*, Lucca, Italy (Online Event), 2021, LNCS 12656, pp. 467–482. Springer.

[42] T. Nguyen, A. Yates, A. Zirikly, B. Desmet, and A. Cohan. Improving the generalizability of depression detection by leveraging clinical questionnaires. In S. Muresan, P. Nakov, and A. Villavicencio, eds., *The 60th Annual Meeting of the Association for Computational Linguistics (ACL 2022)*, Dublin, Ireland, 2022, pp. 8446–8459. ACL.

[43] T.-P. Nguyen and S. Razniewski. Materialized knowledge bases from commonsense transformers. In A. Bosselut, X. Li, B. Yuchen, V. Shwartz, B. P. Majumder, Y. Kumar Lal, R. Rudinger, X. Ren, N. Tandon, and V. Zouhar, eds., *Proceedings of the First Workshop on Commonsense Representation and Reasoning (CSRR 2022)*, Dublin, Ireland, 2022, pp. 36–42. ACL.

[44] T.-P. Nguyen, S. Razniewski, A. Varde, and G. Weikum. Extracting cultural commonsense knowledge at scale. In *WWW '23, ACM Web Conference*, Austin, TX, USA, 2023. ACM. Accepted.

[45] T.-P. Nguyen, S. Razniewski, and G. Weikum. Advanced semantics for commonsense knowledge extraction. In J. Leskovec, M. Grobelnik, M. Najork, J. Tang, and L. Zia, eds., *The Web Conference 2021 (WWW 2021)*, Ljubljana, Slovenia, 2021, pp. 2636–2647. ACM.

[46] R. Pradeep, Y. Liu, X. Zhang, Y. Li, A. Yates, and J. Lin. Squeezing water from a stone: A bag of tricks for further improving cross-encoder effectiveness for reranking. In M. Hagen, S. Verbene, C. Macdonald, C. Seifert, K. Balog, K. Nørvåg, and V. Setty, eds., *Advances in Information Retrieval (ECIR 2022)*, Stavanger, Norway, 2022, LNCS 13185, pp. 655–670. Springer.

[47] S. Razniewski, H. Arnaout, S. Ghosh, and F. Suchanek. On the limits of machine knowledge: Completeness, recall and negation in web-scale knowledge bases. *Proceedings of the VLDB Endowment (Proc. VLDB)*, 14(12):3175–3177, 2021.

[48] S. Razniewski, N. Tandon, and A. S. Varde. Information to wisdom: Commonsense knowledge extraction and compilation. In L. Lewin-Eytan, D. Carmel, E. Yom-Tov, E. Agichtein, and E. Gabrilovich, eds., *WSDM '21, 14th International Conference on Web Search and Data Mining*, Virtual Event, Israel, 2021, pp. 1143–1146. ACM.

[49] S. Singhania, T.-P. Nguyen, and S. Razniewski, eds. *Knowledge Base Construction from Pre-trained Language Models 2022 (LM-KBC 2022)*, Virtual Event, Hanghzou, China, 2022, CEUR Workshop Proceedings 3274. CEUR-WS.

[50] A. Tigunova, P. Mirza, A. Yates, and G. Weikum. Exploring personal knowledge extraction from conversations with CHARM. In L. Lewin-Eytan, D. Carmel, E. Yom-Tov, E. Agichtein, and E. Gabrilovich, eds., *WSDM '21, 14th International Conference on Web Search and Data Mining*, Virtual Event, Israel, 2021, pp. 1077–1080. ACM.

[51] G. H. Torbati, A. Yates, and G. Weikum. You get what you chat: Using conversations to personalize search-based recommendations. In D. Hiemstra, M.-F. Moens, J. Mothe, R. Perego, M. Potthast, and F. Sebastiani, eds., *Advances in Information Retrieval (ECIR 2021)*, Lucca, Italy (Online Event), 2021, LNCS 12656, pp. 207–223. Springer.

[52] H. D. Tran and A. Yates. Dense retrieval with entity views. In M. Al Hasan and L. Xiong, eds., *CIKM '22, 31st ACM International Conference on Information and Knowledge Management*, Atlanta GA USA, 2022, pp. 1955–1964. ACM.

[53] K. H. Tran, A. Ghazimatin, and R. Saha Roy. Counterfactual explanations for neural recommenders. In F. Diaz, C. Shah, T. Suel, P. Castells, R. Jones, T. Sakai, A. Bellogin, and M. Yushioka, eds., *SIGIR '21, 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, Virtual Event, Canada, 2021, pp. 1627–1631. ACM.

[54] G. Weikum. Knowledge graphs 2021: A data odyssey. *Proceedings of the VLDB Endowment (Proc. VLDB)*, 14(12):3233–3238, 2021.

[55] A. Yates, R. Nogueira, and J. Lin. Pretrained transformers for text ranking: BERT and beyond. In F. Diaz, C. Shah, T. Suel, P. Castells, R. Jones, T. Sakai, A. Bellogín, and M. Yushioka, eds., *SIGIR '21, 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, Virtual Event, Canada, 2021, pp. 2666–2668. ACM.

[56] X. Zhang, J. Xin, A. Yates, and J. Lin. Bag-of-words baselines for semantic code search. In R. Lachmy, Z. Yao, G. Durrett, M. Gligoric, J. J. Li, R. Mooney, G. Neubig, Y. Su, H. Sun, and R. Tsarfaty, eds., *The 1st Workshop on Natural Language Processing for Programming (NLP4Prog 2021)*, Bangkog, Thailand (Online), 2021, pp. 88–94. ACL.

[57] X. Zhang, A. Yates, and J. Lin. Comparing score aggregation approaches for document retrieval with pretrained transformers. In D. Hiemstra, M.-F. Moens, J. Mothe, R. Perego, M. Potthast, and F. Sebastiani, eds., *Advances in Information Retrieval (ECIR 2021)*, Lucca, Italy (Online Event), 2021, LNCS 12657, pp. 150–163. Springer.

[58] Z. Zheng, K. Hui, B. He, X. Han, L. Sun, and A. Yates. Contextualized query expansion via unsupervised chunk selection for text retrieval. *Information Processing & Management*, 58(5), Article 102672, 2021.

## Theses

[1] C. X. Chu. *Knowledge Extraction from Fictional Texts.* PhD thesis, Universität des Saarlandes, Saarbrücken, 2022.

[2] J. Fischer. *More than the sum of its parts.* PhD thesis, Universität des Saarlandes, Saarbrücken, 2022.

[3] M. H. Gad-Elrab. *Explainable Methods for Knowledge Graph Refinement and Exploration via Symbolic Reasoning.* PhD thesis, Universität des Saarlandes, Saarbrücken, 2021.

[4] A. Ghazimatin. *Enhancing Explainability and Scrutability of Recommender Systems.* PhD thesis, Universität des Saarlandes, Saarbrücken, 2021.

[5] A. Guimarães. *Data Science Methods for the Analysis of Controversial Social Media Discussions.* PhD thesis, Universität des Saarlandes, Saarbrücken, 2022.

[6] V. T. Ho. *Entities with Quantities: Extraction, Search and Ranking.* PhD thesis, Universität des Saarlandes, Saarbrücken, 2022.

[7] P. Lahoti. *Operationalizing Fairness for Responsible Machine Learning.* PhD thesis, Universität des Saarlandes, Saarbrücken, 2022.

[8] P. Mandros. *Discovering robust dependencies from data.* PhD thesis, Universität des Saarlandes, Saarbrücken, 2021.

[9] A. Marx. *Information-Theoretic Causal Discovery.* PhD thesis, Universität des Saarlandes, Saarbrücken, 2021.

[10] S. Nag Chowdhury. *Text-image synergy for multimodal retrieval and annotation.* PhD thesis, Universität des Saarlandes, Saarbrücken, 2021.

[11] X. Shen. *Deep Latent-Variable Models for Neural Text Generation.* PhD thesis, Universität des Saarlandes, Saarbrücken, 2021.

[12] A. Tigunova. *Extracting Personal Information from Conversations.* PhD thesis, Universität des Saarlandes, Saarbrücken, 2022.

# 32 D6: Visual Computing and Artificial Intelligence

## 32.1 Personnel

**Director**

Prof. Dr. Christian Theobalt

**Senior Researchers and Group Leaders**

Dr. Mohamed Elgharib (from May 2021)
Dr. Vladislav Golyanik (from May 2021)
Dr. Marc Habermann (from November 2021)
Dr. Adam Kortylewski (from February 2022)
Dr. Thomas Leimkühler (from September 2021) (co-affiliated with D4)

**Researchers**

Dr. Mengyu Chu (from March 2020 until April 2022) (co-affiliated with D4)
Dr. Mohamed Elgharib (until April 2021)
Dr. Vladislav Golyanik (until April 2021)
Lingjie Liu, Ph.D. (until February 2023)
Dr. Kripasindhu Sarkar (until January 2022)
Rishabh Dabral, Ph.D. (since February 2022)
Diogo Luvizon, Ph.D. (since October 2021)
Xingang Pan, Ph.D. (since August 2021)
Fangneng Zhan, Ph.D. (since April 2022)

**PhD Students**

Hiroyasu Akada (from October 2022)
Mallikarjun Byrasandra Ramalinga Reddy
Jianchun Chen (from April 2023)
Pulkit Gera (from October 2022)
Gereon Fox
Marc Habermann (until October 2021)
Ikhsanul Habibie
Navami Kairanda (from March 2022)
Linjie Lyu (from June 2021)
Mohit Mendiratta (from October 2021)
Viktor Rudnev (from April 2021)
Pramod Ramesh Rao (from July 2022)
Soshi Shimada

Guoxing Sun (from November 2022)
Ayush Tewari (until July 2022)
Edith Tretschk
Jian Wang (until December 2022)
Wanyue Zhang (from March 2023)
Jiayi Wang (until December 2022)
Heming Zhu (from February 2022)

**Long-term Guests**
Dr. Adam Kortylewski (from February 2022)

**Associated PhD Students**
Anindita Ghosh (IMPRS, co-advised with P. Slusallek (UdS/DFKI))
Noshaba Cheema (IMPRS, co-advised with P. Slusallek (UdS/DFKI))

**Research Engineer**
Oleksandr Sotnychenko

**Head of Real Virtual Lab**
Dr. Marc Habermann

**Secretaries**
Sabine Budde
Ellen Fries (until December 2021)

## 32.2 Visitors

From March 2021 to February 2023, the following researchers visited our group:

| | | |
|---|---|---|
| Zhi Li | 01.12.20–01.12.21 | Xi'an Jiaotong University, China |
| Willi Menapace | 01.05.21–31.05.22 | University of Trento, Trento, Italy |
| Alexander Trevithick | 15.05.21–14.09.21 | University of California, San Diego, USA |
| Marcel Seelbach | 01.10.21–31.03.22 | Universität Siegen, Siegen, Germany |
| Congyi Zhang | 26.10.21–01.11.22 | University of Hong Kong, Hong Kong |
| Ayush Tewari | 18.03.22–25.04.22 | MIT, Cambridge, USA |
| Vincent Sitzmann | 28.03.22–15.04.22 | MIT, Cambridge, USA |
| Lakshika Rathi | 27.05.22–01.08.22 | Indian Institute of Technology, Delhi, India |
| Pascal Fua | 12.07.22–15.07.22 | EPFL, Lausanne, France |
| Hao Li | 28.09.22 | Pinscreen, Los Angeles, USA, MBZUAI, Abu Dhabi, UAE |
| Eduardo Alvarado | 13.10.22–14.10.22 | École Polytechnique, Paris, France |
| Shahram Izadi | 09.11.22–10.11.22 | Google, San Franciso, USA |
| Thabo Beeler | 09.11.22–10.11.22 | Google, Zurich, Switzerland |
| Siddharth Seth | 01.12.22–31.05.23 | India |

## 32.3 Group Organization

The research in the department is structured into the following eleven research areas. Each of the research areas is led by one coordinator or a team of coordinators, which are listed in brackets:

- Marker-less 3D Estimation of Full Human Body and Hand Poses (M. Habermann, V. Golyanik and C. Theobalt)
- Human Performance Capture (C. Theobalt and M. Habermann)
- Neural Rendering and Editing of Human Models (C. Theobalt and L. Liu)
- 3D Reconstruction, Neural Rendering and Editing of Human Faces (M. Elgharib and C. Theobalt)
- Reconstructing and Medelling General Deformable Scenes (V. Golyanik and C. Theobalt)
- Simulation, Image Synthesis and Inverse Rendering (T. Leimkühler and C. Theobalt)
- Neural Scene Representations and Neural Rendering (L. Liu and C. Theobalt)
- Generative Models (X. Pan and C. Theobalt)
- Robust World Perception and Recognition (A. Kortylewski)
- Quantum Visual Computing (V. Golyanik)
- Foundational Methods for Visual Real-World Reconstruction and Artificial Intelligence (V. Golyanik and C. Theobalt)

Importantly, the research areas mentioned above are not defined by the boundaries of subgroups or fixed research teams in the department. The research themes mentioned above rather represent coherent lines of research that form methodical foci of the department as a whole. Research in these areas, however, is highly collaborative and cuts across the boundaries of the research groups and teams in the department. In other words, as it will be further detailed in part three of the report, researchers working in each area form teams in a problem-centric way; teams differ depending on the specific sub-problem in each research area to be investigated.

The organization of the department follows a very flat hierarchy. Christian Theobalt and the senior researchers and group leaders represent an executive committee of the department that regularly meets to discuss strategic questions. For each individual project, there are regular meetings of the involved researchers, which happen on a weekly, sometimes biweekly basis. Further, each subgroup of the department led by a group leader usually has a regular weekly meeting. In addition, there are regular meetings regarding questions of relevance for the entire department, such as on technical developments of the common software stack, as well as the Real Virtual lab.

The entire department meets once a week (Tuesday at 2 pm) in a group meeting to discuss state-of-the-art papers and give updates on individual projects in front of all department members. Further, as a means to foster collaborations across departments, we run the weekly CG Lunch meeting (Thursdays at 11:15 am) together with the Computer Graphics Department, in which ongoing research projects are presented in a cross-departmental way. Members of the Computer Vision and Machine Learning department are also invited to these meetings. In the following, we review in detail the research conducted in the department. We also discuss software and datasets that we make available, describe the Real Virtual Lab facilities that we are building, and explain the scope of the new VIA center at MPI for Informatics, which was established as a result of a new strategic partnership on basic research between MPI for Informatics and Google.

### Long-term Guests

**Dr. Adam Kortylewski.** The *Generative Vision and Robust Learning* group was founded at MPI-INF in February 2022 with the research focus to understand the principles that enable vision models to achieve largely enhanced generalization capabilities in out-of-distribution (OOD) scenarios. We address this challenging research problem mainly from two perspectives:

1) Benchmarking Out-of-Distribution Robustness – One reason for the limited progress in OOD generalization of DNNs is the lack of benchmark datasets. We introduced several datasets for benchmarking OOD robustness, one with on real images with annotations of individual nuisance variables (ECCV'22) and another one using synthetic images in the context of Visual Question Answering (CVPR'23). We also went beyond fixed datasets for testing algorithms, and explored creating adversarial test samples on the fly using generative image models for human pose estimation (CVPR'23).

2) Robust Vision through Neural Analysis-by-Synthesis – We developed deep network architectures that follow an analysis-by-synthesis process and achieve strong robustness in out-of-distribution scenarios for 6D pose estimation (ECCV'22), 3D pose estimation (ECCV'22), classification (submitted to PAMI'23) and 3D face reconstruction (CVPR'23).

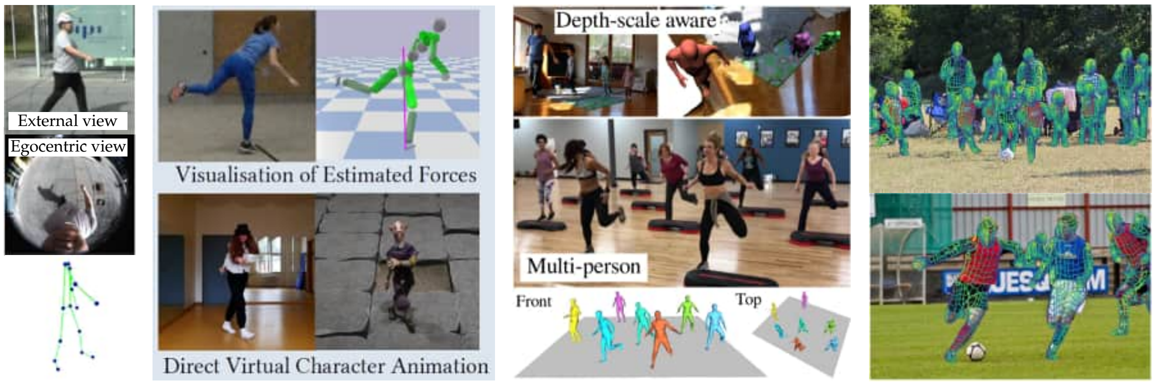
Our work is published in the major vision, graphics and learning conferences. In the reporting period, we published 9 x top tier papers (CVPR/ECCV/ICLR/AAAI) including 2 x Orals/Highlights, as well as 2 x WACV and 2 x EUROGRAPHICS and 2 papers at smaller conferences and workshops.

## 32.4 Marker-less 3D Estimation of Full Human Body and Hand Poses

*Coordinators: Marc Habermann, Vladislav Golyanik, Christian Theobalt*

*Investigators: Soshi Shimada, Lingjie Liu, Jian Wang, Jiayi Wang, Hiroyasu Akada, Zhi Li, Bharat Lal Bhatnagar, Xianghui Xie, Viktor Rudnev, Jalees Nehvi, Rishabh Dabral, Diogo C. Luvizon, Adam Kortylewski, Marc Habermann, Mohamed Elgharib, Vladislav Golyanik, Bernt Schiele, Gerard Pons-Moll, Hans-Peter Seidel, and Christian Theobalt*

Estimating the 3D pose of humans is a fundamental machine perception problem and is the key driver to several downstream applications. While there exist marker-based tracking methods for performing motion capture, they are physically invasive as they require placing




(a) Sparse 3D markers (b) Physics aware MoCap (c) Scene-aware multi-human MoCap (d) Dense pose estimation

Figure 32.1: Recent advances in human motion capture and human pose estimation.

markers on the body. This prohibits several applications that require humans in diverse clothing. To address this issue, we propose several methods to perform *markerless* full-body motion capture.

As shown in Fig. 32.1, the human body can be represented as a set of sparse 3D markers, body joint angles, or dense correspondence maps between images and the 3D surface of the human body, among other representations. In addition to being an inherently ill-posed problem owing to the loss of depth information with RGB inputs, several other challenges exist. For example, the image observations often suffer from heavy occlusions and truncations. Additional difficulties arise when accounting for common observed situations involving humans and scene interactions, multiple humans in the scene, loose clothing, etc. Finally, there also exists the setting of egocentric motion capture, wherein the camera is attached to the body, for instance, as in a VR headset. While egocentric motion capture allows for an immersive user experience, it also introduces additional challenges as the person is only partially visible.

Solutions advancing the state of the art in this field developed by our group focus on four main directions, including kinematic 3D human pose estimation, egocentric camera setups, motion capture with environmental constraints, and reconstruction of hands. In the following, we detail how each direction has been advanced by our group.

### 32.4.1 Kinematic 3D Pose Estimation of Single and Multiple Persons

As discussed earlier, monocular RGB human motion capture is an inherently ill-posed problem owing to the absence of depth information. Introducing human motion dynamics-based constraints can assist in limiting the motion space to physically plausible estimates. To that end, in Shimada et al. [11], we propose the first fully learning-based monocular markerless motion capture approach that integrates explicit physics modelling for improved plausibility of 3D motions. As shown in Fig. 32.1 (b), the method leverages an explicit rigid-body dynamics model and is aware of the contact forces with the ground and prevents foot-floor collisions in a novel optimization layer, allowing physically plausible virtual character animations.

Figure 32.2: Egocentric pose estimation: proposed concept of glasses equipped with two fisheye cameras (a), a virtual human model wearing the glasses (b), and the rendered egocentric views (c).

One ideally would like to obtain 3D human body estimations for multiple people from a lightweight capturing setup, *i.e.*, from a single RGB camera. As illustrated in Fig. 32.1 (c), this would allow every-day users to perform 3D motion capture with a personal smartphone. To tackle this challenging setup, we propose [5] to integrate scene information with human body priors over space and across time from a video sequence. Several types of regressed data modalities obtained from models trained on large-scale data are combined in a new optimization approach. It estimates, for each person in the scene, the 3D body pose as a set of joint angles, proxy shape, the absolute position in the scene, and the scene scale.

Humans can also be represented by a dense pose that establishes correspondence between pixel locations and a 3D human model, as shown in Fig. 32.1 (d). Inferring dense poses from an image or a video requires estimating the UV-maps of each individual, which is a very complex task. Most existing multi-person dense pose estimation methods do so in a top-down manner, which can be prohibitively slow with a large number of persons in the scene. To address this, in Ma et al. [6], we propose a single shot pipeline to estimate dense poses in an efficient manner. With a ResNet-based feature pyramid backbone, the proposed network splits the task into instance mask identification and global IUV-map prediction. The network dynamically generates the instance mask head parameters for each instance which can be used in conjunction with the global IUV-maps to provide per-instance dense pose outputs.

### 32.4.2 Egocentric 3D Human Pose Estimation

Unlike traditional systems with a third-person camera, in the egocentric motion capture setup the camera is usually mounted on the head or the body, allowing the user to move around a large space without restrictions, as illustrated in Fig. 32.2. Among the limitations of existing egocentric methods is the fact that the estimated poses are often confined to the local coordinate system of the fisheye camera. Since this restriction hampers many applications, in Wang et al. [14] we propose a solution to this problem with a new method for monocular egocentric global 3D body pose estimation. The proposed approach uses a spatio-temporal optimization to minimize heatmap reprojection errors, while also enforcing local and global body motion priors learned from a motion capture dataset. This results in accurate and temporally stable global poses, outperforming existing state-of-the-art methods in both quantitative and qualitative measures. Another challenge related to egocentric pose estimation is the fact that the human body is highly occluded and often interacting with the scene. To address this issue, we propose a scene-aware egocentric pose estimation method [16] that predicts the scene depth, including the regions occluded by the human, and regresses the 3D human pose considering human-scene interactions, which results in more plausible predictions.

A significant challenge in estimating egocentric 3D human pose is the lack of egocentric data and the domain gap between synthetic and real egocentric sequences. To address this issue, we propose a new large-scale egocentric dataset called Egocentric Poses in the Wild (EgoPW) [13]. This dataset includes data captured by a head-mounted fisheye camera and an auxiliary external camera, which provides an additional observation of the human body from a third-person perspective during training. Furthermore, we also propose a new egocentric pose estimation method [13] that generates pseudo labels for the EgoPW dataset and uses them to train an egocentric pose estimation network. Experiments show that the proposed approach accurately predicts 3D poses from a single in-the-wild egocentric image.

To mitigate the problem of lack of data for egocentric pose estimation, we further propose UnrealEgo (Fig. 32.2) [1] – a new large-scale naturalistic dataset for egocentric 3D human pose estimation. UnrealEgo is based on an advanced setup of eyeglasses equipped with two fisheye cameras that can be used in unconstrained environments. We design a virtual prototype and attach them to 3D human models for stereo view capture in various 3D environments within the Unreal Engine. As a result, UnrealEgo is the first dataset to provide in-the-wild stereo images with the largest variety of motions among existing egocentric datasets. Furthermore, we propose a new benchmark method with the simple but effective idea of devising a 2D keypoint estimation module for stereo inputs to improve 3D human pose estimation. Our extensive experiments show that our approach outperforms previous state-of-the-art methods qualitatively and quantitatively.

### 32.4.3 3D Motion Capture with Environmental Constraints

Inspired by recent works on physics-based pose estimation, we proposed GraviCap [3] to jointly recover the trajectories of a human and an object undergoing a free-flight trajectory. To do this, we rely on intrinsic camera calibration, image observations like 2D keypoints, and the knowledge about the value of gravitational acceleration. The parabolic trajectory of the object acts as a strong constraint and allows us to estimate the scale of the scene, thereby providing the absolute 3D trajectory of the object and the human in the scene. We achieve this by jointly optimizing the two trajectories and introducing human-object interaction constraints.

While GraviCap is focused on coarse object contacts, HULC [10] investigates fine-grained, surface level contacts with large-scale objects in the environment. To this end, it proposes a new learning-based contact label estimator for humans and the scene environment using a pixel-aligned implicit function. HULC also demonstrates that estimating such contact labels can help to improve the pose estimation plausibility by proposing a novel sampling strategy on a learned pose manifold that avoids scene penetration using a hard constraint. It is


(a) Probabilistic two-hand reconstruction (b) 3D Hand Pose Estimation from an Event Camera

Figure 32.3: Two proposed 3D hand pose estimation methods from RGB images (left) and from event cameras (right).

noteworthy that this task mandates collecting large-scale contact-labelled data, which is done by automatically annotating the GTA-IM dataset for character-scene interactions. However, several human-scene interactions involve non-rigid object deformations. In MoCapDeform [4], we present the first work to integrate such non-rigid scene deformations into the motion capture pipeline, thus jointly solving for human pose and object deformation. This is achieved by performing a raycasting optimization for finding the contact labels in the scene. With MoCapDeform, we also release a new dataset for experimental evaluation containing the ground-truth 3D human motion annotation and the deformed scene mesh.

In BEHAVE [2], we introduce a method to capture human-scene interaction in a *portable* setting with four RGB-D sensors and contribute the largest dataset of human-scene interactions with 3D annotations. Given the noisy point-clouds from RGB-D fusion, we obtain human-object contacts on the basis of the correspondences of the human body and the object using a parametric body model. The object's 6DoF pose is optimized by fitting a pre-scanned template mesh to the noisy point-clouds.

In Wan et al. [12], we propose to forecast human motions for large-scale object manipulations by training a graph neural network, termed HO-GCN. The method introduces Object Dynamics Descriptors – defined by the perturbations in the object pose upon application of unit force in several directions – which serve as a key modality to learn object manipulation. We also contribute a large-scale dataset of human-object interactions, captured using marker-based tracking methods, which can be used for future works on object manipulation.

### 32.4.4 3D Reconstruction of Hands

Human hands estimation is an essential task that enables various applications in augmented and virtual reality, robotics, or sign language translations. As illustrated in Fig. 32.3, our group has also proposed new methods considering the open challenges in the field. In Wang et al. [15], we tackle the hard problem of reconstructing two interacting hands from a monocular RGB image. We have shown in this work that existing deterministic methods and metrics are not suitable to model the ambiguities in the problem, since more than one solution exists. As such, we proposed *HandFlow*, the first method that estimates a distribution of plausible reconstructions of two-hand interactions – see Fig. 32.3 (a). To experimentally verify the advantages of this formulation, we extended existing benchmarks with additional plausible annotations and introduced a distributional metric to quantitatively measure reconstruction quality under ambiguity. Finally, we experimentally show that distribution estimation enables downstream applications such as unambiguous viewpoint selection in multiview systems. In HandVoxNet++ [7], we introduce a depth map-based hand pose estimation method that uses Truncated SDF (TSDF) representation to train a voxel-to-voxel network. We align the voxel surface with the mesh and demonstrate that using the TSDF representation improves 3D hand pose estimation results.

Non-rigid tracking of 3D objects, such as human hands, is a well known problem in computer vision. The deformation or motion of non-rigid objects in 3D has been captured in the form of intensity image sequences using RGB or RGB-D cameras, which is often limited in temporal resolution. Event cameras, unlike conventional cameras, offer a very high temporal resolution in the order of microseconds, which allows one to capture high-speed motion or deformation without resulting in motion blur or data redundancy. This characteristic is particularly useful for human hand pose estimation. In Rudnev et al. [9], we presented the first technique for 3D hand pose estimation from a monocular event stream. As shown in Fig. 32.3 (b), the proposed method runs in real-time at milestone 1000 Hz and can reconstruct significantly faster hand motions than any previous work. In a similar setup based on event cameras, we also proposed the first differentiable simulator of event streams [8]. The target scenario is non-rigid 3D tracking of generic deformable objects such as human hands, isometric surfaces and general watertight meshes, directly from incoming event streams based on an analysis-by-synthesis principle and without using additional modalities like intensity images or large-scale datasets. The method correlates a synthetically generated event stream based on an event generation model using a differentiable renderer with an observed event signal. The method is demonstrated on different types of non-rigid 3D objects, including human hands.

## References


- [1] H. Akada, J. Wang, S. Shimada, M. Takahashi, C. Theobalt, and V. Golyanik. UnrealEgo: A new dataset for robust egocentric 3D human motion capture. In S. Avidan, G. Brostow, M. Cissé, G. M. Farinella, and T. Hassner, eds., *Computer Vision – ECCV 2022*, Tel Aviv, Israel, 2022, LNCS 13666, pp. 1–17. Springer.
- [2] B. L. Bhatnagar, X. Xie, I. Petrov, C. Sminchisescu, C. Theobalt, and G. Pons-Moll. BEHAVE: Dataset and method for tracking human object interactions. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2022)*, New Orleans, LA, USA, 2022, pp. 15914–15925. IEEE.
- [3] R. Dabral, S. Shimada, A. Jain, C. Theobalt, and V. Golyanik. Gravity-aware monocular 3D human-object reconstruction. In *ICCV 2021, IEEE/CVF International Conference on Computer Vision*, Virtual Event, 2021, pp. 12365–12374. IEEE.
- [4] Z. Li, S. Shimada, B. Schiele, C. Theobalt, and V. Golyanik. MoCapDeform: Monocular 3D human motion capture in deformable scenes. In *International Conference on 3D Vision*, Hybrid / Prague, Czechia, 2022, pp. 1–11. IEEE.
- [5] D. Luvizon, M. Habermann, V. Golyanik, A. Kortylewski, and C. Theobalt. *Scene-Aware 3D Multi-Human Motion Capture from a Single Camera*, 2023. arXiv: 2301.05175.

- [6] L. Ma, L. Liu, C. Theobalt, and L. Van Gool. Direct dense pose estimation. In *2021 International Conference on 3D Vision*, Virtual Conference, 2021, pp. 721–730. IEEE.
- [7] J. Malik, A. Elhayek, S. Shimada, S. A. Ali, V. Golyanik, C. Theobalt, and D. Stricker. HandVoxNet++: 3D hand shape and pose estimation using voxel-based neural networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(12):8962–8974, 2021.
- [8] J. Nehvi, V. Golyanik, F. Mueller, H.-P. Seidel, M. Elgharib, and C. Theobalt. Differentiable event stream simulator for non-rigid 3D tracking. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPR 2021)*, Virtual Workshop, 2021, pp. 1302–1311. IEEE.
- [9] V. Rudnev, V. Golyanik, J. Wang, H.-P. Seidel, F. Mueller, M. Elgharib, and C. Theobalt. EventHands: Real-time neural 3D hand pose estimation from an event stream. In *ICCV 2021*, Virtual Event, 2021, pp. 12365–12375. IEEE.
- [10] S. Shimada, V. Golyanik, Z. Li, P. Pérez, W. Xu, and C. Theobalt. HULC: 3D HUman motion capture with pose manifold sampLing and dense Contact guidance. In S. Avidan, G. Brostow, M. Cissé, G. M. Farinella, and T. Hassner, eds., *Computer Vision – ECCV 2022*, Tel Aviv, Israel, 2022, LNCS 13682, pp. 516–533. Springer.
- [11] S. Shimada, V. Golyanik, W. Xu, P. Pérez, and C. Theobalt. Neural monocular 3D human motion capture with physical awareness. *ACM Transactions on Graphics (Proc. ACM SIGGRAPH)*, 40(4), Article 83, 2021.
- [12] W. Wan, L. Yang, L. Liu, Z. Zhang, R. Jia, Y.-K. Choi, J. Pan, C. Theobalt, T. Komura, and W. Wang. Learn to predict how humans manipulate large-sized objects from interactive motions. *IEEE Robotics and Automation Letters*, 7(2):4702–4709, 2022.
- [13] J. Wang, L. Liu, W. Xu, K. Sarkar, D. Luvizon, and C. Theobalt. Estimating egocentric 3D human pose in the wild with external weak supervision. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2022)*, New Orleans, LA, USA, 2022, pp. 13147–13156. IEEE.
- [14] J. Wang, L. Liu, W. Xu, K. Sarkar, and C. Theobalt. Estimating egocentric 3D human pose in global space. In *ICCV 2021, IEEE/CVF International Conference on Computer Vision*, Virtual Event, 2021, pp. 11480–11489. IEEE.
- [15] J. Wang, D. Luvizon, F. Mueller, F. Bernard, A. Kortylewski, D. Casas, and C. Theobalt. HandFlow: Quantifying view-dependent 3D ambiguity in two-hand reconstruction with normalizing flow. In *International Symposium on Vision, Modeling, and Visualization (VMV 2022)*, Konstanz, Germany, 2022, pp. 99–106. Eurographics Association.
- [16] J. Wang, D. Luvizon, W. Xu, L. Liu, K. Sarkar, and C. Theobalt. Scene-aware egocentric 3D human pose estimation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2023)*, Vancouver, Canada, 2023. IEEE. Accepted.


## 32.5 Human Performance Capture

*Coordinators: Christian Theobalt and Marc Habermann*

*Investigators: Ayush Tewari, Yue Jiang, Yue Li, Ikhsanul Habibie, Soshi Shimada, Lingjie Liu, Marc Habermann, Vladislav Golyanik, and Christian Theobalt*

While *motion capture* recovers the skeletal pose, *human performance capture* focuses on obtaining the dense 3D geometry of the entire human, potentially also including the deforming


Figure 32.4: Methods for human performance capture from video streams [2] learn a digital character representation, which captures, both, the deforming geometry as well as the surface appearance. In addition to capturing the human performance from video, we also investigate recovering the human performance from a sparse set of body-worn inertial measurement units [6].

clothing, from (sparse) sensing devices. Given the recent developments in AR and VR as well as the ever increasing need for more immersive communication across the globe, accurate and expressive capture of all aspects of the human has become more and more important. Here, human performance capture tries to solve one of the fundamental challenges as these methods try to port the performance of a real human captured from commodity sensing devices, e.g. a single RGB camera, at high detail into the digital world. While this research problem is far from being solved with many open challenges, our works published during the reporting period solved major questions resulting in faster runtime, improved physical plausibility, better 3D accuracy and surface detail, as well as improved expressiveness.

### 32.5.1 Pose and Geometry Recovery from Sparse Sensor Measurements

In Habermann et al. [3], we capture the dense deformation of a pre-scanned 3D human mesh solely from a single RGB video. At the core, two convolutional neural networks predict the skeletal pose and the surface deformation, which is parameterized as an embedded graph, from a single image. In stark contrast to previous work, we show that the surface deformation can be solely learned from multi-view video instead of ground truth 4D scans. Our results demonstrate state-of-the-art accuracy and robustness to in-the-wild captured data.

The aforementioned work, however, models the surface of the human as a single and connected geometry, which is a simplified model as clothed humans typically have multiple layers of geometrically disconnected surfaces, i.e. the human body and clothing pieces. Therefore, in Li et al. [5], we investigate the joint tracking of the naked human body as well as a layer of clothing. To faithfully model the interaction between body and clothing, physical cloth simulation is directly integrated into the deep learning of clothing deformations. We demonstrate that this greatly improves physical plausibility of clothing deformation and also captures the cloth-body interactions.

In addition to improving the surface fidelity, we also investigated improving the expressiveness and completeness of the capturing process. In Zhou et al. [7], we jointly capture the body pose, hand gestures, and facial expressions of a naked human body model from a single RGB image. To achieve real-time performance, we found that deep features learned for the body pose estimation process can be re-used for regressing the hand pose. This not only improves the accuracy, but also saves computational costs. Later, in Jiang et al. [4], we further improved the completeness by replacing the naked human body model with a 3D deformable template, which can also capture cloth deformations. Moreover, we improved the surface tracking by projecting deep pixel-aligned features onto the coarsely deformed human mesh, which are then passed to a graph convolutional network that predicts the final vertex displacements. This effectively reduces the receptive field of the neural architecture, in a way that surface details can be learned on a local level, increasing accuracy and robustness.

All the previously discussed methods assume a single RGB camera as the sensing device. However, we also investigated the capture of the human performance from a sparse set of body-worn inertial measurement units (IMUs). Since the sparse sensing results in an inherently ambiguous setting, in Yi et al. [6] we explicitly model the physics of human motion using the *equation of motion* as a prior. Importantly, despite recovering the skeletal motion of the human, we can for the first time also estimate physical forces acting on the joints and on the ground from a sparse set of IMUs.

### 32.5.2 Pose-conditioned Geometry and Appearance Modeling

We also investigated new ways to obtain parametric and controllable models of the human, which capture the deforming surface and appearance as a function of the skeletal pose and the parameters of the virtual camera. In Habermann et al. [2], we introduced a pose- and view-dependent character model, which faithfully models the geometry and textural appearance of the surface using deep neural networks. Interestingly, the neural models, which learn a mapping from pose and view-point to deformed geometry and appearance, can be solely trained on multi-view videos.

We further improved the surface estimation and the level of details of the appearance by introducing a hybrid representation for the human [1]. This representation consists of an explicit and deformable mesh model as well as a neural radiance field. The results demonstrate that the combination of both representations can achieve a synergistic effect enabling high quality renderings of humans at a resolution of 4K while the underlying geometry also captures high-frequency details such as clothing wrinkles.

### References


- [1] M. Habermann, L. Liu, W. Xu, G. Pons-Moll, M. Zollhöfer, and C. Theobalt. *HDHumans: A Hybrid Approach for High-fidelity Digital Humans*, 2022. arXiv: 2210.12003.
- [2] M. Habermann, L. Liu, W. Xu, M. Zollhöfer, G. Pons-Moll, and C. Theobalt. Real-time deep dynamic characters. *ACM Transactions on Graphics (Proc. ACM SIGGRAPH)*, 40(4), Article 94, 2021.
- [3] M. Habermann, W. Xu, M. Zollhöfer, G. Pons-Moll, and C. Theobalt. A deeper look into DeepCap. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(4):4009–4002, 2023.

- [4] Y. Jiang, M. Habermann, V. Golyanik, and C. Theobalt. HiFECap: Monocular high-fidelity and expressive capture of human performances. In *33rd British Machine Vision Conference (BMVC 2022)*, London, UK, 2022, Article 826. BMVA Press.
- [5] Y. Li, M. Habermann, B. Thomaszewski, S. Coros, T. Beeler, and C. Theobalt. Deep physics-aware inference of cloth deformation for monocular human performance capture. In *2021 International Conference on 3D Vision*, Virtual Conference, 2021, pp. 373–384. IEEE.
- [6] X. Yi, Y. Zhou, M. Habermann, S. Shimada, V. Golyanik, C. Theobalt, and F. Xu. Physical Inertial Poser (PIP): Physics-aware real-time human motion tracking from sparse inertial sensors. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2022)*, New Orleans, LA, USA, 2022, pp. 13157–13168. IEEE.
- [7] Y. Zhou, M. Habermann, I. Habibie, A. Tewari, C. Theobalt, and F. Xu. Monocular real-time full body capture with inter-part correlations computer vision and pattern recognition. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2021)*, Virtual Conference, 2021, pp. 4809–4820. IEEE.


## 32.6 Neural Rendering and Editing of Human Models

*Coordinators: Christian Theobalt and Lingjie Liu*

*Investigators: Kripasindhu Sarkar, Lingjie Liu, Christian Theobalt, Marc Habermann, Viktor Rudnev, Vladislav Golyanik, Mohamed Elgharib and Hans-Peter Seidel.*

Photo-realistic rendering of humans with controls over scene properties, such as body pose, appearance, and viewpoints, is an important problem in computer graphics. Approaching this with classical computer graphics pipelines is a tedious process, often requiring production-quality geometry and appearance human templates, natural human performances captured by dense camera arrays, and sophisticated global illumination rendering methods. Our recent works combine the traditional graphics pipeline with new neural network-based image formulation models, or even neural dynamic scene representations. These methods can synthesize photo-realistic human animations with diverse types of controls in a lightweight setup.

We start with approaches requiring identity-specific (single or multi-view) data. In Kappel et al. [2], we presented a new framework for high-fidelity and temporally-consistent human motion transfer with natural pose-dependent non-rigid deformations for loose garments. The proposed framework performs image generation in three stages, i.e., synthesizing human shape, structure, and appearance, and is trained using a stack of recurrent deep neural networks that generate intermediate representations from 2D poses and their temporal derivatives. This approach outperforms state-of-the-art techniques in terms of video realism, and provides artistic control of results through the manipulation of individual framework stages. In Liu et al. [3], we proposed Neural Actor (NA), which utilizes a coarse body model as the proxy to un-warp the surrounding 3D space into a canonical pose; see Fig. 32.5. A neural radiance field is then used to learn pose-dependent geometric deformations and pose- and view-dependent appearance effects in the canonical space from multi-view video input. To synthesize novel views of high-fidelity dynamic geometry and appearance, the method uses 2D texture maps defined on the body model as latent variables for predicting residual

Figure 32.5: Neural Actor [3] presents a new template-guided neural implicit representation for photo-realistic free-viewpoint rendering of moving human actors with pose control. Here we show the synthesized results of Neural Actor for novel views and novel poses, with the corresponding posed mesh models shown at the lower right.

deformations and dynamic appearance. Experiments demonstrate that NA achieves better quality than the methods in the state of the art on both playback as well as on novel pose synthesis, and can even generalize well to new poses that starkly differ from the training data.

Several of our techniques proposed during the reporting period operate on a single image of a human. In this setup, we proposed StylePoseGAN [4], which enables photo-realistic re-rendering of a human from a single image with explicit control over body pose, shape, and appearance. The proposed approach extends a non-controllable generator to accept conditioning of pose and appearance separately. Our method disentangles pose, appearance, and body parts to avoid artifacts, blurring, unrealistic distortions, and severe changes of textures. The method achieves state-of-the-art image generation fidelity, outperforming existing single-image re-rendering methods, and opens up further applications such as garment transfer, motion transfer, virtual try-on, head swap, and appearance interpolation. In Yoon et al. [5], we propose a new method that synthesizes human animation from a single image, aiming to address the limitations of existing methods that result in visual artifacts and inconsistencies. Our method uses a compositional neural network to predict silhouette, garment labels, and textures, and utilizes a unified representation of appearance and labels in UV coordinates to generate the appearance in response to pose changes. The proposed approach outperforms existing methods in terms of synthesis quality, temporal coherence, and generalization ability.

We also investigated a setting assuming a egocentric fisheye camera. Our EgoRenderer [1] can render full-body neural avatars of a person captured by a wearable egocentric fisheye camera, allowing photorealistic novel views of the actor in arbitrary motions from arbitrary virtual camera viewpoints. The system overcomes unique challenges, such as the top-down view and large lens distortion, by decomposing the rendering process into several steps, including texture synthesis, pose construction, and neural image translation. Experimental evaluations show that EgoRenderer generates realistic free-viewpoint avatars of a person captured by an egocentric camera and outperforms several baselines.

### References

- [1] T. Hu, K. Sarkar, L. Liu, M. Zwicker, and C. Theobalt. EgoRenderer: Rendering human avatars from egocentric camera images. In *ICCV 2021, IEEE/CVF International Conference on Computer Vision*, Virtual Event, 2021, pp. 14508–14518. IEEE.
- [2] M. Kappel, V. Golyanik, M. Elgharib, J.-O. Henningson, H.-P. Seidel, S. Castillo, C. Theobalt, and M. A. Magnor. High-fidelity neural human motion transfer from monocular video computer vision and pattern recognition. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2021)*, Virtual Conference, 2021, pp. 1541–1550. IEEE.
- [3] L. Liu, M. Habermann, V. Rudnev, K. Sarkar, J. Gu, and C. Theobalt. Neural actor: Neural free-view synthesis of human actors with pose control. *ACM Transactions on Graphics (Proc. ACM SIGGRAPH Asia)*, 41(6), Article 219, 2022.
- [4] K. Sarkar, V. Golyanik, L. Liu, and C. Theobalt. *Style and Pose Control for Image Synthesis of Humans from a Single Monocular View*, 2021. arXiv: 2102.11263.
- [5] J. S. Yoon, L. Liu, V. Golyanik, K. Sarkar, H. S. Park, and C. Theobalt. Pose-guided human animation from a single image in the wild. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2021)*, Virtual Conference, 2021, pp. 15034–15043. IEEE.


## 32.7 3D Reconstruction, Neural Rendering and Editing of Human Faces

*Coordinators: Mohamed Elgharib and Christian Theobalt*

*Investigators: Mallikarjun B R, Ayush Tewari, Gereon Fox, Pramod Rao, Tarun Yenamandra, Congyi Zhang, Xingang Pan, Lingjie Liu, Mohamed Elgharib, and Christian Theobalt*

The human face is a central element of our visual communication and hence its digitization is of utmost importance in visual computing. We have developed several new methods for the 3D neural reconstruction and rendering of the human face. Our work includes methods that utilize explicit scene representations [5, 4, 1], as well as methods that utilize learned neural implicit scene representations [8, 9, 6, 7]. We show several advantages that learned neural implicits can bring to the digitization of the human face, including the ability of processing the entire human head and the ability of producing 3D-consistent results. Next, we will discuss in more detail methods for the 3D modeling and reconstruction of the human face (see Sec. 32.7.1) as well as methods for neural rendering and editing (see Sec. 32.7.2). Finally, in Sec. 32.7.3 we will review our efforts to combat potential misuse of facial edits.

### 32.7.1 Parametric Models of the Human Face and its 3D Reconstruction

We developed several methods for modeling and reconstructing the human face from image data. While our earlier works use explicit means for representation the scene [5, 1], later works used learned implicit representations [8, 9] (please see Fig. 32.6).

The majority of 3D face reconstruction techniques rely on 3D morphable models, which commonly decompose the facial deformations into identity geometry, expressions, and skin

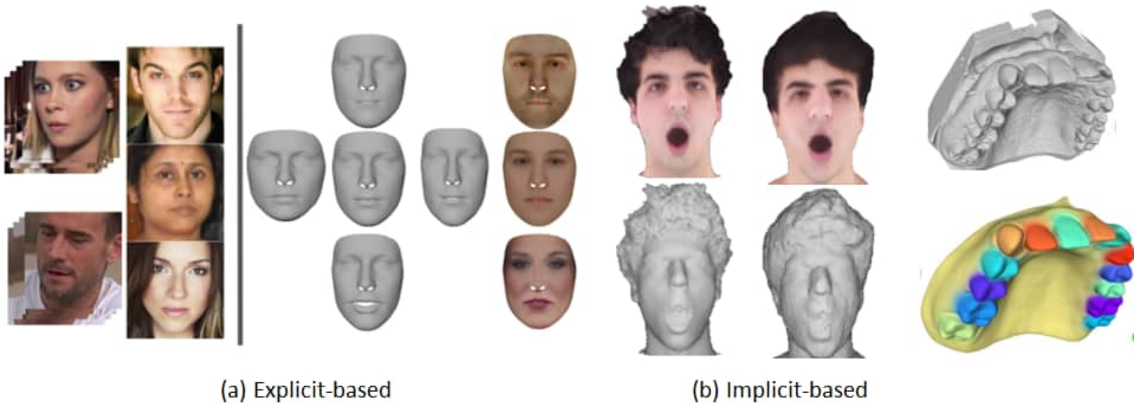



Figure 32.6: We developed methods for modeling and reconstructing the human face using explicit [5] (a) and implicit [9] (b) means of scene representation.

reflectance. However, these models are trained on a limited number of 3D scans since obtaining 3D scans at a large-scale is hard. Hence, these approaches are not generalizable across different identities and expressions. In B R et al. [5] we presented a novel approach that learns comprehensive 3D models of face identity geometry, albedo, and expression entirely from in-the-wild 2D images and videos. By leveraging the vast amounts of available data and adopting a self-supervised learning-based strategy, our approach can generate face models that surpass the range of existing methods. We use a new network design and loss functions to ensure a disentangled parameterization of identity, albedo, and expression basis, learned from unlabelled in-the-wild data. This is the first time this has been achieved. Moreover, our approach enables monocular reconstruction in real-world conditions at high-quality. In another work [1], we presented a new 3D face reconstruction method using a single input image. Our work is the first to enable monocular reconstruction of detailed face geometry, spatially varying face reflectance and complex scene illumination at very high speed on the basis of semantically meaningful scene parameters. Here, a new CNN-based approach is proposed which can be trained in a self-supervised way on unlabelled image data. We also used an end-to-end differentiable ray tracing image formation which, in contrast to earlier rasterization-based models, provides a more accurate light-geometry interaction and is able to synthesize images with complex illumination accounting for self-shadows. To the best of our knowledge, this is the first time a differentiable ray tracer is used for neural-network-based face reconstruction in an inverse-rendering setup.

While our earlier works in modeling and reconstruction of the human face uses explicit scene representations, in the later works we adapted implicit scene representations. To this end, we presented i3DMM [8], the first 3D morphable model of the human head using a neural-network-based implicit representation. Our model captures different aspects of the human heads such as identity, texture, facial expressions, as well as the entire head region, including face interior and scalp‘s hair. The highlights of our model are, it is the first 3D morphable model that represents head shape along with the scalp’s hair, and it is learned directly from captured 3D scans and not from registered meshes like existing

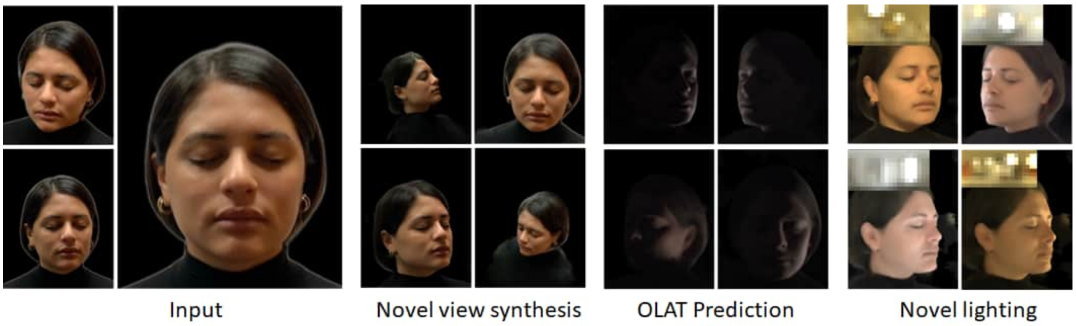



Figure 32.7: Our method VoRF [7] models the face using a continuous volumetric neural radiance field. It learns the one-light-at-a-time (OLAT) images and enable novel viewpoint synthesis and relighting.

models. In [9] we presented the first 3D morphable model of the human teeth and gum using a learning-based implicit representation. Our model is based on a component-wise representation for each tooth and the gum, together with a learnable latent code for each of such components. Both Yenamandra et al. [8] and Zhang et al. [9] enable novel applications such as interpolation in the learnable latent space. This demonstrates the ability of neural implicit scene representations in modelling and reconstructing the human head and its various components.

### 32.7.2 Face Editing and Relighting

Modeling the reflectance field of a face is needed to simulate various lighting effects of the face, such as diffuse, specular, inter-reflection, and self-shadowing. Existing methods for estimating face reflectance from a single image often assume that faces are diffuse, with only a few approaches incorporating a specular component. In [4], we propose a novel neural representation for face reflectance that enables us to estimate all the reflectance components that contribute to the final appearance from a single monocular image. Rather than modeling each reflectance component separately, our neural representation allows us to generate a set of basis faces in a geometrically invariant space, which is parameterized by the input light direction, viewpoint, and face geometry. We train our method on a dataset of 300 individuals captured under 150 different lighting conditions from eight viewpoints using a light stage. Our new learning-based approach reconstructs the reflectance field of a face from a single image, and can render the face from any viewpoint in any lighting condition. It is clear that the size of the light stage dataset plays an important role in the overall reconstruction and relighting quality. However, capturing a large number of identities in a light stage is challenging and not easily scalable to hundreds or thousands of subjects. Thus in [3] we investigated how to relight faces photorealistically while using a limited number of light stage identities. The core novelty here lies in exploiting the latent space of strong generative models such as StyleGAN. By performing reflectance learning in this latent space we show that it is possible to learn a high quality facial relighting method using as few as 15 identities. In addition, learning reflectance in such latent space allows rendering the full human head

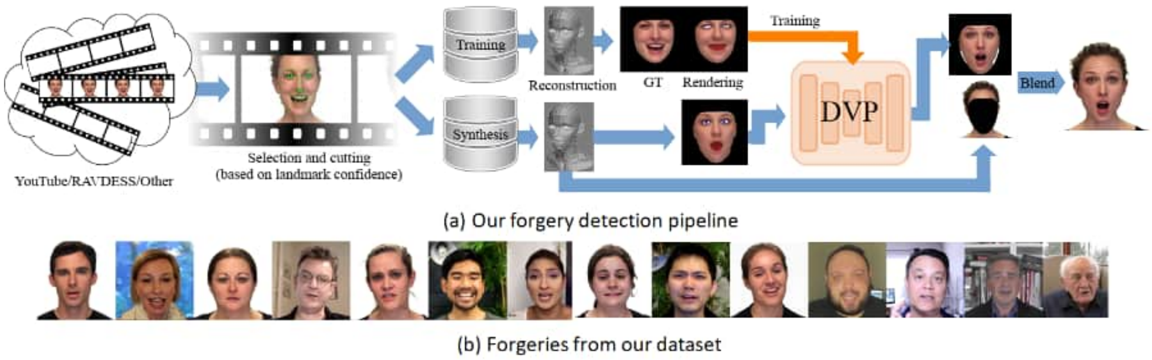



Figure 32.8: We developed a method [2] for detecting highly videorealistic simulated facial avatars (a). We also publically released a novel dataset, VideoForensicsHQ, to foster future research in this important problem (b).

with the scalp's hair and varying expressions given a single image. This is even the case when all the light stage subjects have closed eyes and covered hair.

In [7] we investigated using implicit representations for facial relighting. Our method, known as VoRF, models human head as a continuous volumetric field via NeRF, and learns a prior of human heads under diverse natural illumination conditions. It is trained in an auto-decoder style over a diverse set of head shapes and appearances, enabling generalization to new subjects. VoRF models the face reflectance using One-Light-at-A-Time (OLAT) images that capture self-shadows, specularities, and sub-surface scattering properties. It leverages intermediate features of the face prior network to render consistent HDR OLATs under novel viewpoints, enabling relighting at arbitrary viewpoints with desired environment maps. Results show the ability of relighting and editing the camera viewpoint using as few as a single input image. Our approach also outperforms related methods visually and numerically against groundtruth. Please see Fig. 32.7 for an overview. Another way to achieve facial relighting is to obtain the explicit material properties of faces. Using a light stage setup here could be demanding. Hence in [6], we presented a new method for unsupervised inverse rendering that only uses unpaired images for training. Given a pretrained face GAN, we devised an exploration-and-exploitation algorithm to generate pseudo multi-view and multi-lighting images, which provide sufficient cues to recover the geometry and the material properties including albedo, specular intensity, and shininess. The recovered precise geometry and materials enable photorealistic non-Lambertian relighting effects, which are achieved without using any light stage data.

### 32.7.3 Face Imagery Manipulation Detection

Model- and learning-based approaches to face video synthesis have reached high levels of visual realism. Some of these methods allow facial expressions and facial identities to be modified and/or transferred at very high fidelity. Reacting to concerns that these could be misused to modify videos in unethical ways, the research community has developed techniques to detect forgeries, for generic content as well as specifically for faces. To compare the effectiveness of detection methods it is vital to evaluate them on benchmark datasets. However, when a forgery detector achieves a high detection accuracy on a dataset, we must wonder: Does this mean that the detector is very good, or does it mean that the synthesized results in the dataset are just too easy to detect? Based on the observation that the synthetic results in existing benchmark datasets of forged face videos seem to be easy to spot for the human eye we have formulated the hypothesis that the accuracy of existing face video forgery detection methods depends on visual artefacts that humans would be able to spot with the naked eye. If this hypothesis is correct, detector performance should drop as soon as synthesized results are missing such artefacts. The artefacts in question include temporal jitter, implausible lighting, unnatural smoothness and blending boundaries, occurring as part of the synthesis process. In the course of our investigation [2], we make two main contributions: First, we present VideoForensicsHQ, a benchmark dataset of high quality face video manipulations, designed to not include said artefacts. Our evaluation of existing detectors on VideoForensicsHQ shows that their performance leaves room for improvement. Second, making use of this room, we present a novel family of learning-based detectors that examine combinations of color, low-level noise and temporal correlations. We find these to perform better than previous methods on high-quality synthesized results and to even generalize to unseen synthesis methods. For an overview of this work, please refer to Fig. 32.8.

## References


- [1] A. Dib, C. Thebault, J. Ahn, P.-H. Gosselin, C. Theobalt, and L. Chevallier. Towards high fidelity monocular face reconstruction with rich reflectance using self-supervised learning and ray tracing. In *ICCV 2021, IEEE/CVF International Conference on Computer Vision*, Virtual Event, 2021, pp. 12799–12809. IEEE.
- [2] G. Fox, W. Liu, H. Kim, H.-P. Seidel, M. Elgharib, and C. Theobalt. VideoForensicsHQ: Detecting high-quality manipulated face videos. In *IEEE International Conference on Multimedia and Expo (ICME 2021)*, Shenzhen, China (Virtual), 2021, pp. 1–6. IEEE.
- [3] Mallikarjun B R, A. Tewari, A. Dib, T. Weyrich, B. Bickel, H.-P. Seidel, H. Pfister, W. Matusik, L. Chevallier, M. Elgharib, and C. Theobalt. PhotoApp: Photorealistic appearance editing of head portraits. *ACM Transactions on Graphics (Proc. ACM SIGGRAPH)*, 40(4), Article 44, 2021.
- [4] Mallikarjun B R, A. Tewari, T.-H. Oh, T. Weyrich, B. Bickel, H.-P. Seidel, H. Pfister, W. Matusik, M. Elgharib, and C. Theobalt. Monocular reconstruction of neural face reflectance fields. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2021)*, Virtual Conference, 2021, pp. 4789–4798. IEEE.
- [5] Mallikarjun B R, A. Tewari, H.-P. Seidel, M. Elgharib, and C. Theobalt. Learning complete 3D morphable face models from images and videos. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2021)*, Virtual Conference, 2021, pp. 3360–3370. IEEE.
- [6] X. Pan, A. Tewari, L. Liu, and C. Theobalt. GAN2X: Non-Lambertian inverse rendering of image GANs. In *International Conference on 3D Vision*, Hybrid / Prague, Czechia, 2022, pp. 711–721. IEEE. Accepted.
- [7] P. Rao, Mallikarjun B R, G. Fox, T. Weyrich, B. Bickel, H. Pfister, W. Matusik, A. Tewari, C. Theobalt, and M. Elgharib. VoRF: Volumetric Relightable Faces. In *33rd British Machine Vision Conference (BMVC 2022)*, London, UK, 2022, Article 708. BMVA Press.

- [8] T. Yenamandra, A. Tewari, F. Bernard, H.-P. Seidel, M. Elgharib, D. Cremers, and C. Theobalt. i3DMM: Deep implicit 3D morphable model of human heads. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2021)*, Virtual Conference, 2021, pp. 12798–12808. IEEE.
- [9] C. Zhang, M. Elgharib, G. Fox, M. Gu, C. Theobalt, and W. Wang. An implicit parametric morphable dental model. *ACM Transactions on Graphics (Proc. ACM SIGGRAPH ASIA)*, 41(6), Article 217, 2022.


## 32.8 Reconstructing and Modeling General Deformable Scenes

*Coordinators: Vladislav Golyanik and Christian Theobalt*

*Investigators: Edith Tretschk, Navami Kairanda, Mallikarjun B R, Soshi Shimada, Erik C.M. Johnson, Rishabh Dabral, Adam Kortylewski, Marc Habermann, Mohamed Elgharib, Christian Theobalt, and Vladislav Golyanik*

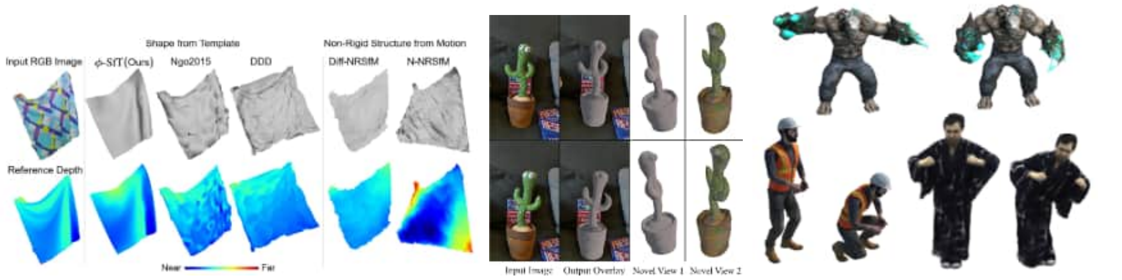
Our department also investigates general deformable objects beyond humans. In the reporting period, works in this area focused on scene-specific reconstruction with modeling techniques that have recently become available: 3D neural scene representations and differentiable physics simulation. Most of our work considers the monocular setting where a single input video is given. In particular, we led a state-of-the-art report [4] that systematically reviews all types of dense monocular non-rigid 3D reconstruction techniques for faces, hands, human bodies, animals, and general objects (together with the fundamentals of the field and a discussion of open challenges and social impliciations), whereas previous surveys addressed only small parts of this report.

We next discuss reconstruction methods developed by our department during the reporting period. Fig. 32.9 contains exemplary results. We have made code for all of our papers in this section publicly available.

In $\phi$-SfT [2], we propose an optimization-based method to reconstruct a temporally coherent sequence of 3D shapes from a monocular RGB video, given a single initial 3D template in advance. In contrast to previous Shape-from-Template (SfT) techniques that predominantly use geometric and simplified deformation models, $\phi$-SfT uses differentiable physics simulation that explicitly accounts for the physical fold-formation process and ensures high physical fidelity and realism for the surface evolution. Our adaptive gradient-based optimization recovers the deformation parameters such as force and material properties.

We also propose Non-Rigid NeRF (NR-NeRF) [5], one of the first works to take NeRF beyond the static setting to dynamic scenes by adding a deformation network on top of a static NeRF, which leads to view-consistent space warping. Although it works without additional input annotations and only requires camera parameters for its monocular input video, it can photorealistically reconstruct general real-world scenes undergoing small motions and then render them under novel views. It opens up a new area of general dynamic reconstruction aside from the classical SfT and Non-Rigid Structure from Motion (NRSfM).

In Unbiased4D [1], we shift from NR-NeRF's focus on novel-view synthesis to surface reconstruction. To this end, we propose to combine NR-NeRF-style ray bending with an SDF network for the geometry. Optionally, Unbiased4D can also handle challenging large




(a) $\phi$-SfT [2] (b) Unbiased 4D [1] (c) MoNeRF [3]

Figure 32.9: Results from three of our proposed methods for monocular reconstruction of general deformable objects.

deformations using coarse geometric proxies combined with a novel scene-flow loss. For this work, we also release a benchmark dataset for general, large-scale deforming scenes.

Finally, we propose MoNeRF [3], a volumetric 4D scene representation from multi-view data that is used in a "monocularized" fashion: only one view per timestamp has to be recorded, stored, and provided to the method during training, which accelerates training and significantly decreases the required storage space. Qualitatively, MoNeRF outperforms previous methods assuming the monocularized setting. In many applications, the demonstrated visual accuracy is sufficient, and we show that training with multiple views would be superfluous. MoNeRF's deformation module decouples spatial and temporal information, and the canonical scene leverages hash-encoded neural radiance fields. This leads to much faster training and reduces jitter artifacts for novel views from a static virtual camera.

## References


- [1] E. C. M. Johnson, M. Habermann, S. Shimada, V. Golyanik, and C. Theobalt. Unbiased 4D: Monocular 4D reconstruction with a neural deformation model. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW 2023)*, Vancouver, Canada, 2023. Accepted.

- [2] N. Kairanda, E. Tretschk, M. Elgharib, C. Theobalt, and V. Golyanik. $\phi$-SfT: Shape-from-template with a physics-based deformation model. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2022)*, New Orleans, LA, USA, 2022, pp. 3938–3948. IEEE.

- [3] M. Kappel, V. Golyanik, S. Castillo, C. Theobalt, and M. A. Magnor. *Fast Non-Rigid Radiance Fields from Monocularized Data*, 2022. arXiv: 2212.01368.

- [4] E. Tretschk, N. Kairanda, Mallikarjun B R, R. Dabral, A. Kortylewski, B. Egger, M. Habermann, P. Fua, C. Theobalt, and V. Golyanik. State of the art in dense monocular non-rigid 3D reconstruction. *Computer Graphics Forum (Proc. EUROGRAPHICS)*. Accepted 2023.

- [5] E. Tretschk, A. Tewari, V. Golyanik, M. Zollhofer, C. Lassner, and C. Theobalt. Non-rigid neural radiance fields: Reconstruction and novel view synthesis of a dynamic scene from monocular video. In *ICCV 2021, IEEE/CVF International Conference on Computer Vision*, Virtual Event, 2021, pp. 12939–12950. IEEE.

## 32.9 Simulation, Image Synthesis and Inverse Rendering

*Coordinators: Thomas Leimkühler and Christian Theobalt*

*Investigators: Mallikarjun B R, Bin Chen, Mengyu Chu, Marc Habermann, Adam Kortylewski, Thomas Leimkühler, Lingjie Liu, Linjie Lyu, Abhimitra Meka, Karol Myszkowski, Xingang Pan, Hans-Peter Seidel, Ayush Tewari, Chao Wang, Rhaleb Zayer, Fangneng Zhan and Christian Theobalt*

Computer Graphics and Computer Vision are two core disciplines in visual computing and can be considered the inverse processes of each other: Computer Graphics aims at synthesizing images from scene descriptions, while the goal of Computer Vision is to infer scene descriptions from images. Our research does justice to this intimate relationship by investigating computational light transport both from the synthesis and the analysis side, while focussing on data-driven models.

Taking an even wider point of view, we consider multimodal image synthesis and editing. As information exists in various modalities in the real world, effective interaction and fusion among multimodal information plays a key role for the creation and perception of multimodal data. In our survey [9], we comprehensively contextualize the advancements of recent multimodal image synthesis and editing approaches and formulate taxonomies according to data modalities and model architectures. We introduce different guidance modalities, provide extensive descriptions of approaches, benchmarks and metrics, before delivering insights about the current research challenges and possible directions for future investigations.

Simulation and reconstruction of complex physical processes pose a particularly challenging problem for both Computer Graphics and Computer Vision. In a joint line of research with D4, we have investigated data-driven approaches for quantitative control and image-based reconstruction of fluids. In particular, we have proposed an algorithm to connect visual quantities, such as smoke density, with realistic motions learned from a training dataset. Our approach [2] allows powerful control over fluid simulations by embedding physical quantities into the latent space of a generative model. Further, we have developed a method [1] to reconstruct dynamic fluids from sparse multiview videos. By combining physics-informed neural networks with neural radiance fields, we achieve high-quality fluid reconstructions of scenes with unknown lighting and arbitrary obstacles. More information on these two projects can be found in the corresponding D4 section.

### 32.9.1 Image Decomposition and Inverse Rendering

The color of each pixel in a real-world image is the result of complex interactions between light and material. Recovering these constituents solely from images is referred to as inverse rendering and is important in the context of several applications such as scene understanding, virtual reality, and controllable image synthesis. We have advanced the state of the art in image decomposition and inverse rendering along several axes.

We have proposed the first approach for the decomposition of a monocular color video into direct and indirect illumination components in real time [6]. We retrieve the contribution made to the scene appearance by object reflectance, light sources and inter-reflections. Our approach works for regular videos and produces temporally coherent decomposition layers at

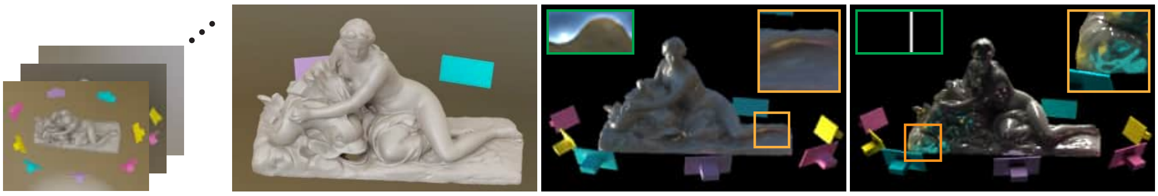



Figure 32.10: Neural Radiance Transfer Fields: Our method takes multiple views of a scene under one unknown illumination condition as input and allows novel-view synthesis and relighting (corresponding environment maps in green insets) with intricate multi-bounce illumination (orange insets).

real-time frame rates. At the core of our approach are several sparsity priors that enable the estimation of the per-pixel direct and indirect illumination layers based on a small set of jointly estimated base reflectance colors. The resulting variational decomposition problem uses a new formulation based on sparse and dense sets of non-linear equations that we solve efficiently using a novel alternating data-parallel optimization strategy. Our evaluation reveals improvements over the state of the art in this field, in both quality and runtime performance. In addition, we demonstrate various real-time appearance editing applications for videos with consistent illumination.

Further, inspired by classical pre-computed radiance transfer, we have demonstrated the recovery of a Neural Radiance Transfer Field [5] from a set of real images of the scene under a single unknown lighting condition, enabling free-viewpoint relighting with realistic global illumination (Fig. 32.10). Our novel synthetic one-light-at-a-time supervision strategy enables learning a good approximation of pre-computed radiance transfer from inputs under very limited unknown lighting conditions. The proposed method significantly outperforms the state of the art in inverse rendering, in particular in terms of relighting quality and the capability of global-illumination synthesis.

Recovering high-dynamic-range (HDR) radiance values from ordinary low-dynamic-range (LDR) images is a challenging inverse problem, in particular in the presence of overexposed image regions. With GlowGAN [8] we have introduced a novel data-driven paradigm to learn HDR imagery from unstructured LDR image collections. The key idea is to train a generative adversarial network (GAN) to synthesize HDR images which, when projected to LDR under various exposures, are indistinguishable from real LDR images. Besides HDR generation, we demonstrate the new application of unsupervised inverse tone mapping enabled by GlowGAN, significantly advancing the state of the art.

### 32.9.2 Differentiable Rendering

Although the quality of rendered images has significantly improved over decades of research in Computer Graphics, the differentiability of the rendering process remains an active area of research. A differentiable renderer needs to be able to compute gradients with respect to the image formation process, and hence allows the propagation of cues from 2D images towards scene parameters, such as geometry, material, and illumination. Such an ability is essential for combining graphics models with data-driven approaches.

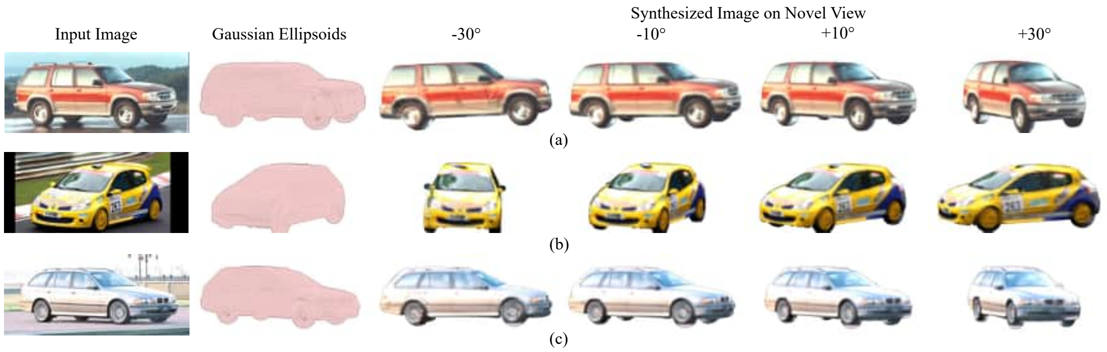



Figure 32.11: Texture sampling and re-rendering a novel view with VoGE. The inputs include a single RGB image and Gaussian ellipsoids with corresponding pose. The results are produced without any training or information about symmetry.

We have introduced a method for efficient differentiable visibility and soft shadow computation [4] that can be easily integrated into a differentiable rasterizer. Unlike existing differentiable renderers which either only model direct illumination without shadows, or are comparably time-consuming due to a Monte Carlo sampling strategy, our method allows for a significantly more efficient differentiable shadow computation. Our novel component is based on a spherical-harmonics approximation of the scene illumination and visibility, where the occluding surface is approximated by spheres. We show that this greatly facilitates solving inverse problems such as recovering texture, illumination, rigid pose, and geometric deformations from images.

Choosing a suitable scene representation is the key to differentiable rendering. With a focus on explicit unstructured representations, we have developed two methods that allow to solve a wide variety of inverse problems. First, we have proposed VoGE [7], which utilizes volumetric Gaussian reconstruction kernels as geometric primitives. Our rendering pipeline uses ray tracing to determine the nearest primitives and blends them as mixtures based on their volume density distributions along the rays. For real-time rendering, we propose an approximate closed-form solution for the volume density aggregation and a coarse-to-fine rendering strategy, accelerated by a custom CUDA implementation. Our experiments reveal that VoGE outperforms state-of-the-art counterparts when applied to various vision tasks (Fig. 32.11), e.g., object pose estimation, shape/texture fitting, and occlusion reasoning. Second, we have developed a neural point-based approach for the reconstruction of reflections. Curved reflectors pose a particularly challenging problem for image-based rendering, which we address using our Neural Point Catacaustics [3]. At its core, our method employs a neural warp field to displace a point cloud as a function of camera position, such that non-linear reflection flows can be estimated through our differentiable rendering pipeline. More information on this project can be found in the corresponding D4 section.

## References


- [1] M. Chu, L. Liu, Q. Zheng, E. Franz, H.-P. Seidel, C. Theobalt, and R. Zayer. Physics informed neural fields for smoke reconstruction with sparse data. *ACM Transactions on Graphics*, 41(4), Article 119, 2022.

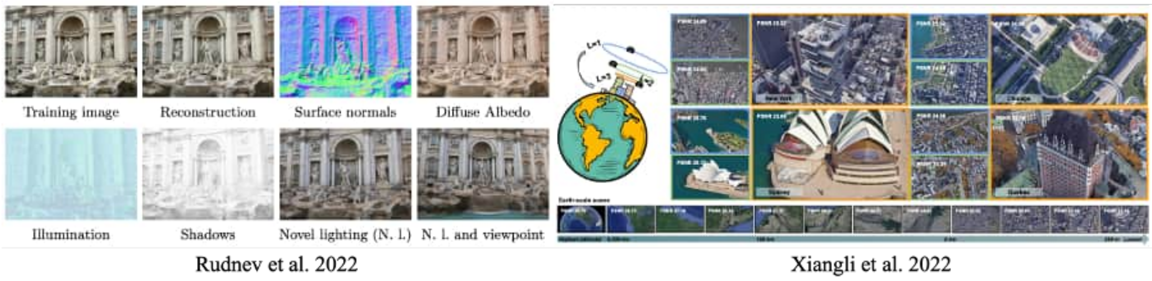



Figure 32.12: 3D scenes can be represented as neural radiance fields. In recent works, we have shown how they can be leveraged for relighting scenes observed under uncontrolled settings [2]. Moreover, we also demonstrate how NeRFs can scale to large scenes, e.g. an entire city [9].


- [2] M. Chu, N. Thuerey, H.-P. Seidel, C. Theobalt, and R. Zayer. Learning meaningful controls for fluids. *ACM Transactions on Graphics (Proc. ACM SIGGRAPH)*, 40(4), Article 100, 2021.
- [3] G. Kopanas, T. Leimkühler, G. Rainer, C. Jambon, and G. Drettakis. Neural point catacaustics for novel-view synthesis of reflections. *ACM Transactions on Graphics (Proc. ACM SIGGRAPH Asia)*, 41(6), Article 201, 2022.
- [4] L. Lyu, M. Habermann, L. Liu, Mallikarjun B R, A. Tewari, and C. Theobalt. Efficient and differentiable shadow computation for inverse problems. In *ICCV 2021, IEEE/CVF International Conference on Computer Vision*, Virtual Event, 2021, pp. 13087–13096. IEEE.
- [5] L. Lyu, A. Tewari, T. Leimkühler, M. Habermann, and C. Theobalt. Neural radiance transfer fields for relightable novel-view synthesis with global illumination. In S. Avidan, G. Brostow, M. Cissé, G. M. Farinella, and T. Hassner, eds., *Computer Vision – ECCV 2022*, Tel Aviv, Israel, 2022, LNCS 13677, pp. 153–169. Springer.
- [6] A. Meka, M. Shafiei, M. Zollhöfer, C. Richardt, and C. Theobalt. Real-time global illumination decomposition of videos. *ACM Transactions on Graphics*, 40(3), Article 22, 2021.
- [7] A. Wang, P. Wang, J. Sun, A. Kortylewski, and A. Yuille. *VoGE: A Differentiable Volume Renderer using Gaussian Ellipsoids for Analysis-by-Synthesis*, 2022. arXiv: 2205.15401.
- [8] C. Wang, A. Serrano, X. Pan, B. Chen, H.-P. Seidel, C. Theobalt, K. Myszkowski, and T. Leimkühler. *GlowGAN: Unsupervised Learning of HDR Images from LDR Images in the Wild*, 2022. arXiv: 2211.12352.
- [9] F. Zhan, Y. Yu, R. Wu, J. Zhang, S. Lu, L. Liu, A. Kortylewski, C. Theobalt, and E. Xing. *Multimodal Image Synthesis and Editing: A Survey*, 2022. arXiv: 2112.13592.


## 32.10 Neural Scene Representations and Neural Rendering

*Coordinators: Lingjie Liu and Christian Theobalt*

*Investigators: Ayush Tewari, Edith Tretschk, Lingjie Liu, Mohamed Elgharib, Viktor Rudnev, Xingang Pan, Vladislav Golyanik and Christian Theobalt*

Modeling and representing 3D scenes as well as the rendering process, which projects the scenes onto a 2D image plane, has been researched over decades and defines essentially the heart of Computer Graphics. For example, meshes are one specific scene representation, but there are many others such as pointclouds, and implicit surface representations to only name a few. Moreover, inverse graphics refers to the process of recovering scene parameters, e.g. geometry and appearance, by differentiably rendering the current scene estimate and comparing it to some observation, e.g. multi-view images, in an analysis-by-synthesis manner. Now, *neural scene representations* and *neural rendering* methods revisit the classical (inverse) Computer Graphics pipeline and propose new ways to combine classical explicit scene representations and rendering approaches with neural network-based approaches in end-to-end architectures. Since this has been an active research over the last years, we recently presented a survey [4, 5], which provides a comprehensive overview. In the reporting period, we contributed several new methods advancing the state of the art in neural scene representation and neural rendering.

#### Neural Surface Reconstruction and Rendering of Objects

Reconstructing and rendering single objects from a collection of multi-view images has been studied over decades resulting in multi-view stereo algorithms achieving impressive reconstruction quality. However, those algorithms heavily rely on correspondence matching and even the best photometric reconstruction methods still lack photorealism.

Thus, in Wang et al. [7], we present a new neural implicit scene representation and propose an unbiased volume rendering formulation enabling high-quality surface reconstruction from multi-view imagery. In detail, the scene geometry is represented as a signed distance field (SFD) paramterized by a coordinate-based multi-layer perceptron. However, in order to volume render an SDF, it has to be converted into a density. In this work, it is demonstrated how such a conversion can be achieved and more importantly how an unbiased surface rendering can be mathematically proven. As a result, the reconstructed results by far exceed the quality of classical multi-view stereo and competing neural methods.

The main downside of the previous formulation is the high computational cost and long training time. To enhance the efficiency of neural surface reconstruction, we present Voxurf [8], a voxel-based method that maintains color-geometry dependency. The geometry is modeled as hierarchical features and the scene is learned from coarse to fine, which in combination drastically improves convergence speed. Compared to NeuS, Voxurf achieves a 20x training speedup while showing higher 3D reconstruction quality.

Moreover, implicit signed distance-based surface representations typically require sphere tracing when one wants to render them onto the 2D image plane, which is time-consuming as for every ray multiple MLP evaluations are required. In Yenamandra et al. [10], we propose a directional distance network, which for every point within a sphere around the actual object can predict the distance along a certain direction. This removes the need for sphere tracing during rendering time and, thus, improves the overall rendering speed. Importantly, naively training a directional distance field would lead to 3D inconsistencies. Therefore, we propose a dedicated supervision scheme and network formulation, which maintains 3D consistency while enabling the efficient directional distance representation.

In addition to regular cameras, there are also alternative sensing devices such as event cameras, which can potentially capture images at almost infite frame rates while having a significantly reduced data bandwith. In Rudnev et al. [3], we presented the first method for dense photorealistic RGB novel view synthesis of a static scene from event camera streams. We used Neural Radiance Fields as the base model, and thanks to the proposed combination of event supervision, volumetric rendering and event-specific regularising techniques, EventNeRF outperforms the baselines in the rendered image quality, fast motion handling, low-illumination handling, and training data storage requirements. Thus, this paper extends the spectrum of practical event-based techniques with a 3D representation learning approach.

Furthermore, there is also research concerning the conditioning of neural radiance fields on (sparse) reference images or their respective features. Here, for a 3D point in the scene, features are acquired in all reference views by projecting the 3D point onto the individual images. However, in case the 3D point is occluded in some of the views, the retrieved features are wrongly assigned leading to artifacts in the novel view rendering. Therefore, in Liu et al. [1], we propose to explicitly model the occlusion by predicting a occlusion probability for a 3D point with respect to all reference views. Results demonstrate that this drastically improves the novel view rendering results in case of occlusions.

### Neural 3D Reconstruction and Rendering of Large-Scale Scenes

While the previous subsection focused on neural scene representations and rendering methods for single objects, this section discusses works in which we model larger scales, e.g. entire rooms or even large real word scenes.

In Wang et al. [6], we tackle the problem of indoor scene reconstruction using an SDF-based neural scene representation. To guide the geometry reconstruction, the SDF can be supervised on regressed per-view normal maps. However, such normal maps are typically oversmoothed and contain artifacts. Thus, naively supervising on those would degrade reconstruction quality. As a remedy, in this work, an adaptive normal checking prior is proposed, which can effectively discard those outliers.

Moving from indoor to outdoor settings, we also presented the first method for outdoor scene relighting based on Neural Radiance Fields [2]. In contrast to the prior art, our technique allows simultaneous direct editing of both scene illumination and camera viewpoint using only a collection of outdoor photos shot in uncontrolled settings. To evaluate our method we also collected a large-scale dataset of photo collections with eight different architectural landmarks in Saarbrücken, where for each site, we captured photos from multiple viewpoints and illumination conditions. For each recording session, we also recorded ground-truth environment maps together with a colour-calibration chequerboard to allow accurate numerical evaluations on real data for the first time in literature.

Finally, despite spatially constrained outdoor scenes, we also investigated the modeling of very large scenes using neural representations. While most existing works consider scenes of a single scale, we also pushed forward the scale limit that a NeRF could model in the project BungeeNeRF [9]. With a progressive architecture, it is capable of modelling extremely multi-scale scene in different level-of-details and, for the first time, is able to render from the scale of a building to the whole planet.

## References

• [1] Y. Liu, S. Peng, L. Liu, Q. Wang, P. Wang, C. Theobalt, X. Zhou, and W. Wang. Neural rays for occlusion-aware image-based rendering. In *IEEE/CVF Conference on Computer Vision and*

*Pattern Recognition (CVPR 2022)*, New Orleans, LA, USA, 2022, pp. 7814–7823. IEEE.

- [2] V. Rudnev, M. Elgharib, W. A. P. Smith, L. Liu, V. Golyanik, and C. Theobalt. NeRF for outdoor scene relighting. In S. Avidan, G. Brostow, M. Cissé, G. M. Farinella, and T. Hassner, eds., *Computer Vision – ECCV 2022*, Tel Aviv, Israel, 2022, LNCS 13676, pp. 615–631. Springer.
- [3] V. Rudnev, M. Elgharib, C. Theobalt, and V. Golyanik. *EventNeRF: Neural Radiance Fields from a Single Colour Event Camera*, 2022. arXiv: 2206.11896.
- [4] A. Tewari, O. Fried, J. Thies, V. Sitzmann, S. Lombardi, Z. Xu, T. Simon, M. Nießner, E. Tretschk, L. Liu, B. Mildenhall, P. Srinivasan, R. Pandey, S. Orts-Escolano, S. Fanello, M. Guo, G. Wetzstein, J.-Y. Zhu, C. Theobalt, M. Agrawala, D. B. Goldman, and M. Zollhöfer. Advances in neural rendering. In *SIGGRAPH 2021 Courses (ACM SIGGRAPH 2021)*, Virtual Event, USA, 2021, Article 1. ACM.
- [5] A. Tewari, J. Thies, B. Mildenhall, P. Srinivasan, E. Tretschk, Y. Wang, C. Lassner, V. Sitzmann, R. Martin-Brualla, S. Lombardi, T. Simon, C. Theobalt, M. Nießner, J. T. Barron, G. Wetzstein, M. Zollhöfer, and V. Golyanik. Advances in neural rendering. *Computer Graphics Forum (Proc. EUROGRAPHICS)*, 41(2):703–735, 2022.
- [6] J. Wang, P. Wang, X. Long, C. Theobalt, T. Komura, L. Liu, and W. Wang. NeuRIS: Neural reconstruction of indoor scenes using normal priors. In S. Avidan, G. Brostow, M. Cissé, G. M. Farinella, and T. Hassner, eds., *Computer Vision – ECCV 2022*, Tel Aviv, Israel, 2022, LNCS 13692, pp. 139–155. Springer.
- [7] P. Wang, L. Liu, Y. Liu, C. Theobalt, T. Komura, and W. Wang. NeuS: Learning neural implicit surfaces by volume rendering for multi-view reconstruction. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P. S. Liang, and J. Wortman Vaughan, eds., *Advances in Neural Information Processing Systems 34 (NeurIPS 2021)*, Virtual, 2021, pp. 27171–27183. Curran Assoicates, Inc.
- [8] T. Wu, J. Wang, X. Pan, X. Xu, C. Theobalt, Z. Liu, and D. Lin. *Voxurf: Voxel-based Efficient and Accurate Neural Surface Reconstruction*, 2022. arXiv: 2208.12697.
- [9] Y. Xiangli, L. Xu, X. Pan, N. Zhao, A. Rao, C. Theobalt, B. Dai, and D. Lin. BungeeNeRF: Progressive neural radiance field for extreme multi-scale scene rendering. In S. Avidan, G. Brostow, M. Cissé, G. M. Farinella, and T. Hassner, eds., *Computer Vision – ECCV 2022*, Tel Aviv, Israel, 2022, LNCS 13692, pp. 106–122. Springer.
- [10] T. Yenamandra, A. Tewari, N. Yang, F. Bernard, C. Theobalt, and D. Cremers. *HDSDF: Hybrid Directional and Signed Distance Functions for Fast Inverse Rendering*, 2022. arXiv: 2203.16284.

## 32.11 Generative Models

*Coordinators: Xingang Pan and Christian Theobalt*

*Investigators: Gereon Fox, Kripasindhu Sarkar, Ikhsanul Habibie, Dushyant Mehta, Lingjie Liu, Mallikarjun B R, Xingang Pan, Noshaba Cheema, Rishabh Dabral, Muhammad Hamza Mughal, Mohamed Elgharib, Vladislav Golyanik, Gerard Pons-Moll, Hans-Peter Seidel, and Christian Theobalt*

Apart from modeling and rendering of existing scenes, a key problem in visual computing and artificial intelligence (AI) is the creation of new content that looks like plausible variants of real-world scenes that may not actually exist in reality. Towards the goal of leveraging AI to simplify and democratize visual content creation, our group has conducted

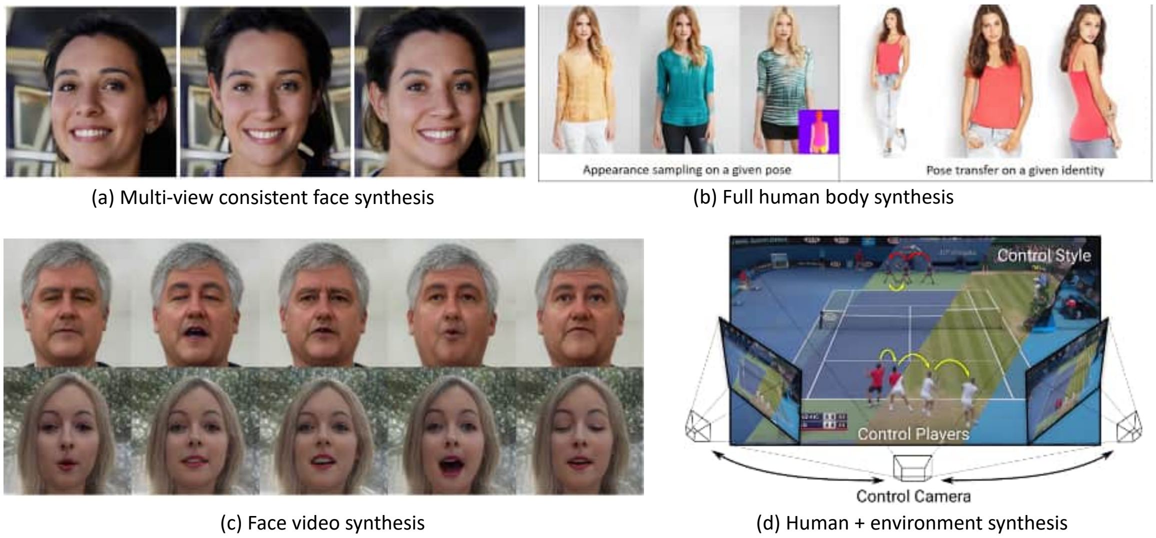



Figure 32.13: Generative models of humans as well as humans and environments.

pioneering research on generative models. Specifically, we have proposed new generative models of humans, which can generate novel digital humans in unseen environments under arbitrary poses at very high quality. To support more flexible motion or gesture synthesis, we also studied neural synthesis, which is conditioned on user inputs such as text and speech.

### 32.11.1 Generative Models of Humans and Humans+Environments

Synthesizing photorealistic humans is an important problem in Computer Graphics and has broad applications such as AR/VR, VFX, and gaming. Our work in the reporting period studied generative models to learn the distribution of human appearance from large-scale datasets. As shown in Fig. 32.13, these generative models can synthesize different and diverse humans as well as their interaction with the environment.

Part of our research particularly focuses on the photorealistic generation of the human *face*. We demonstrated several methods that can synthesize different virtual viewpoints of the human face in a multi-view consistent manner while only using unstructured 2D images for training. Fig. 32.13 (a) shows the results of StyleNeRF [6], which is a hybrid architecture that can synthesize high-resolution images while maintaining multi-view consistency. We achieved this by modeling a low-resolution feature map using neural radiance fields (NeRFs) [13] and then progressively applying upsampling in 2D to increase the resolution. We also studied the disentanglement of 3D shape and appearance in Disentangled3D [15]. This is the first 3D generative model that allows us to modify the shape without affecting appearance or vice versa. The disentanglement of different face parts (e. g., eyes, nose, mouth) is also investigated in gCoRF [11], which supports sampling each part independently. While human faces were the predominant scene type studied in these works, these methods are also applicable to model other types of objects.

We also proposed a generative model, HumanGAN [14], for synthesizing images of dressed humans providing control over pose, local body part appearance, and garment style, as shown in Fig. 32.13 (b). This model addresses the limitations of existing generative adversarial networks that do not allow convenient control over semantically relevant individual parts of the image and cannot draw samples that differ in partial aspects. The proposed model encodes part-based latent appearance vectors in a normalized pose-independent space and warps them to different poses, which allows it to preserve body and clothing appearance under varying posture. The model outperforms task-specific baselines in terms of realism and output resolution.

While the aforementioned methods only synthesize a static human, in StyleVideoGAN [2], we introduced a method to generate plausible animations of unseen talking heads. Given a minimal training set of only ten minutes of footage of a single person, StyleVideoGAN learns to generate motion trajectories of arbitrary duration for this person, as illustrated in Fig. 32.13 (c). In addition, leveraging the StyleGAN latent space [9, 10], StyleVideoGAN can transfer these motion trajectories to unseen, randomly generated faces that do not exist in reality, or even to reconstructions of real people. We achieve this by embedding the training footage into the latent space of a pre-trained StyleGAN, which not only makes our training procedure considerably more efficient than those of previous methods, but also allows us to exploit the geometric properties of the StyleGAN latent space.

Humans often interact with environments. In Playable Environments [12], we present a method that synthesizes both a human and its environment, while also supporting user interaction, as shown in Fig. 32.13 (d). With a single image at inference time, our method allows the user to move objects in 3D while generating a video by providing a sequence of desired actions, offering an experience similar to playing a video game. The actions are learnt in an unsupervised manner from monocular videos only. Based on a 3D neural radiance field representation [13], the camera can be controlled to render the desired viewpoint. Playable Environments enables several creative applications not attainable by prior video synthesis works, including playable 3D video generation, stylization, and manipulation.

### 32.11.2 Synthesis from Text and Speech

In many applications, it is desirable to synthesize content conditioned on specific user inputs, especially for motion and gesture synthesis. As illustrated in Fig. 32.14, our group has studied motion and gesture synthesis from text and speech, which are intuitive and informative user inputs.

We presented a learning-based method for generating animated 3D pose sequences depicting multiple sequential or superimposed actions that are described in long, compositional sentences [3]. Specifically, we propose a hierarchical two-stream sequential model to explore a finer joint-level mapping between natural language sentences and the corresponding 3D pose sequences of the motions. Our method has a two manifold representations of motion – one for upper and one for lower body movements. We evaluated our proposed model on the publicly available KIT Motion-Language Dataset containing 3D pose data with human-annotated sentences and showed that our model advances the state-of-the-art text-based motion synthesis in objective evaluations by a margin of 50%.

In IMoS [5, 4], we presented the first framework to synthesize the full-body motion of


Figure 32.14: Motion synthesis from text and speech with our MoFusion approach [1].

virtual human characters performing specified actions with 3D objects placed within their reach. Our system is conditioned both on the person's action (intention) as well as the object, and outputs diverse sequences of full-body motions. To achieve the objective, we designed an intent-driven full-body motion generator, which uses a pair of decoupled conditional variational autoregressor networks for synthesizing the arm movements and the movements of the rest of the body. Our method also performs optimization for the positions of the objects with six degrees of freedom (6DoF) such that they plausibly fit within the hands of the synthesized characters. We evaluated our method on the GRAB dataset and showed that our model advances the state of the art in action-conditioned motion synthesis in both qualitative and objective evaluations.

Many studies have shown that gestures are a crucial non-verbal element in conversations between humans. Incorporating such gestural information when synthesizing the motion of a talking virtual 3D avatar is a vital ingredient to ensure a lifelike generative model. Unfortunately, designing a system that can automatically map human speech input into 3D body gesture is a challenging task due to the ambiguous nature of the problem. To approach this challenge, in Habibie et al. [8], we proposed a Generative Adversarial Network (GAN) based speech-to-gesture system that measures the synthesis quality of the 3D gestures on the basis of real paired audio-gesture data. This strategy prevents the model from learning a speech-to-gesture mapping that produces unnatural and out-of-sync 3D motion results. In addition to a novel algorithm, this work also introduces a new large-scale speech-gesture dataset containing 3D annotations for more that 33 hours of talking subjects in an in-the-wild setting.

In Habibie et al. [7], we proposed a controllable speech-to-gesture synthesis model. Compared to the earlier work of Habibie et al. [8] that performs direct regression, the new approach allows users to also manipulate the generated outcome based on various control signals, such as the desired hand height, velocity, extent, and even keyword-based motion queries. This is achieved by leveraging the recently popular Motion Matching algorithm to query the most similar speech-gesture features from the database and combining it with a learning-based system using adversarial training to improve the initial queried synthesis results. Since it relies on database matching to find the optimal results, gesture manipulation can be achieved by restricting the search space without the need for dedicated training models for each signal type.

This is followed up with the proposal of MoFusion [1], which introduces a framework for performing multimodal conditional motion synthesis (Fig. 32.14). MoFusion leverages the power of Denoising Diffusion Probabilistic Models (DDPM) to learn the distribution of human motion while also allowing for diversity in sampling for tasks like text-to-motion and music-to-dance synthesis. Further, it proposes a variable weighting schedule to integrate plausibility constraints into the diffusion framework.

## References


• [1] R. Dabral, M. H. Mughal, V. Golyanik, and C. Theobalt. *MoFusion: A Framework for Denoising-Diffusion-based Motion Synthesis*, 2022. arXiv: 2212.04495.

• [2] G. Fox, A. Tewari, M. Elgharib, and C. Theobalt. StyleVideoGAN: A temporal generative model using a pretrained StyleGAN. In *The 32nd British Machine Vision Conference (BMVC 2021)*, Virtual Conference, 2021. BMVA Press.

• [3] A. Ghosh, N. Cheema, C. Oguz, C. Theobalt, and P. Slusallek. Synthesis of compositional animations from textual descriptions. In *ICCV 2021, IEEE/CVF International Conference on Computer Vision*, Virtual Event, 2021, pp. 1376–1386. IEEE.

• [4] A. Ghosh, N. Cheema, C. Oguz, C. Theobalt, and P. Slusallek. Text-based motion synthesis with a hierarchical two-stream RNN. In *ACM SIGGRAPH 2021 Posters*, Virtual Event, USA, 2021, Article 42. ACM.

• [5] A. Ghosh, R. Dabral, V. Golyanik, C. Theobalt, and P. Slusallek. *IMoS: Intent-Driven Full-Body Motion Synthesis for Human-Object Interactions*, 2022. arXiv: 2212.07555.

• [6] J. Gu, L. Liu, P. Wang, and C. Theobalt. StyleNeRF: A style-based 3D aware generator for high-resolution image synthesis. In *International Conference on Learning Representations (ICLR 2022)*, Virtual, 2022, pp. 1–25. OpenReview.net.

• [7] I. Habibie, M. Elgharib, K. Sarkar, A. Abdullah, S. Nyatsanga, M. Neff, and C. Theobalt. A motion matching-based framework for controllable gesture synthesis from speech. In M. Nandigjav, N. J. Mitra, and A. Hertzmann, eds., *Proceedings SIGGRAPH 2022 Conference Papers Proceedings (ACM SIGGRAPH 2022)*, Vancouver, Canada, 2022, Article 346. ACM.

• [8] I. Habibie, W. Xu, D. Mehta, L. Liu, H.-P. Seidel, G. Pons-Moll, M. Elgharib, and C. Theobalt. Learning speech-driven 3D conversational gestures from video. In *Proceedings of the 21st ACM International Conference on Intelligent Virtual Agents (IVA 2021)*, Virtual Event, Japan, 2021, pp. 101–108. ACM.

[9] T. Karras, S. Laine, and T. Aila. A style-based generator architecture for generative adversarial networks. *CoRR*, abs/1812.04948, 2018.

[10] T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, and T. Aila. Analyzing and improving the image quality of stylegan. *CoRR*, abs/1912.04958, 2019.

• [11] Mallikarjun B R, A. Tewari, X. Pan, M. Elgharib, and C. Theobalt. gCoRF: Generative compositional radiance fields. In *International Conference on 3D Vision*, Hybrid / Prague, Czechia, 2022, pp. 567–576. IEEE.

- [12] W. Menapace, S. Lathuilière, A. Siarohin, C. Theobalt, S. Tulyakov, V. Golyanik, and E. Ricci. Playable environments: Video manipulation in space and time. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2022)*, New Orleans, LA, USA, 2022, pp. 3574–3583. IEEE.

[13] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. In *European Conference on Computer Vision (ECCV)*, 2020.

- [14] K. Sarkar, L. Liu, V. Golyanik, and C. Theobalt. HumanGAN: A generative model of humans images. In *2021 International Conference on 3D Vision*, Virtual, 2021, pp. 258–267. IEEE.

- [15] A. Tewari, Mallikarjun B R, X. Pan, O. Fried, M. Agrawala, and C. Theobalt. Disentangled3D: Learning a 3D generative model with disentangled geometry and appearance from monocular images. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2022)*, New Orleans, LA, USA, 2022, pp. 1506–1515. IEEE.


## 32.12 Robust World Perception and Recognition

*Coordinator: Adam Kortylewski*

*Investigators: Adam Kortylewski, Fangneng Zhan, Christian Theobalt*

Over the last decade, we have seen a tremendous increase in the performance of computer vision systems due to advances in deep learning. However, Deep Neural Networks (DNNs) are still far from reaching human-level performance at visual recognition tasks. The most important limitation of DNNs is that they fail to give reliable predictions in unseen or adverse viewing conditions, which would not fool a human observer. They are unreliable when objects are partially occluded, seen in an unusual pose or context, or in bad weather. This lack of robustness in DNNs is generally acknowledged but largely remains unsolved.

### 32.12.1 Benchmarking Out-of-Distribution Robustness

One reason for the limited progress in OOD generalization of DNNs is the lack of benchmark datasets that are specifically designed to measure OOD robustness. With OOD-CV [14] we introduced the very first dataset for benchmarking OOD robustness on real images with annotations of individual nuisance variables and labels for several vision tasks (image classification, detection and 3D pose estimation). This work was selected as oral presentation at ECCV 2022. Moreover, an extended version with a three times larger dataset and many additional experiments and analyses is now under review at PAMI [13]. Based on this dataset, we also organized the “Workshop and Challenge on Out-of-Distribution Generalization in Computer Vision” with more than 40 teams participating in the challenge.

In the context of Visual Question Answering (VQA), current methods are typically developed on standard benchmarks with the implicit assumption that testing data comes from the same underlying distribution as training data. To enable out-of-distribution benchmarking of VQA models, we introduced SuperCLEVR [6], a virtual benchmark where different factors in VQA domain shifts can be isolated to study domain shifts in these factors independently. Based on our study of existing models, we show that disentangling reasoning and perception,

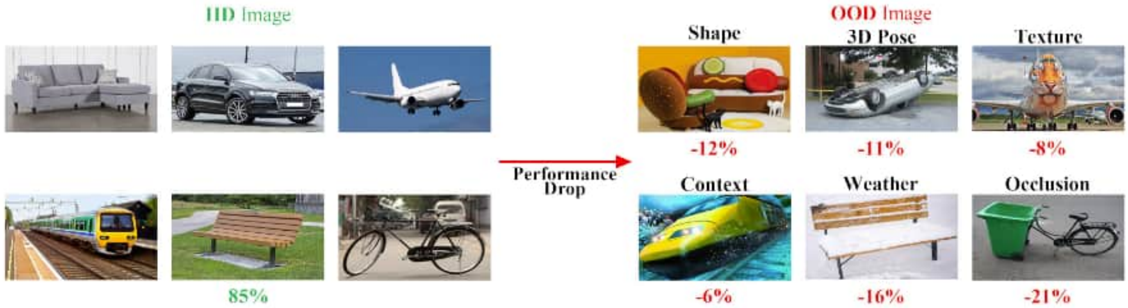



Figure 32.15: Our benchmark [14] made it possible, for the first time, to study the robustness of image classification, object detection, 3D pose estimation and 6D pose estimation to OOD shifts in individual nuisance variables, including OOD changes in shape, pose, texture, context, weather and partial occlusion. We observed that, for example, ResNet50 achieves about 85% accuracy when tested on images that are similarly distributed as the training data (IID). However, the performance deteriorates significantly when individual nuisance factors in the test images break the IID assumption.

and introducing a probabilistic statistical reasoning process, form a strong VQA model that is more robust to domain shifts.

Going beyond fixed datasets for testing algorithms, we built on related work [9] and explored the concept of Adversarial Examination (AE), a strategy to systematically identify the weak points of AI algorithms, in an adversarial manner in order to find weaknesses in the model before deploying it in critical scenarios. AE involves creating test samples on the fly using generative image models that can be controlled in a fine-grained and interpretable manner. New samples are created based on the response of the AI algorithm to previous samples using an efficient search strategy in the latent space of the generative model. Samples that lead to failures can in turn be used to enhance the performance and robustness of the model.

In particular, we extended the concept of AE to human pose and shape estimation (HPS) [7], using a simulator that was controlled in a fine-grained manner to explore the manifold of images of human pose, e.g. by varying poses, shapes, and clothes. We introduced a strategy for exploring the high-dimensional parameter space with a multi-agent reinforcement learning system, in which the agents collaborate to explore different parts of the parameter space. We showed that our PoseExaminer discovered a variety of limitations in current state-of-the-art models that are relevant in real-world scenarios but are missed by current benchmarks. For example, it found large regions of realistic human poses that are not predicted correctly, as well as reduced performance for humans with skinny and corpulent body shapes.

### 32.12.2 Robust Vision through Neural Analysis-by-Synthesis

In contrast to computer vision, human vision is highly robust and generalizes well under occlusion or environmental changes. Cognitive studies suggest that the robustness of the human visual perception arises from the analysis-by-synthesis process [11]. In the past, we developed a number of deep network architectures that follow an analysis-by-synthesis process

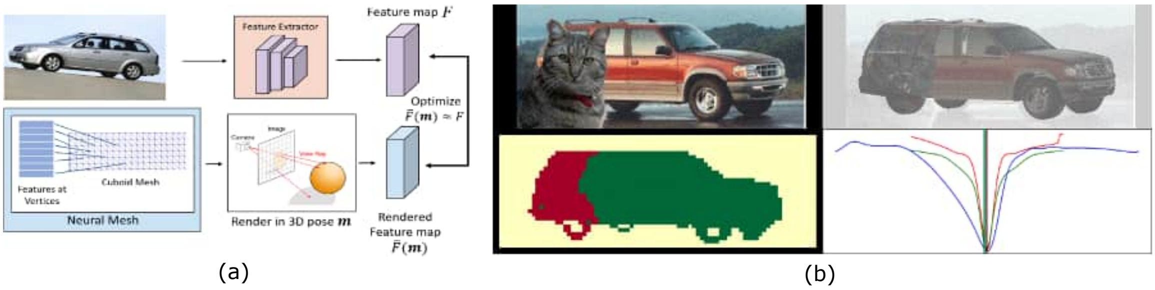



Figure 32.16: (a): Our general scheme of "feature-level analysis-by-synthesis" for the task of pose estimation [8]. An input image is first processed into a feature map F. The 3D pose $m$ of a neural mesh is optimized through differentiable rendering to reconstruct the input features F. (b): Top-left: the input image; Top-right: A mesh superimposed on the input image in the predicted 3D pose. Bottom-left: The occluder localization result, where yellow is background, green is the non-occluded area of the object and red is the occluded area as predicted by our model. Bottom- right: The loss landscape for each of the three camera parameters respectively. The colored vertical lines demonstrate the final prediction and the ground-truth parameter is at center of x-axis.

and achieve strong robustness in out-of-distribution scenarios for 3D pose estimation [10], and have in the last year extended these to 6D pose estimation [8]. The model represents objects as meshes and learns a generative model of the neural feature activations at each mesh vertex (Figure 32.16-a). The key advantage of performing analysis-by-synthesis on neural network features is that these can be trained to be invariant to instance-specific details, which makes the inference process much more efficient and robust compared to image-level generative models (Figure 32.16-b). The key to inducing large convergence basins in the optimization process is a contrastive learning framework that we introduced in the context of keypoint detection [1] to learn features that are invariant to instance-specific details (such as changes in the shape and texture), as well as to changes in the 3D pose and scale of the object.

Moreover, analysis-by-synthesis enables neural networks to learn object-centric representations that enable predicting the object shapes even when they are occluded, i.e. enable amodal object segmentation. In [12] we introduce a model for the detection of partially occluded cell nuclei, where we replace the fully connected classifier in neural networks with a Bayesian generative model of the neural network features. The model is trained from non-occluded images using bounding box annotations, but is applied to generalize out-of-task to object segmentation and to generalize out-of-distribution to segment occluded cell nuclei.

### 32.12.3 Part-based Representations

It is natural to represent objects in terms of their parts. This has the potential to improve the performance of algorithms for object recognition, segmentation or pose estimation but can also help enhance interpretability of deep neural networks as well as robustness when combined with robust outlier models [5].

One main obstacle for research on part-based models is the lack of datasets with per-pixel part annotations. To help address this problem, we introduced PartImageNet [4], a large, high-quality dataset with part segmentation annotations, consisting of 158 classes from ImageNet with approximately 24,000 images. PartImageNet offers part-level annotations on a general set of classes including non-rigid, articulated objects, and can be utilized for many vision tasks including object and part segmentation, few-shot learning and part discovery.

In our work CORL [3] we demonstrated the effectiveness of part-based compositional representations for few-shot learning. In particular, we proposed a neural network architecture that explicitly represents objects as a dictionary of shared components and their spatial composition for few-shot classification. Our results demonstrated the value of this interpretable compositional learning framework on a variety of datasets, where we achieved competitive performance while being more interpretable.

In the context of fine-grained visual classification (FGVC) we proposed a novel transformer-based framework TransFG [2] that leverages the innate multi- head self-attention mechanism in a novel part selection module to find the discriminative regions and compute their spatial relations. This was the first study exploring the potential of vision transformers in the context of fine-grained visual classification, where we achieved state-of-the-art performance.

## References


- [1] Y. Bai, A. Wang, A. Kortylewski, and A. Yuille. CoKe: Contrastive learning for robust keypoint detection. In *2023 IEEE Winter Conference on Applications of Computer Vision (WACV 2023)*, Waikoloa Village, HI, USA, 2023, pp. 65–74. IEEE.
- [2] J. He, J. Chen, S. Liu, A. Kortylewski, C. Yang, Y. Bai, and C. Wang. TransFG: A transformer architecture for fine-grained recognition. In *Proceedings of the 36th AAAI Conference on Artificial Intelligence*, Virtual Conference, 2022, pp. 852–860. AAAI.
- [3] J. He, A. Kortylewski, and A. Yuille. CORL: Compositional representation learning for few-shot classification. In *2023 IEEE Winter Conference on Applications of Computer Vision (WACV 2023)*, Waikoloa Village, HI, USA, 2023, pp. 3879–3888. IEEE.
- [4] J. He, S. Yang, S. Yang, A. Kortylewski, X. Yuan, J. Chen, S. Liu, C. Yang, Q. Yu, and A. L. Yuille. PartImageNet: A large, high-quality dataset of parts. In S. Avidan, G. Brostow, M. Cissé, G. M. Farinella, and T. Hassner, eds., *Computer Vision – ECCV 2022*, Tel Aviv, Israel, 2022, LNCS 13668, pp. 128–145. Springer.

[5] A. Kortylewski, J. He, Q. Liu, and A. L. Yuille. Compositional convolutional neural networks: A deep architecture with innate robustness to partial occlusion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 8940–8949.

- [6] Z. Li, X. Wang, E. Stengel-Eskin, A. Kortylewski, W. Ma, B. Van Durme, and A. Yuille. *Super-CLEVR: A Virtual Benchmark to Diagnose Domain Robustness in Visual Reasoning*, 2022. arXiv: 2212.00259.
- [7] Q. Liu, A. Kortylewski, and A. Yuille. PoseExaminer: Automated testing of out-of-distribution robustness in human pose and shape estimation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2023)*, Vancouver, Canada, 2023. IEEE. Accepted.
- [8] W. Ma, A. Wang, A. L. Yuille, and A. Kortylewski. Robust category-level 6D pose estimation with coarse-to-fine rendering of neural features. In S. Avidan, G. Brostow, M. Cissé, G. M. Farinella, and T. Hassner, eds., *Computer Vision – ECCV 2022*, Tel Aviv, Israel, 2022, LNCS 13669, pp. 492–508. Springer.

[9] M. Shu, C. Liu, W. Qiu, and A. Yuille. Identifying model weakness with adversarial examiner. In *Proceedings of the AAAI conference on artificial intelligence*, 2020.

[10] A. Wang, A. Kortylewski, and A. Yuille. Nemo: Neural mesh models of contrastive features for robust 3d pose estimation. In *International Conference on Learning Representations (ICLR 2021)*, 2021.

[11] A. Yuille and D. Kersten. Vision as bayesian inference: analysis by synthesis? *Trends in cognitive sciences*, 10(7):301–308, 2006.

• [12] Y. Zhang, A. Kortylewski, Q. Liu, S. Park, B. Green, E. Engle, G. Almodovar, R. Walk, S. Soto-Diaz, J. Taube, A. Szalay, and A. Yuille. A light-weight interpretable model for nuclei detection and weakly-supervised segmentation. In Y. Huo, B. A. Millis, Y. Zhou, X. Wang, A. P. Harrison, and Z. Xu, eds., *Medical Optical Imaging and Virtual Microscopy Image Analysis (MOVI 2022)*, Singapore, 2022, LNCS 13578, pp. 145–155. Springer.

• [13] B. Zhao, J. Wang, W. Ma, A. Jesslen, S. Yang, S. Yu, O. Zendel, C. Theobalt, A. Yuille, and A. Kortylewski. *OOD-CV-v2: An extended Benchmark for Robustness to Out-of-Distribution Shifts of Individual Nuisances in Natural Images*, 2023. arXiv: 2304.10266.

• [14] B. Zhao, S. Yu, W. Ma, M. Yu, S. Mei, A. Wang, J. He, A. L. Yuille, and A. Kortylewski. OOD-CV: A benchmark for robustness to out-of-distribution shifts of individual nuisances in natural images. In S. Avidan, G. Brostow, M. Cissé, G. M. Farinella, and T. Hassner, eds., *Computer Vision – ECCV 2022*, Tel Aviv, Israel, 2022, LNCS 13668, pp. 163–180. Springer.

## 32.13 Quantum Visual Computing (QVC)

*Coordinator: Vladislav Golyanik*

*Investigators: Maximilian Krahn, Harshil Bhatia, Edith Tretschk, Christian Theobalt and Vladislav Golyanik*

The last years were accompanied by the transition of quantum algorithm design from a theoretical to an applied discipline. In particular, quantum computer vision and visual computing are nowadays recognized in the research community as new and important research fields, and the interest in them is growing.

### 32.13.1 Quantum Hardware and Specifics of the Field

Visual computing is an applied research area, and our focus lies on adiabatic quantum computers (AQC) since modern realizations of the latter (*i.e.,* D-Wave Advantage with >5000 qubits) allow experimental evaluation of the developed methods on real quantum hardware (also with problems of sizes relevant in practical applications). AQC require the target problem to be provided in the form of a quadratic unconstrained binary optimization (QUBO) problem. In comparison, gate-based quantum computers offer a universal set of operations on qubits but can, as of 2023, effectively execute only short (in terms of the total number of operations) algorithms with a few qubits.

During the reporting period, we investigated the randomness properties of superconducting qubits used in the D-Wave quantum annealers [2]. This work was motivated by 1) Our endeavour to better understand low-level aspects of the target quantum hardware we use

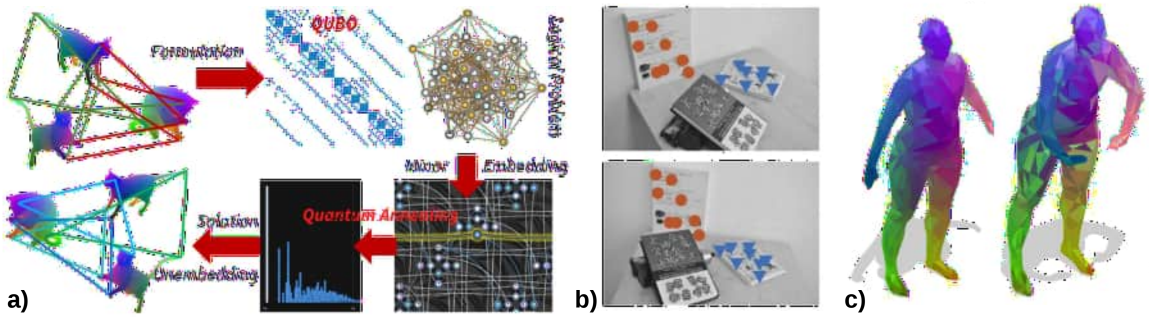



Figure 32.17: (a): A common scheme of "one-sweep" methods relying on QC, shown here with the example of permutation synchronization [3]; (b): Independent motions (disks and triangles) segmented by the approach of Arrigoni *et al.* [1]; (c): Results of Q-Match [7] for aligning two human shapes; the same colours on both meshes encode the correspondences.

and 2) Multiple applications of random number generation in visual computing such as randomized sampling, Monte Carlo rendering and initialization of neural network weights. Even though straightforward in theory, generating truly random numbers on modern quantum hardware is far from trivial. We observed that multiple aspects have to be considered to successfully generate truly random numbers, such as digital-to-analogue quantization errors, flux errors, temperature errors, and spin bath polarization. We also analysed how qubit bias errors propagate into the output bias.

Regarding the method classification, most existing QCV techniques can be categorized into two types, *i.e.,* so-called "one-sweep" methods and hybrid (iterative) approaches. The one-sweep methods first prepare a problem on the CPU in the form admissible to the quantum hardware (QUBO); see Fig. 32.17-(a). Next, the objective is sampled on a quantum annealer, and candidate solutions to the target problem corresponding to different energy levels are obtained. In contrast, hybrid methods iterate between the preparation of the objective on the CPU and sampling it on a quantum annealer. Iterating between a CPU and quantum processing unit (QPU) is often necessary because the entire problem has to be decomposed to fit on the quantum hardware or because the method itself is iterative and requires a sequence of updates towards an optimal solution. In the following, we first summarise "one-sweep" synchronization approaches and then discuss our hybrid techniques.

## 32.13.2 "One-Sweep" QVC Approaches

Our Q-Sync [3] is the first approach for permutation synchronization using AQC. Its objective is to consolidate input pairwise (and, perhaps, noisy) matches into globally-consistent correspondences for multiple sets of input features (extracted on images or shapes). The inputs and the solution are both encoded as sets of permutations, either pairwise or global. The global permutations (representing the solution) are obtained after minimising a QUBO objective with linear constraints imposing the correct structure of the permutation matrices. Q-Sync obtains state-of-the-art results for small problems that can be mapped on modern quantum hardware. The reviewed paper motivated several further works described below.

Our work by Arrigoni and colleagues introduced the first quantum approach for motion segmentation called *QuMoSeg* [1]. The goal of motion segmentation is to classify input two-frame matches between points in multiple images into different independent motions (Fig. 32.17-(b). Applications of motion segmentation include structure from motion and vehicle motion estimation. QuMoSeg formulates the problem in the framework of synchronization (in this case, of relative segmentations) in the AQC admissible form (QUBO). Compared to Q-Sync [3], QuMoSeg can solve larger problems, since the size of the QUBO weight matrix grows linearly in the number of the input points (if assuming a known number of points per motion in every image, even slightly larger problems can be mapped to an AQC). QuMoSeg reaches competitive accuracy on a wide range of problems, both when applying simulated and quantum annealing to sample the objectives.

### 32.13.3 Iterative QVC Approaches

Seelbach Benkner *et al.* proposed Q-Match [7], *i.e.,* a new hybrid approach for aligning two meshes using cyclic alpha expansion; see Fig. 32.17-(c). Q-Match encodes correspondences with permutation matrices. However, in contrast to previous works that impose linear constraints to maintain the structure of the permutations [3], we guarantee in Q-Match valid permutations per construction in every update step. Aligning comparably large meshes (*e.g.,* with 500 points) is possible with Q-Match by decomposing the problem into smaller QUBO sub-problems (with up to 50 vertices to be aligned). Q-Match significantly outperforms the previous quantum state of the art and achieves accuracy comparable to classical methods.

Next, we developed a new version of the Frank-Wolfe (FW) algorithm that leverages AQC [8]. FW is an iterative method for constrained convex optimization that updates the solution in every iteration toward the optimum of the linearized objective [4]. Quantum Frank-Wolfe (Q-FW) is a hybrid solver tailored for binary optimization problems subject to linear (in)equality constraints. Its main idea is to reformulate QUBO with constraints as a copositive program, which is then solved using FW iteration while satisfying the constraints. This has the advantage that the regularization weight of the linear term does not need to be set explicitly. The original constrained quadratic binary optimization problem is transformed into a sequence of QUBOs that are solved on an AQC. Q-FW outperforms previous quantum methods for graph matching [5] and permutation synchronization [3].

### 32.13.4 QUBO Learning for Quantum Annealers

So far, obtaining QUBO forms for computer vision problems (and other domains) required problem-specific knowledge and analytical derivations [5, 3, 7, 1]. The latter inevitably impose restrictions on the choice of the solution encodings for the target problems. In contrast to existing works, our most recent paper proposes to learn QUBO forms from data [6]. Thus, our QuAnt approach by Seelbach Benkner *et al.* can be trained to instantiate QUBOs for different problems such as graph matching, point cloud alignment and transformation estimation. The achieved experimental results are competitive with the existing quantum state of the art, while the method has the advantages of compact encodings and higher accuracy on noisy inputs. QuAnt also generalises better on the test sets than the baseline trained to directly regress the solutions, which further demonstrates the usefulness of integrating a combinatorial quantum annealing solver into a neural network as a custom neural network layer.

### References


- [1] F. Arrigoni, W. Menapace, M. Seelbach Benkner, E. Ricci, and V. Golyanik. Quantum motion segmentation. In S. Avidan, G. Brostow, M. Cissé, G. M. Farinella, and T. Hassner, eds., *Computer Vision – ECCV 2022*, Tel Aviv, Israel, 2022, LNCS 13689, pp. 506–523. Springer.
- [2] H. Bhatia, E. Tretschk, C. Theobalt, and V. Golyanik. Generation of truly random numbers on a quantum annealer. *IEEE Access*, 10:112832–112844, 2022.
- [3] T. Birdal, V. Golyanik, C. Theobalt, and L. Guibas. Quantum permutation synchronization. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2021)*, Virtual Conference, 2021, pp. 13122–13133. IEEE. Accepted.

[4] M. Frank and P. Wolfe. An algorithm for quadratic programming. *Naval Research Logistics Quarterly*, 3(1-2):95–110, 1956.

[5] M. Seelbach Benkner, V. Golyanik, C. Theobalt, and M. Moeller. Adiabatic quantum graph matching with permutation matrix constraints. In *International Conference on 3D Vision (3DV)*, 2020.

- [6] M. Seelbach Benkner, M. Krahn, E. Tretschk, Z. Lähner, M. Moeller, and V. Golyanik. QuAnt: Quantum annealing with learnt couplings. In *Eleventh International Conference on Learning Representations (ICLR 2023)*, Kigali, Rwanda, 2023. OpenReview.net. Accepted.
- [7] M. Seelbach Benkner, Z. Lähner, V. Golyanik, C. Wunderlich, C. Theobalt, and M. Moeller. Q-Match: Iterative shape matching via quantum annealing. In *ICCV 2021, IEEE/CVF International Conference on Computer Vision*, Virtual Event, 2021, pp. 7566–7576. IEEE.
- [8] A. Yurtsever, T. Birdal, and V. Golyanik. Q-FW: A hybrid classical-quantum Frank-Wolfe for quadratic binary optimization. In S. Avidan, G. Brostow, M. Cissé, G. M. Farinella, and T. Hassner, eds., *Computer Vision – ECCV 2022*, Tel Aviv, Israel, 2022, LNCS 13683, pp. 352–369. Springer.


## 32.14 Foundational Methods for Visual Real-World Reconstruction and Artificial Intelligence

*Coordinators: Vladislav Golyanik, Christian Theobalt*

*Investigators: Maximilian Krahn, Lingjie Liu, Vladislav Golyanik and Christian Theobalt*

This section reports on foundational methods for visual computing, reconstruction and machine learning published during the reporting period. These techniques investigate low-level computer vision aspects and often provide building blocks for more complex frameworks. Several of the reviewed techniques that we proposed during the reporting period address correspondence and data association problems. In the following paragraphs, we discuss graph and point set matching, keypoint detection and processing, depth estimation and 3D reconstruction of archaeological sherds.

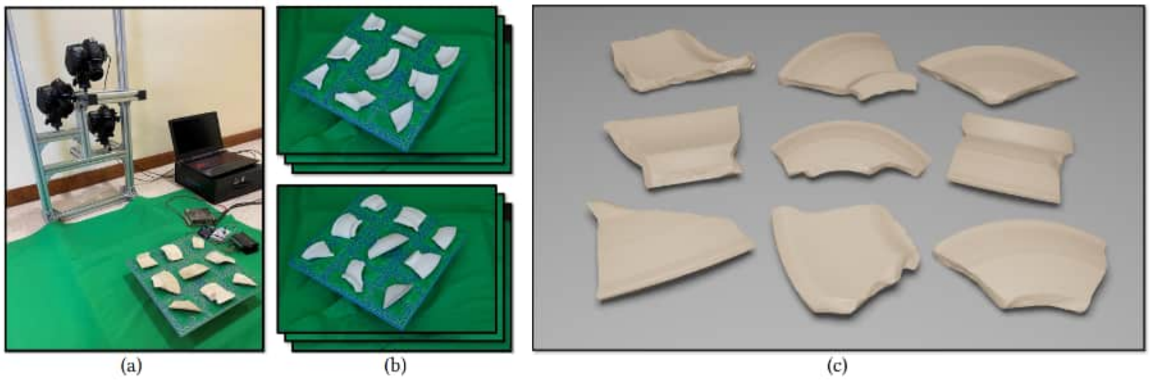


Figure 32.18: Overview of the FIRES capture setup (a, b) and the 3D reconstructions of shards (c) [7]. FIRES is a new system for in-the-field 3D reconstruction of ceramic vessel fragments.

We considered matching and clustering of graphs [3]. In contrast to the majority of the previous methods in the literature, the proposed approach addresses both these problems jointly. Another distinctive property of our approach is the rephrasing of graph matching as rigid point set registration operating on spectral graph embeddings. We conclude that convex semi-definite relaxation of the latter problem leads to an increased efficiency compared to previous graph matching techniques relying on convex relaxations, which is demonstrated experimentally on widely-used datasets. The tests also showed that the developed method with constraints (coupling both problems) can resolve difficult cases that can be ambiguous for methods addressing graph matching and clustering disjointly.

We also looked at the problem of point set registration. In contrast to graphs, points in point sets (which are equivalent to "nodes" in graphs) are given in explicit coordinates and relations between the points (which are "edges" in graphs) are unknown. Our Fast Gravitational Approach (FGA) is a new physically-inspired method for point set registration relying on gravitational particle dynamics [1]. It is specifically designed to operate on possibly noisy and non-uniformly sampled point clouds with clustered outliers (*i.e.,* in which meaningful shape parts do not have valid correspondences in other input shapes). The experiments on point clouds obtained with depth cameras and LIDARs confirm the high registration accuracy of FGA in the above-mentioned cases. FGA can consider alignment cues encoded as particle masses, and the acceleration in FGA is achieved thanks to an octree that allows combining simulated forces exerted by groups of particles.

We further proposed a new method for image correspondence pruning called Laplacian Motion Coherence Network (LMCNet) [4]. LMCNet uses a novel formulation to capture global motion coherence from sparse putative correspondences, which is challenging due to their sparsity and uneven distributions. The proposed network combines global motion coherence and local coherence to detect inlier correspondences and outperforms existing methods in relative camera pose estimation and correspondence pruning of dynamic scenes.

We also developed a new solution to the problem of semi-supervised landmark detection. One of our works proposes a model-agnostic shape-regulated self-training framework for this problem, that takes advantage of the global shape constraint of anatomical landmarks [2]. The proposed framework includes a PCA-based shape model to adjust and eliminate abnormal pseudo labels, and a region attention loss to help the network focus on structure-consistent regions around pseudo labels. The experiments show that the proposed method achieves notable improvements in accuracy on three medical image datasets compared to previous approaches.

Several of our approaches address depth map estimation under different assumptions. Next, our new method for multi-view depth estimation from a single video takes into account the geometric and temporal coherence among the frames [6]. It uses an Epipolar Spatio-Temporal (EST) transformer to explicitly associate geometric and temporal correlation with multiple estimated depth maps and a compact hybrid network consisting of a 2D context-aware network and a 3D matching network to reduce computational cost. The experiments show that the proposed method achieves higher accuracy in depth estimation and significant speedup than the compared methods. Long *et al.* proposed a method for single-image depth estimation, which uses surface normal constraints to effectively correlate depth estimation with geometric consistency by adaptively determining reliable local geometry from a set of randomly sampled candidates [5]. The experiments show that the proposed method outperforms the compared methods in terms of accuracy, efficiency, and robustness.

FIRES is a new system for 3D reconstruction of ceramic vessel fragments or sherds [7] that meets well practical demand of archaeological field research; see Fig. 32.18. Sherds carry rich historical information and, hence, need to be accurately reconstructed for digital analysis. FIRES consists of two main components: A custom-built image acquisition device (*i.e.,* three consumer-grade cameras installed on an aluminium frame) and an automatic multi-view 3D reconstruction pipeline with automatic front-and-back matching and registration to obtain the complete 3D models of fragments from the acquired images. The proposed system has higher throughput (data acquisition efficiency per hour) compared to previous methods. Beyond the archaeological domain, our FIRES paper introduced general concepts allowing for scanning compartmentalized 3D real-world objects and structures.

## References


- [1] S. A. Ali, K. Kahraman, C. Theobalt, D. Stricker, and V. Golyanik. Fast gravitational approach for rigid point set registration with ordinary differential equations. *IEEE Access*, 9:79060–79079, 2021.
- [2] R. Chen, Y. Ma, L. Liu, N. Chen, Z. Cui, G. Wei, and W. Wang. Semi-supervised anatomical landmark detection via shape-regulated self-training. *Neurocomputing*, 471:335–345, 2022.
- [3] M. Krahn, F. Bernard, and V. Golyanik. Convex joint graph matching and clustering via semidefinite. In *2021 International Conference on 3D Vision*, Virtual, 2021, pp. 1216–1226. IEEE.
- [4] Y. Liu, L. Liu, C. Lin, Z. Dong, and W. Wang. Learnable motion coherence for correspondence pruning. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2021)*, Nashville, TN, US (Virtual), 2021, pp. 3236–3245. IEEE.
- [5] X. Long, C. Lin, L. Liu, W. Li, C. Theobalt, R. Yang, and W. Wang. Adaptive surface normal constraint for depth estimation. In *ICCV 2021, IEEE/CVF International Conference on Computer Vision*, Virtual Event, 2021, pp. 12829–12838. IEEE.

- [6] X. Long, L. Liu, W. Li, C. Theobalt, and W. Wang. Multi-view depth estimation using epipolar spatio-temporal networks. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2021)*, Virtual Conference, 2021, pp. 8254–8263. IEEE.
- [7] J. Wang, C. Zhang, P. Wang, X. Li, P. J. Cobb, C. Theobalt, and W. Wang. *FIRES: Fast Imaging and 3D Reconstruction of Archaeological Sherds*, 2022. arXiv: 2211.06897.


## 32.15 Software and Datasets

In the reseaerch area of our department, reference datasets for training and evaluation are essential. Over many years, Christian Theobalt and his research team have created and shared with the research community lare corpora of datasets that are widely used. The team also has released codebases and reference implementations of the algorithms in many of its research papers. We make all the codebases and data availale through a central assets website, which can be reached from the main website of the Visual Computing and Artificial Intelligence Department. Many of the codebases and datasets we have produced and made available have not only stimulated further research but also have become established benchmarks in the field. We made available resources from all areas of our research, but are most recognized for the datasets we have created in the area of marker-less capture, modeling, and animation of humans, human faces, human hands, and humans in interaction with the enviroment. The following are example resources we added to our assets page during the reporting period. Examples of the projects with released datasets are DynaCap [6] (multi-view recordings of five human performances), NeRF-OSR [10] (footage of eight outdoor sites captured from various viewpoints using a DSLR camera along with spherical environmental maps), EventHands [10] (rendered synthetic images, synthetic events, synthetic 3D hand shapes and synthetic 3D hand poses for training neural networks to regress 3D hand shapes from events) and $\phi$-SfT [3] (multiple 3D states of deformable clothes to study the shape-from-template problem). Together with D2 (Gerard Pons-Moll), we also released the BEHAVE dataset of captured human-object interactions [2]. The following are examples of released source codes for papers from the reporting period: MoCapDeform [5], UnrealEgo [1], NeRF-OSR [9], PIP [13], $\phi$-SfT [3], Playable Environments [8], Neural Radiance Transfer Fields [7], Neural Point Catacaustics [4], Non-Rigid NeRF [12], EventHands [10], and Q-Match [11].

In this context, it is important to note that the new Real Virtual Lab facilities (Sec. 32.16) that we are building are not only essential for us to greatly advance the state-of-the-art by our own research. The labs will also be a key factor for us to enable us to continue our tradition to make datasets of highest quality available for open academic research of the research community at large.

In addition to making the results of our basic research broadly available, some of our research results have empowered commercial applications. Christian Theobalt is co-founder of the Captury GmbH which is selling a world-leading technology for markerless full-body motion capture which is the outcome of many years of basic research in his team. Further, face capture and neural face rendering techniques developed in the department were important elements in the development of a groundbreaking new tecnology for neural visual dubbing and neural visual editing of faces in movies developed by the company Flawless.AI. The department and Flawless.AI have been collaborating on basic research for neural face capture and modeling.

## References


- [1] H. Akada, J. Wang, S. Shimada, M. Takahashi, C. Theobalt, and V. Golyanik. UnrealEgo: A new dataset for robust egocentric 3D human motion capture. In S. Avidan, G. Brostow, M. Cissé, G. M. Farinella, and T. Hassner, eds., *Computer Vision – ECCV 2022*, Tel Aviv, Israel, 2022, LNCS 13666, pp. 1–17. Springer.
- [2] B. L. Bhatnagar, X. Xie, I. Petrov, C. Sminchisescu, C. Theobalt, and G. Pons-Moll. BEHAVE: Dataset and method for tracking human object interactions. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2022)*, New Orleans, LA, USA, 2022, pp. 15914–15925. IEEE.
- [3] N. Kairanda, E. Tretschk, M. Elgharib, C. Theobalt, and V. Golyanik. $\phi$-SfT: Shape-from-template with a physics-based deformation model. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2022)*, New Orleans, LA, USA, 2022, pp. 3938–3948. IEEE.
- [4] G. Kopanas, T. Leimkühler, G. Rainer, C. Jambon, and G. Drettakis. Neural point catacaustics for novel-view synthesis of reflections. *ACM Transactions on Graphics (Proc. ACM SIGGRAPH Asia)*, 41(6), Article 201, 2022.
- [5] Z. Li, S. Shimada, B. Schiele, C. Theobalt, and V. Golyanik. MoCapDeform: Monocular 3D human motion capture in deformable scenes. In *International Conference on 3D Vision*, Hybrid / Prague, Czechia, 2022, pp. 1–11. IEEE.
- [6] L. Liu, M. Habermann, V. Rudnev, K. Sarkar, J. Gu, and C. Theobalt. Neural actor: Neural free-view synthesis of human actors with pose control. *ACM Transactions on Graphics (Proc. ACM SIGGRAPH Asia)*, 41(6), Article 219, 2022.
- [7] L. Lyu, A. Tewari, T. Leimkühler, M. Habermann, and C. Theobalt. Neural radiance transfer fields for relightable novel-view synthesis with global illumination. In S. Avidan, G. Brostow, M. Cissé, G. M. Farinella, and T. Hassner, eds., *Computer Vision – ECCV 2022*, Tel Aviv, Israel, 2022, LNCS 13677, pp. 153–169. Springer.
- [8] W. Menapace, S. Lathuilière, A. Siarohin, C. Theobalt, S. Tulyakov, V. Golyanik, and E. Ricci. Playable environments: Video manipulation in space and time. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2022)*, New Orleans, LA, USA, 2022, pp. 3574–3583. IEEE.
- [9] V. Rudnev, M. Elgharib, W. A. P. Smith, L. Liu, V. Golyanik, and C. Theobalt. NeRF for outdoor scene relighting. In S. Avidan, G. Brostow, M. Cissé, G. M. Farinella, and T. Hassner, eds., *Computer Vision – ECCV 2022*, Tel Aviv, Israel, 2022, LNCS 13676, pp. 615–631. Springer.
- [10] V. Rudnev, V. Golyanik, J. Wang, H.-P. Seidel, F. Mueller, M. Elgharib, and C. Theobalt. EventHands: Real-time neural 3D hand pose estimation from an event stream. In *ICCV 2021*, Virtual Event, 2021, pp. 12365–12375. IEEE.
- [11] M. Seelbach Benkner, Z. Lähner, V. Golyanik, C. Wunderlich, C. Theobalt, and M. Moeller. Q-Match: Iterative shape matching via quantum annealing. In *ICCV 2021, IEEE/CVF International Conference on Computer Vision*, Virtual Event, 2021, pp. 7566–7576. IEEE.

- [12] E. Tretschk, A. Tewari, V. Golyanik, M. Zollhofer, C. Lassner, and C. Theobalt. Non-rigid neural radiance fields: Reconstruction and novel view synthesis of a dynamic scene from monocular video. In *ICCV 2021, IEEE/CVF International Conference on Computer Vision*, Virtual Event, 2021, pp. 12939–12950. IEEE.
- [13] X. Yi, Y. Zhou, M. Habermann, S. Shimada, V. Golyanik, C. Theobalt, and F. Xu. Physical Inertial Poser (PIP): Physics-aware real-time human motion tracking from sparse inertial sensors. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2022)*, New Orleans, LA, USA, 2022, pp. 13157–13168. IEEE.

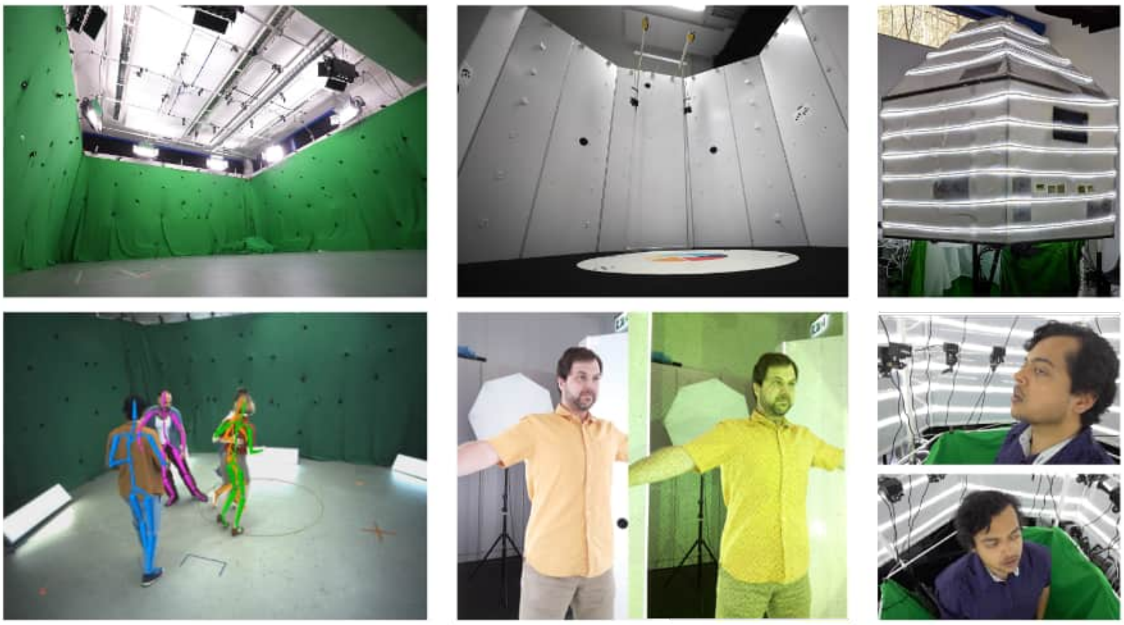

## 32.16 Real Virtual Lab

We are currently building a unique lab space, which we call *Real Virtual Lab (RVL)* that will host world leading capture and display facilities for our research. In its final stage of construction, the lab will feature 350 sqm of space. Currently, the lab comprises 3 rooms hosting worldwide unique and state-of-the-art capture systems. By end of 2024, the lab will be further expanded by a lab hall extension to the MPI for Informatics. In particular, we designed and built several multi-camera systems allowing us to capture real world scenes at very high fidelity. Such data not only provides ground truth measurements in order to validate our algorithms, but it also provides the necessary training data for the neural-explicit methods, which we are developing. In the following, we explain some of the setups in more depth.

**Multi-view Video Studio.** In 2021, we finished the construction of our new multi-view video studio, which offers 36sqm of space allowing us to capture individual people or even multiple persons at the same time (see Fig. 32.19 a)). The system features 120 machine vision cameras that can stream 4K videos at 50fps. Due to the immense amount of data, we designed a dedicated streaming mechanism, which connects the cameras with 10 servers. For further data processing, there are also 8 additional GPU servers that reconstruct for example dense 3D geometry from the raw multi-view images. Over the reporting period, we released multiple datasets with this setup, for example the *DynaCap* dataset, which is one of the most challenging benchmarks for neural human rendering. The multi-view video studio is a versatile facility, which in addition to the main camera system also hosts complementary camera systems, for example for real-time marker-based and marker-less motion capture.

**Body and Object Scanner.** Our body scanner features a dense 140 RGB camera and light projector system to phrotogrammetically reconstruct shape and texture at high detail (see Fig. 32.19 b)). In combination with our recent works on neural implicit geometry recovery, it enables us to reconstruct high-fidelity 3D geometry and appearance of static humans, but also of large objects. For instance, we recorded the upper-body dataset for the i3DMM project [8], which includes 64 subjects performing ten different facial expressions.

**Face Capture Rig.** We further designed a multi-camera setup for recording the human head in motion (see Fig. 32.19 c)). The setup is equipped with 24 video cameras capable of capturing at 4K resolution. The cameras are mounted on a sphere-like structure, which



a) b) c)

Figure 32.19: The Real Virtual Lab hosts multiple state-of-the-art multi-camera and scanning systems, including a large multi-view video studio (a)), a large static scanner (b)), and a face capture rig (c)).

is sitting on several metal supporting stands and which also features a diffuse LED based backlighting system. The cameras are hardware synced which enables capturing multiple synchronized videos from the different viewpoints. Recently, we have applied our methods for photorealistic free-view point rendering of head avatars leveraging the data from this hardware setup.

**Planned Further Extensions** These setups have enabled us to investigate fundamental research questions with a new level of data quality, but there are certain additional modalities of data for which we will build additional specialized setups as part of the further expansion of the laboratory. Precisely, we are currently building a large 3.8 meter multi-spectral lightstage, which is a spherical dome hosting controllable lights and cameras at the same time, enabling time-multiplexed recording of multi-view video under controlled illumination. It allows precise dynamic geometry capture, as well as measurement of the light transport and open entirely new possibilities to exploring neuro-explicit models for static and dynamic scenes that take into account detailed light-surface-camera interactions.

Further on, we are extending the laboratory space with a new lab hall building that will feature a ceiling height of more than 6m and thus enable us to build new larger scale setups which cannot be constructed in the rooms available in the current MPI for Informatics building.

Figure 32.20: Opening of the VIA center in November 2022 – from left to right: P. Druschel (MPI-SWS, Section Chair CPT Section of MPG); Christian Theobalt (MPI-INF), Maximilian Prugger (Dept. Secretary General, MPG), Shahram Izadi (Senior Director and Head of AR, Google), Anke Rehlinger (Minister-President, Saarland); Jakob von Weizsäcker (State Secretary of Science, Saarland); Manfred Schmitt (President of Saarland University), Lutz Mache (Public Policy and Government Relations Manager at Google).

## 32.17 The Saarbrücken Research Center for Visual Computing, Interaction and Artificial Intelligence – VIA

The MPI for Informatics and Google have agreed on a long term strategic partnership to conduct basic foundational research in the area of visual computing, interaction and artificial intelligence. As a result of this partnership, the Saarbrücken Research Center for Visual Computing, Interaction and Artificial Intelligence (VIA) was established at the MPI for Informatics with funding from Google. The center was formally opened with an opening ceremony in November 2022. VIA will be directed by Christian Theobalt from the MPI for Informatics. At the VIA center, Google and MPI for Informatics will conduct joint basic research projects. Results will be openly published. With the strategic partnership between MPI for Informatics and Google, a world leading institution of academic research in computer science and a world leading industrial research partner in computer science join forces. the center is in its initial phase, in which it receives funding from Google for five years. It is an entity of the MPI for Informatics, researchers working in the center are employees of MPI for Informatics. From the beginning, the intention is to further extend and expand the collaboration.

The research of the Saarbrücken VIA center will address frontier research challenges in computer graphics, computer vision, and human machine interaction, at the intersection to artificial intelligence and machine learning. It complements and strengthens the research already conducted at MPI for Informatics. The center will investigate profoundly advanced algorithms that, on the one hand, empower advanced functional capabilities in the aforementioned areas, and at the same time provide an advanced level of accountability in the form of enhanced explainability, robustness and trustworthiness. Example directions of research will be: mixed and extended reality, AI for visual computing and simulation, advanced computer vision for real world perception and understanding, foundational algorithms for virtual humans, AI-empowered interaction, immersive spaces and immersive telepresence, neural capture and neural rendering, and human-centric computing and interaction, to name a few.

Research in computer graphics, computer vision, HCI and artificial intelligence also has a long standing tradition of scientific excellence on Saarland Informatics Campus as a whole, which is underlined by the establishment of the new center. Researchers in the new Saarbrücken VIA center will closely collaborate with this long standing excellent research ecosystem in computer graphics, computer vision, HCI and artificial intelligence conducted at partner institutions in Saarbücken, for example the Max-Planck-Institute for Software Systems, Saarland University, the CISPA Helmholtz center, or the German Research Center for Artificial Intelligence (DFKI). The new center is therefore a means for MPI for Informatics to further strengthen and expand its long standing collaborative research in visual computing, HCI and AI with institutions on campus.

## 32.18 Academic Activities

### 32.18.1 Journal Positions

*Christian Theobalt is on the editorial board of*

- *Visual Informatics* (since 2016),
- *IEEE Computer Graphics & Applications* (since 2018),
- *IEEE Transaction on Pattern Analysis and Machine Intelligence* (since 2019),
- *Elsevier Computers and Graphics* (since 2020).

### 32.18.2 Conference and Workshop Positions

#### Membership in Program and Organizing Committees

*Mohamed Elgharib:*

- PC member – *Computer Vision and Pattern Recognition (CVPR)*, virtual, June 2021.
- PC member – *Computer Vision and Pattern Recognition (CVPR)*, June 2023.

*Vladislav Golyanik:*

- Co-Organizer, PC member – *Quantum Computer Vision and Machine Learning Workshop*, June 2023.

- PC member – *Computer Vision and Pattern Recognition (CVPR) Workshop on Event-based Vision*, virtual, June 2021.
- PC member – *Computer Vision and Pattern Recognition (CVPR) Workshop CV4Animals*, virtual, June 2021.
- PC member – *International Conference on Computer Vision (ICCV)*, virtual, October 2021.
- PC member – *Computer Vision and Pattern Recognition (CVPR)*, June 2022.
- PC member – *Computer Vision and Pattern Recognition (CVPR)*, June 2023.
- PC member – *International Conference on Computer Vision (ICCV)*, October 2023.

*Marc Habermann:*

- PC member – Computer Vision and Pattern Recognition (CVPR), virtual, June 2021.
- PC member – *Computer Vision and Pattern Recognition (CVPR) Workshop CV4Animals*, virtual, June 2021.
- PC member – *International Conference on Computer Vision (ICCV)*, virtual, October 2021.
- PC member – *Computer Vision and Pattern Recognition (CVPR)*, June 2022.
- PC member – *European Conference on Computer Vision (ECCV)*, October 2022.
- PC member – *Computer Vision and Pattern Recognition (CVPR)*, June 2023.
- PC member – *International Conference on Computer Vision (ICCV)*, October 2023.

*Adam Kortylewski:*

- Area Chair – *Advances in Neural Information Processing Systems (NeurIPS)*, 2023.
- PC member – *Workshop on Generative Models for Computer Vision*, June 2023.
- PC member – *Workshop on Out-Of-Distribution Generalization in Computer Vision*, October 2022.
- PC member – *3rd Robust Vision Challenge 2022*, October 2022
- PC member – *3rd Workshop on Adversarial Robustness in the Real World*, October 2022.
- PC member – *Workshop on Deep Learning for Geometric Computing*, June 2022.
- PC member – *Workshop on Socially Responsible Machine Learning*, May 2022.
- PC member – *Computer Vision and Pattern Recognition (CVPR)*, virtual, June 2021.
- PC member – *International Conference on Computer Vision (ICCV)*, virtual, October 2021.
- PC member – *Computer Vision and Pattern Recognition (CVPR)*, June 2022.
- PC member – *European Conference on Computer Vision (ECCV)*, October 2022.
- PC member – *Computer Vision and Pattern Recognition (CVPR)*, June 2023.
- PC member – *International Conference on Computer Vision (ICCV)*, October, 2023.

*Thomas Leimkühler:*

- PC member – *EUROGRAPHICS Short Papers*, May 2021.
- PC member – *High Performance Graphics*, virtual, July 2021.
- PC member – *High Performance Graphics*, virtual, July 2022.
- PC member – *EUROGRAPHICS Short Papers*, May 2023.
- PC member – *Computer Vision and Pattern Recognition (CVPR)*, June 2023.
- PC member – *International Conference on Computer Vision (ICCV)*, October 2023.

*Lingjie Liu:*

- PC member – *Computer Vision and Pattern Recognition (CVPR)*, virtual, June 2021.
- PC member – *International Conference on Computer Vision (ICCV)*, virtual, October 2021.
- PC member – *Computer Vision and Pattern Recognition (CVPR)*, June 2022.
- PC member – *European Conference on Computer Vision (ECCV)*, October 2022.
- PC member – *Computer Vision and Pattern Recognition (CVPR)*, June 2023.
- PC member – *International Conference on Computer Vision (ICCV)*, October 2023.
- PC member – *ACM SIGGRAPH Technical Papers*, Virtual, virtual, August 2022.
- PC member – *Pacific Graphics Technical Papers*, October 2022.
- PC member – *ACM SIGGRAPH Technical Papers*, virtual, August 2023.

*Diogo Luvizon:*

- PC member – *International Conference on Computer Vision (ICCV)*, October 2023.
- PC member – *Computer Vision and Pattern Recognition (CVPR)*, June 2023.
- PC member – *European Conference on Computer Vision (ECCV)*, October 2022.
- PC member – *Computer Vision and Pattern Recognition (CVPR)*, June 2022.
- PC member – *International Conference on Computer Vision (ICCV)*, virtual, October 2021.
- PC member – *Computer Vision and Pattern Recognition (CVPR)*, virtual, June 2021.

*Christian Theobalt:*

- PC member – *SIGGRAPH Asia*, Dec, 2023.
- PC member – *IEEE International Conference on Computer Vision (ICCV)*, Oct, 2023.
- PC member – *German Conference on Pattern Recognition*, Sept, 2023.
- Conflict of Interest Coordinator – *ACM SIGGRAPH*, Aug, 2023.
- PC member – *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2023.
- PC member – *Workshop on New Trends in Image Restoration and Enhancement (NTIRE), collocated with CVPR*, June, 2023.

- Co-Organizer – *Workshop on Generative Models for Computer Vision, collocated with CVPR*, June 2023.
- Co-Chair – *EUROGRAPHICS State of the Art Reports*, May, 2023.
- Co-Organizer – *Workshop on 3D Morphable Models ans Beyond*, hybrid, Schloss Dagstuhl, November 2022.
- PC member – *ECCV Workshop on Advances in Image Manipulaton (AIM)*, Oct, 2022.
- PC member – *Vision, Modeling and Visualization*, Sept, 2022.
- PC member – *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2022.
- PC member – *Workshop on New Trends in Image Restoration and Enhancement (NTIRE), collocated with CVPR*, June, 2022.
- Area Chair – *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, virtual, June 2021.
- Member – *Member EUROGRAPHICS Awards Committee*, since 2021.

*Fangneng Zhan:*

- PC member – *Computer Vision and Pattern Recognition (CVPR)*, virtual, June 2021.
- PC member – *International Conference on Computer Vision (ICCV)*, virtual, October 2021.
- PC member – *Computer Vision and Pattern Recognition (CVPR)*, June 2022.
- PC member – *European Conference on Computer Vision (ECCV)*, October 2022.
- PC member – *Computer Vision and Pattern Recognition (CVPR)*, June 2023.
- PC member – *CVPR Workshop on Generative Models for Computer Vision*, June 2023.
- PC member – *International Conference on Computer Vision (ICCV)*, October 2023.

All members of the department further serve regularly as reviewers for the top conferences and journals in computer graphics, computer vision ad artifical intelligence.

### 32.18.3 Invited Talks and Tutorials

*Mohamed Elgharib:*

- *Neural Reconstruction and Rendering*, Invited Speaker, Aalto University, December 2022.
- *Neural Reconstruction and Rendering*, Invited Speaker, University of York, January 2023.

*Vladislav Golyanik:*

- *3D Computer Vision: From a Classical to a Quantum Perspective*, SIC Lecture Series speaker, Saarland Informatics Campus, September 2021.
- *Trends and Applications of Computer Vision*, Invited lecturer, University of Trento (virtual), November 2021.

- *Overview of 4D and Quantum Vision Research Group*, Speaker at reading group session (Snap Research), virtual, December 2021.
- *Advances in Neural Rendering*, EUROGRAPHICS 2022 STAR tutorial, Reims, France, April 2022.
- *Advances in Quantum Computer Vision*, Invited Speaker, Workshop on Quantum Information, Saarland Informatics Campus, December 2022.

*Marc Habermann:*

- *Real-time Deep Dynamic Characters*, Invited Speaker at Max Planck Institute for Perceiving Systems, Virtual, June 2021.
- *Real-time Deep Dynamic Characters*, Invited Speaker at ETH (AIT), Virtual, July 2021.
- *Real-time Deep Dynamic Characters*, Invited Speaker at Google Zuerich, Virtual, September 2021.
- *Human Performance Capture and Synthesis*, Invited Speaker at Adobe London, Virtual, July 2022.
- *Human Performance Capture and Synthesis*, Invited Speaker at IIIT Hyderabad, Virtual, November 2022.

*Adam Kortylewski:*

- *Generative Computer Vision*, TrustML Young Scientist Seminars, RIKEN, Tokyo, Japan, virtual, March 2023.
- *Robust Computer Vision through Neural Analysis-by-Synthesis*, 2nd International Workshop on Practical Deep Learning in the Wild, AAAI 2023, Washington, USA, February, 2023.
- *Robust Computer Vision through Neural Analysis-by-Synthesis*, Microsoft Research, Redmond, USA, October, 2022.
- *Robust Deep Networks for Computer Vision*, SIC Lecture Series speaker, Saarland Informatics Campus, May, 2022.
- *Computer Vision does not generalize, 3DMMs and beyond can help*, Dagstuhl Seminar on 3D Morphable Models and Beyond,Castle Dagstuhl, Germany, Mar, 2022.

*Thomas Leimkühler:*

- *Exploring the Continuum of Image Synthesis Algorithms*, SIC Lecture Series speaker, Saarland Informatics Campus, August 2021.

*Lingjie Liu:*

- *Fast Rendering of Neural Radiance Fields*, Invited Speaker at SIGGRAPH 2021 Course on Advances in Neural Rendering, August 2021.
- *Neural Actor: Neural Free-view Synthesis of Human Actors with Pose Control*, Invited Speaker at 3DV 2021 Course on Advances in Neural Rendering, November 2021.

- *Neural Rendering of Human Actors*, Invited Speaker at Nanyang Technological University, June 2021.
- *Neural Scene Representations and Neural Rendering*, Invited Speaker at the University of Edinburgh, Facebook AI Research, Google, Peking University, Siemens Healthineers, Baidu Research, Adobe Research, July – October 2021.
- *Neural Representation and Neural Rendering of 3D Real-world Scenes*, Invited Speaker at the University of Toronto, the University of Maryland, the University of Pennsylvania, Imperial College London, University of Southern California, UCSB, Rice University, National University of Singapore, HKUST, The University of North Carolina at Chapel Hill, ETH Zürich, Nanyang Technological University, February – April 2022.

*Diogo Luvizon:*

- *Novel View Synthesis from a Single Image through Multiplane Image Representation*, Invited Speaker, IMAGINE/LIGM École des Ponts Paris Tech, Virtual, January 2022.

*Xingang Pan:*

- *Neural Generation and Rendering of the 3D Visual World*, NAP2023 Research Seminars, Nanyang Technological University, Singapore, February 2023.
- *Image Editing and 3D Reconstruction based on Deep Generative Models*, Invited Speaker at HUAWEI Technologies Co., Ltd, Virtual, April 2022.
- *A Shading-Guided Generative Implicit Model for Shape-Accurate 3D-Aware Image Synthesis*, Invited Speaker at Tsinghua University AI Time, Virtual, March 2022.
- *Recovering the Hidden Visual Dimension via Deep Generative Models*, Invited Speaker at HUAWEI Noah's Ark Lab, Virtual, September 2021.

*Christian Theobalt:*

- *Maschinelles Lernen in der Computergrafik und Bilderkennung – Neuartige Algorithmen, um die reale Welt zu erfassen und darzustellen*, Annual Assembly, GDNAE (Society of German Natural Scientists and Physicians, Academy of Sciences), Sept 2022.
- *Capturing the Real World in Motion: New Ways to Unite Graphics, Vision and Machine Learning*, Join CAP-RFIAP Conference (French National Pattern Recognition and Computer Vision Conference), July, 2022.
- *Neural Methods for Reconstruction ad Rendering of Real World Scenes*, BiGmax Workshop (Big-Data-Driven Material Science, Bochum University, April 2022.
- *Neural Methods for Reconstruction and Rendering of Real World Scenes*, Invited Talk, AI Seminar, virtual, Carnegie Mellon University, Feb 2022.
- *Neural Rendering and Video-based Animation of Human Actors*, Speaker at SIGGRAPH 2021 Course on Advances in Neural Rendering, Aug 2021.
- *Synthesis of Portrait Images with 3D Control*, NTIRE Workshop collocated with CVPR, virtual, June 2021.

*Fangneng Zhan*

- *On the Gauge Transformation of Neural Fields*, Invited Speaker, The AI Talks (NTU, NUS), Virtual, March 2023.

### 32.18.4 Other Academic Activities

*Vladislav Golyanik:*

- Member of the Doctoral Thesis Defence Committee, MPI for Informatics and Saarland University, February 2022.
- IMPRS-TRUST PhD Program Evaluator, 2021-2023.
- ELLIS PhD Program Evaluator, November 2021.
- ELLIS PhD Program Evaluator, November 2022.
- ELLIS member (since January 2023).

*Marc Habermann:*

- ELLIS PhD Program Evaluator, November 2021.
- ELLIS PhD Program Evaluator, November 2022.
- Member of the Doctoral Thesis Defence Committee, University of North Carolina at Chapel Hill, February 2023.

*Christan Theobalt:*

- Since 2023 – Chair of EUROGRAPHICS Awards Committee.
- Since Nov 2021 – Director, Saarbrücken Research Center for Visual Computing, Interaction and Artificial Intelligence (VIA) – result of a strategic partnership between MPI for Informatics ang Google.
- Managing Director, Max-Planck-Institute for Informatics (July 2021-July 2023).
- Since 2021 – Member of EUROGRAPHICS Awards Committee.
- Since 2021 – Member of Selection Committee – Fraunhofer / Max Planck Joint Research Projects.
- Co-Director of new ELLIS program on Learning for Graphics and Vision.
- Principal Investigator – ELLIS Unit SAM, Saarbrücken.
- Panel Chair (PE6) for ERC Consolidator Grants (2021).
- Member of Advisory Board – Bundeswettbewerb Informatik (BWINF, German Federal Computer Science High School Student Competition / Olympiad) (until 2023).
- Advisor to the IT Incubator of Max Planck Society/Saarland University.
- Co-Founder of the Captury GmbH (www.thecaptury.com).
- Reviewer for ERC Starting/Advanced Grants.
- Reviewer for funding proposals submitted to German Ntl. Science Foundation (DFG), Austrian Science Fund (FWF), Swiss National Science Foundation (SNF).

- External Reviewer for PhD theses (only for the period of 2021-2023): Gaspard Zoss (ETH, 2021), Jae Shin Yoon (U. Minnesota, 2022), Adrian Spurr (ETH, 2022).
- Graduate Student Selection Committees of the International Max-Planck Research School for Computer Science and Trust (IMPRS-CS, IMPRS-TRUST), and the Saarbrückecken Graduate School of Computer Science.

## 32.19 Teaching Activities

### Summer Semester 2021

**Seminars:**
Computer Vision and Machine Learning for Computer Graphics (CVMLCG)(C. Theobalt, M. Elgharib, V. Golyanik)

### Summer Semester 2022

**Seminars:**
Computer Vision and Machine Learning for Computer Graphics (CVMLCG)(C. Theobalt, M. Habermann, T. Leimkühler)

### Master and Bachelor Theses

- Mohit Mendiratta: Egocentric Videoconferencing, 2021.
- Maximilian Krahn: Convex Joint Graph Matching and Clustering via Semidefinite Relaxations, 2021.
- Jalees Nehvi: Differentiable Event Stream Simulator for Non-Rigid 3D Tracking, 2021.
- Navami Kairanda: Shape-from-Template with a Physics-Based Deformation Model, 2022.
- Hiroyasu Akada: UnrealEgo: A New Dataset for Robust Egocentric 3D Human Motion Capture, 2022 (co-supervision with Keio University, Japan).
- André Jonas: 3D Tracking of Two Hands from an Event Camera (co-supervision with TU Kaiserslautern)
- Artur Jesslen: Multi-View 3D Hand Reconstruction (co-supervision with EPFL, Switzerland), 2022.
- Eric C. M. Johnson: Monocular 4D Reconstruction with a Neural Deformation Model, 2022.
- Zhouyingcheng Liao: Video-based Neural Character Skinning, 2022.
- Pramod Rao: VoRF: Volumetric Reflectance Fields, 2022.

### Courses and Tutorials

- Advances in Neural Rendering (SIGGRAPH 2021 Course)
- Tutorial on the Advances in Neural Rendering (3DV 2021)

- Advances in Neural Rendering (EUROGRAPHICS 2022 STAR)

## 32.20 Dissertations, Habilitations, Awards

### 32.20.1 Dissertations

#### Completed and Defended

Ayush Tewari: Self-Supervised Reconstruction and Synthesis of Faces, 26.07.2021.

Marc Habermann: Real-Time Human Performance Capture and Synthesis, 29.10.2021.

### 32.20.2 Offers for Faculty Positions

- Mengyu Chu (co-affiliated with D4): Tenure-track Assistant Professor, Peking University, China, 2022.
- Lingjie Liu: Tenure-Track Assistant Professor, University of Pennsylvania, Philadelphia, USA, 2022.
- Xingang Pan: Tenure-Track Assistant Professor, Nanyang Technological University, Singapore, 2023.

### 32.20.3 Awards

- Christian Theobalt: *Fellow of EUROGRAPHICS*, EUROGRAPHICS, 2021.
- Franziska Müller: *Dr. Eduard Martin Award*, Saarland University, 2021.
- Ayush Tewari: *Otto-Hahn-Medal*, Max-Planck-Society, 2021.
- Adam Kortylewski *DFG Emmy Noether Group*, DFG, 2022.
- Marc Habermann: *EG Thesis Award*, Eurographics, 2022.
- Marc Habermann: *DAGM MVTec Dissertation Award 2022*, DAGM, 2022.
- Marc Habermann: *Otto-Hahn-Medal*, Max-Planck-Society, 2022
- Ikhsanul Habibie, Weipeng Xu, Dushyant Mehta, Lingjie Liu, Hans-Peter Seidel, Gerard Pons-Moll, Mohamed Elgharib, Christian Theobalt: *Best Paper Award*, IVA, 2021.
- Xinyu Yi, Yuxiao Zhou, Marc Habermann, Soshi Shimada, Vladislav Golyanik, Christian Theobalt, Feng Xu: *Best Paper Award Finalist*, CVPR, 2022.
- Zhi Li, Soshi Shimada, Bernt Schiele, Christian Theobalt and Vladislav Golyanik: *Best Student Paper Award*, 3DV, 2022.
- Jiayi Wang, Diogo Luvizon, Franziska Müller, Florian Bernard, Adam Kortylewski, Dan Casas and Christian Theobalt: *Best Paper Award Honorable Mention*, VMV, 2022.

- Pramod Rao, Mallikarjun B R, Gereon Fox, Tim Weyrich, Bernd Bickel, Hanspeter Pfister, Wojciech Matusik, Ayush Tewari, Christian Theobalt and Mohamed Elgharib: *Best Paper Award Honorable Mention*, BMVC, 2022.

## 32.21 Grants and Cooperations

- Christian Theobalt, ERC Consolidator Grant 4DRepLy.
- Adam Kortylewski, Emmy Noether Starting Grant.
- Saarbrücken Research Center for Visual Computing, Interaction, and Artificial Intelligence (VIA) – result of a new strategic partnership on basic research between MPI for Informatics and Google (funding support from Google).
- Cooperation with with Patrick Perez (Valeo): Physics-based Pose Estimation (funding support by Valeo.AI).
- Cooperation with Facebook Reality Labs: General Deformable Scene Capture (funding support by Facebook Reality Labs).
- Cooperation with Facebook Reality Labs: Inverse Rendering (funding support by Facebook Reality Labs).
- Cooperation with Flawless.AI: Neural Face Modeling.
- Cooperation with Facebook Reality Labs: Human Performance Capture.
- Cooperation with Philipp Slusallek (Saarland University): Full-Body Motion Synthesis for Human-Object Interactions.
- Cooperation with Michael Möller and Zorah Lähner (The University of Siegen): Quantum Computer Vision and Shape Alignment.
- Cooperation with Alp Yurtsever (Umeå University), Tolga Birdal (ICL) and Leonidas Guibas (Stanford University): Quantum Computer Vision
- Cooperation with Federica Arrigoni (Polytechnic University of Milan), Willi Menapace and Elisa Ricci (University of Trento): Motion Segmentation with Quantum Annealers
- Cooperation with Willi Menapace, Elisa Ricci (University of Trento), Sergey Tulyakov, Aliaksandr Siarohin (Snap Research) and Stéphane Lathuilière (Telecom Paris): Controllable Video Generation.
- Cooperation with Hiroyasu Akada and Masaki Takahashi (Keio University): Egocentric 3D Human Motion Capture
- Cooperation with Pascal Fua (EPFL) and Bernhard Egger (University of Erlangen-Nuremberg): Dense Monocular Non-Rigid 3D Reconstruction.

- Cooperation with Alan Yuille (Johns Hopkins University): Multiple projects in the context of robust computer vision with neuro-explicit models.
- Cooperation with Thomas Vetter (University of Basel) and Bernhard Egger (University of Erlangen-Nuremberg): Robust face reconstruction.
- Cooperation with Alex Szalay (Johns Hopkins University) and Alan Yuille (Johns Hopkins University): Interpretable medical image analysis with compositional models.
- Cooperation with Nataniel Ruiz (Boston University), Cihang Xie (USC), Alan Yuille (Johns Hopkins University) and Stan Sclaroff: Simulated Adversarial Testing of Face Recognition Models.
- Cooperation with Gerard Pons-Moll (U. Tübingen, MPI for Informatics): multiple projects in human capture and modeling.
- Cooperation with Stanford University (Gordon Wetzstein): Neural rendering.
- Cooperation with MIT (Ayush Tewari): neural face modeling, generative modeling.
- Cooperation with Technicolor Rennes (Abdallah Dib, Louis Chevallier): Face capture and modeling.
- Cooperation with Youngjoong Kwon and Henry Fuchs (University of North Carolina at Chapel Hill): Neural human rendering.
- Cooperation with Yue Li (ETH), Bernhard Thomaszewski (ETH), Stelian Coros (ETH), and Thabo Beeler (Google): Physics-aware human performance capture.
- Cooperation with Benoit Guillard and Pascal Fua (EPFL): Band-limited shape learning.
- Cooperation with Wojciech Matusik (MIT CSAIL), Ayush Tewari (MIT CSAIL), Hanspeter Pfister (Harvard University), Tim Weyrich (Friedrich-Alexander-Universität Erlangen-Nürnberg and University College London) and Bernd Bickel (IST-Austria): Facial Relighting.
- Cooperation with Will Smith (University of York): Outdoor Scene Relighting.
- Cooperation with Congyi Zhang (The University of Hong Kong) and Min Gu (The University of Hong Kong) and Wenping Wang (Texas A&M University): 3D Morphable Dental Model.
- Cooperation with Wenping Wang (The University of Hong Kong / Texas A&M University): Multiple projects on neural rendering ad modeling.
- Cooperation with Tarun Yenamandra (Technical University of Munich), Florian Bernard (Technical University of Munich) and Daniel Cremers (Technical University of Munich): 3D Implicit Morphable Face Model.
- Cooperation with Ahsan Abdullah (UC Davis), Simbarashe Nyatsanga (UC Davis) and Michael Neff (UC Davis): Motion Matching for Gesture Synthesis from Speech.

– Cooperation with Moritz Kappel, Jann-Ole Henningson, Susana Castillo (Technische Universität Braunschweig) and Marcus Magnor (Technische Universität Braunschweig): Neural Human Motion Transfer.

– Cooperation with Justus Thies (MPI for Intelligent Systems): Multiple projects on motion simulation and neural face modeling.

## 32.22 Publications

### Journal articles and book chapters


[1] S. A. Ali, K. Kahraman, C. Theobalt, D. Stricker, and V. Golyanik. Fast gravitational approach for rigid point set registration with ordinary differential equations. *IEEE Access*, 9:79060–79079, 2021.

[2] H. Bhatia, E. Tretschk, C. Theobalt, and V. Golyanik. Generation of truly random numbers on a quantum annealer. *IEEE Access*, 10:112832–112844, 2022.

[3] R. Chen, Y. Ma, L. Liu, N. Chen, Z. Cui, G. Wei, and W. Wang. Semi-supervised anatomical landmark detection via shape-regulated self-training. *Neurocomputing*, 471:335–345, 2022.

[4] M. Chu, L. Liu, Q. Zheng, E. Franz, H.-P. Seidel, C. Theobalt, and R. Zayer. Physics informed neural fields for smoke reconstruction with sparse data. *ACM Transactions on Graphics*, 41(4), Article 119, 2022.

[5] M. Chu, N. Thuerey, H.-P. Seidel, C. Theobalt, and R. Zayer. Learning meaningful controls for fluids. *ACM Transactions on Graphics (Proc. ACM SIGGRAPH)*, 40(4), Article 100, 2021.

[6] M. Habermann, L. Liu, W. Xu, M. Zollhöfer, G. Pons-Moll, and C. Theobalt. Real-time deep dynamic characters. *ACM Transactions on Graphics (Proc. ACM SIGGRAPH)*, 40(4), Article 94, 2021.

[7] M. Habermann, W. Xu, M. Zollhöfer, G. Pons-Moll, and C. Theobalt. A deeper look into DeepCap. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(4):4009–4002, 2023.

[8] G. Kopanas, T. Leimkühler, G. Rainer, C. Jambon, and G. Drettakis. Neural point catacaustics for novel-view synthesis of reflections. *ACM Transactions on Graphics (Proc. ACM SIGGRAPH Asia)*, 41(6), Article 201, 2022.

[9] L. Liu, M. Habermann, V. Rudnev, K. Sarkar, J. Gu, and C. Theobalt. Neural actor: Neural free-view synthesis of human actors with pose control. *ACM Transactions on Graphics (Proc. ACM SIGGRAPH Asia)*, 41(6), Article 219, 2022.

[10] J. Malik, A. Elhayek, S. Shimada, S. A. Ali, V. Golyanik, C. Theobalt, and D. Stricker. HandVoxNet++: 3D hand shape and pose estimation using voxel-based neural networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(12):8962–8974, 2021.

[11] Mallikarjun B R, A. Tewari, A. Dib, T. Weyrich, B. Bickel, H.-P. Seidel, H. Pfister, W. Matusik, L. Chevallier, M. Elgharib, and C. Theobalt. PhotoApp: Photorealistic appearance editing of head portraits. *ACM Transactions on Graphics (Proc. ACM SIGGRAPH)*, 40(4), Article 44, 2021.

[12] A. Meka, M. Shafiei, M. Zollhöfer, C. Richardt, and C. Theobalt. Real-time global illumination decomposition of videos. *ACM Transactions on Graphics*, 40(3), Article 22, 2021.

[13] S. Shimada, V. Golyanik, W. Xu, P. Pérez, and C. Theobalt. Neural monocular 3D human motion capture with physical awareness. *ACM Transactions on Graphics (Proc. ACM SIGGRAPH)*, 40(4), Article 83, 2021.

[14] A. Tewari, J. Thies, B. Mildenhall, P. Srinivasan, E. Tretschk, Y. Wang, C. Lassner, V. Sitzmann, R. Martin-Brualla, S. Lombardi, T. Simon, C. Theobalt, M. Nießner, J. T. Barron, G. Wetzstein, M. Zollhöfer, and V. Golyanik. Advances in neural rendering. *Computer Graphics Forum (Proc. EUROGRAPHICS)*, 41(2):703–735, 2022.

[15] W. Wan, L. Yang, L. Liu, Z. Zhang, R. Jia, Y.-K. Choi, J. Pan, C. Theobalt, T. Komura, and W. Wang. Learn to predict how humans manipulate large-sized objects from interactive motions. *IEEE Robotics and Automation Letters*, 7(2):4702–4709, 2022.

[16] C. Zhang, M. Elgharib, G. Fox, M. Gu, C. Theobalt, and W. Wang. An implicit parametric morphable dental model. *ACM Transactions on Graphics (Proc. ACM SIGGRAPH ASIA)*, 41(6), Article 217, 2022.

## Conference articles

[1] H. Akada, J. Wang, S. Shimada, M. Takahashi, C. Theobalt, and V. Golyanik. UnrealEgo: A new dataset for robust egocentric 3D human motion capture. In S. Avidan, G. Brostow, M. Cissé, G. M. Farinella, and T. Hassner, eds., *Computer Vision – ECCV 2022*, Tel Aviv, Israel, 2022, LNCS 13666, pp. 1–17. Springer.

[2] F. Arrigoni, W. Menapace, M. Seelbach Benkner, E. Ricci, and V. Golyanik. Quantum motion segmentation. In S. Avidan, G. Brostow, M. Cissé, G. M. Farinella, and T. Hassner, eds., *Computer Vision – ECCV 2022*, Tel Aviv, Israel, 2022, LNCS 13689, pp. 506–523. Springer.

[3] Y. Bai, A. Wang, A. Kortylewski, and A. Yuille. CoKe: Contrastive learning for robust keypoint detection. In *2023 IEEE Winter Conference on Applications of Computer Vision (WACV 2023)*, Waikoloa Village, HI, USA, 2023, pp. 65–74. IEEE.

[4] B. L. Bhatnagar, X. Xie, I. Petrov, C. Sminchisescu, C. Theobalt, and G. Pons-Moll. BEHAVE: Dataset and method for tracking human object interactions. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2022)*, New Orleans, LA, USA, 2022, pp. 15914–15925. IEEE.

[5] T. Birdal, V. Golyanik, C. Theobalt, and L. Guibas. Quantum permutation synchronization. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2021)*, Virtual Conference, 2021, pp. 13122–13133. IEEE. Accepted.

[6] R. Dabral, S. Shimada, A. Jain, C. Theobalt, and V. Golyanik. Gravity-aware monocular 3D human-object reconstruction. In *ICCV 2021, IEEE/CVF International Conference on Computer Vision*, Virtual Event, 2021, pp. 12365–12374. IEEE.

[7] A. Dib, C. Thebault, J. Ahn, P.-H. Gosselin, C. Theobalt, and L. Chevallier. Towards high fidelity monocular face reconstruction with rich reflectance using self-supervised learning and ray tracing. In *ICCV 2021, IEEE/CVF International Conference on Computer Vision*, Virtual Event, 2021, pp. 12799–12809. IEEE.

[8] G. Fox, W. Liu, H. Kim, H.-P. Seidel, M. Elgharib, and C. Theobalt. VideoForensicsHQ: Detecting high-quality manipulated face videos. In *IEEE International Conference on Multimedia and Expo (ICME 2021)*, Shenzhen, China (Virtual), 2021, pp. 1–6. IEEE.

[9] G. Fox, A. Tewari, M. Elgharib, and C. Theobalt. StyleVideoGAN: A temporal generative model using a pretrained StyleGAN. In *The 32nd British Machine Vision Conference (BMVC 2021)*, Virtual Conference, 2021. BMVA Press.

[10] A. Ghosh, N. Cheema, C. Oguz, C. Theobalt, and P. Slusallek. Synthesis of compositional animations from textual descriptions. In *ICCV 2021, IEEE/CVF International Conference on Computer Vision*, Virtual Event, 2021, pp. 1376–1386. IEEE.

[11] A. Ghosh, N. Cheema, C. Oguz, C. Theobalt, and P. Slusallek. Text-based motion synthesis with a hierarchical two-stream RNN. In *ACM SIGGRAPH 2021 Posters*, Virtual Event, USA, 2021, Article 42. ACM.

[12] J. Gu, L. Liu, P. Wang, and C. Theobalt. StyleNeRF: A style-based 3D aware generator for high-resolution image synthesis. In *International Conference on Learning Representations (ICLR 2022)*, Virtual, 2022, pp. 1–25. OpenReview.net.

[13] I. Habibie, M. Elgharib, K. Sarkar, A. Abdullah, S. Nyatsanga, M. Neff, and C. Theobalt. A motion matching-based framework for controllable gesture synthesis from speech. In M. Nandigjav, N. J. Mitra, and A. Hertzmann, eds., *Proceedings SIGGRAPH 2022 Conference Papers Proceedings (ACM SIGGRAPH 2022)*, Vancouver, Canada, 2022, Article 346. ACM.

[14] I. Habibie, W. Xu, D. Mehta, L. Liu, H.-P. Seidel, G. Pons-Moll, M. Elgharib, and C. Theobalt. Learning speech-driven 3D conversational gestures from video. In *Proceedings of the 21st ACM International Conference on Intelligent Virtual Agents (IVA 2021)*, Virtual Event, Japan, 2021, pp. 101–108. ACM.

[15] J. He, J. Chen, S. Liu, A. Kortylewski, C. Yang, Y. Bai, and C. Wang. TransFG: A transformer architecture for fine-grained recognition. In *Proceedings of the 36th AAAI Conference on Artificial Intelligence*, Virtual Conference, 2022, pp. 852–860. AAAI.

[16] J. He, A. Kortylewski, and A. Yuille. CORL: Compositional representation learning for few-shot classification. In *2023 IEEE Winter Conference on Applications of Computer Vision (WACV 2023)*, Waikoloa Village, HI, USA, 2023, pp. 3879–3888. IEEE.

[17] J. He, S. Yang, S. Yang, A. Kortylewski, X. Yuan, J. Chen, S. Liu, C. Yang, Q. Yu, and A. L. Yuille. PartImageNet: A large, high-quality dataset of parts. In S. Avidan, G. Brostow, M. Cissé, G. M. Farinella, and T. Hassner, eds., *Computer Vision – ECCV 2022*, Tel Aviv, Israel, 2022, LNCS 13668, pp. 128–145. Springer.

[18] T. Hu, K. Sarkar, L. Liu, M. Zwicker, and C. Theobalt. EgoRenderer: Rendering human avatars from egocentric camera images. In *ICCV 2021, IEEE/CVF International Conference on Computer Vision*, Virtual Event, 2021, pp. 14508–14518. IEEE.

[19] Y. Jiang, M. Habermann, V. Golyanik, and C. Theobalt. HiFECap: Monocular high-fidelity and expressive capture of human performances. In *33rd British Machine Vision Conference (BMVC 2022)*, London, UK, 2022, Article 826. BMVA Press.

[20] E. C. M. Johnson, M. Habermann, S. Shimada, V. Golyanik, and C. Theobalt. Unbiased 4D: Monocular 4D reconstruction with a neural deformation model. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW 2023)*, Vancouver, Canada, 2023. Accepted.

[21] N. Kairanda, E. Tretschk, M. Elgharib, C. Theobalt, and V. Golyanik. $\phi$-SfT: Shape-from-template with a physics-based deformation model. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2022)*, New Orleans, LA, USA, 2022, pp. 3938–3948. IEEE.

[22] M. Kappel, V. Golyanik, M. Elgharib, J.-O. Henningson, H.-P. Seidel, S. Castillo, C. Theobalt, and M. A. Magnor. High-fidelity neural human motion transfer from monocular video computer vision and pattern recognition. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2021)*, Virtual Conference, 2021, pp. 1541–1550. IEEE.

[23] M. Krahn, F. Bernard, and V. Golyanik. Convex joint graph matching and clustering via semidefinite. In *2021 International Conference on 3D Vision*, Virtual, 2021, pp. 1216–1226. IEEE.

[24] Y. Li, M. Habermann, B. Thomaszewski, S. Coros, T. Beeler, and C. Theobalt. Deep physics-aware inference of cloth deformation for monocular human performance capture. In *2021 International Conference on 3D Vision*, Virtual Conference, 2021, pp. 373–384. IEEE.

[25] Z. Li, S. Shimada, B. Schiele, C. Theobalt, and V. Golyanik. MoCapDeform: Monocular 3D human motion capture in deformable scenes. In *International Conference on 3D Vision*, Hybrid / Prague, Czechia, 2022, pp. 1–11. IEEE.

[26] Y. Liu, L. Liu, C. Lin, Z. Dong, and W. Wang. Learnable motion coherence for correspondence pruning. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2021)*, Nashville, TN, US (Virtual), 2021, pp. 3236–3245. IEEE.

[27] Y. Liu, S. Peng, L. Liu, Q. Wang, P. Wang, C. Theobalt, X. Zhou, and W. Wang. Neural rays for occlusion-aware image-based rendering. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2022)*, New Orleans, LA, USA, 2022, pp. 7814–7823. IEEE.

[28] X. Long, C. Lin, L. Liu, W. Li, C. Theobalt, R. Yang, and W. Wang. Adaptive surface normal constraint for depth estimation. In *ICCV 2021, IEEE/CVF International Conference on Computer Vision*, Virtual Event, 2021, pp. 12829–12838. IEEE.

[29] X. Long, L. Liu, W. Li, C. Theobalt, and W. Wang. Multi-view depth estimation using epipolar spatio-temporal networks. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2021)*, Virtual Conference, 2021, pp. 8254–8263. IEEE.

[30] L. Lyu, M. Habermann, L. Liu, Mallikarjun B R, A. Tewari, and C. Theobalt. Efficient and differentiable shadow computation for inverse problems. In *ICCV 2021, IEEE/CVF International Conference on Computer Vision*, Virtual Event, 2021, pp. 13087–13096. IEEE.

[31] L. Lyu, A. Tewari, T. Leimkühler, M. Habermann, and C. Theobalt. Neural radiance transfer fields for relightable novel-view synthesis with global illumination. In S. Avidan, G. Brostow, M. Cissé, G. M. Farinella, and T. Hassner, eds., *Computer Vision – ECCV 2022*, Tel Aviv, Israel, 2022, LNCS 13677, pp. 153–169. Springer.

[32] L. Ma, L. Liu, C. Theobalt, and L. Van Gool. Direct dense pose estimation. In *2021 International Conference on 3D Vision*, Virtual Conference, 2021, pp. 721–730. IEEE.

[33] W. Ma, A. Wang, A. L. Yuille, and A. Kortylewski. Robust category-level 6D pose estimation with coarse-to-fine rendering of neural features. In S. Avidan, G. Brostow, M. Cissé, G. M. Farinella, and T. Hassner, eds., *Computer Vision – ECCV 2022*, Tel Aviv, Israel, 2022, LNCS 13669, pp. 492–508. Springer.

[34] Mallikarjun B R, A. Tewari, T.-H. Oh, T. Weyrich, B. Bickel, H.-P. Seidel, H. Pfister, W. Matusik, M. Elgharib, and C. Theobalt. Monocular reconstruction of neural face reflectance fields. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2021)*, Virtual Conference, 2021, pp. 4789–4798. IEEE.

[35] Mallikarjun B R, A. Tewari, X. Pan, M. Elgharib, and C. Theobalt. gCoRF: Generative compositional radiance fields. In *International Conference on 3D Vision*, Hybrid / Prague, Czechia, 2022, pp. 567–576. IEEE.

[36] Mallikarjun B R, A. Tewari, H.-P. Seidel, M. Elgharib, and C. Theobalt. Learning complete 3D morphable face models from images and videos. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2021)*, Virtual Conference, 2021, pp. 3360–3370. IEEE.

[37] W. Menapace, S. Lathuilière, A. Siarohin, C. Theobalt, S. Tulyakov, V. Golyanik, and E. Ricci. Playable environments: Video manipulation in space and time. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2022)*, New Orleans, LA, USA, 2022, pp. 3574–3583. IEEE.

[38] J. Nehvi, V. Golyanik, F. Mueller, H.-P. Seidel, M. Elgharib, and C. Theobalt. Differentiable event stream simulator for non-rigid 3D tracking. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPR 2021)*, Virtual Workshop, 2021, pp. 1302–1311. IEEE.

[39] X. Pan, A. Tewari, L. Liu, and C. Theobalt. GAN2X: Non-Lambertian inverse rendering of image GANs. In *International Conference on 3D Vision*, Hybrid / Prague, Czechia, 2022, pp. 711–721. IEEE. Accepted.

[40] P. Rao, Mallikarjun B R, G. Fox, T. Weyrich, B. Bickel, H. Pfister, W. Matusik, A. Tewari, C. Theobalt, and M. Elgharib. VoRF: Volumetric Relightable Faces. In *33rd British Machine Vision Conference (BMVC 2022)*, London, UK, 2022, Article 708. BMVA Press.

[41] V. Rudnev, M. Elgharib, W. A. P. Smith, L. Liu, V. Golyanik, and C. Theobalt. NeRF for outdoor scene relighting. In S. Avidan, G. Brostow, M. Cissé, G. M. Farinella, and T. Hassner, eds., *Computer Vision – ECCV 2022*, Tel Aviv, Israel, 2022, LNCS 13676, pp. 615–631. Springer.

[42] V. Rudnev, V. Golyanik, J. Wang, H.-P. Seidel, F. Mueller, M. Elgharib, and C. Theobalt. EventHands: Real-time neural 3D hand pose estimation from an event stream. In *ICCV 2021*, Virtual Event, 2021, pp. 12365–12375. IEEE.

[43] K. Sarkar, L. Liu, V. Golyanik, and C. Theobalt. HumanGAN: A generative model of humans images. In *2021 International Conference on 3D Vision*, Virtual, 2021, pp. 258–267. IEEE.

[44] M. Seelbach Benkner, Z. Lähner, V. Golyanik, C. Wunderlich, C. Theobalt, and M. Moeller. Q-Match: Iterative shape matching via quantum annealing. In *ICCV 2021, IEEE/CVF International Conference on Computer Vision*, Virtual Event, 2021, pp. 7566–7576. IEEE.

[45] S. Shimada, V. Golyanik, Z. Li, P. Pérez, W. Xu, and C. Theobalt. HULC: 3D HUman motion capture with pose manifold sampLing and dense Contact guidance. In S. Avidan, G. Brostow, M. Cissé, G. M. Farinella, and T. Hassner, eds., *Computer Vision – ECCV 2022*, Tel Aviv, Israel, 2022, LNCS 13682, pp. 516–533. Springer.

[46] A. Tewari, Mallikarjun B R, X. Pan, O. Fried, M. Agrawala, and C. Theobalt. Disentangled3D: Learning a 3D generative model with disentangled geometry and appearance from monocular images. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2022)*, New Orleans, LA, USA, 2022, pp. 1506–1515. IEEE.

[47] E. Tretschk, A. Tewari, V. Golyanik, M. Zollhofer, C. Lassner, and C. Theobalt. Non-rigid neural radiance fields: Reconstruction and novel view synthesis of a dynamic scene from monocular video. In *ICCV 2021, IEEE/CVF International Conference on Computer Vision*, Virtual Event, 2021, pp. 12939–12950. IEEE.

[48] J. Wang, L. Liu, W. Xu, K. Sarkar, D. Luvizon, and C. Theobalt. Estimating egocentric 3D human pose in the wild with external weak supervision. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2022)*, New Orleans, LA, USA, 2022, pp. 13147–13156. IEEE.

[49] J. Wang, L. Liu, W. Xu, K. Sarkar, and C. Theobalt. Estimating egocentric 3D human pose in global space. In *ICCV 2021, IEEE/CVF International Conference on Computer Vision*, Virtual Event, 2021, pp. 11480–11489. IEEE.

[50] J. Wang, D. Luvizon, F. Mueller, F. Bernard, A. Kortylewski, D. Casas, and C. Theobalt. HandFlow: Quantifying view-dependent 3D ambiguity in two-hand reconstruction with normalizing flow. In *International Symposium on Vision, Modeling, and Visualization (VMV 2022)*, Konstanz, Germany, 2022, pp. 99–106. Eurographics Association.

[51] J. Wang, P. Wang, X. Long, C. Theobalt, T. Komura, L. Liu, and W. Wang. NeuRIS: Neural reconstruction of indoor scenes using normal priors. In S. Avidan, G. Brostow, M. Cissé, G. M. Farinella, and T. Hassner, eds., *Computer Vision – ECCV 2022*, Tel Aviv, Israel, 2022, LNCS 13692, pp. 139–155. Springer.

[52] P. Wang, L. Liu, Y. Liu, C. Theobalt, T. Komura, and W. Wang. NeuS: Learning neural implicit surfaces by volume rendering for multi-view reconstruction. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P. S. Liang, and J. Wortman Vaughan, eds., *Advances in Neural Information Processing Systems 34 (NeurIPS 2021)*, Virtual, 2021, pp. 27171–27183. Curran Assoicates, Inc.

[53] Y. Xiangli, L. Xu, X. Pan, N. Zhao, A. Rao, C. Theobalt, B. Dai, and D. Lin. BungeeNeRF: Progressive neural radiance field for extreme multi-scale scene rendering. In S. Avidan, G. Brostow, M. Cissé, G. M. Farinella, and T. Hassner, eds., *Computer Vision – ECCV 2022*, Tel Aviv, Israel, 2022, LNCS 13692, pp. 106–122. Springer.

[54] T. Yenamandra, A. Tewari, F. Bernard, H.-P. Seidel, M. Elgharib, D. Cremers, and C. Theobalt. i3DMM: Deep implicit 3D morphable model of human heads. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2021)*, Virtual Conference, 2021, pp. 12798–12808. IEEE.

[55] X. Yi, Y. Zhou, M. Habermann, S. Shimada, V. Golyanik, C. Theobalt, and F. Xu. Physical Inertial Poser (PIP): Physics-aware real-time human motion tracking from sparse inertial sensors. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2022)*, New Orleans, LA, USA, 2022, pp. 13157–13168. IEEE.

[56] J. S. Yoon, L. Liu, V. Golyanik, K. Sarkar, H. S. Park, and C. Theobalt. Pose-guided human animation from a single image in the wild. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2021)*, Virtual Conference, 2021, pp. 15034–15043. IEEE.

[57] A. Yurtsever, T. Birdal, and V. Golyanik. Q-FW: A hybrid classical-quantum Frank-Wolfe for quadratic binary optimization. In S. Avidan, G. Brostow, M. Cissé, G. M. Farinella, and T. Hassner, eds., *Computer Vision – ECCV 2022*, Tel Aviv, Israel, 2022, LNCS 13683, pp. 352–369. Springer.

[58] Y. Zhang, A. Kortylewski, Q. Liu, S. Park, B. Green, E. Engle, G. Almodovar, R. Walk, S. Soto-Diaz, J. Taube, A. Szalay, and A. Yuille. A light-weight interpretable model for nuclei detection and weakly-supervised segmentation. In Y. Huo, B. A. Millis, Y. Zhou, X. Wang, A. P. Harrison, and Z. Xu, eds., *Medical Optical Imaging and Virtual Microscopy Image Analysis (MOVI 2022)*, Singapore, 2022, LNCS 13578, pp. 145–155. Springer.

[59] B. Zhao, S. Yu, W. Ma, M. Yu, S. Mei, A. Wang, J. He, A. L. Yuille, and A. Kortylewski. OOD-CV: A benchmark for robustness to out-of-distribution shifts of individual nuisances in natural images. In S. Avidan, G. Brostow, M. Cissé, G. M. Farinella, and T. Hassner, eds., *Computer Vision – ECCV 2022*, Tel Aviv, Israel, 2022, LNCS 13668, pp. 163–180. Springer.

[60] Y. Zhou, M. Habermann, I. Habibie, A. Tewari, C. Theobalt, and F. Xu. Monocular real-time full body capture with inter-part correlations computer vision and pattern recognition. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2021)*, Virtual Conference, 2021, pp. 4809–4820. IEEE.

## Technical reports and preprints

[1] R. Dabral, M. H. Mughal, V. Golyanik, and C. Theobalt. *MoFusion: A Framework for Denoising-Diffusion-based Motion Synthesis*, 2022. arXiv: 2212.04495.

[2] A. Dib, C. Thebault, J. Ahn, P.-H. Gosselin, C. Theobalt, and L. Chevallier. *Towards High Fidelity Monocular Face Reconstruction with Rich Reflectance using Self-supervised Learning and Ray Tracing*, 2021. arXiv: 2103.15432.

[3] A. Ghosh, R. Dabral, V. Golyanik, C. Theobalt, and P. Slusallek. *IMoS: Intent-Driven Full-Body Motion Synthesis for Human-Object Interactions*, 2022. arXiv: 2212.07555.

[4] M. Habermann, L. Liu, W. Xu, G. Pons-Moll, M. Zollhöfer, and C. Theobalt. *HDHumans: A Hybrid Approach for High-fidelity Digital Humans*, 2022. arXiv: 2210.12003.

[5] M. Kappel, V. Golyanik, S. Castillo, C. Theobalt, and M. A. Magnor. *Fast Non-Rigid Radiance Fields from Monocularized Data*, 2022. arXiv: 2212.01368.

[6] Z. Li, X. Wang, E. Stengel-Eskin, A. Kortylewski, W. Ma, B. Van Durme, and A. Yuille. *Super-CLEVR: A Virtual Benchmark to Diagnose Domain Robustness in Visual Reasoning*, 2022. arXiv: 2212.00259.

[7] D. Luvizon, M. Habermann, V. Golyanik, A. Kortylewski, and C. Theobalt. *Scene-Aware 3D Multi-Human Motion Capture from a Single Camera*, 2023. arXiv: 2301.05175.

[8] V. Rudnev, M. Elgharib, C. Theobalt, and V. Golyanik. *EventNeRF: Neural Radiance Fields from a Single Colour Event Camera*, 2022. arXiv: 2206.11896.

[9] K. Sarkar, V. Golyanik, L. Liu, and C. Theobalt. *Style and Pose Control for Image Synthesis of Humans from a Single Monocular View*, 2021. arXiv: 2102.11263.

[10] A. Wang, P. Wang, J. Sun, A. Kortylewski, and A. Yuille. *VoGE: A Differentiable Volume Renderer using Gaussian Ellipsoids for Analysis-by-Synthesis*, 2022. arXiv: 2205.15401.

[11] C. Wang, A. Serrano, X. Pan, B. Chen, H.-P. Seidel, C. Theobalt, K. Myszkowski, and T. Leimkühler. *GlowGAN: Unsupervised Learning of HDR Images from LDR Images in the Wild*, 2022. arXiv: 2211.12352.

[12] J. Wang, C. Zhang, P. Wang, X. Li, P. J. Cobb, C. Theobalt, and W. Wang. *FIRES: Fast Imaging and 3D Reconstruction of Archaeological Sherds*, 2022. arXiv: 2211.06897.

[13] T. Wu, J. Wang, X. Pan, X. Xu, C. Theobalt, Z. Liu, and D. Lin. *Voxurf: Voxel-based Efficient and Accurate Neural Surface Reconstruction*, 2022. arXiv: 2208.12697.

[14] T. Yenamandra, A. Tewari, N. Yang, F. Bernard, C. Theobalt, and D. Cremers. *HDSDF: Hybrid Directional and Signed Distance Functions for Fast Inverse Rendering*, 2022. arXiv: 2203.16284.

[15] F. Zhan, Y. Yu, R. Wu, J. Zhang, S. Lu, L. Liu, A. Kortylewski, C. Theobalt, and E. Xing. *Multimodal Image Synthesis and Editing: A Survey*, 2022. arXiv: 2112.13592.

# 33 RG1: Automation of Logic

## 33.1 Personnel

**Head of Group**
Prof. Dr. Christoph Weidenbach

**Researchers**
Dr. Martin Bromberger
Dr. Sibylle Möhle (August 2022–)
Dr. Hamid Rahkooy (–October 2022)
Dr. Uwe Waldmann

**Ph.D. Students**
Martin Desharnais (September 2021–)
Fajar Haifani (–March 2023)
Hendrik Leidinger
Lorenz Leutgeb (October 2021–)
Simon Schwarz (October 2022–)

**Long-term Guests**
Dr. Jasmin Christian Blanchette
Dr. Mathias Fleury
Dr. Thomas Sturm
Dr. Sophie Tourret

**Secretaries**
Jennifer Müller

## 33.2 Visitors

From March 2021 to February 2023, the following researchers visited our group:

| | | |
|---|---|---|
| Deepak Kapur | 15.10.2021–11.01.2022 | University of New Mexico |
| Jannik Vierling | 18.05.2022–19.05.2022 | TU Wien |
| Stephan Schulz | 09.03.2023 | DHBW Stuttgart |

## 33.3 Group Organization

The four PhD students are co-supervised by two senior members of the group, including our long-term guests. There are weekly weakly meetings with the PhD students where the long-term guests typically participate via Zoom. In addition, there are regular weekly project related meetings, in particular for software development.

### Long-term Guests

**Jasmin Blanchette.** Jasmin Blanchette is head of the Chair of Theoretical Computer Science at Ludwig-Maximilians-Universität München. His research focuses on the use and development of proof assistants and automatic theorem provers (ATPs).

On the proof assistant front, Blanchette developed or codeveloped three tools for the Isabelle/HOL proof assistant. First, the Nitpick tool generates counterexamples using a SAT solver under the hood. By discovering flaws early, it saves users time. Second, the Sledgehammer tool uses automatic theorem provers as backends to discharge Isabelle proof obligations. By providing automatic proofs to straightforward but tedious proof obligations, it is a timesaver. Third, the (co)datatype package provides definitional mechanisms for introducing (co)inductive datatypes and (co)recursive functions safely and conveniently.

On the automatic theorem proving front, Blanchette and his colleagues developed lambda-superposition, an extension of the standard superposition calculus for first-order logic to higher-order logic. Lambda-superposition is implemented in two provers, E and Zipperposition, which dominate the higher-order theorem division of the CADE ATP System Competition since 2020.

Finally, Blanchette also applies proof assistants to verify the correctness of the logical calculi underlying ATPs. He verified the correctness of a resolution prover and contributed to the development of the fastest verified SAT solver.

**Thomas Sturm.** Thomas Sturm studied computer science at the University of Passau, Germany, where he held positions as a researcher and faculty member (Privatdozent) after finishing his studies in 1995. In 2008 he was awarded a Ramón y Cajal Fellowship by the Spanish Ministry of Science and Innovation and moved to the University of Cantabria in Santander. In 2011 he joined the MPI for Informatics in Saarbrücken, Germany, where he headed a research group for Arithmetic Reasoning. In 2016 he was appointed as a research director at French CNRS. Sturm has worked as a visiting researcher at various international institutes and research facilities including SRI International in Menlo Park, Zuse Institute Berlin, Fujitsu Laboratories Japan, Lomonosov Moscow State University, and Forschungszentrum Jülich. Sturm is an editor at the Journal of Symbolic Computation (Elsevier) and at Mathematics in Computer Science (Springer).

Sturm's research interests span the domains of exact and efficient computation, computer algebra, logic, and formal reasoning. This includes the development of effective quantifier elimination methods and decision procedures for various algebraic theories, their efficient implementation, and their application in the sciences and in engineering. On the foundational side, he is working on methods from tropical algebra in a liberal sense with strong focus on the real numbers as the domain of interest. His methods are used in SMT solving and in mathematical biology. On the applied side, one focus is interdisciplinary research on specialized decision methods for the qualitative analysis of dynamic properties of reaction systems in chemistry and systems biology.

**Sophie Tourret.** Sophie Tourret is a senior guest researcher at MPI for Informatics in RG1. These last two years, she has collaborated with Christoph Weidenbach for the supervision of Fajar Haifani, a PhD student in RG1 during that time period. She has also collaborated with Christoph Weidenbach and Jasmin Blanchette in the supervision of Martin Desharnais, who is still a PhD student in RG1. She usually works in Inria Nancy, France and she has been travelling regularly, approximately once a month, to MPI to visit Fajar and Martin. While the work with Martin has not yet led to publications, the one with Fajar has been most recently published at IJCAR 2022 and the DL workshop 2022. The latter publication led to Fajar receiving the best student paper award at the DL workshop.

In terms of community services, these last two years, Sophie Tourret has pursued her work as a board member and newsletter editor of the Association for Automated Reasoning. She has also been elected to become a trustee for CADE, the flagship conference of automated reasoning (with IJCAR). She has been involved in the organisation of IJCAR as a workshop chair and she has otherwise been a pc member in the following conferences: IJCAI-PRICAI 2021, CADE 2021, IJCAI-ECAI 2022, CPP 2022, IJCAR 2022, IJCAI 2023, CADE 2023; and in the following workshops: PAAR 2022, SMT 2022.

## 33.4 Foundations of Automated Reasoning

*Coordinator: Christoph Weidenbach*

Our main focus in the reporting period has been the further development of the clause learning from simple models (SCL) paradigm and the development of explanation techniques for first-order reasoning, going beyond proofs and models.

We extended the SCL calculus for plain first-order logic to equational reasoning SCL(EQ) (Sect. 33.4.1) and to the exhaustive exploration of partial models (Sect. 33.4.2). In addition, we unified the reasoning strategies suggested in previous work on SCL [3, 1] with the recent before mentioned two extensions on equational reasoning and exhaustive model exploration [2]. One of our goals is to establish SCL as a framework for integrating other first-order calculi. To this end we showed that SCL can in fact simulate superposition in first-order logic without equality (Sect. 33.4.3). For reasoning with theories, we showed that for certain fragments, an effective model can be constructed for a clause set saturated by (hierarchic) superposition (Sect. 33.4.4). A next step will be to relate the model construction to the SCL partial model construction as well.

In explaining first-order reasoning beyond proofs and models one question is which clauses are actually needed for a refutation. Here, we developed the concept of syntactic and semantic relevance (Sect. 33.4.5) and provide reasoning technology for their computation. Interestingly, we generalized the almost 60 years old result on set-of-support completeness this way. Another important question is of an abductive nature: which clauses have to be added to a clause set for a certain conjecture, that is not logically implied, to hold. There is already considerable literature on this topic. We added the fine-grained concept of connection-minimal abduction (Sect. 33.4.6), that in particular rules out abductive answers that are not *closely connected* to the conjecture. For the description logic $\mathcal{EL}$ we established an effective reasoning pipeline, computing complete sets of connection-minimal hypotheses implying a given conjecture.

## References


- [1] M. Bromberger, A. Fiori, and C. Weidenbach. Deciding the Bernays-Schoenfinkel fragment over bounded difference constraints by simple clause learning over theories. In F. Henglein, S. Shoham, and V. Yakir, eds., *Verification, Model Checking, and Abstract Interpretation (VMCAI 2021)*, Copenhagen, Denmark (Online), 2021, LNCS 12597, pp. 511–533. Springer.
- [2] M. Bromberger, S. Schwarz, and C. Weidenbach. *SCL(FOL) Revisited*, 2023. arXiv: 2302.05954.

[3] A. Fiori and C. Weidenbach. SCL clause learning from simple models. In P. Fontaine, ed., *Automated Deduction – CADE 27*, Natal, Brazil, 2019, LNAI 11716, pp. 233–249. Springer.


### 33.4.1 SCL for First-Order Logic with Equality

*Investigators: Hendrik Leidinger, Christoph Weidenbach*

Developing a sound, complete and efficient calculus for first-order logic with equality providing explicit model building is a challenge. A number of researchers have tried along this line, but without success [3, 9, 4] in the sense that a syntactic implementation of superposition is still the current prime calculus for first-order logic with equality. The success of superposition relies on its strong redundancy notions, in particular the availability of unit rewriting. Therefore, the calculus should at least support unit rewriting, because this has turned out to be essential in the context of first-order equational reasoning. The explicit model should be represented in a way that false and propagating clauses can be efficiently detected, similar to CDCL (Conflict Driven Clause Learning) [11, 13, 15] and all other approaches to first-order reasoning with explicit model guidance [1, 5, 2, 8, 9, 14]. Finally, learned clauses should be non-redundant, following the classical superposition definition of redundancy. This holds for CDCL [16], SCL (Clause Learning from Simple models) [8, 7] and SCL(T) [6]. With our new calculus SCL(EQ) [10], we have been able to incorporate all of these features. Our basic idea is again to have a model built on ground literals (equations). This supports the efficient computation of a convergent system. With respect to such a convergent ground system false and propagating clauses can be effectively detected.

In order to efficiently detect false and propagating clauses, an efficient implementation of the congruence closure algorithm [12] is required. SCL(EQ) occasionally requires (finitely) many ground instances of a non-ground literal to be propagated. To this end we are currently working on a generalization of congruence closure supporting constrained literals, whereby the constraints always only allow a finite number of ground instantiations. This should significantly reduce the number of merge and deduction steps at the expense of constraint satisfiability checks. Constrained congruence closure and its implementation is a work in progress.

## References


[1] G. Alagi and C. Weidenbach. NRCL – a model building approach to the Bernays-Schönfinkel fragment. In C. Lutz and S. Ranise, eds., *Frontiers of Combining Systems (FroCoS 2015)*, Wrocław, Poland, 2015, LNAI 9322, pp. 69–84. Springer.

[2] P. Baumgartner, A. Fuchs, and C. Tinelli. Lemma learning in the model evolution calculus. In *Logic for Programming, Artificial Intelligence, and Reasoning: 13th International Conference,*

*LPAR 2006, Phnom Penh, Cambodia, November 13-17, 2006. Proceedings 13*, 2006, pp. 572–586. Springer.

[3] P. Baumgartner, U. Furbach, and B. Pelzer. The hyper tableaux calculus with equality and an application to finite model computation. *Journal Log. Comput.*, 20(1):77–109, 2010.

[4] M. P. Bonacina, U. Furbach, and V. Sofronie-Stokkermans. On first-order model-based reasoning. In N. Martí-Oliet, P. C. Ölveczky, and C. L. Talcott, eds., *Logic, Rewriting, and Concurrency – Essays dedicated to José Meseguer on the Occasion of His 65th Birthday*, 2015, LNCS 9200, pp. 181–204. Springer.

[5] M. P. Bonacina and D. A. Plaisted. Semantically-guided goal-sensitive reasoning: model representation. *Journal of Automated Reasoning*, 56:113–141, 2016.

• [6] M. Bromberger, A. Fiori, and C. Weidenbach. Deciding the Bernays-Schoenfinkel fragment over bounded difference constraints by simple clause learning over theories. In F. Henglein, S. Shoham, and V. Yakir, eds., *Verification, Model Checking, and Abstract Interpretation (VMCAI 2021)*, Copenhagen, Denmark (Online), 2021, LNCS 12597, pp. 511–533. Springer.

• [7] M. Bromberger, S. Schwarz, and C. Weidenbach. *SCL(FOL) Revisited*, 2023. arXiv: 2302.05954.

[8] A. Fiori and C. Weidenbach. SCL clause learning from simple models. In P. Fontaine, ed., *Automated Deduction – CADE 27*, Natal, Brazil, 2019, LNAI 11716, pp. 233–249. Springer.

[9] K. Korovin. Inst-Gen–a modular approach to instantiation-based automated reasoning. *Programming Logics: Essays in Memory of Harald Ganzinger*, 7797:239–270, 2013.

• [10] H. Leidinger and C. Weidenbach. SCL(EQ): SCL for first-order logic with equality. In J. Blanchette, L. Kovács, and D. Pattinson, eds., *Automated Reasoning (IJCAR 2022)*, Haifa, Israel, 2022, LNAI 13385, pp. 228–247. Springer.

[11] M. W. Moskewicz, C. F. Madigan, Y. Zhao, L. Zhang, and S. Malik. Chaff: Engineering an efficient SAT solver. In *Proceedings of the 38th annual Design Automation Conference*, 2001, pp. 530–535.

[12] G. Nelson and D. C. Oppen. Fast decision procedures based on congruence closure. *Journal of the ACM*, 27(2):356–364, 1980.

[13] R. Nieuwenhuis, A. Oliveras, and C. Tinelli. Solving SAT and SAT modulo theories: From an abstract Davis–Putnam–Logemann–Loveland procedure to DPLL(T). *Journal of the ACM (JACM)*, 53(6):937–977, 2006.

[14] R. Piskac, L. de Moura, and N. Bjørner. Deciding effectively propositional logic using DPLL and substitution sets. *Journal of Automated Reasoning*, 44:401–424, 2010.

[15] J. P. M. Silva and K. A. Sakallah. GRASP – a new search algorithm for satisfiability. In *International Conference on Computer Aided Design, ICCAD*, 1996, pp. 220–227. IEEE Computer Society Press.

[16] C. Weidenbach. Automated reasoning building blocks. In R. Meyer, A. Platzer, and H. Wehrheim, eds., *Correct System Design*, Oldenburg, Germany, 2015, LNCS 9360, pp. 172–188. Springer.

### 33.4.2 Exploring Partial Models with SCL

*Investigators: Martin Bromberger, Simon Schwarz, and Christoph Weidenbach*

The family of SCL (Clause Learning from Simple Models) calculi [5, 1, 4] learns clauses with respect to a partial model assumption, similar to CDCL (Conflict Driven Clause Learning).

The partial model always consists of ground first-order literals and is built by decisions and propagations. In contrast to propositional logic where propagation chains are always finite, in first-order logic they can become infinite. Therefore, the SCL family does not require exhaustive propagation and the size of the partial model is always finitely bounded. Any partial model not leading to a conflict constitutes a model for the respective finitely bounded ground clause set. We show that all potential partial models can be explored as part of the SCL calculus for first-order logic without equality and that any overall model is an extension of a partial model considered [2, 3]. Furthermore, SCL turns into a semi-decision procedure for first-order logic by extending the finite bound for any partial model not leading to a conflict.

### References


- [1] M. Bromberger, A. Fiori, and C. Weidenbach. Deciding the Bernays-Schoenfinkel fragment over bounded difference constraints by simple clause learning over theories. In F. Henglein, S. Shoham, and V. Yakir, eds., *Verification, Model Checking, and Abstract Interpretation (VMCAI 2021)*, Copenhagen, Denmark (Online), 2021, LNCS 12597, pp. 511–533. Springer.
- [2] M. Bromberger, S. Schwarz, and C. Weidenbach. Exploring partial models with SCL. In B. Konev, C. Schon, and A. Steen, eds., *Practical Aspects of Automated Reasoning 2022 (PAAR 2022)*, Haifa, Israel, 2022, CEUR Workshop Proceedings 3201. CEUR-WS.org.
- [3] M. Bromberger, S. Schwarz, and C. Weidenbach. Exploring partial models with SCL. In R. Piskac and A. Voronkov, eds., *LPAR 2023: 24th International Conference on Logic for Programming, Artificial Intelligence and Reasoning*, 2023. Accepted for publication.
- [4] M. Bromberger, S. Schwarz, and C. Weidenbach. *SCL(FOL) Revisited*, 2023. arXiv: 2302.05954.

[5] A. Fiori and C. Weidenbach. SCL clause learning from simple models. In P. Fontaine, ed., *Automated Deduction – CADE 27*, Natal, Brazil, 2019, LNAI 11716, pp. 233–249. Springer.


### 33.4.3 SCL Can Simulate Non-Redundant Superposition Clause Learning

*Investigators: Martin Bromberger, Chaahat Jain, and Christoph Weidenbach*

Superposition [1, 2, 11] is currently considered as the prime calculus for first-order logic reasoning where all leading first-order theorem provers implement a variant thereof [8, 12, 9, 13]. More recently, the family of SCL calculi (Clause Learning from Simple Models, or just Simple Clause Learning) [7, 3, 5, 10, 6] was introduced. We have shown that SCL(FOL) can simulate the derivation of non-redundant clauses by superposition for first-order logic without equality [4]. In general, superposition can derive redundant clauses. So for our simulation result, we have to restrict superposition to a strategy that only derives non-redundant clauses, namely the strategy used in the completeness proof for superposition, based on a ground model construction operator [1]. This also means that the simulation result is restricted to the ground case. The main differences between superposition and SCL for first-order logic without equality are: (i) superposition assumes a fixed ordering on literals where the ordering in SCL is dynamic and evolves out of the satisfiability of clauses, (ii) superposition performs single superposition left and factoring inferences where SCL typically performs several such inferences to derive a single learned clause. We mitigate these differences for our simulation by defining a new strategy SCL-SUP for SCL that forces SCL to construct partial models based on a fixed literal ordering. The partial models constructed in this way are equivalent to the partial models constructed by the superposition model operator. Moreover, SCL-SUP guarantees that any inferences skipped by SCL are in fact obsolete because they are either made redundant by an inference shared by SCL and superposition or unnecessary in SCL due to more flexibility with regard to factorization. From this perspective the SCL calculus can be viewed as a generalization of the superposition calculus.

## References


[1] L. Bachmair and H. Ganzinger. Rewrite-based equational theorem proving with selection and simplification. *Journal of Logic and Computation*, 4(3):217–247, 1994. Revised version of Max-Planck-Institut für Informatik technical report, MPI-I-91-208, 1991.

[2] L. Bachmair and H. Ganzinger. Resolution theorem proving. In A. Robinson and A. Voronkov, eds., *Handbook of Automated Reasoning*, vol. I, ch. 2, pp. 19–99. Elsevier, 2001.

• [3] M. Bromberger, A. Fiori, and C. Weidenbach. Deciding the Bernays-Schoenfinkel fragment over bounded difference constraints by simple clause learning over theories. In F. Henglein, S. Shoham, and V. Yakir, eds., *Verification, Model Checking, and Abstract Interpretation (VMCAI 2021)*, Copenhagen, Denmark (Online), 2021, LNCS 12597, pp. 511–533. Springer.

• [4] M. Bromberger, C. Jain, and C. Weidenbach. SCL(FOL) can simulate non-redundant superposition clause learning. In B. Pientka and C. Tinelli, eds., *CADE-29: 29th international Conference on Automated Deduction*, 2023. Accepted for publication.

• [5] M. Bromberger, S. Schwarz, and C. Weidenbach. Exploring partial models with SCL. In B. Konev, C. Schon, and A. Steen, eds., *Practical Aspects of Automated Reasoning 2022 (PAAR 2022)*, Haifa, Israel, 2022, CEUR Workshop Proceedings 3201. CEUR-WS.org.

• [6] M. Bromberger, S. Schwarz, and C. Weidenbach. *SCL(FOL) Revisited*, 2023. arXiv: 2302.05954.

[7] A. Fiori and C. Weidenbach. SCL clause learning from simple models. In P. Fontaine, ed., *Automated Deduction – CADE 27*, Natal, Brazil, 2019, LNAI 11716, pp. 233–249. Springer.

[8] K. Korovin. iProver – An instantiation-based theorem prover for first-order logic (system description). In A. Armando, P. Baumgartner, and G. Dowek, eds., *Automated Reasoning, 4th International Joint Conference, IJCAR 2008, Sydney, Australia, August 12-15, 2008, Proceedings*, 2008, LNCS 5195, pp. 292–298. Springer.

[9] L. Kovács and A. Voronkov. First-order theorem proving and Vampire. In N. Sharygina and H. Veith, eds., *Computer Aided Verification – 25th International Conference, CAV 2013, Saint Petersburg, Russia, July 13-19, 2013. Proceedings*, 2013, LNCS 8044, pp. 1–35. Springer.

• [10] H. Leidinger and C. Weidenbach. SCL(EQ): SCL for first-order logic with equality. In J. Blanchette, L. Kovács, and D. Pattinson, eds., *Automated Reasoning (IJCAR 2022)*, Haifa, Israel, 2022, LNAI 13385, pp. 228–247. Springer.

[11] R. Nieuwenhuis and A. Rubio. Paramodulation-based theorem proving. In A. Robinson and A. Voronkov, eds., *Handbook of Automated Reasoning*, vol. I, ch. 7, pp. 371–443. Elsevier, 2001.

[12] S. Schulz, S. Cruanes, and P. Vukmirović. Faster, higher, stronger: E 2.3. In P. Fontaine, ed., *Proc. of the 27th CADE, Natal, Brasil*, 2019, no. 11716 in LNAI, pp. 495–507. Springer.

[13] C. Weidenbach, D. Dimova, A. Fietzke, M. Suda, and P. Wischnewski. SPASS version 3.5. In R. A. Schmidt, ed., *22nd International Conference on Automated Deduction (CADE-22)*, 2009, LNAI, pp. 140–145. Springer.

### 33.4.4 Explicit Model Construction for Saturated Constrained Horn Clauses

*Investigators: Martin Bromberger, Lorenz Leutgeb, and Christoph Weidenbach*

Constrained Horn Clauses (CHCs) combine logical formulas with constraints over various domains, e.g. linear real arithmetic, linear integer arithmetic, equalities of uninterpreted functions [10]. This formalism has gained widespread attention in recent years due to its applications in a variety of fields, including program analysis and verification: safety, general invariants [5], liveness and termination [22, 12], complexity and resource analysis [20], intermediate representation [14], and software testing [21]. Technical controls, so called *Supervisors*, like an engine electronic control unit, or a lane change assistant in a car [7, 6] can be modelled, run, and proven safe. The success of applying CHCs to program verification has inspired the development of multiple solvers, see [4] and references therein.

Superposition is a decision procedure for various first-order logic fragments, e.g., [1, 15, 16], even if they are extended with theories [19, 13, 8]. The decision results for SCL (Simple Clause Learning) reasoning [8] are also based on the completeness of hierarchic superposition. Clause sets finitely saturated by hierarchic superposition do not offer an explicit model representation, rather the guarantee that all non-redundant inferences have been performed without deriving a contradiction. Still, this concept is powerful: There are saturated CHCs of the Bernays-Schönfinkel fragment extended with linear arithmetic, where it is undecidable whether a simple ground fact is a consequence out of the saturated set [11, 17]. The proof for refutational completeness of hierarchic superposition even implies the existence of a model if the clause set is saturated and does not contain the empty clause [18, 3, 2]. However, the definition of the model from the proof is based on an infinite axiomatization of the background theory. It is therefore not possible to use the proof to construct an explicit model with a finite representation.

In [9] we present an automated model building approach that yields an explicit and finite model representations for finitely saturated Horn clause sets of linear arithmetic combined with the Bernays-Schönfinkel fragment. It enables effective evaluation of clauses with respect to the model and therefore supports explanation and bug finding in case of failed refutations. The approach we present does not exploit features of linear arithmetic beyond equality and the existence of a well-founded order for the theories universe. The results may therefore be adapted to other constraint domains.

## References


[1] L. Bachmair, H. Ganzinger, and U. Waldmann. Superposition with simplification as a decision procedure for the monadic class with equality. In G. Gottlob, A. Leitsch, and D. Mundici, eds., *Computational Logic and Proof Theory, Third Kurt Gödel Colloquium, KGC'93*, 1993, LNCS, pp. 83–96. Springer.

[2] L. Bachmair, H. Ganzinger, and U. Waldmann. Refutational theorem proving for hierarchic first-order theories. *Applicable Algebra in Engineering, Communication and Computing (Aaecc)*, 5(3/4):193–212, 1994.

[3] P. Baumgartner and U. Waldmann. Hierarchic superposition revisited. In C. Lutz, U. Sattler, C. Tinelli, A.-Y. Turhan, and F. Wolter, eds., *Description Logic, Theory Combination, and All That*, LNCS 11560, pp. 15–56. Springer, Berlin, 2019.

[4] N. Bjørner, A. Gurfinkel, K. L. McMillan, and A. Rybalchenko. Horn clause solvers for program verification. In L. D. Beklemishev, A. Blass, N. Dershowitz, B. Finkbeiner, and W. Schulte, eds., *Fields of Logic and Computation II – Essays Dedicated to Yuri Gurevich on the Occasion of His 75th Birthday*, 2015, LNCS 9300, pp. 24–51. Springer.

[5] N. S. Bjørner, K. L. McMillan, and A. Rybalchenko. On solving universally quantified Horn clauses. In F. Logozzo and M. Fähndrich, eds., *Static Analysis – 20th International Symposium, SAS 2013, Seattle, WA, USA, June 20-22, 2013. Proceedings*, 2013, LNCS 7935, pp. 105–125. Springer.

• [6] M. Bromberger, I. Dragoste, R. Faqeh, C. Fetzer, L. González, M. Krötzsch, M. Marx, H. K. Murali, and C. Weidenbach. A sorted Datalog hammer for supervisor verification conditions modulo simple linear arithmetic. In D. Fisman and G. Rosu, eds., *Tools and Algorithms for the Construction and Analysis of Systems (TACAS 2022)*, Munich, Germany, 2022, LNCS 13243, pp. 480–501. Springer.

• [7] M. Bromberger, I. Dragoste, R. Faqeh, C. Fetzer, M. Krötzsch, and C. Weidenbach. A Datalog hammer for supervisor verification conditions modulo simple linear arithmetic. In B. Konev and G. Reger, eds., *Frontiers of Combining Systems (FroCoS 2021)*, Birmingham, UK, 2021, LNAI 12941, pp. 3–24. Springer.

• [8] M. Bromberger, A. Fiori, and C. Weidenbach. Deciding the Bernays-Schoenfinkel fragment over bounded difference constraints by simple clause learning over theories. In F. Henglein, S. Shoham, and V. Yakir, eds., *Verification, Model Checking, and Abstract Interpretation (VMCAI 2021)*, Copenhagen, Denmark (Online), 2021, LNCS 12597, pp. 511–533. Springer.

• [9] M. Bromberger, L. Leutgeb, and C. Weidenbach. *Explicit Model Construction for Saturated Constrained Horn Clauses*, 2023. arXiv: 2305.05064.

[10] E. De Angelis, F. Fioravanti, J. P. Gallagher, M. V. Hermenegildo, A. Pettorossi, and M. Proietti. Analysis and transformation of constrained Horn clauses for program verification. *Theory Pract. Log. Program.*, 22(6):974–1042, 2022.

[11] P. J. Downey. Undecidability of Presburger Arithmetic with a Single Monadic Predicate Letter. Technical report, Center for Research in Computer Technology, Harvard University, 1972.

[12] G. Fedyukovich, Y. Zhang, and A. Gupta. Syntax-guided termination analysis. In *CAV 2018*, 2018, LNCS 10981, pp. 124–143. Springer.

[13] A. L. Fietzke and C. Weidenbach. Superposition as a decision procedure for timed automata. *Mathematics in Computer Science*, 6(4):409–425, 2013.

[14] G. Gange, J. A. Navas, P. Schachte, H. Søndergaard, and P. J. Stuckey. Horn clauses as an intermediate representation for program analysis and transformation. *Theory Pract. Log. Program.*, 15(4-5):526–542, 2015.

[15] H. Ganzinger and H. de Nivelle. A superposition decision procedure for the guarded fragment with equality. In *14th LICS, 1999*, 1999, pp. 295–303. IEEE Computer Society.

[16] T. Hillenbrand and C. Weidenbach. Superposition for bounded domains. In M. P. Bonacina and M. Stickel, eds., *Automated Reasoning and Mathematics*, LNCS 7788, pp. 68–100. Springer, Berlin, 2013.

[17] M. Horbach, M. Voigt, and C. Weidenbach. *The Universal Fragment of Presburger Arithmetic with Unary Uninterpreted Predicates is Undecidable*, 2017. arXiv: 1703.01212.

[18] E. Kruglov. *Superposition Modulo Theory.* PhD thesis, Universität des Saarlandes, Saarbrücken, 2013.

[19] E. Kruglov and C. Weidenbach. Superposition decides the first-order logic fragment over ground theories. *Mathematics in Computer Science*, 6(4):427–456, 2012.

[20] P. López-García, L. Darmawan, M. Klemen, U. Liqat, F. Bueno, and M. V. Hermenegildo. Interval-based resource usage verification by translation into Horn clauses and an application to energy consumption. *Theory Pract. Log. Program.*, 18(2):167–223, 2018.

[21] F. Mesnard, É. Payet, and G. Vidal. Concolic testing in CLP. *Theory Pract. Log. Program.*, 20(5):671–686, 2020.

[22] F. Spoto, F. Mesnard, and É. Payet. A termination analyzer for java bytecode based on path-length. *ACM Trans. Program. Lang. Syst.*, 32(3):8:1–8:70, 2010.

### 33.4.5 Relevance for First-Order Logic Clause Sets

*Investigators: Fajar Haifani, Sophie Tourret, Christoph Weidenbach*

Given an unsatisfiable clause set, one could ask which clauses are necessary for any possible refutation (syntactically relevant), usable for some refutation (syntactically semi-relevant), or unusable (syntactically irrelevant). We formalize this notion of relevance for first-order logic [2] and demonstrate how this notion can be useful for explaining an entailment with respect to some axiom set [1]. The syntactic characterization of relevance is accompanied by a semantic characterization via conflict literals (contradictory simple facts) [3]. From an unsatisfiable clause set, we can always derive a pair of conflict literals. A relevant clause is necessary to derive any conflict literal, a semi-relevant clause is necessary to derive some conflict literal, and an irrelevant clause is not useful in deriving any conflict literals. It sheds some further light on why an entailment holds beyond known notions such as a minimal unsatisfiable subset. In order to test if a clause is (syntactically) semi-relevant, we sharpened and generalized the well-known completeness result for the set-of-support (SOS) resolution strategy: resolution equipped with the SOS strategy is refutationally complete on a clause set $N$ and SOS $M$ if and only if there is a resolution refutation from $N$ $M$ using a clause in $M$ [2].

#### References

[1] F. Haifani, P. Koopmann, S. Tourret, and C. Weidenbach. On a notion of relevance. In S. Borgwardt and T. Meyer, eds., *Proceedings of the 33rd International Workshop on Description Logics (DL 2020)*, Rhodes, Greece (Virtual Event), 2020, CEUR Workshop Proceedings 2663, Article 10. ceur-ws.org.

• [2] F. Haifani, S. Tourret, and C. Weidenbach. Generalized completeness for SOS resolution and its application to a new notion of relevance. In A. Platzer and G. Sutcliffe, eds., *Automated Deduction – CADE 28*, Virtual Event, 2021, LNAI 12699, pp. 327–343. Springer.

• [3] F. Haifani and C. Weidenbach. Semantic relevance. In J. Blanchette, L. Kovács, and D. Pattinson, eds., *Automated Reasoning (IJCAR 2022)*, Haifa, Israel, 2022, LNAI 13385, pp. 208–227. Springer.

### 33.4.6 Connection-Minimal Abduction for the EL Description Logic

*Investigators: Fajar Haifani, Sophie Tourret, Christoph Weidenbach, in cooperation with Patrick Koopmann (TU Dresden)*

Abductive reasoning helps in finding extensions of a knowledge base that imply some desired consequence. For example, abduction is useful for repairing incomplete knowledge bases as well as to explain a possibly unexpected observation. We particularly focus on TBox abduction in the $\mathcal{E}L$ description logic which is rather lightweight but prevalent in practice. The $\mathcal{E}L$ description logic corresponds via translation to a decidable first-order logic fragment, including the translation of potential models. The solution space for abductive reasoning can be huge or even infinite. So, different kinds of minimality notions can help to separate the chaff from the grain. We argue that the existing ones are insufficient in some settings and hence introduce connection minimality [3, 4, 2, 1]. This criterion accepts an abductive hypotheses only when it is *related* to the desired consequence. In addition, we provide a first-order reasoning pipeline to compute the connection-minimal hypotheses in a sound and complete way. The key technique relies on the notion of prime implicates. While the negation of a single prime implicate is already a first-order hypothesis, a connection-minimal hypothesis typically requires a combination of several prime implicates. Termination is obtainable by restricting the depth of the terms in the prime implicates. We also present an evaluation on ontologies from the medical domain by implementing a prototype with SPASS as a prime implicate generation engine.

#### References


- [1] F. Haifani, P. Koopmann, and S. Tourret. Abduction in EL via translation to FOL. In R. A. Schmidt, C. Wernhard, and Y. Zhao, eds., *Second-Order Quantifier Elimination and Related Topics (SOQE 2021)*, Online Event, 2021, CEUR Workshop Proceedings 3009, pp. 46–58. CEUR-WS.

- [2] F. Haifani, P. Koopmann, S. Tourret, and C. Weidenbach. *Connection-minimal Abduction in EL via Translation to FOL – Technical Report*, 2022. arXiv: 2205.08449.

- [3] F. Haifani, P. Koopmann, S. Tourret, and C. Weidenbach. Connection-minimal abduction in El via translation to FOL (extended abstract). In O. Arieli, M. Homola, J. C. Jung, and M.-L. Mugnier, eds., *Description Logics 2022 (DL 2022)*, Haifa, Israel, 2022, CEUR Workshop Proceedings 3263. CEUR-WS.org.

- [4] F. Haifani, P. Koopmann, S. Tourret, and C. Weidenbach. Connection-minimal abduction in $\mathcal{EL}$ via translation to FOL. In J. Blanchette, L. Kovács, and D. Pattinson, eds., *Automated Reasoning (IJCAR 2022)*, Haifa, Israel, 2022, LNAI 13385, pp. 188–207. Springer.

## 33.5 Arithmetic

*Coordinator: Thomas Sturm*

### 33.5.1 Parametric Geometric Analysis of Steady State Regimes

*Investigators: Hamid Rahkooy, Thomas Sturm*

In [2] we studied toricity of steady state regimes of biological models with polynomial vector fields. We proposed a generalization of the notion of toricity, in terms of the geometry of the variety instead of the syntactic shape of generators of the ideal. From an algorithmic point of view, our approach over the real numbers was purely based on logic. All relevant tests could be reduced to the decision of existential sentences in real closed fields, which corresponds to the SMT theory `QF_NRA`. In subsequent work [3], we transferred our logic-based approach to the complex numbers using Hilbert's Nullstellensatz as a decision procedure, which requires radical memberships tests based on Gröbner bases.

The next natural step is now the investigation of the same problems with parametric reaction rates. This is well motivated, as reaction rates are either measured with limited precision, or estimated often only by order of magnitude. Relevant biological findings should be robust under variation of those parameters; in his excellent textbook Feinberg points out: *The network itself will be our object of study, not the network endowed with a particular set of rate constants* [1, p.19].

Our generalization over the complex numbers [5] requires the careful use of comprehensive Gröbner bases and corresponding techniques [6]. Over the real numbers [4] the presence of parameters exceeds the SMT framework, and we make use of real quantifier elimination methods. We successfully analyze various biological models from the literature. In benchmark series with $n$-site phosphorylation networks we could (for $n = 5$) process models with up to 54 species and 30 parametric rate constants, which amounts to the elimination of 54 real quantifiers in an 84-dimensional space, arriving at concise scientifically interpretable conditions in the parameters.

### References


[1] M. Feinberg. *Foundations of Chemical Reaction Network Theory*, Applied Mathematical Sciences 202. Springer, 2019.

• [2] D. Grigoriev, A. Iosif, H. Rahkooy, T. Sturm, and A. Weber. Efficiently and effectively recognizing toricity of steady state varieties. *Mathematics in Computer Science*, 15:199–232, 2021.

[3] H. Rahkooy and T. Sturm. First-order tests for toricity. In F. Boulier, M. England, T. M. Sadykov, and E. V. Vorozhtsov, eds., *Computer Algebra in Scientific Computing (CASC 2020)*, Linz, Austria (Virtual Event), 2020, LNCS 12291, pp. 510–527. Springer.

• [4] H. Rahkooy and T. Sturm. Parametric toricity of steady state varieties of reaction networks. In F. Boulier, M. England, T. M. Sadykov, and E. V. Vorozhtsov, eds., *Computer Algebra in Scientific Computing (CASC 2021)*, Sochi, Russia, 2021, LNCS 12865, pp. 314–333. Springer.

• [5] H. Rahkooy and T. Sturm. Testing binomiality of chemical reaction networks using comprehensive Gröbner systems. In F. Boulier, M. England, T. M. Sadykov, and E. V. Vorozhtsov, eds., *Computer Algebra in Scientific Computing (CASC 2021)*, Sochi, Russia, 2021, LNCS 12865, pp. 334–352. Springer.

[6] V. Weispfenning. Comprehensive Gröbner bases. *J. Symb. Comput.*, 14(1):1–29, 1992.

### 33.5.2 Signature-based Gröbner Basis Computation

*Investigators: Hamid Rahkooy, Thomas Sturm in cooperation with Alexander Demin (Univ. HSE, Moscow)*

In the course of an internship project, we have realized a REDUCE package F5 for the computation of Gröbner Bases based on Faugère's F5 algorithm [3, 2]. It has been launched as a part of the regular REDUCE distribution on SourceForge. The implementation uses multivariate rational functions as the coefficient field. The package is compatible with existing REDUCE term orderings, as used with the GROEBNER package and the CGB package.

From an Automated Reasoning viewpoint this offers a decision procedure for algebraically closed fields, which includes the complex numbers. The code has been designed with possible future generalization to the formal treatment of the coefficients as parameters and comprehensive Gröbner Bases in mind, which, in combination with other components of REDUCE, namely Redlog, would yield quantifier elimination and decision procedures for algebraically closed fields, which includes the complex numbers, as well as for real closed fields, which includes the real numbers. For further details see [1].

#### References

• [1] A. Demin, H. Rahkooy, and T. Sturm. F5: A REDUCE package for signature-based Gröbner basis computation. In *CASC 2022 – Computer Algebra in Scientific Computing*, Gebze, Turkey, 2022. HAL.

[2] C. Eder and J. Perry. F5C: A variant of Faugère's F5 algorithm with reduced Gröbner bases. *J. Symb. Comput.*, 45(12):1442–1458, 2010.

[3] J. C. Faugère. A new efficient algorithm for computing Gröbner bases without reduction to zero (F5). In *Proc. ISSAC 2002*, pp. 75–83. ACM, 2002.

### 33.5.3 Generation and Public Provision of Formal Specifications of Biomodels

*Investigators: Thomas Sturm in cooperation with Christoph Lüders, (Univ. Kassel), Ovidiu Radulescu (Univ. Montpellier)*

Within our research line on Biological Systems we are applying automated reasoning to problems such as sustained oscillations and Hopf bifurcations, multi-stationarity, multi-scale model reduction, dynamical invariants, and structural properties of steady state varieties. Compared to numerical analysis and simulation, our approach provides not only quantitative but also qualitative results about network dynamics, to some extent in parametric settings. Our focus has been on reaction networks in the sense of chemical reaction network theory [3]. Such networks are usually stored and exchanged in the Systems Biology Markup Language (SBML), a free, open and standardized XML-based format [4]. However, on the one hand, formal methods do not utilize the full information contained in SBML models. For instance, SBML was designed with a focus on network simulation and supports corresponding concepts like events and initial assignments, which are not natural from our point of view. On the other hand, our automated reasoning is based on symbolic computation, which operates on formal objects like polynomials, exact rational numbers, and ordinary differential equations (ODEs), which are not readily available in SBML. For instance, ODEs describing differential network kinetics can be modeled as pieces of code to be used with numerical solvers instead of mathematical expressions accessible for formal methods.

The rigorous construction of suitable formal models requires joint competence and combined efforts from computer science, mathematics, and biology. Not only resolving but even recognizing issues is a challenge with every single model considered. Software tool chains must be integrated with human interaction. Emerging from combined efforts within the SYMBIONT project [1, 2], we have picked up the challenge and launched a web service ODEbase at odebase.org that provides reliable input data for formal methods to the research community. We identify the following benefits:

1. Interdisciplinary competence: Researchers get access to biologically adequate translations of existing relevant biomodels instead of ad hoc solutions.

2. Availability: ODEbase models used and cited in the literature can be conveniently reviewed on the basis of the original data and re-used in follow-up publications.

3. Canonical reference: ODEbase provides an unambiguous mapping of the, in general, too liberal SBML names for species concentrations and parameters to common mathematical notation. This facilitates comparability of results between publications.

4. Benchmarking: Beyond its primary purpose, ODEbase is perfectly suited to generate benchmark sets for novel algorithms and software in the field.

At the time of writing, ODEbase comprises 660 models, mostly from the renowned biomodels.net online database [7]. For further details on the project see [5, 6].

## References


[1] F. Boulier, F. Fages, O. Radulescu, S. S. Samal, A. Schuppert, W. Seiler, T. Sturm, S. Walcher, and A. Weber. The SYMBIONT project: Symbolic methods for biological networks. *Faculty of 1000 Research*, 7, Article 1341, 2018.

[2] F. Boulier, F. Fages, O. Radulescu, S. S. Samal, A. Schuppert, W. Seiler, T. Sturm, S. Walcher, and A. Weber. The SYMBIONT project: Symbolic methods for biological networks. *ACM Communications in Computer Algebra*, 52(3):67–70, 2018.

[3] M. Feinberg. *Foundations of Chemical Reaction Network Theory*, Applied Mathematical Sciences 202. Springer, 2019.

[4] M. Hucka et al. The systems biology markup language (SBML): A medium for representation and exchange of biochemical network models. *Bioinformatics*, 19(4):524–531, 2003.

• [5] C. Lüders, T. Sturm, and O. Radulescu. *ODEbase: A Repository of ODE Systems for Systems Biology*, 2022. arXiv: 2201.08980.

• [6] C. Lüders, T. Sturm, and O. Radulescu. ODEbase: A repository of ODE systems for systems biology. *Bioinformatics Advances*, 2(2), Article vbac027, 2022.

[7] L. Novere et al. BioModels database: A free, centralized database of curated, published, quantitative kinetic models of biochemical and cellular systems. *Nucleic acids res.*, 34(suppl_1):D689–D691, 2006.

### 33.5.4 Reduction of Reaction Network Kinetics to Multiple Timescales

*Investigators: Thomas Sturm, in cooperation with Niclas Kruff and Sebastian Walcher (Aachen, Germany), Christoph Lüders (Bonn, Germany), Ovidiu Radulescu (Montpellier, France)*

In the previous biennial report we had discussed work in progress on interdisciplinary work in computer science, mathematics, and systems biology, which is concerned with the reduction of a system of ordinary differential equations into several simpler subsystems, each corresponding to a certain order of magnitude of velocities of differential variables. To our knowledge this is the first mathematically rigorous approach for reaction networks that allows for multiple time scales. Previous work either did not give any formal guarantees on the obtained results, or was limited to only two different time scales. Our computations are based on massive SMT solving over various theories, including `QF_LRA` for tropicalizations, `QF_NRA` for testing Hurwitz conditions on Eigenvalues, and `QF_LIA` for finding sufficient differentiability conditions for hyperbolic attractivity of critical manifolds. Gröbner reduction techniques are used for final algebraic simplification.

This work has been finished during the reporting period, and the results have been published as a journal article [1].

#### References

• [1] N. Kruff, C. Lüders, O. Radulescu, T. Sturm, and S. Walcher. Algorithmic reduction of biological networks with multiple time scales. *Mathematics in Computer Science*, 15:499–534, 2021.

## 33.6 Saturation Theorem Proving

*Coordinators: Sophie Tourret, Uwe Waldmann*

Saturation calculi like resolution or superposition have been a long time research topic of our group. During the report period, we have worked in particular on extending superposition provers to higher-order logics, on an abstract view on saturation calculi with splitting, on efficient subsumption tests for BS(LRA) clauses, and on the (in-)completeness of destructive equality resolution in the superposition calculus.

### 33.6.1 Extension of a High-Performance Prover to Higher-Order Logic

*Investigators: Jasmin Blanchette, in cooperation with Petar Vukmirović (Vrije Universiteit Amsterdam) and Stephan Schulz (DHBW Stuttgart)*

Most users of proof assistants want more proof automation. Some proof assistants discharge goals by translating them to first-order logic and invoking an efficient prover on them, but much is lost in translation. Instead, we decided to extend first-order provers with native support for higher-order features. Building on our earlier extension of the E [2] prover to $\lambda$-free higher-order logic [6], we now extended E to full higher-order logic. We call this new prover $\lambda$E. This work has been accepted at TACAS 2023.

This new extension of E is based on λ-superposition [1], a proof calculus we developed together with other researchers. The calculus is designed for reasoning about higher-order logic and is implemented in the experimental superposition prover Zipperposition [1]. This OCaml prover is not nearly as efficient as E, which is written in C; nevertheless, it has won the higher-order division of the CASC prover competition [4, 5] in 2020, 2021, and 2022.

We expected to obtain even better performance with λE. In λE's implementation, we used the extensive experience with Zipperposition to choose a set of effective rules that could easily be retrofitted into an originally first-order prover. Another guiding principle was gracefulness: Our changes should not impact the strong first-order performance of E.

We evaluated λE on a selection of proof assistants benchmarks as well as all higher-order theorems in the TPTP library (the problems used at CASC) [3]. λE outperformed all other higher-order provers on the proof assistant benchmarks; on the TPTP benchmarks, it ended up second only to the cooperative version of Zipperposition, which employs E as a backend. An arguably fairer comparison without the backend puts λE in first place for both benchmark suites.

## References


- [1] A. Bentkamp, J. Blanchette, S. Tourret, P. Vukmirović, and U. Waldmann. Superposition with lambdas. *Journal of Automated Reasoning*, 65:893–940, 2021.

[2] S. Schulz, S. Cruanes, and P. Vukmirovic. Faster, higher, stronger: E 2.3. In P. Fontaine, ed., *Automated Deduction – CADE 27 – 27th International Conference on Automated Deduction, Natal, Brazil, August 27-30, 2019, Proceedings*, 2019, LNCS 11716, pp. 495–507. Springer.

[3] G. Sutcliffe. The TPTP problem library and associated infrastructure – from CNF to TH0, TPTP v6.4.0. *J. Autom. Reason.*, 59(4):483–502, 2017.

[4] G. Sutcliffe. The 10th IJCAR automated theorem proving system competition – CASC-J10. *AI Commun.*, 34(2):163–177, 2021.

- [5] G. Sutcliffe and M. Desharnais. The CADE-28 automated theorem proving system competition – CASC-28. *AI Communications*, 34(4):259–276, 2021.

- [6] P. Vukmirović, J. Blanchette, S. Cruanes, and S. Schulz. Extending a brainiac prover to lambda-free higher-order logic. *International Journal on Software Tools for Technology Transfer (STTT)*, 24:67–87, 2022.


### 33.6.2 Techniques for Higher-Order Superposition

*Investigators: Jasmin Blanchette and Sophie Tourret, in cooperation with Petar Vukmirović, Alexander Bentkamp, Visa Nummelin (Vrije Universiteit Amsterdam), and Simon Cruanes (Aesthetic Integration)*

The λ-superposition calculus, which extends superposition to higher-order logic, introduces new challenges such as infinitely branching inference rules, new possibilities such as reasoning about Booleans, and the need to curb the explosion of specific higher-order rules. We introduced techniques [5, 6] that address these issues and extensively evaluated their implementation in the Zipperposition [1] theorem prover. The new techniques cover most parts of a modern higher-order theorem prover, from preprocessing to additional calculus rules to heuristics to backend integration. Largely thanks to their use, Zipperposition won the higher-order division of the last three CASC competitions [3, 4]. Specifically, we introduced the following techniques.

Many higher-order problems extensively use symbol definitions to simplify their representation. We studied several ways to exploit the definitions.

By working on formulas rather than clauses, tableau techniques take a more holistic view of a higher-order problem. Through its support for delayed clausification and, more generally, calculus-level formula manipulation, $\lambda$-superposition enables us to simulate most successful tableau techniques in a saturating prover.

The main drawback of $\lambda$-superposition compared with combinatory superposition [2], a rival, is that $\lambda$-superposition relies on rules that enumerate possibly infinite sets of unifiers. We designed a mechanism that interleaves infinitely branching inferences with the standard saturation process.

Using first-order backends to finish the proof is common practice in higher-order reasoning. We described how to achieve a balance between allowing native higher-order reasoning and delegating reasoning to a first-order backend.

## References


• [1] A. Bentkamp, J. Blanchette, S. Tourret, P. Vukmirović, and U. Waldmann. Superposition with lambdas. *Journal of Automated Reasoning*, 65:893–940, 2021.

[2] A. Bhayat and G. Reger. A combinator-based superposition calculus for higher-order logic. In N. Peltier and V. Sofronie-Stokkermans, eds., *Automated Reasoning – 10th International Joint Conference, IJCAR 2020, Paris, France, July 1-4, 2020, Proceedings, Part I*, 2020, LNCS 12166, pp. 278–296. Springer.

[3] G. Sutcliffe. The 10th IJCAR automated theorem proving system competition – CASC-J10. *AI Commun.*, 34(2):163–177, 2021.

• [4] G. Sutcliffe and M. Desharnais. The CADE-28 automated theorem proving system competition – CASC-28. *AI Communications*, 34(4):259–276, 2021.

• [5] P. Vukmirović, A. Bentkamp, J. Blanchette, S. Cruanes, V. Nummelin, and S. Tourret. Making higher-order superposition work. In A. Platzer and G. Sutcliffe, eds., *Automated Deduction – CADE 28*, Virtual Event, 2021, LNAI 12699, pp. 415–432. Springer.

• [6] P. Vukmirović, A. Bentkamp, J. Blanchette, S. Cruanes, V. Nummelin, and S. Tourret. Making higher-order superposition work. *Journal of Automated Reasoning*, 66:541–564, 2022.


### 33.6.3 A Unifying Splitting Framework

*Investigators: Jasmin Blanchette and Sophie Tourret, in cooperation with Gabriel Ebner (Vrije Universiteit Amsterdam)*

The AVATAR architecture, which is based on a satisfiability (SAT) solver, is a very successful approach for splitting clauses in a superposition prover. Voronkov reported that an AVATAR-enabled Vampire could solve 421 TPTP problems that had never been solved before by any system [3, Sect. 9], a mind-boggling number. But is AVATAR refutationally complete? With our work, we provided a qualified yes answer [2]. But before we could answer this question, we needed to mathematize splitting.

Our starting point was the *saturation framework* by Waldmann et al. [4, 5], based on the work of Bachmair and Ganzinger [1]. It covers a wide array of techniques, but "the main missing piece of the framework is a generic treatment of clause splitting" [4, p. 332]. We provided that missing piece, in the form of a *splitting framework*, and used it to show the completeness of an AVATAR-like architecture.

Our framework has five layers, linked by refinement. The first layer consists of a *base calculus*, such as resolution or superposition, which must be refutationally complete. From a base calculus, our framework can be used to derive the second layer, which we call the *splitting calculus*. This extends the base calculus with splitting and inherits the base's completeness.

A major component of AVATAR is the SAT solver. A *model-guided prover* operates on states that combine a SAT model and a clause set. This layer is also dynamically complete.

The fourth layer introduces AVATAR's *locking* mechanism. With locking, a clause can be temporarily disabled by another clause. Here we made a first discovery: AVATAR-style locking compromises completeness and must be curtailed.

Finally, the fifth layer is an *AVATAR-based prover*. This refines the locking model-guided prover of the fourth layer with the given clause procedure, which saturates a clause set by distinguishing between active and passive clauses. Here we made another discovery: Selecting clauses fairly is not enough to guarantee completeness. We need a stronger criterion.

There are also implications for other splitting architectures. We applied our framework to splitting without backtracking, labeled splitting, and SMT with quantifiers. This gave us a solid basis for comparison as well as some new theoretical results.

## References


[1] L. Bachmair and H. Ganzinger. Resolution theorem proving. In J. A. Robinson and A. Voronkov, eds., *Handbook of Automated Reasoning (in 2 volumes)*, pp. 19–99. Elsevier and MIT Press, 2001.

• [2] G. Ebner, J. Blanchette, and S. Tourret. A unifying splitting framework. In A. Platzer and G. Sutcliffe, eds., *Automated Deduction – CADE 28*, Virtual Event, 2021, LNAI 12699, pp. 344–360. Springer.

[3] A. Voronkov. AVATAR: the architecture for first-order theorem provers. In A. Biere and R. Bloem, eds., *Computer Aided Verification – 26th International Conference, CAV 2014, Held as Part of the Vienna Summer of Logic, VSL 2014, Vienna, Austria, July 18-22, 2014. Proceedings*, 2014, LNCS 8559, pp. 696–710. Springer.

[4] U. Waldmann, S. Tourret, S. Robillard, and J. Blanchette. A comprehensive framework for saturation theorem proving. In N. Peltier and V. Sofronie-Stokkermans, eds., *Automated Reasoning (IJCAR 2020)*, Paris, France (Virtual Conference), 2020, LNAI 12166, pp. 316–334. Springer.

• [5] U. Waldmann, S. Tourret, S. Robillard, and J. Blanchette. A comprehensive framework for saturation theorem proving. *Journal of Automated Reasoning*, 66:499–539, 2022.

### 33.6.4 An Efficient Subsumption Test Pipeline for BS(LRA) Clauses

*Investigators: Martin Bromberger, Lorenz Leutgeb, and Christoph Weidenbach*

Deletion of subsumed formulas is crucial for saturation theorem proving. Implementations are only competitive if clauses that are redundant, and therefore do not inform progress, are deleted to prevent further redundant calculations and to free up capacity of valuable main memory. Although the SCL calculi family [1, 5, 3] does not require forward subsumption tests, backward subsumption remains important. Various heuristics for efficient detection of subsumed clauses have been reported and benchmarked for first-order logic (FOL) [7, 6]. However, these schemes are not applicable when FOL is combined with a background theory, such as linear real arithmetic (LRA) [2]: In contrast to FOL, where purely syntactic criteria are sufficient to check for subsumption, deciding subsumption in the presence of LRA constraints always requires a semantic component. We addressed this challenge by presenting the *sample point heuristic* [4], which takes into account LRA constraints. The main idea is to store a solution per constrained clause, which can later be used in attempts to quickly falsify subsumption. We can obtain these sample points almost for free, by coupling our heuristic with tautology deletion, another important and common measure for redundancy elimination. This is possible because a failed check for tautology deletion corresponds to a satisfiable solution of the non-tautological clause. As a proof of concept, we implemented a saturation theorem prover for the Bernays-Schönfinkel fragment of FOL with LRA. We implemented a pipeline design combining other simple heuristics and the sample point heuristic, which is extensible and allows for an easy comparison of the cost-effectivenes of all heuristics. We showed that the sample point heuristic is very specific, i.e. 94% out of all true negatives are predicted correctly, and effective (speed-up by a factor of 124), thus greatly improving the performance of subsumption checking.

#### References

[1] G. Alagi and C. Weidenbach. NRCL – a model building approach to the Bernays-Schönfinkel fragment. In C. Lutz and S. Ranise, eds., *Frontiers of Combining Systems (FroCoS 2015)*, Wrocław, Poland, 2015, LNAI 9322, pp. 69–84. Springer.

[2] E. Althaus, E. Kruglov, and C. Weidenbach. Superposition modulo linear arithmetic SUP(LA). In S. Ghilardi and R. Sebastiani, eds., *Frontiers of Combining Systems (FroCos 2009)*, Trento, Italy, 2009, LNAI 5749, pp. 84–99. Springer.

• [3] M. Bromberger, A. Fiori, and C. Weidenbach. Deciding the Bernays-Schoenfinkel fragment over bounded difference constraints by simple clause learning over theories. In F. Henglein, S. Shoham, and V. Yakir, eds., *Verification, Model Checking, and Abstract Interpretation (VMCAI 2021)*, Copenhagen, Denmark (Online), 2021, LNCS 12597, pp. 511–533. Springer.

• [4] M. Bromberger, L. Leutgeb, and C. Weidenbach. An efficient subsumption test pipeline for BS(LRA) clauses. In J. Blanchette, L. Kovács, and D. Pattinson, eds., *Automated Reasoning (IJCAR 2022)*, Haifa, Israel, 2022, LNAI 13385, pp. 147–168. Springer.

[5] A. Fiori and C. Weidenbach. SCL clause learning from simple models. In P. Fontaine, ed., *Automated Deduction – CADE 27*, Natal, Brazil, 2019, LNAI 11716, pp. 233–249. Springer.

[6] P. Graf. *Term Indexing*, LNCS 1053. Springer, 1996.

[7] T. Tammet. Towards efficient subsumption. In *CADE-15, 1998*, 1998, LNCS 1421, pp. 427–441. Springer.

### 33.6.5 Destructive Equality Resolution in the Superposition Calculus

*Investigator: Uwe Waldmann*

Bachmair's and Ganzinger's Superposition Calculus [1] comes with an abstract redundancy concept that describes under which circumstances clauses can be simplified away or deleted during a saturation without destroying the refutational completeness of the calculus. Typical concrete simplification and deletion techniques that are justified by the abstract redundancy concept are tautology deletion, subsumption deletion, and demodulation, and with a more refined definition of redundancy [2] joinability and connectedness can be covered as well.

The outlier is *destructive equality resolution* (DER), that is, the replacement of a clause $x \approx t \quad C$ with $x \; / \; \mathrm{vars}(t)$ by $C \; x \quad t$ . This operation is for instance implemented in the E prover [3], and it is known to be useful in practice: It increases the number of problems that E can solve and it also reduces E's runtime per solved problem. Little is known, however, about how it affects the refutational completeness of the calculus, except for the special case that $t$ is also a variable, where DER is equivalent to selecting the literal $x \approx t$ so that equality resolution becomes the only possible inference with this clause.

We have shown that naively adding DER to the standard redundancy criterion destroys the refutational completeness of the calculus, and that this holds even if we impose a particular strategy on simplification steps (say, that simplifications must be performed eagerly and that demodulation and subsumption have a higher precedence than DER). The counterexample to refutational completeness is highly unrealistic, though, as it relies on the deletion of a clause that is redundant, but that no reasonable superposition prover would ever detect as redundant.

The question remains therefore, whether DER destroys the refutational completeness of the calculus *in practice*. Are there alternative clause orderings and associated redundancy criteria that are refutationally complete together with the Superposition Calculus and that justify DER as well as (all/most) commonly implemented deletion and simplification techniques, say, tautology detection, demodulation, or subsumption?

The key problem that one encounters here is that the ordering requirements of the superposition inference rule and the DER rule are diametrically opposed. DER appears to be highly incompatible with the usual clause orderings that one obtains as multiset extensions of some literal ordering. As a first step, we were able to overcome this problem and developed a clause ordering and an associated redundancy criterion that is compatible with the inference rules of the Superposition Calculus and justifies DER (with some restrictions in the non-Horn case), propositional subsumption, and demodulation to subterms. The question whether this can be extended to full DER for the non-Horn case, general demodulation, and first-order subsumption, or whether there is a practical counterexample to refutational completeness is still open though.

### References

[1] L. Bachmair and H. Ganzinger. Rewrite-based equational theorem proving with selection and simplification. *Journal of Logic and Computation*, 4(3):217–247, 1994. Revised version of Max-Planck-Institut für Informatik technical report, MPI-I-91-208, 1991.

[2] A. Duarte and K. Korovin. Ground joinability and connectedness in the superposition calculus. In J. Blanchette, L. Kovács, and D. Pattinson, eds., *Automated Reasoning – 11th International Joint Conference, IJCAR 2022, Haifa, Israel, August 8-10, 2022, Proceedings*, 2022, LNCS 13385, pp. 169–187. Springer.

[3] S. Schulz. E—a brainiac theorem prover. *AI Commun.*, 15(2–3):111–126, 2002.

## 33.7 Formalizing Logic

*Coordinator: Sophie Tourret*

Our work in formalizing logical calculi has two important aspects. First, we formally prove correctness of our pen and paper proofs on properties of our calculi. Second, we use the proof obligations generated during formalisation as benchmarks for our automated reasoning calculi.

### 33.7.1 Refining a Modular Saturation Framework into Variants of the Given-Clause Loop in Isabelle/HOL

*Investigators: Jasmin Blanchette, Sophie Tourret, in cooperation with Qi Qiu (Univ. Lyon, France)*

We built on the framework for proving the completeness of automatic theorem provers based on resolution, superposition, or other saturation calculi that we had previously created and formalized [6, 8], to formally verify in Isabelle/HOL[5] four variants of the given-clause loop, the main loop at the core of saturation-based theorem provers. The Otter and iProver loops extend the existing given-clause loop formalization with more specific rules corresponding respectively to what happens in the OTTER prover [4] and the more recent iProver prover [3]. Their formalization integrates fairness criteria ensuring termination. The DISCOUNT and Zipperposition loops similarly extend the lazy given-clause loop formalization as in the DISCOUNT [2] and Zipperposition prover [7]. A specificity of the recent Zipperposition loop is that it is compatible with superposition in higher-order logic, in that it can deal with inferences producing infinitely many resolvents via dovetailing the choice of a given-clause formula with the production of resolvents from scheduled inferences.

This work brings us one refinement closer to executable verified saturation provers and it also clarifies some fine points previously misunderstood in the literature concerning the most recent and complex variant of the given-clause loop, namely the Zipperposition loop. The formalisation itself is available on the Archive of Formal Proofs [1] and a paper describing this work is currently submitted at CADE-29.

## References


[1] J. Blanchette, Q. Qiu, and S. Tourret. Given clause loops. *Archive of Formal Proofs*, 2023.

[2] J. Denzinger, M. Kronenburg, and S. Schulz. DISCOUNT—a distributed and learning equational prover. *J. Autom. Reason.*, 18(2):189–198, 1997.

[3] A. Duarte and K. Korovin. Implementing superposition in iProver (system description). In N. Peltier and V. Sofronie-Stokkermans, eds., *IJCAR 2020, Part I*, 2020, LNCS 12167, pp. 388–397. Springer.

[4] W. W. McCune. OTTER 3.0 Reference Manual and Guide. Technical report, Argonne National Lab, 1994.

[5] T. Nipkow, L. C. Paulson, and M. Wenzel. *Isabelle/HOL: A Proof Assistant for Higher-Order Logic*, LNCS 2283. Springer, 2002.

[6] S. Tourret and J. Blanchette. A modular Isabelle framework for verifying saturation provers. In C. Hriţcu and A. Popescu, eds., *CPP '21, 10th ACM SIGPLAN International Conference on Certified Programs and Proofs*, Virtual, Denmark, 2021, pp. 224–237. ACM.

[7] P. Vukmirović, A. Bentkamp, J. Blanchette, S. Cruanes, V. Nummelin, and S. Tourret. Making higher-order superposition work. *Journal of Automated Reasoning*, 66:541–564, 2022.

[8] U. Waldmann, S. Tourret, S. Robillard, and J. Blanchette. A comprehensive framework for saturation theorem proving. *Journal of Automated Reasoning*, 66:499–539, 2022.


### 33.7.2 An Isabelle/HOL Formalization of the SCL(FOL) Calculus

*Investigators: Martin Bromberger, Martin Desharnais, Christoph Weidenbach*

The SCL ("Clause Learning from Simple Models" or simply "Simple Clause Learning") family of calculi lifts a conflict-driven clause learning (CDCL) approach to first-order logic: SCL(FOL) is for first-order logic without equality [9, 7], SCL(T) is for first-order logic with theories [5], SCL(EQ) is for first-order logic with equality [10], and HSCL is for exhaustive partial models exploration in first-order logic without equality [6].

We developed an Isabelle/HOL formalization [8, 4] of SCL(FOL) based on and developed in parallel to [7]. The main results are soundness, non-redundancy of learned clauses, termination, and refutational completeness. Compared to the pen-and-paper presentation, we simplified the calculus (e.g., no more explicit tracking of decision levels), generalized the calculus (e.g., the Backtrack rule), strengthened existing theorems (e.g., the dynamic non-redundancy theorem), and proved new ones (e.g., the static non-subsumption theorem). We discovered a previously unknown bug in the Backtrack rule and the fix was contributed back to [7].

We followed the basic ideas of the formalization of rule-based, model-driven CDCL calculi for propositional logic [3, 2] but, compared to propositional logic, first-order logic adds a number of challenges: the extra term level requires the consideration of variables, substitutions, groundings, and the concept of factorization. In order to preserve completeness, propagation of ground literals must not be exhaustive anymore, resulting in a level-wise exploration w.r.t. a bounding atom. Inside this bound, we showed that the calculus always terminates. If one level does not suffice for refutational completeness, the bound can be increased and exploration can continue. We proved the existence of a bound sufficient to derive , which guarantees that only finitely many levels need to be explored.

Our work is part of the IsaFoL (Isabelle Formalization of Logic) [1] effort. It is available from the *Archive of Formal Proofs* (AFP) and builds heavily upon many other entries of the AFP. We contributed many lemmas and definitions back to both the Isabelle distribution and the used AFP entries.

## References


[1] J. C. Blanchette. Formalizing the metatheory of logical calculi and automatic provers in Isabelle/HOL (invited talk). In A. Mahboubi and M. O. Myreen, eds., *CPP'19, 8th ACM SIGPLAN International Conference on Certified Programs and Proofs*, Cascais, Portugal, 2019, pp. 1–13. ACM.

[2] J. C. Blanchette, M. Fleury, P. Lammich, and C. Weidenbach. A verified SAT solver framework with learn, forget, restart, and incrementality. *Journal of Automated Reasoning*, 61(1-4):333–365, 2018.

[3] J. C. Blanchette, M. Fleury, and C. Weidenbach. A verified SAT solver framework with learn, forget, restart, and incrementality. In N. Olivetti and A. Tiwari, eds., *Automated Reasoning (IJCAR 2016)*, Coimbra, Portugal, 2016, LNAI 9706, pp. 25–44. Springer.

• [4] M. Bromberger, M. Desharnais, and C. Weidenbach. An Isabelle/HOL formalization of the SCL(FOL) calculus. In B. Pientka and C. Tinelli, eds., *CADE-29: 29th international Conference on Automated Deduction*, 2023. Accepted for publication.

• [5] M. Bromberger, A. Fiori, and C. Weidenbach. Deciding the Bernays-Schoenfinkel fragment over bounded difference constraints by simple clause learning over theories. In F. Henglein, S. Shoham, and V. Yakir, eds., *Verification, Model Checking, and Abstract Interpretation (VMCAI 2021)*, Copenhagen, Denmark (Online), 2021, LNCS 12597, pp. 511–533. Springer.

• [6] M. Bromberger, S. Schwarz, and C. Weidenbach. Exploring partial models with SCL. In B. Konev, C. Schon, and A. Steen, eds., *Practical Aspects of Automated Reasoning 2022 (PAAR 2022)*, Haifa, Israel, 2022, CEUR Workshop Proceedings 3201. CEUR-WS.org.

• [7] M. Bromberger, S. Schwarz, and C. Weidenbach. *SCL(FOL) Revisited*, 2023. arXiv: 2302.05954.

[8] M. Desharnais. A formalization of the SCL(FOL) calculus: Simple clause learning for first-order logic. *Archive of Formal Proofs*, 2023. `https://isa-afp.org/entries/Simple_Clause_Learning.html`, Formal proof development.

[9] A. Fiori and C. Weidenbach. SCL clause learning from simple models. In P. Fontaine, ed., *Automated Deduction – CADE 27*, Natal, Brazil, 2019, LNAI 11716, pp. 233–249. Springer.

• [10] H. Leidinger and C. Weidenbach. SCL(EQ): SCL for first-order logic with equality. In J. Blanchette, L. Kovács, and D. Pattinson, eds., *Automated Reasoning (IJCAR 2022)*, Haifa, Israel, 2022, LNAI 13385, pp. 228–247. Springer.


### 33.7.3 Seventeen Provers Under the Hammer

*Investigators: Martin Desharnais, Jasmin Blanchette, in cooperation with Petar Vukmirović (Vrije Universiteit Amsterdam), and Makarius Wenzel*

One of the main success stories of automatic theorem provers has been their integration into proof assistants. Such integrations, or "hammers," increase proof automation and hence user productivity. There is ample evidence that hammers can be effective. The landmark

Judgment Day study by Böhme and Nipkow [5] showed for the first time that a combination of three automatic provers (E [9], SPASS [14] and Vampire [7]) could discharge about half of the goals that arise in typical Isabelle/HOL [8] developments. Smaller evaluations in later papers (e.g., [4, 12, 2, 10]) cover more provers, but they mostly predate the new generation of higher-order provers: cvc5 [1], Ehoh [13], Leo-III [11], Vampire [3], and Zipperposition [13].

We use Isabelle/HOL's Sledgehammer and Mirabelle tools to find out how useful modern provers are at proving formulas in higher-order logic. Our evaluation [6] follows in the steps of Judgment Day, but instead of three provers we use 17, including modern SMT solvers and higher-order provers. The evaluation is first and foremost an assessment of the provers, their features, and of various encoding schemes (e.g., of types and $\lambda$-abstractions) on Sledgehammer-generated problems. In particular, we wanted to determine whether native implementations of various features perform better than encodings, which is a reasonable thing to hope for.

Specifically, we found that native support for monomorphic types was beneficial, but the current support for polymorphism is disappointing. Native support for features such as linear arithmetic and higher-order logic helps some provers and is detrimental to others. Overall, we see that simply implementing a feature in a prover is often not enough; to be useful, the feature must be fine-tuned based on benchmarks. Our work offers an alternative yardstick for comparing modern provers, next to the benchmarks and competitions emerging from the TPTP World and SMT-LIB.

## References


[1] H. Barbosa, A. Reynolds, D. E. Ouraoui, C. Tinelli, and C. W. Barrett. Extending SMT solvers to higher-order logic. In P. Fontaine, ed., *CADE-27*, 2019, LNCS 11716, pp. 35–54. Springer.

• [2] A. Bentkamp, J. Blanchette, S. Tourret, and P. Vukmirović. Superposition for full higher-order logic. In A. Platzer and G. Sutcliffe, eds., *Automated Deduction – CADE 28*, Virtual Event, 2021, LNAI 12699, pp. 396–412. Springer.

[3] A. Bhayat and G. Reger. A combinator-based superposition calculus for higher-order logic. In N. Peltier and V. Sofronie-Stokkermans, eds., *IJCAR 2020, Part I*, 2020, LNCS 12166, pp. 278–296. Springer.

[4] J. C. Blanchette, S. Böhme, and L. C. Paulson. Extending Sledgehammer with SMT solvers. *J. Autom. Reason.*, 51(1):109–128, 2013.

[5] S. Böhme and T. Nipkow. Sledgehammer: Judgement Day. In J. Giesl and R. Hähnle, eds., *IJCAR 2010*, 2010, LNCS 6173, pp. 107–121. Springer.

• [6] M. Desharnais, P. Vukmirović, J. Blanchette, and M. Wenzel. Seventeen provers under the hammer. In J. Andronick and L. de Moura, eds., *13th International Conference on Interactive Theorem Proving (ITP 2022)*, Haifa, Israel, 2022, Leibniz International Proceedings in Informatics 237, Article 8. Schloss Dagstuhl.

[7] L. Kovács and A. Voronkov. First-order theorem proving and Vampire. In N. Sharygina and H. Veith, eds., *Computer Aided Verification – 25th International Conference, CAV 2013, Saint Petersburg, Russia, July 13-19, 2013. Proceedings*, 2013, LNCS 8044, pp. 1–35. Springer.

[8] T. Nipkow, L. C. Paulson, and M. Wenzel. *Isabelle/HOL: A Proof Assistant for Higher-Order Logic*, LNCS 2283. Springer, 2002.

[9] S. Schulz, S. Cruanes, and P. Vukmirović. Faster, higher, stronger: E 2.3. In P. Fontaine, ed., *Proc. of the 27th CADE, Natal, Brasil*, 2019, no. 11716 in LNAI, pp. 495–507. Springer.

• [10] H.-J. Schurr, M. Fleury, and M. Desharnais. Reliable reconstruction of fine-grained proofs in a proof assistant. In A. Platzer and G. Sutcliffe, eds., *Automated Deduction – CADE 28*, Virtual Event, 2021, LNAI 12699, pp. 450–467. Springer.

[11] A. Steen and C. Benzmüller. Extensional higher-order paramodulation in Leo-III. *J. Autom. Reason.*, 65(6):775–807, 2021.

[12] N. Sultana, J. C. Blanchette, and L. C. Paulson. LEO-II and Satallax on the Sledgehammer test bench. *J. Applied Logic*, 11(1):91–102, 2013.

• [13] P. Vukmirović, A. Bentkamp, J. Blanchette, S. Cruanes, V. Nummelin, and S. Tourret. Making higher-order superposition work. *Journal of Automated Reasoning*, 66:541–564, 2022.

[14] C. Weidenbach, D. Dimova, A. Fietzke, M. Suda, and P. Wischnewski. SPASS version 3.5. In R. A. Schmidt, ed., *22nd International Conference on Automated Deduction (CADE-22)*, 2009, LNAI, pp. 140–145. Springer.

## 33.8 Software and Applications

*Coordinator: Martin Bromberger*

Software development is an important step in automated reasoning research because we develop reasoning techniques in order to solve real world problems. For this purpose, we either develop our own state-of-the-art software systems or we develop prototypes as proofs of concept so our techniques are reimplemented and used by others. This prototyping is especially important because the logics we consider are complex, e.g., NP-complete or even undecidable. This means our techniques are only useful in practice if they behave more efficiently on real world problems than on random problems. As a result, a pure complexity analysis is often not sufficient to predict the efficiency of our techniques. We rely instead on experimentation through implementation in order to verify advances in the theory of automated reasoning. The maturity of our software efforts ranges from prototypical experimentation (Sects. 33.8.2, 33.8.3, 33.8.4, 33.8.5), to reasonably robust libraries and state-of-the-art systems (Sect. 33.8.1). We also apply our software on real world problems. For instance, we used the Datalog hammer presented in Sect. 33.8.3 to debug and verify a supervisor for lane change assistants and the results presented in Sect. 33.8.5 led to the first fully-automated *expected amortised cost analysis* of self-adjusting data structures.

### 33.8.1 The SPASS Workbench

*Investigators: Martin Bromberger, Mathias Fleury, Tobias Gehl, Hendrik Leidinger, Lorenz Leutgeb, Harish K. Murali, Simon Schwarz, and Christoph Weidenbach*

In recent years, we have shifted our focus from one monolithic first-order theorem prover to a workbench of tools and libraries, where the classic SPASS superposition prover is simply one component. The development of the SPASS workbench is bottom-up and lower level tools are often reused as part of higher level solvers.

On the lower level, we have implemented a very efficient CNF transformation that supports non-standard logical operators, such as if-then-else, and includes the small clause normal form approach [3, 4, 5]. It is therefore the first CNF implementation that both supports non-standard logical operators and advanced renaming techniques. The second low level tool that we have developed is the SAT solver SPASS-SAT. It is based on the CDCL paradigm [1], but with a rigorous application of certain redundancy rules based on results from [6]. Although SPASS-SAT is not competitive on today's SAT benchmarks, it behaves very well in the CDCL(T) framework that we use for instance in the CDCL(LA) solver SPASS-SATT.

On the higher level, we have our classical first-order theorem prover SPASS and our CDCL(LA) solver SPASS-SATT [2]. Although we have stopped the development of SPASS, it continues to be used by several research groups in the world. Its binaries are still available for download[1] and there even is a web interface for online use[2].

We also started developing a new prototype: SPASS-SPL (Sect. 33.8.2), a theorem prover for the verification of supervisors (e.g. engine control units) that combines arithmetic and classical first-order reasoning.

### References

[1] A. Biere, M. Heule, H. van Maaren, and T. Walsh, eds. *Handbook of Satisfiability*, Frontiers in Artificial Intelligence and Applications 185. IOS Press, 2009.

[2] M. Bromberger, M. Fleury, S. Schwarz, and C. Weidenbach. SPASS-SATT: A CDCL(LA) solver. In P. Fontaine, ed., *Automated Deduction – CADE 27*, Natal, Brazil, 2019, LNAI 11716, pp. 111–122. Springer.

[3] A. Nonnengart and C. Weidenbach. Computing small clause normal forms. In A. Robinson and A. Voronkov, eds., *Handbook of Automated Reasoning*, vol. 1, pp. 335–367. Elsevier, Amsterdam, 2001.

[4] P. Schmelzeisen. Propositional CNF-Transformation Revisited. Bachelor's thesis, Saarland University, 2016.

[5] S. Schwarz. Propositional CNF-Transformation Again Revisited. Bachelor's thesis, Saarland University, 2019.

[6] C. Weidenbach. Automated reasoning building blocks. In R. Meyer, A. Platzer, and H. Wehrheim, eds., *Correct System Design*, Oldenburg, Germany, 2015, LNCS 9360, pp. 172–188. Springer.

### 33.8.2 SPASS-SPL: a SUPERLOG Solver

*Investigators: Martin Bromberger, Tobias Gehl, Hendrik Leidinger, Lorenz Leutgeb, Harish K. Murali, Simon Schwarz, and Christoph Weidenbach*

We are working on ways to model, execute and verify *supervisors* [6], which are components in technical systems that control system functionality. An example for a supervisor is the electronic control unit (ECU) of a combustion engine. In this context we have been looking for a decision procedure of a logical fragment that has sufficient expressiveness to model supervisor functionality and properties; at the same time the fragment should run efficiently,

[1] https://webspass.spass-prover.org/

[2] http://www.spass-prover.org/classic-spass-theorem-prover/

i.e., generating consequences out of ground facts (inputs), and, finally, it should be push button verifiable (decidable).

The theorem prover SPASS-SPL is our current prototype for SupERLog reasoning. We have implemented in SPASS-SPL three different approaches that solve SupERLog formulas. The first approach is an extension of unit-resulting resolution [8] for first-order reasoning that uses SPASS-SATT [3] for theory reasoning. The amount of newly learned clauses is kept manageable by eliminating redundant clauses through tautology elimination and forward subsumption. Thanks to our sample point heuristic (Sect. 33.6.4) SPASS-SPL can simplify the majority of subsumption checks because it quickly and correctly determines that they will fail [4]. This approach works for Bernays–Schönfinkel clauses modulo linear arithmetic constraints (BS(LA)) and is refutationally complete for the Horn Bernays-Schönfinkel fragment modulo linear arithmetic (HBS(LA)).

Our second approach, is called the Datalog hammer (Sect. 33.8.3) and it focuses on a decidable fragment of BS(LA), the Horn Bernays-Schönfinkel fragment over simple linear arithmetic (HBS(SLA)). The approach combines ideas from quantifier elimination and instantiation for arithmetic with the reasoning strength of Datalog rule engines for first-order horn clauses [10].

The third approach is based on our various SCL calculi [5, 2] and its implementation is at the moment still under development. SCL combines techniques from SAT (propositional satisfiability) and SMT (satisfiability modulo theories) solving [9, 1, 7] and first-order theorem proving. In Sect. 33.8.4, we describe the lifting of one such technique, the two-watched literals scheme, in detail. SCL(T) will allow SPASS-SPL to solve non-Horn Bernays–Schönfinkel clause sets modulo arithmetic constraints and it will grant it explicit model generation capabilities and non-redundancy guarantees [2]. The latter makes the expensive redundancy checks obsolete that our current version of SPASS-SPL has to perform.

## References

[1] R. J. Bayardo Jr. and R. Schrag. Using CSP look-back techniques to solve real-world SAT instances. In B. Kuipers and B. L. Webber, eds., *Proceedings of the Fourteenth National Conference on Artificial Intelligence and Ninth Innovative Applications of Artificial Intelligence Conference, AAAI 97, IAAI 97, July 27-31, 1997, Providence, Rhode Island, USA.*, 1997, pp. 203–208.

• [2] M. Bromberger, A. Fiori, and C. Weidenbach. Deciding the Bernays-Schoenfinkel fragment over bounded difference constraints by simple clause learning over theories. In F. Henglein, S. Shoham, and V. Yakir, eds., *Verification, Model Checking, and Abstract Interpretation (VMCAI 2021)*, Copenhagen, Denmark (Online), 2021, LNCS 12597, pp. 511–533. Springer.

[3] M. Bromberger, M. Fleury, S. Schwarz, and C. Weidenbach. SPASS-SATT: A CDCL(LA) solver. In P. Fontaine, ed., *Automated Deduction – CADE 27*, Natal, Brazil, 2019, LNAI 11716, pp. 111–122. Springer.

• [4] M. Bromberger, L. Leutgeb, and C. Weidenbach. An efficient subsumption test pipeline for BS(LRA) clauses. In J. Blanchette, L. Kovács, and D. Pattinson, eds., *Automated Reasoning (IJCAR 2022)*, Haifa, Israel, 2022, LNAI 13385, pp. 147–168. Springer.

• [5] M. Bromberger, S. Schwarz, and C. Weidenbach. *SCL(FOL) Revisited*, 2023. arXiv: 2302.05954.

[6] R. Faqeh, C. Fetzer, H. Hermanns, J. Hoffmann, M. Klauck, M. A. Köhl, M. Steinmetz, and C. Weidenbach. Towards dynamic dependable systems through evidence-based continuous certification. In T. Margaria and B. Steffen, eds., *Leveraging Applications of Formal Methods, Verification and Validation: Engineering Principles*, Rhodes, Greece (Virtual Event), 2020, LNCS 12477, pp. 416–439. Springer.

[7] R. Nieuwenhuis, A. Oliveras, and C. Tinelli. Solving SAT and SAT modulo theories: From an abstract Davis–Putnam–Logemann–Loveland procedure to DPLL(T). *Journal of the ACM (JACM)*, 53(6):937–977, 2006.

[8] R. Overbeek, J. McCharen, and L. Wos. Complexity and related enhancements for automated theorem-proving programs. *Computers & Mathematics with Applications*, 2(1):1–16, 1976.

[9] J. P. M. Silva and K. A. Sakallah. GRASP – a new search algorithm for satisfiability. In *International Conference on Computer Aided Design, ICCAD*, 1996, pp. 220–227. IEEE Computer Society Press.

[10] J. Urbani, C. J. H. Jacobs, and M. Krötzsch. Column-oriented Datalog materialization for large knowledge graphs. In D. Schuurmans and M. P. Wellman, eds., *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, February 12-17, 2016, Phoenix, Arizona, USA*, 2016, pp. 258–264. AAAI Press.


### 33.8.3 A Datalog Hammer for Supervisor Verification Conditions Modulo Simple Linear Arithmetic

*Investigators: Martin Bromberger, Harish K. Murali, and Christoph Weidenbach, in cooperation with Irina Dragoste, Rasha Faqeh, Christof Fetzer, Larry González, Markus Krötzsch, and Maximilian Marx (Techn. Univ. Dresden)*

As already elaborated in Sect. 33.8.2, we are working on ways to model, execute, and verify *supervisors*. The logics we use to model supervisors and their properties are called SupERLogs—(Sup)ervisor (E)ffective(R)easoning (Log)ics—and are instances of function-free first-order logic extended with integer arithmetic [6].

One such SupERLog that is sufficient to model an (ECU) and many other supervisors is the Horn Bernays-Schönfinkel fragment over simple linear arithmetic (HBS(SLA)). HBS clause sets are also called Datalog programs and HBS(SLA) can be seen as an extension of Datalog programs that allows simple arithmetic inequalities (i.e. constant variable bounds $x \leq c$) in the bodies of rules. The fragment is also equivalent to constrained Horn clauses (CHC), where the theory constraints are restricted to simple linear arithmetic (SLA).

While supervisor safety conditions formalized as existentially quantified properties can often already be automatically verified [1], conjectures about invariants formalized as universally quantified properties are a further challenge. We have shown that supervisor safety conditions and invariants can be automatically proven by a Datalog hammer [4, 3]. Analogous to the Sledgehammer project [2] of Isabelle [8] translating higher-order logic conjectures to first-order logic (modulo theories) conjectures, our Datalog hammer translates HBS(SLA) conjectures into pure Datalog programs, i.e. HBS clause sets over a finite set of first-order constants. The translation from HBS(SLA) to HBS preserves validity and satisfiability and it is inspired by the test point method for quantifier elimination in arithmetic [7]. It works because a finite number of test points is sufficient to decide whether a universal/existential conjecture is a consequence of an HBS(SLA) clause set.

We have implemented the Datalog hammer as an extension to the SPASS-SPL system (option -d). The output of the Datalog hammer is then solved by an integration of VLog into SPASS-SPL via the VLog API [5]. With the resulting combination, we were able to verify supervisor code for two examples: a lane change assistant in a car and an electronic control unit of a supercharged combustion engine. Our experiments show, that our performance on existential conjectures is at the same level as SMT (satisfiability modulo theories) and CHC solvers and that SPASS-SPL is the only solver that can prove and disprove universal conjectures. Moreover, we actually used SPASS-SPL to debug the prototype supervisor for lane change assistants during its development. The lane change examples in [3] are based on versions generated during this debugging process where SPASS-SPL found the following bugs: (i) it did not always return a result, (ii) it declared actions as both safe and unsafe at the same time, and (iii) it declared actions as safe although they would lead to collisions. The supervisor is now fully verified.

### Sorted Datalog Hammer:

In a recent paper [3], we have presented several new improvements to our Datalog hammer: we have generalized it to mixed real-integer arithmetic and finite first-order sorts; we extended the class of acceptable inequalities beyond variable bounds and positively grounded inequalities; and we significantly reduced the size of the hammer output by a soft typing discipline. We call the result the sorted Datalog hammer. It not only allows us to handle more complex supervisor code and to model already considered supervisor code more concisely, but it also improves our performance on real world benchmark examples.

## References


[1] N. Bjørner, A. Gurfinkel, K. L. McMillan, and A. Rybalchenko. Horn clause solvers for program verification. In L. D. Beklemishev, A. Blass, N. Dershowitz, B. Finkbeiner, and W. Schulte, eds., *Fields of Logic and Computation II – Essays Dedicated to Yuri Gurevich on the Occasion of His 75th Birthday*, 2015, LNCS 9300, pp. 24–51. Springer.

[2] S. Böhme and T. Nipkow. Sledgehammer: Judgement Day. In J. Giesl and R. Hähnle, eds., *IJCAR 2010*, 2010, LNCS 6173, pp. 107–121. Springer.

• [3] M. Bromberger, I. Dragoste, R. Faqeh, C. Fetzer, L. González, M. Krötzsch, M. Marx, H. K. Murali, and C. Weidenbach. A sorted Datalog hammer for supervisor verification conditions modulo simple linear arithmetic. In D. Fisman and G. Rosu, eds., *Tools and Algorithms for the Construction and Analysis of Systems (TACAS 2022)*, Munich, Germany, 2022, LNCS 13243, pp. 480–501. Springer.

• [4] M. Bromberger, I. Dragoste, R. Faqeh, C. Fetzer, M. Krötzsch, and C. Weidenbach. A Datalog hammer for supervisor verification conditions modulo simple linear arithmetic. In B. Konev and G. Reger, eds., *Frontiers of Combining Systems (FroCoS 2021)*, Birmingham, UK, 2021, LNAI 12941, pp. 3–24. Springer.

[5] D. Carral, I. Dragoste, L. González, C. Jacobs, M. Krötzsch, and J. Urbani. VLog: A rule engine for knowledge graphs. In C. Ghidini et al., ed., *Proc. 18th Int. Semantic Web Conf. (ISWC'19, Part II)*, 2019, LNCS 11779, pp. 19–35. Springer.

[6] R. Faqeh, C. Fetzer, H. Hermanns, J. Hoffmann, M. Klauck, M. A. Köhl, M. Steinmetz, and C. Weidenbach. Towards dynamic dependable systems through evidence-based continuous

certification. In T. Margaria and B. Steffen, eds., *Leveraging Applications of Formal Methods, Verification and Validation: Engineering Principles*, Rhodes, Greece (Virtual Event), 2020, LNCS 12477, pp. 416–439. Springer.

[7] R. Loos and V. Weispfenning. Applying linear quantifier elimination. *The Computer Journal*, 36(5):450–462, 1993.

[8] T. Nipkow, L. C. Paulson, and M. Wenzel. *Isabelle/HOL: A Proof Assistant for Higher-Order Logic*, LNCS 2283. Springer, 2002.

### 33.8.4 A Two-Watched Literal Scheme for First-Order Logic

*Investigators: Martin Bromberger, Tobias Gehl, Lorenz Leutgeb, and Christoph Weidenbach*

We are currently working on a prototype implementation of our SCL(FOL) calculus [2]. SCL(FOL) is heavily inspired by CDCL SAT solvers. One indispensable part of any CDCL SAT solver implementation is the two-watched literal scheme [4, 3]. It is used by CDCL SAT solvers to efficiently detect which clauses can propagate or are falsified with respect to an explicit partial model assumption called trail. Propagation here means the clause is false except for one undefined literal.

With respect to propagation and falsification, there is no need to do any update on the status of a clause, as long as it has two undefined literals or it is true in the current model assumption. This leads to the two-watched literal scheme. In any clause two literals are watched and the clause is indexed in a so-called watched list for each of its watched literals. If CDCL adds a new literal $L$ to the trail, then it is sufficient to check those clauses for propagations and conflicts where one of the watched literals is the complement of $L$. This means the two-watched literal scheme greatly reduces the number of clauses we need to check for propagations and conflicts.

As part of our prototype implementation, we have lifted the propositional two-watched literal scheme to first-order logic without equality [1]. Our trail consists of ground first-order literals and clauses are full first-order clauses containing implicitly universally quantified variables. Again we want to detect propagating and false clauses by the two-watched literal scheme by only considering the watched literals for trail extensions. For first-order logic, the two-watched literals scheme gets more sophisticated because of variable instantiation. The first extension concerns universally quantified variables. Variable instantiation may result in merging literals and different clauses may produce identical instances by variable instantiation. For example, the first-order clause $R(x, y) \quad R(a, z) \quad R(u, b)$ where $a$, $b$ are constants and $x$, $y$, $z$, $u$ are variables represents already the propagating ground instance $R(a, b)$. Furthermore, updates on the trail induce further instances of clauses. The second extension removes exhaustive propagation as a requirement of the two-watched literals scheme. Even in a first-order setting without non-constant function symbols, the trail may grow exponentially with respect to the maximal arity of a predicate symbol. Thus exhaustive propagation cannot be afforded in first-order logic, in general. On the other hand any CDCL style calculus typically breaks if a decided (guessed) literal immediately results in a false clause, a conflict. Our solution to this is a one step propagation look-ahead and by separating the trail from a set of potentially propagating literals. Despite these extensions, we managed to keep the main properties of the scheme: first, only watched literals need to be considered for trail changes, and, second, after backtracking no updates are needed except for the newly learned clause.

### References


• [1] M. Bromberger, T. Gehl, L. Leutgeb, and C. Weidenbach. A two-watched literal scheme for first-order logic. In B. Konev, C. Schon, and A. Steen, eds., *Practical Aspects of Automated Reasoning 2022 (PAAR 2022)*, Haifa, Israel, 2022, CEUR Workshop Proceedings 3201. CEUR-WS.org.

• [2] M. Bromberger, S. Schwarz, and C. Weidenbach. *SCL(FOL) Revisited*, 2023. arXiv: 2302.05954.

[3] M. W. Moskewicz, C. F. Madigan, Y. Zhao, L. Zhang, and S. Malik. Chaff: Engineering an efficient SAT solver. In *Proceedings of the 38th annual Design Automation Conference*, 2001, pp. 530–535.

[4] J. P. M. Silva and K. A. Sakallah. GRASP – a new search algorithm for satisfiability. In *International Conference on Computer Aided Design, ICCAD*, 1996, pp. 220–227. IEEE Computer Society Press.


## 33.8.5 Automated Expected Amortised Cost Analysis of Probabilistic Data Structures

*Investigators: Lorenz Leutgeb, in cooperation with Georg Moser (Universität Innsbruck), and Florian Zuleger (Technische Universität Wien)*

Sleator and Tarjan introduced the notion of *amortised* complexity in order to better argue about the performance of self-adjusting data structures such as splay trees [6]. Analysing these data structures requires sophisticated potential functions, which typically contain logarithmic expressions. Possibly for these reasons, and despite the recent progress in automated resource analysis, they have so far eluded automation.

In [5] we present the first fully-automated *expected amortised cost analysis* of self-adjusting data structures, that is, of *randomised splay trees* [1], *randomised splay heaps*, and *randomised meldable heaps* [2], which so far have only (semi-)manually been analyzed in the literature. It extends our previous work [4, 3] on the fully-automated *amortised cost analysis* of deterministic self-adjusting data structures, such as *splay heaps*, *splay trees*, and *pairing heaps*.

Our analysis follows the physicist's method of amortized analysis via potentials and is stated as a type-and-effect system for a first-order functional programming language. It makes use of logarithmic potential functions and is the first such system to exhibit *logarithmic amortised complexity*. It supports sampling over discrete distributions, non-deterministic choice and a ticking operator. The latter allows for the specification of fine-grained cost models.

We state two soundness theorems based on two different—but strongly related—typing rules of ticking, which account differently for the cost of non-terminating computations. Finally we provide a prototype implementation able to fully automatically analyze the aforementioned case studies. Since the type system heavily depends on constraints in linear real arithmetic, its implementation is based on satisfiability modulo theory (SMT) solving.

## References


[1] S. Albers and M. Karpinski. Randomized splay trees: Theoretical and experimental results. *IPL*, 81(4):213–221, 2002.

[2] A. Gambin and A. Malinowski. Randomized meldable priority queues. In *SOFSEM '98: Theory and Practice of Informatics, 25th Conference on Current Trends in Theory and Practice of Informatics, Jasná, Slovakia, November 21-27, 1998, Proceedings*, 1998, LNCS 1521, pp. 344–349. Springer.

[3] M. Hofmann, L. Leutgeb, D. Obwaller, G. Moser, and F. Zuleger. Type-based analysis of logarithmic amortised complexity. *MSCS*, 32(6):794–826, 2022.

[4] L. Leutgeb, G. Moser, and F. Zuleger. ATLAS: automated amortised complexity analysis of self-adjusting data structures. In *CAV 2021*, 2021, LNCS 12760, pp. 99–122. Springer.

• [5] L. Leutgeb, G. Moser, and F. Zuleger. Automated expected amortised cost analysis of probabilistic data structures. In S. Shoham and Y. Vizel, eds., *Computer Aided Verification (CAV 2022)*, Haifa, Israel, 2022, LNCS 13372, pp. 70–91. Springer.

[6] D. D. Sleator and R. E. Tarjan. Self-adjusting binary search trees. *JACM*, 32(3):652–686, 1985.


# 33.9 Academic Activities

## 33.9.1 Journal Positions

*Christoph Weidenbach:*

– *Journal of Automated Reasoning* (Editor)

*Jasmin Blanchette:*

– *Journal of Automated Reasoning* (Editor-in-chief)
– Special issue on *IJCAR 2022*, to be published (Guest co-editor)

*Thomas Sturm:*

– *Journal of Symbolic Computation* (Editor)
– *Mathematics in Computer Science* (Editor)
– Special issue on *CASC 2019* (Guest co-editor)
– Special issue on *CASC 2020* (Guest co-editor)
– Special issue on *CASC 2021* (Guest co-editor)

## 33.9.2 Conference and Workshop Positions

### Membership in program and organization committees

*Christoph Weidenbach:*

– *FroCoS 2021, 13th International Symposium on Frontiers of Combining Systems*, Birmingham, UK, September 2021 (PC member),

- *IJCAR 2022, 11th International Joint Conference on Automated Reasoning*, Haifa, Israel, August 2022 (PC member),
- *LPAR 2023, 24th International Conference on Logic for Programming, Artificial Intelligence and Reasoning*, Manizales, Colombia, June 2023 (PC member),
- *Summer School 2021: Verification Technology, Systems & Applications*, Liège, Belgium, October 2021 (Co-organizer),
- *Summer School 2022: Verification Technology, Systems & Applications*, Saarbrücken, Germany, September 2022 (Co-organizer).

*Jasmin Blanchette:*

- *NFM 2021, NASA Formal Methods: 13th International Symposium*, Virtual Event, May 2021 (PC member),
- *ITP 2021, Interactive Theorem Proving*, Virtual Event, June/July 2021 (PC member),
- *CADE-28, 28th International Conference on Automated Deduction*, Virtual Event, July 2021 (PC member),
- *CAV 2021, 33rd International Conference on Computer-Aided Verification*, Virtual Event, July 2021 (PC member),
- *FMM 2021, Fifth Workshop on Formal Mathematics for Mathematicians*, Virtual Event, July 2021 (PC chair and co-organizer),
- *AITP 2021, 6th Conference on Artificial Intelligence and Theorem Proving*, Aussois, France, September 2021 (PC member),
- *FroCoS 2021, 13th International Symposium on Frontiers of Combining Systems*, Birmingham, UK, September 2021 (PC member),
- *TACAS, 28th International Conference on Tools and Algorithms for the Construction and Analysis of Systems*, Munich, Germany, April 2022 (PC member),
- *IJCAR 2022, 11th International Joint Conference on Automated Reasoning*, Haifa, Israel, August 2022 (PC chair),
- *ITP 2022, Interactive Theorem Proving*, Haifa, Israel, August 2022 (PC member),
- *DT-2022, Deduktionstreffen*, Virtual Event, September 2022 (PC member),
- *AITP 2022, 7th Conference on Artificial Intelligence and Theorem Proving*, Aussois, France, September 2022 (PC member),
- *LPAR 2023, 24th International Conference on Logic for Programming, Artificial Intelligence and Reasoning*, Manizales, Colombia, June 2023 (PC member),
- *ITP 2023, Interactive Theorem Proving*, Białystok, Poland, July/August 2023 (PC member).

*Martin Bromberger:*

- *SMT 2022, 20th International Workshop on Satisfiability Modulo Theories*, Haifa, Israel, August 2022 (PC member),

- *LPAR 2023, 24th International Conference on Logic for Programming, Artificial Intelligence and Reasoning*, Manizales, Colombia, June 2023 (Workshop and tutorial chair),
- *SC$^2$ 2023, International Workshop on Satisfiability Checking and Symbolic Computation*, Tromsø, Norway, July 2023 (PC member),
- *SMT-COMP 2023, International Satisfiability Modulo Theories (SMT) Competition*, Rome, Italy, July 2023 (Co-organizer).

*Martin Desharnais:*

- *CASC-28: CADE ATP System Competition of the 28th International Conference on Automated Deduction*, Virtual Event, July 2021 (Co-organizer)
- *CASC-J11: CADE ATP System Competition part of the 11th International Joint Conference on Automated Reasoning*, Haifa, Israel, August 2022 (Co-organizer).

*Hamid Rahkooy:*

- *SC$^2$ 2022, 7th International Workshop on Satisfiability Checking and Symbolic Computation*, Haifa, Israel, August 2022 (PC member),
- *CASC 2022, 24th International Workshop on Computer Algebra in Scientific Computing*, Gebze, Turkey, August 2022 (PC member).

*Thomas Sturm:*

- *ISSAC 2021, 46th International Symposium on Symbolic and Algebraic Computation*, Saint Petersburg, Russia and virtually, July 2021 (PC member),
- *CASC 2021, 23rd Conference on Computer Algebra in Scientific Computing*, Sochi, Russia and virtually, September 2021 (PC member),
- *International interdisciplinary workshop on Symbolic Methods for Biological Networks*, Bonn, Germany, March 2022 (Co-organizer),
- *CASC 2022, 24th International Workshop on Computer Algebra in Scientific Computing*, Gebze, Turkey, August 2022 (PC member),
- *SC$^2$ 2023, 8th International Workshop on Satisfiability Checking and Symbolic Computation*, Tromsø, Norway, July 2023 (Co-chair),
- *CASC 2023, 25th International Workshop on Computer Algebra in Scientific Computing*, Havana, Cuba, August/September 2023 (PC member).

*Sophie Tourret:*

- *CADE-28, 28th International Conference on Automated Deduction*, Virtual Event, July 2021 (Publicity chair and PC member),
- *IJCAI-21, 30th International Joint Conference on Artificial Intelligence*, Virtual Event, August 2021 (PC member),
- *CPP 2022, Certified Programs and Proofs*, Philadelphia, PA, USA, January 2022 (PC member),

- *IJCAI-ECAI 2022, 31st International Joint Conference on Artificial Intelligence*, Vienna, Austria, July 2022 (PC member),
- *IJCAR 2022, 11th International Joint Conference on Automated Reasoning*, Haifa, Israel, August 2022 (Workshop co-chair and PC member),
- *SMT 2022, 20th International Workshop on Satisfiability Modulo Theories*, Haifa, Israel, August 2022 (PC member),
- *PAAR 2022, 8th Workshop on Practical Aspects of Automated Reasoning*, Haifa, Isreal, August 2022 (PC member),
- *CADE-29, 29th International Conference on Automated Deduction*, Rome, Italy, July 2023 (PC member),
- *IJCAI 2023, 32nd International Joint Conference on Artificial Intelligence*, Macao, China, August 2023 (PC member).

*Uwe Waldmann:*

- *CADE-28, 28th International Conference on Automated Deduction*, Virtual Event, July 2021 (PC member),
- *IJCAR 2022, 11th International Joint Conference on Automated Reasoning*, Haifa, Israel, August 2022 (PC member),
- *DT-2022, Deduktionstreffen*, Virtual Event, September 2022 (PC member),
- *CADE-29, 29th International Conference on Automated Deduction*, Rome, Italy, July 2023 (PC member).

### Membership in steering and other committees

*Christoph Weidenbach:*

- Member of the Steering Committee for the French-German Computer Science Cooperation Agreement between INRIA, CNRS, University of Metz, University of Nancy 1, University of Nancy 2, Institut National Polytechnique de Lorraine at Nancy, Saarland University, University of Kaiserslautern, Fraunhofer Institute for Experimental Software Engineering (IESE) Kaiserslautern, Max Planck Institute for Informatics, Max Planck Institute for Software Systems, DFKI.
- Member of the Selection Committee of the Saarbrücken Graduate School in Computer Science.
- Chairman of the Steering Committee "Bundeswettbewerb Informatik"
- Trustee and President of CADE, Conference on Automated Deduction (until 2022)
- Member of the Steering Committee of IJCAR (International Joint Conference on Automated Reasoning)

*Thomas Sturm:*

- Advisory role as a project partner in the UK EPSRC Project EP/T015748/1 & EP/T015713/1 "Pushing Back the Doubly-Exponential Wall of Cylindrical Algebraic Decomposition" (since 2021)

*Sophie Tourret:*

- Member of the Steering Committee of *CADE, International Conference on Automated Deduction* (elected 2022),
- Member of the Steering Committee of *PAAR, Workshop on Practical Aspects of Automated Reasoning*,
- Member of the Hiring Committee Inria Nancy – Grand Est 2022,
- AAR board member and editor of the AAR newsletter since 2017.

*Uwe Waldmann:*

- Member of the Selection Committee for the McCune Award 2019 & 2020.

### 33.9.3 Invited Talks and Tutorials

*Christoph Weidenbach:*

- *Algorithm Design for Solving Hard Problems*, Colloquium, University of New Mexico (Online), December 7, 2022.

*Martin Bromberger:*

- *A Datalog hammer for supervisor verification conditions modulo simple linear arithmetic*, Research Seminar Logic and AI, TU Dresden, March 31, 2022.

*Martin Desharnais:*

- *Seventeen Provers under the Hammer*, Vampire Seminar, Vienna (Online), April 28, 2022.

*Thomas Sturm:*

- *Real Quantifier Elimination by Virtual Substitution*, Invited talk, Workshop "Trends in Arithmetic Theories" at the 49th EATCS International Colloquium on Automata, Languages and Programming (ICALP). Paris, July 4, 2022.

*Sophie Tourret:*

- *In the reviewer's eye*, ETAPS 2023 mentoring workshop, Paris, April 23, 2023.

## 33.10 Teaching Activities

### Winter Semester 2021/2022

- Automated Reasoning (U. Waldmann)

### Summer Semester 2022

- Competitive Programming (M. Bläser, K. Bringmann, M. Bromberger, C. Weidenbach)
- Automated Reasoning II (U. Waldmann)

#### Winter Semester 2022/2023

- Automated Reasoning (C. Weidenbach, M. Bromberger, S. Möhle, S. Schwarz

#### Master and Bachelor Theses

- Christian Baldus: A fast algorithm for generalized propositional CNF transformation, Bachelor's thesis, 2021 (Supervisor: C. Weidenbach)
- Simon Schwarz: Exploring Partial Models with SCL, Master thesis, 2022 (Supervisor: M. Bromberger)

## 33.11 Dissertations, Awards

### 33.11.1 Dissertations

- Fajar Haifani: *On a Notion of Abduction and Relevance for First-Order Logic Clause Sets*, March 9, 2023.

### 33.11.2 Awards

- Fajar Haifani, Patrick Koopmann, Sophie Tourret, Christoph Weidenbach: *Best Student Paper Award* at 35th International Workshop on Description Logics for *Connection-minimal Abduction in EL via Translation to FOL*.
- Jasmin Blanchette: *Dutch Prize for ICT Research 2022*.
- Sophie Tourret: *Distinguished PC Member* at IJCAI-ECAI 2022 (quality of reviews ranked in the top 3%).

## 33.12 Grants and Cooperations

#### Transregional Collaborative Research Centre 248 "Foundations of Perspicuous Software Systems"

The Transregional Collaborative Research Centre 248 "Foundations of Perspicuous Software Systems" aims at enabling comprehension in a cyber-physical world with the human in the loop.

- Starting date: January 2019.
- Duration: 4 years.
- Funding: DFG Transregional Collaborative Research Centre.
- MPI-INF investigators: Christoph Weidenbach (head), Martin Bromberger, Fajar Haifani.
- Partners: TU Dresden, Saarland University, Max Planck Institute for Software Systems.

### VeriDis: Modeling and Verification of Distributed Algorithms and Systems

The VeriDis project aims to exploit and further develop the advances and integration of interactive and automated theorem proving applied to the area of concurrent and distributed systems. The goal of the project is to assist algorithm and system designers to carry out formally proved developments, where proofs of relevant properties as well as bugs can be found fully automatically.

- Starting date: September 2011.
- Duration: 12 years.
- Funding: On separate request, MPI-INF and Inria Nancy, additional sponsoring by Inria Nancy.
- MPI-INF investigators: Christoph Weidenbach (head), Martin Bromberger, Martin Desharnais, Fajar Haifani, Hendrik Leidinger, Lorenz Leutgeb, Sibylle Möhle, Hamid Rahkooy, Simon Schwarz, Uwe Waldmann
- Partners: Stephan Merz (Inria Nancy, France).

### SYMBIONT – Symbolic Methods for Biological Networks

SYMBIONT is an interdisciplinary project ranging from mathematics via computer science to systems biology and systems medicine. The project has a clear focus on fundamental research on mathematical methods, and prototypes in software, which are in turn benchmarked against models from computational biology databases.

The principal approach of SYMBIONT is to combine symbolic methods with model reduction methods for the analysis of biological networks. In order to cope more effectively with the parameter uncertainty problem we impose an entirely new paradigm replacing thinking about single instances with thinking about orders of magnitude.

Our computational methods are diverse and involve various branches of mathematics such as tropical geometry, real algebraic geometry, theories of singular perturbations, invariant manifolds and symmetries of differential systems.

- Starting date: July 2018.
- Duration: 46 months.
- Funding: Bilateral ANR, France (ANR-17-CE40-0036) and DFG, Germany (391322026).
- MPI-INF investigators: Thomas Sturm (site coordinator), Hamid Rahkooy.
- Partners: University of Lille 1, CNRS, Inria & University of Lorraine, University of Montpellier, Inria Saclay, RWTH Aachen, University of Bonn, University of Kassel.

### Matryoshka

The goal of the Matryoshka project is to make interactive verification more cost-effective by increasing the level of automation of proof assistants. We will enrich first-order calculi (superposition and SMT) with higher-order reasoning in a careful manner, to preserve their desirable properties. With higher-order superposition and higher-order SMT in place, we will develop highly automatic provers building on modern superposition provers and SMT solvers, following a novel stratified architecture. These new provers will then be integrated in proof assistants and will be available as backends to more specialized verification tools.

- Starting date: March 2017.
- Duration: 5 years.
- Funding: European Research Council (ERC) Starting Grant 2016, No. 713999.
- Principal investigator: Jasmin Blanchette; Senior collaborators at MPI-INF: Sophie Tourret, Uwe Waldmann.
- Partners: P. Fontaine (Nancy, France), S. Schulz (Stuttgart, Germany).

## 33.13 Publications

### Books, proceedings, special issues

[1] M. England, W. Koepf, T. Sadykov, W. M. Seiler, and T. Sturm, eds. *Special Issue: Selected Papers of Computer Algebra in Scientific Computing (CASC 2019)*, *Mathematics in Computer Science*, 15(2), 2021. Springer.

### Journal articles and book chapters

[1] I. Akbarbaglu, H. P. Aghababa, and H. Rahkooy. On the algebraic structures in $\mathrm{A}_\Phi(G)$. *Mediterranian Journal of Mathematics*, 19(3), Article 131, 2022.

[2] A. Bentkamp, J. Blanchette, S. Cruanes, and U. Waldmann. Superposition for lambda-free higher-order logic. *Logical Methods in Computer Science*, 17(2), Article 1, 2021.

[3] A. Bentkamp, J. Blanchette, V. Nummelin, S. Tourret, P. Vukmirović, and U. Waldmann. Mechanical mathematicians. *Communications of the ACM*, 66(4):80–90, 2023.

[4] A. Bentkamp, J. Blanchette, S. Tourret, and P. Vukmirović. Superposition for higher-order logic. *Journal of Automated Reasoning*, 67(1), Article 10, 2023.

[5] A. Bentkamp, J. Blanchette, S. Tourret, P. Vukmirović, and U. Waldmann. Superposition with lambdas. *Journal of Automated Reasoning*, 65:893–940, 2021.

[6] J. Blanchette, Q. Qiu, and S. Tourret. Given clause loops. *Archive of Formal Proofs*, 2023.

[7] F. Frohn and C. Fuhs. A calculus for modular loop acceleration and non-termination proofs. *International Journal on Software Tools for Technology Transfer*, 24:691–715, 2022.

[8] D. Grigoriev, A. Iosif, H. Rahkooy, T. Sturm, and A. Weber. Efficiently and effectively recognizing toricity of steady state varieties. *Mathematics in Computer Science*, 15:199–232, 2021.

[9] N. Kruff, C. Lüders, O. Radulescu, T. Sturm, and S. Walcher. Algorithmic reduction of biological networks with multiple time scales. *Mathematics in Computer Science*, 15:499–534, 2021.

[10] C. Lüders, T. Sturm, and O. Radulescu. ODEbase: A repository of ODE systems for systems biology. *Bioinformatics Advances*, 2(2), Article vbac027, 2022.

[11] W. M. Seiler, M. Seiß, and T. Sturm. A logic based approach to finding real singularities of implicit ordinary differential equations. *Mathematics in Computer Science*, 15:333–352, 2021.

[12] G. Sutcliffe and M. Desharnais. The CADE-28 automated theorem proving system competition – CASC-28. *AI Communications*, 34(4):259–276, 2021.

[13] S. Tourret and C. Weidenbach. A posthumous contribution by Larry Wos: Excerpts from an unpublished column. *Journal of Automated Reasoning*, 66:575–584, 2022.

[14] M. Voigt. Decidable $^*$ $^*$ first-order fragments of linear rational arithmetic with uninterpreted predicates. *Journal of Automated Reasoning*, 65:357–423, 2021.

[15] P. Vukmirović, A. Bentkamp, J. Blanchette, S. Cruanes, V. Nummelin, and S. Tourret. Making higher-order superposition work. *Journal of Automated Reasoning*, 66:541–564, 2022.

[16] P. Vukmirović, J. Blanchette, S. Cruanes, and S. Schulz. Extending a brainiac prover to lambda-free higher-order logic. *International Journal on Software Tools for Technology Transfer (STTT)*, 24:67–87, 2022.

[17] U. Waldmann, S. Tourret, S. Robillard, and J. Blanchette. A comprehensive framework for saturation theorem proving. *Journal of Automated Reasoning*, 66:499–539, 2022.


## Conference articles


[1] A. Bentkamp, J. Blanchette, S. Tourret, and P. Vukmirović. Superposition for full higher-order logic. In A. Platzer and G. Sutcliffe, eds., *Automated Deduction – CADE 28*, Virtual Event, 2021, LNAI 12699, pp. 396–412. Springer.

[2] J. Blanchette. Lambda-superposition: From theory to trophy. In B. Pientka and C. Tinelli, eds., *CADE-29: 29th international Conference on Automated Deduction*, 2023. Accepted for publication.

[3] J. Blanchette, Q. Qiu, and S. Tourret. Verified given clause procedures. In B. Pientka and C. Tinelli, eds., *CADE-29: 29th international Conference on Automated Deduction*, 2023. Accepted for publication.

[4] M. Bromberger, M. Desharnais, and C. Weidenbach. An Isabelle/HOL formalization of the SCL(FOL) calculus. In B. Pientka and C. Tinelli, eds., *CADE-29: 29th international Conference on Automated Deduction*, 2023. Accepted for publication.

[5] M. Bromberger, I. Dragoste, R. Faqeh, C. Fetzer, L. González, M. Krötzsch, M. Marx, H. K. Murali, and C. Weidenbach. A sorted Datalog hammer for supervisor verification conditions modulo simple linear arithmetic. In D. Fisman and G. Rosu, eds., *Tools and Algorithms for the Construction and Analysis of Systems (TACAS 2022)*, Munich, Germany, 2022, LNCS 13243, pp. 480–501. Springer.

[6] M. Bromberger, I. Dragoste, R. Faqeh, C. Fetzer, M. Krötzsch, and C. Weidenbach. A Datalog hammer for supervisor verification conditions modulo simple linear arithmetic. In B. Konev and G. Reger, eds., *Frontiers of Combining Systems (FroCoS 2021)*, Birmingham, UK, 2021, LNAI 12941, pp. 3–24. Springer.

[7] M. Bromberger, A. Fiori, and C. Weidenbach. Deciding the Bernays-Schoenfinkel fragment over bounded difference constraints by simple clause learning over theories. In F. Henglein, S. Shoham, and V. Yakir, eds., *Verification, Model Checking, and Abstract Interpretation (VMCAI 2021)*, Copenhagen, Denmark (Online), 2021, LNCS 12597, pp. 511–533. Springer.

[8] M. Bromberger, T. Gehl, L. Leutgeb, and C. Weidenbach. A two-watched literal scheme for first-order logic. In B. Konev, C. Schon, and A. Steen, eds., *Practical Aspects of Automated Reasoning 2022 (PAAR 2022)*, Haifa, Israel, 2022, CEUR Workshop Proceedings 3201. CEUR-WS.org.

[9] M. Bromberger, C. Jain, and C. Weidenbach. SCL(FOL) can simulate non-redundant superposition clause learning. In B. Pientka and C. Tinelli, eds., *CADE-29: 29th international Conference on Automated Deduction*, 2023. Accepted for publication.

[10] M. Bromberger, L. Leutgeb, and C. Weidenbach. An efficient subsumption test pipeline for BS(LRA) clauses. In J. Blanchette, L. Kovács, and D. Pattinson, eds., *Automated Reasoning (IJCAR 2022)*, Haifa, Israel, 2022, LNAI 13385, pp. 147–168. Springer.

[11] M. Bromberger, S. Schwarz, and C. Weidenbach. Exploring partial models with SCL. In B. Konev, C. Schon, and A. Steen, eds., *Practical Aspects of Automated Reasoning 2022 (PAAR 2022)*, Haifa, Israel, 2022, CEUR Workshop Proceedings 3201. CEUR-WS.org.

[12] M. Bromberger, S. Schwarz, and C. Weidenbach. Exploring partial models with SCL. In R. Piskac and A. Voronkov, eds., *LPAR 2023: 24th International Conference on Logic for Programming, Artificial Intelligence and Reasoning*, 2023. Accepted for publication.

[13] A. Demin, H. Rahkooy, and T. Sturm. F5: A REDUCE package for signature-based Gröbner basis computation. In *CASC 2022 – Computer Algebra in Scientific Computing*, Gebze, Turkey, 2022. HAL.

[14] M. Desharnais, P. Vukmirović, J. Blanchette, and M. Wenzel. Seventeen provers under the hammer. In J. Andronick and L. de Moura, eds., *13th International Conference on Interactive Theorem Proving (ITP 2022)*, Haifa, Israel, 2022, Leibniz International Proceedings in Informatics 237, Article 8. Schloss Dagstuhl.

[15] G. Ebner, J. Blanchette, and S. Tourret. A unifying splitting framework. In A. Platzer and G. Sutcliffe, eds., *Automated Deduction – CADE 28*, Virtual Event, 2021, LNAI 12699, pp. 344–360. Springer.

[16] F. Frohn, M. Hark, and J. Giesl. Termination of polynomial loops. In D. Pichardie and M. Sighireanu, eds., *Static Analysis (SAS 2020)*, Chicago, IL, USA (Online Event), 2020, LNCS 12389, pp. 89–112. Springer.

[17] F. Haifani, P. Koopmann, and S. Tourret. Abduction in EL via translation to FOL. In R. A. Schmidt, C. Wernhard, and Y. Zhao, eds., *Second-Order Quantifier Elimination and Related Topics (SOQE 2021)*, Online Event, 2021, CEUR Workshop Proceedings 3009, pp. 46–58. CEUR-WS.

[18] F. Haifani, P. Koopmann, S. Tourret, and C. Weidenbach. Connection-minimal abduction in El via translation to FOL (extended abstract). In O. Arieli, M. Homola, J. C. Jung, and M.-L. Mugnier, eds., *Description Logics 2022 (DL 2022)*, Haifa, Israel, 2022, CEUR Workshop Proceedings 3263. CEUR-WS.org.

[19] F. Haifani, P. Koopmann, S. Tourret, and C. Weidenbach. Connection-minimal abduction in $\mathcal{EL}$ via translation to FOL. In J. Blanchette, L. Kovács, and D. Pattinson, eds., *Automated Reasoning (IJCAR 2022)*, Haifa, Israel, 2022, LNAI 13385, pp. 188–207. Springer.

[20] F. Haifani, S. Tourret, and C. Weidenbach. Generalized completeness for SOS resolution and its application to a new notion of relevance. In A. Platzer and G. Sutcliffe, eds., *Automated Deduction – CADE 28*, Virtual Event, 2021, LNAI 12699, pp. 327–343. Springer.

[21] F. Haifani and C. Weidenbach. Semantic relevance. In J. Blanchette, L. Kovács, and D. Pattinson, eds., *Automated Reasoning (IJCAR 2022)*, Haifa, Israel, 2022, LNAI 13385, pp. 208–227. Springer.

[22] H. Leidinger and C. Weidenbach. SCL(EQ): SCL for first-order logic with equality. In J. Blanchette, L. Kovács, and D. Pattinson, eds., *Automated Reasoning (IJCAR 2022)*, Haifa, Israel, 2022, LNAI 13385, pp. 228–247. Springer.

[23] L. Leutgeb, G. Moser, and F. Zuleger. Automated expected amortised cost analysis of probabilistic data structures. In S. Shoham and Y. Vizel, eds., *Computer Aided Verification (CAV 2022)*, Haifa, Israel, 2022, LNCS 13372, pp. 70–91. Springer.

[24] V. Nummelin, A. Bentkamp, S. Tourret, and P. Vukmirović. Superposition with first-class Booleans and inprocessing clausification. In A. Platzer and G. Sutcliffe, eds., *Automated Deduction – CADE 28*, Virtual Event, 2021, LNAI 12699, pp. 378–395. Springer.

[25] H. Rahkooy and T. Sturm. Parametric toricity of steady state varieties of reaction networks. In F. Boulier, M. England, T. M. Sadykov, and E. V. Vorozhtsov, eds., *Computer Algebra in Scientific Computing (CASC 2021)*, Sochi, Russia, 2021, LNCS 12865, pp. 314–333. Springer.

[26] H. Rahkooy and T. Sturm. Testing binomiality of chemical reaction networks using comprehensive Gröbner systems. In F. Boulier, M. England, T. M. Sadykov, and E. V. Vorozhtsov, eds., *Computer Algebra in Scientific Computing (CASC 2021)*, Sochi, Russia, 2021, LNCS 12865, pp. 334–352. Springer.

[27] H. Rahkooy and C. Vargas Montero. A graph theoretical approach for testing binomiality of reversible chemical reaction networks. In *SYNASC 2020, 22nd International Symposium on Symbolic and Numeric Algorithms for Scientific Computing*, Timisoara, Romania, 2020, pp. 101–108. IEEE.

[28] H.-J. Schurr, M. Fleury, and M. Desharnais. Reliable reconstruction of fine-grained proofs in a proof assistant. In A. Platzer and G. Sutcliffe, eds., *Automated Deduction – CADE 28*, Virtual Event, 2021, LNAI 12699, pp. 450–467. Springer.

[29] S. Tourret and J. Blanchette. A modular Isabelle framework for verifying saturation provers. In C. Hriţcu and A. Popescu, eds., *CPP '21, 10th ACM SIGPLAN International Conference on Certified Programs and Proofs*, Virtual, Denmark, 2021, pp. 224–237. ACM.

[30] P. Vukmirović, A. Bentkamp, J. Blanchette, S. Cruanes, V. Nummelin, and S. Tourret. Making higher-order superposition work. In A. Platzer and G. Sutcliffe, eds., *Automated Deduction – CADE 28*, Virtual Event, 2021, LNAI 12699, pp. 415–432. Springer.


## Technical reports and preprints


[1] A. Bentkamp, J. Blanchette, S. Tourret, P. Vukmirović, and U. Waldmann. *Superposition with Lambdas*, 2021. arXiv: 2102.00453.

[2] J. Blanchette and P. Vukmirović. *SAT-Inspired Higher-Order Eliminations*, 2022. arXiv: 2208.07775.

[3] M. Bromberger, I. Dragoste, R. Faqeh, C. Fetzer, L. González, M. Krötzsch, M. Marx, H. K. Murali, and C. Weidenbach. *A Sorted Datalog Hammer for Supervisor Verification Conditions Modulo Simple Linear Arithmetic*, 2022. arXiv: 2201.09769.

[4] M. Bromberger, I. Dragoste, R. Faqeh, C. Fetzer, M. Krötzsch, and C. Weidenbach. *A Datalog Hammer for Supervisor Verification Conditions Modulo Simple Linear Arithmetic*, 2021. arXiv: 2107.03189.

[5] M. Bromberger, L. Leutgeb, and C. Weidenbach. *Explicit Model Construction for Saturated Constrained Horn Clauses*, 2023. arXiv: 2305.05064.

[6] M. Bromberger, S. Schwarz, and C. Weidenbach. *SCL(FOL) Revisited*, 2023. arXiv: 2302.05954.

[7] F. Haifani, P. Koopmann, S. Tourret, and C. Weidenbach. *Connection-minimal Abduction in EL via Translation to FOL – Technical Report*, 2022. arXiv: 2205.08449.

[8] H. Leidinger and C. Weidenbach. *SCL(EQ): SCL for First-Order Logic with Equality*, 2022. arXiv: 2205.08297.

[9] C. Lüders, T. Sturm, and O. Radulescu. *ODEbase: A Repository of ODE Systems for Systems Biology*, 2022. arXiv: 2201.08980.

[10] H. Rahkooy and T. Sturm. *Parametric Toricity of Steady State Varieties of Reaction Networks*, 2021. arXiv: 2105.10853.

[11] H. Rahkooy and T. Sturm. *Testing Binomiality of Chemical Reaction Networks Using Comprehensive Gröbner Systems*, 2021. arXiv: 2107.01706.

## Theses

[1] F. Haifani. *On a Notion of Abduction and Relevance for First-Order Logic Clause Sets*. PhD thesis, Universität des Saarlandes, Saarbrücken, 2023.

# 34 RG2: Network and Cloud Systems

## 34.1 Personnel

**Head of Group**
Yiting Xia, Ph.D.

**Researchers**
Jialong Li, Ph.D. (from November 2021)

**Ph.D. Students**
Yiming Lei (from September 2021)

**Secretaries**
Iris Wagner

## 34.2 Visitors

From March 2021 to February 2023, the following researchers visited our group:

| | | |
|---|---|---|
| Federico De Marchi | 01.03.22–present | Saarland University |
| Rui Pan | 01.02.22–30.08.22 | University of Wisconsin–Madison |
| Zhengqing Liu | 01.04.22–30.09.22 | Ecole Polytechnique, Paris |
| Vadim Farutin | 01.10.22–present | Saarland University |

## 34.3 Group Organization

At the moment, the research group consists of one tenure-track faculty, one postdoctoral researcher, one Ph.D. student, and a secretary.

In October 2020, Yiting Xia, Ph.D., joined MPI-INF as a tenure-track faculty. Before joining MPI, she was a research scientist at Facebook. She received her Ph.D. degree in Computer Science from Rice University, Texas, USA in 2018 and got her Bachelor's Degree in Telecommunications Engineering with First Class Honors from a joint program between Beijing University of Posts and Telecommunications (BUPT) and Queen Mary University of London (QMUL) in 2011. Yiting Xia is working in the field of network infrastructure for cloud computing. Her research follows a cross-layer approach and covers broad topics for optimizing the network stack, including switch hardware, network protocols, software systems, and cloud applications. Her current research topics include network architecture for next-generation cloud, optical network interconnects for growing cloud traffic and network-accelerated machine learning systems.

Jialong Li, Ph.D., joined the group as a postdoctoral researcher in November 2021. Before that, he received his Bachelor's degree and Ph.D. degree in Electronic Engineering from Tsinghua University, China in 2016 and 2021, respectively. His PhD thesis was on optical access networks and metro networks. His current research interests include data center networks, optical networks, and optical communications.

Yiming Lei joined the group as a Ph.D. student in October 2021. He received the Bachelor's degree in Computer Science from Bejing University of Posts and Telecommunications in 2021. His research interests include computer networks, cloud systems, domain-specific architectures, and data-centric computing.

Iris Wagner is a joint administrative assistant of the Internet Architecture group (D3) and Network and Cloud Systems group (RG2).

## 34.4 Optical Data Center Networks

### 34.4.1 Low-Latency Routing in Optical Data Center Networks

*Investigators: Yiting Xia, Jialong Li, and Yiming Lei, in cooperation with Federico De Marchi (Saarland University), Raj Joshi (National University of Singapore), and Balakrishnan Chandrasekaran (Vrije Universiteit Amsterdam)*

The growth of data center networks (DCNs) have largely benefited from the Moore's law for networking—the bandwidth of electrical switches doubles every two years at the same cost and power. As this bandwidth scaling is slowing down, the networking community has started exploring high-radix passive optical network interconnects, which have lower per-port cost and consume less power than electrical switches. The latest optical-DCN designs deliver up to 4-times the bandwidth and consume only 23%–26% power of a cost-equivalent electrical DCN.

A typical optical DCN fabric comprises of a number of optical switches that interconnect electrical top-of-rack switches (ToRs) and end servers (refer Fig. 34.1). The fabric uses circuit switching to establish dedicated optical circuits that are time-shared amongst the different ToR pairs. The delays incurred in establishing the circuits, however, substantially affect latency-sensitive traffic (or "mice" flows) that then use these circuits. In this paper, we focus on minimizing the impact of these delays on latency-sensitive traffic.

Prior attempts at solving this problem either use an electrical-optical hybrid network and send mice flows over the "always-on" electrical network, or reduce the circuit-establishment delays using novel optical switching hardware. The electrical-optical dual-fabric, however, doubles the deployment and maintenance costs of DCNs, while the latter requires extensive customizations to commercial network devices and the standard network stack, e.g., to adapt to the 1ns optical switching speed in Sirius. In contrast to such prior work, we propose a simple solution that leverages programmable switches: a routing algorithm with the specific objective of accelerating mice flows.

The idea of leveraging routing to accelerate latency-sensitive flows has been used in prior work, albeit within a narrow scope. Opera pursued a meticulous co-design of the optical network topology and routing for guaranteeing that mice flows *always* have (multi-hop) optical paths available via intermediate ToRs. Similar to prior designs, Opera assumes, however, that packets must be buffered on end servers, as optical switches are bufferless. As

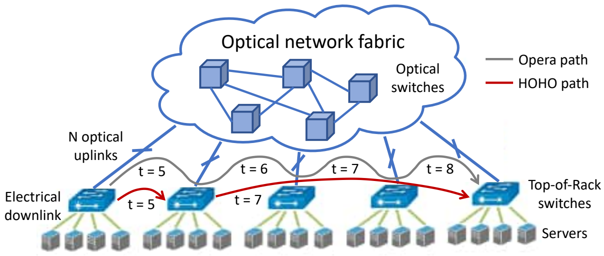



Figure 34.1: Illustration of an optical DCN where ToRs connect to an optical network fabric. If it takes one unit of time to traverse one hop, a packet leaving the source ToR at $t = 5$ (the notation shows departure times) reaches the target at $t = 9$ in Opera, which only supports non-stop paths. In HOHO, in contrast, the same packet waits at the first hop from $t = 6$ to $t = 7$ to choose the best path and arrive earlier at $t = 8$.

soon as an optical path is available, packets hop on that path and ride it until the destination. They cannot hop off at intermediate ToRs even if a different optical path later offers an earlier arrival time (at the destination). Routing in Opera is, hence, sub-optimal: It searches for **non-stop paths**, rather than **the fastest paths**. *We offer support for packets to "hop-off" at ToRs by rethinking packet buffering on ToRs.*

Buffering at ToRs was deemed impossible due to the limited packet buffer on switches, the difficulty in synchronizing switches to coordinate with optical circuit configurations, and the lack of processing logic for scheduling packet transmissions at precise times. Recent programmable switches offer rich functionalities to clear these technical obstacles. Switches can, for example, provide temporal buffering for a small number of packets, be time-synchronized at nanosecond-level precision, and provide time-based scheduled packet transmission via calendar queues.

We exploit the recent technological innovations in programmable switches to realize a novel routing algorithm for minimizing the delays experienced by latency-sensitive flows and summarize our contributions as follows [1]: (a) we present a Hop-On Hop-Off (HOHO) routing algorithm that provides the fastest paths—packets can "hop on" and "hop off" at intermediate ToRs to select the best optical paths that minimize their arrival time at the destination; (b) we prove the optimality and robustness of the HOHO algorithm, and sketch its implementation on programmable switches, including the time synchronization, routing lookup, and packet buffering mechanisms; (c) in our packet-level simulations with real DCN traffic, HOHO reduces the flow completion times (FCTs) of latency-sensitive flows by up to 35% and reduces the average path length by 15% compared to Opera. HOHO uses at most 7 queues per egress port and a packet buffer of about 3.24MB, which is far below the capacity limit of commercial switch ASICs.

## References

[1] J. Li, Y. Lei, F. De Marchi, R. Joshi, B. Chandrasekaran, and Y. Xia. Hop-On Hop-Off routing: A fast tour across the optical data center network for latency-sensitive flows. In *APNet '22, 6th Asia-Pacific Workshop on Networking*, 2022.

### 34.4.2 A General Framework for Fast-Switched Optical Data Center Networks

*Investigators: Yiting Xia, Jialong Li, and Yiming Lei, in cooperation with Federico De Marchi (Saarland University), Zhengqing Liu (Ecole Polytechnique, Paris), Raj Joshi (National University of Singapore), and Balakrishnan Chandrasekaran (Vrije Universiteit Amsterdam)*

The last 15 years has witnessed the emergence and development of optical data center networks (DCNs). A series of optical DCN architectures have been proposed to leverage the bandwidth, power, and cost advantages of optical interconnects. Compared to electrical interconnects in traditional DCNs, optical interconnects use circuit switching to establish dedicated optical circuits between end points and shift the circuits across "time slices" to create time-shared networks. Circuit reconfiguration incurs a "switching delay" determined by the specific optical switching technology adopted by the architecture.

This journey started from *slow-switched* optical DCNs, with tens of milliseconds of switching delays. Limited by the switching speed, this type of optical network has to work in tandem with an electrical network to avoid network partitioning, e.g., either augmenting the electrical DCN with on-demand circuits to offload heavy traffic, or serving as "patch panels" for electrical switches and reconfigure the network topology on a seconds to hours granularity. For example, Jupiter—Google's DCN fabric—has achieved 5× capacity increase, 41% power reduction, and 30% cost reduction after deploying slow-switched optical interconnects in the network core. These optical interconnects provide large port counts to interconnect electrical switches and reconfigure the DCN topology when needed, e.g., at device upgrade and failure times, or once a few hours as the DCN traffic evolves.

*Fast-switched* optical DCNs, whose switching delays vary between several nanoseconds to tens of microseconds, have been increasingly recognized in recent years driven by the prevalence of mice flows in DCN applications. These architectures shift circuits continuously to route traffic in the optical domain on an all-optical network fabric. The removal of the electrical network further reduces cost compared to slow-switched optical DCNs, but at the same time deviates from the all-to-all connectivity assumed by conventional DCN designs. How to build the networked system to adapt to the transient circuits, with merely microsecond-scale durations under fast switching, is still largely unknown.

We foresee challenges from multiple fronts for the implementation and eventual deployment of fast-switched optical DCNs. (1) How should network devices be time-synchronize network-wide at sub-microsecond or even nanosecond accuracy to keep traffic in sync with the rapidly reconfigured circuits, preferably in-band over the optical network fabric now that the electrical network is out of the way? (2) As the circuit duration drops to the same scale as the DCN RTT and delays on the host stack, how should the Top-of-Rack switch (ToR) and host systems be designed to maintain good performance? (3) Even if implemented, each optical architecture is a closed ecosystem with heavily coupled optical hardware and networked system, how to upgrade the network from one architecture to another after deployment?

In this project, we address these challenges with OpenOptics, a general framework to make fast-switched optical DCNs practically realizable. *OpenOptics aims to do to fast-switched optical DCNs what OpenFlow did to traditional networks.* It allows specific optical hardware to be integrated into the general framework in a plug-and-play manner to have a workable end-to-end system, and cloud applications can run without changes as if on traditional DCNs. As optical technologies advance, different optical architectures can be realized straightforwardly on top of OpenOptics, and the system can remain intact when the DCN fabric is upgraded to newer optical hardware. By decoupling the software system from the optical hardware, we make the niche area of optical DCNs more accessible to network researchers. We will open-source OpenOptics to encourage real-world testing and education of fast-switched optical DCNs.

The enabler of generality in OpenOptics is HOHO routing, a unified routing algorithm we published at a workshop to apply to different fast-switched optical DCN architectures. Most fast-switched optical DCN architectures use a pre-defined optical schedule, i.e., a repetitive sequence of circuit connections over time slices, to avoid expensive real-time traffic estimation and circuit planning under the short time slice durations. HOHO routing takes advantage of this fact to abstract each architecture by its optical schedule. It takes the optical schedule as input and computes offline the lowest-latency paths (proven to be optimal) for mice flows. We replace the specific routing algorithm of each architecture with HOHO routing, which produces better paths for mice flows and preserves the direct paths between source and destination ToRs for elephant flows.

With the unified HOHO routing, we can unify the ToR and host systems across architectures as well. We implement the runtime system for the offline HOHO algorithm on ToRs and hosts, using P4 on Intel Tofino2 switches and VMA on Mellanox NICs. We bear the aforementioned challenges in mind and embrace a systematic design by testing the boundaries of these commercial tools for fast-switched optical DCNs. Specifically, we realize network-wide in-band ToR synchronization based on profiled synchronization errors between Tofino2 switches; we implement HOHO routing on ToRs with careful measurements of system delays on Tofino2 switches per the critical steps; and we build an application-agnostic host network with a fair judgement of the overheads of kernel and kernel-bypass options.

Our micro-benchmark evaluation of OpenOptics performance shows that our in-band ToR synchronization can keep the synchronization errors under 15ns, our ToR system achieves zero packet loss with 99.93% achievable network utilization, and our host system sends 99.4% packets inside the scheduled time slices. We demonstrate the generality of OpenOptics by realizing Mordia, RotorNet, and Opera—three fast-switched optical DCN architectures—on top of it. Case studies running Memcached and Gloo applications on them show that the tail flow completion times for mice flows in OpenOptics is only 16% worse than that of an electrical DCN.

## 34.5 Network-Accelerated Machine Learning Systems

### 34.5.1 Network Abstraction for Distributed Deep Learning Training

*Investigators: Yiting Xia, Jialong Li, and Yiming Lei, in cooperation with Rui Pan (Princeton University), Zhiqiang Xie (Stanford University), and Binhang Yuan (ETH Zürich)*

Recent years have witnessed the rapid development of deep learning: each leap in the model quality comes with increased scales of neural networks, from AlexNet with 61M parameters in 2012 to MT-NLG with 530B parameters in 2022. Various parallel strategies have been adopted by distributed deep learning training (DDLT) frameworks to accommodate the ever-growing model sizes. As a result, communication among distributed workers, especially over a shared, highly dynamic network with competing training jobs, has become a notable bottleneck of the training process.

The networking community has a long history of resolving bandwidth contentions with flow scheduling, from individual flow scheduling to Coflow scheduling. Surprisingly, despite the popularity of DDLT applications, we have found no flow scheduling solution supporting the diverse DDLT paradigms in GPU clusters! Our analysis suggests two reasons.

The *first reason* is due to the challenge of defining *a global optimization goal* across training jobs. The various DDLT paradigms implement drastically different workflows, which may translate into incompatible network requirements causing network-wide optimization to diverge. As such, communication optimizations for DDLT focus on data parallelism only, and most work conduct per-job optimization, with estimations of the available bandwidth.

Pioneering explorations for flow scheduling in DDLT, also limited to data parallelism, faced exactly this problem. Particularly, CadentFlow identified multiple performance metrics, e.g., weights, deadlines, and priorities, which may pull the optimization from different directions; MLNet proposed to schedule flows by priorities, but how to set priorities to reflect application needs is unknown.

The *second reason* is the lack of *network abstraction* for DDLT. The Coflow network abstraction for traditional cluster applications falls short in DDLT. Coflow defines a collection of semantically-related flows and minimizes the completion time of the last flow. This goal motivates the optimizer to schedule the flows to finish at the same time. Oftentimes in DDLT, though, the followup computations consuming the flow data do not start at the same time.

Taking pipeline parallelism for example, each worker computes on sequential micro-batches and sends the results to a successor worker. Consecutive workers have data dependencies, and the computations per worker follow the order of input data. For high training throughput, GPU workers must be well coordinated to preserve the pipeline throughout the training lifetime. Delay or reordering of data may increase GPU idleness and reduce training efficiency. To match this strict computation pattern, data flows across micro-batches should (ideally) finish in a staggered manner. Formulating the flows as a Coflow tends to finish them simultaneously, making the duration of this computation phase even longer than bandwidth fair sharing!

Through extensive workflow analysis, we generalize this observation to other DDLT paradigms: regardless of the great diversity, *each DDLT paradigm has a unique, pre-defined computation pattern that regulates the finish times of flow transmissions.* These computation patterns, which are essentially computation dependencies (i.e., DAG) and times, are prevalent in distributed applications. Yet, the repetitiveness of DDLT jobs, e.g., similar or identical computations across training layers and iterations, makes it possible to extract the patterns through computation profiling and convey the application-level guidelines to network flows.

Following this insight, we aspire to fill the gap of flow scheduling in DDLT. We propose the *EchelonFlow network abstraction* to finish flows according to strict DDLT computation patterns [1], and with EchelonFlow comes our *global optimization goal* of minimizing communication time while maintaining the computation patterns, like preserving the arrangement of an echelon formation. *EchelonFlow is the first network abstraction for flow scheduling in diverse DDLT paradigms.* It is also extensible to future DDLT paradigms, as long as their computation patterns can be profiled.

Our contributions in this project are from four fronts. (1) We formally define EchelonFlow and formulate a global optimization goal for it. (2) We prove important properties of EchelonFlow. Particularly, EchelonFlow scheduling can minimize completion times of mainstream DDLT paradigms, and EchelonFlow is a superset of Coflow. (3) Through case studies, we show the expressiveness of EchelonFlow by presenting popular DDLT paradigms with the EchelonFlow abstraction. (4) We sketch the system implementation to discuss the practicality of EchelonFlow scheduling.

## References


• [1] R. Pan, Y. Lei, J. Li, Z. Xie, B. Yuan, and Y. Xia. Efficient flow scheduling in distributed deep learning training with echelon formation. In *HotNets '22, 21st ACM Workshop on Hot Topics in Networks*, Austin, TX, USA, 2022, pp. 93–100. ACM.


### 34.5.2 Network-Aware GPU Sharing for Distributed Deep Learning

*Investigators: Yiting Xia, Jialong Li, and Yiming Lei, in cooperation with Vadim Farutin (Saarland University)*

In recent years, the advent of distributed training jobs has significantly hastened the deployment of GPU clusters. However, the prevalent strategy for running training jobs on these clusters is exclusive, which restricts one GPU to be assigned to a single training job. This monopolistic approach leads to low utilization of GPU clusters. Empirical data obtained from Alibaba's production GPU clusters indicates that a mere 10% of the GPUs manage to achieve more than 80% GPU utilization, highlighting the considerable potential for enhancing GPU utilization.

With the aim of improving GPU utilization, recent research has explored the utilization of GPU resources in time-sharing way for training workloads, such as Gandiva, AntMan, and Salus. These approaches primarily focus on implementing GPU time-sharing among multiple jobs, considering the network as a black box. Generally, GPUs require network transmission after computation (Forward/Backward propagation). These approaches assume that computation is the bottleneck in DNN training and do not take network transmission time into account. However, as the complexity of models and datasets rapidly improves, network transmission has become a crucial factor in DNN training and cannot be ignored when applying GPU sharing.

In contrast to the above approaches, when scheduling training jobs, Muri considers four types of resources, including CPU, storage, and network resources. Muri assumes that at any given time, a training job can utilize only one type of resource, resulting in the idleness of other resources. Muri's key idea involves grouping multiple jobs on the same GPU and using four types of resources in an interleaving manner. As a result, Muri achieves a lower job completion time (JCT) and higher resource utilization in comparison to other baselines.

Despite achieving higher resource utilization, Muri faces several problems.

**Problem 1: Muri assumes a fixed network transmission time.** Muri assumes a fixed network transmission time by profiling resource usage duration in advance. This assumption is impractical since the network transmission time varies due to competition for bandwidth among multiple training jobs. Additionally, the network transmission time may differ in a new deployment environment with different bandwidth capacities. Although Muri considers network transmission time when applying resource sharing, the assumption of a fixed network transmission time is not applicable in practice.

**Problem 2: Muri does not consider device placement when mapping jobs to GPUs.** Muri assumes that each GPU occupied by the same training job will send/receive the same amount of data, and therefore only decides which jobs should be mapped to the same GPU without considering which partitions of these grouped jobs should be placed together. However, if this assumption does not hold, we need to consider where to place the partitions since the placement of partitions plays a crucial role in determining the JCT. A simple observation is that assigning two partitions with heavy-loaded traffic on the same GPU will result in a long network transmission time. With the rapid development of DNN training models, more training jobs will not follow this assumption. Therefore, besides grouping jobs, it is imperative to consider where to place partitions, taking into account the communication patterns and data traffic of each training job.

**Problem 3: Muri only groups jobs using the same number of GPUs.** Muri only groups jobs using the same number of GPUs, thereby missing opportunities to realize higher GPU utilization. The new design should aim to group any number of jobs, regardless of their GPU requirements.

To address the aforementioned problems, we propose the Network-Aware GPU Sharing (NAGS) design for DNN training jobs. The NAGS design introduces a new concept called sharing group, which accommodates an arbitrary number of training jobs with any number of GPUs. Two metrics, namely, sharing factor and sharing utilization, are used to evaluate the sharing group's performance. The sharing factor measures the degradation in JCT due to GPU sharing, while the sharing utilization measures the GPU resource utilization.

NAGS tries to answer this research question: *For training jobs with arbitrary amounts of sending/receiving data, how can we group jobs and place partitions to minimize JCT degradation and maximize GPU resource utilization?*

To answer this question, we first propose two theorems and provide the corresponding proofs. Based on these theorems, we derive the minimum sharing factor and the optimal partition placement for any given set of jobs. Furthermore, we prove that constructing sharing groups optimally is an NP-hard problem. To overcome this challenge, we develop heuristic methods to construct three types of sharing groups.

Notably, the NAGS design is not dependent on strong assumptions. Only two main assumptions are required in this work: (1) profiling the traffic matrix of each job in advance, and (2) applying the "big switch" network model. Preliminary results validate the superiority of NAGS over Muri and other baselines.

## 34.6 A Reliable Network Management Workflow System at Scale

*Investigators: Yiting Xia in cooperation with Kuo-Feng Hsu, Jiarong Xing, Ang Chen (Rice University), Yan Cai, Yanping Li, and Ying Zhang (Meta)*

Managing complex networks is challenging, particularly at scale. A planet-scale cloud/content provider's network can contain devices from dozens of vendors, many tens of network element roles, and tens of thousands of circuits in operation, which altogether contribute to thousands of configuration changes every day. Besides, network management involves a diverse set of tasks, ranging from service upgrades, device updates, network expansion, to feature deployment— each requiring a disparate workflow of operations. The combined effect of the network scale and task diversity has made network management a risk-ridden procedure: as changes often come with risk, it is up to the network management system to monitor the network health, detect runtime failures, and assist with recovery when possible.

**Reliability** is, hence, the primary goal of network management given its error-prone nature. We present the network management system in Meta and how it evolves through two generations to enhance reliability.

Our first generation, Netgram, is a network-specific workflow system. In Netgram, a workflow is a Python program written in a pipeline of stages, where each stage performs several execution steps and passes intermediate results to the next. A stage can invoke pre-defined sub-procedures for common network operations, which are flexibly implemented as any executables, such as Ansible, Python, or CLI scripts. Netgram is a significant leap from the former practice of 'method of procedures' (MOPs), which uses loose text to document management steps and rules of thumbs, and thus can be easily misinterpreted by individual operators and often requires manual translation into scripts or configlets. Netgram effectively eliminates manual efforts and streamlines the management tasks in an automated and trackable manner.

After four years operating Netgram in production, however, we learned the lesson that workflow programs have few restrictions or validations, so they could still cause severe damages to the network, especially when applied automatically at scale. Just like how rogue Shell scripts may wipe an entire disk, a problematic workflow program can exert arbitrary influence to the network. We analyzed failure tickets caused by Netgram and interviewed network engineers for operational pain points. From the results, we broke down the reliability goal into four tangible requirements of the workflow system.

**Safety**. Unconstrained workflows could result in unexpected network outages, e.g., draining too much capacity and leaving the network severely congested. A workflow system for network management should provide network operators with a confined set of allowable operations and simple interfaces to carry out validations.

**Consistency**. A network management task may re- quire changes of multiple configurations on a device or changes on multiple devices. Sometimes it requires changes applied to a set of devices in a pre-defined order. A workflow system should guarantee that all necessary changes are all made successfully so that the network is in a consistent and correct state.

**Efficiency**. Many changes need to be deployed as quickly as possible, e.g., to mitigate failures, to balance traffic, and to deploy a security patch. Thus, scheduling and executing workflows efficiently to minimize network vulnerable time is another requirement.

**Resilience**. Workflows may fail for various reasons, e.g., wrong device states, conflicts with other workflows, failure of manual steps (such as maintenance in the field), etc. Partially executed workflows should be undone according to the specific order of each counter-operation, which is not as simple as reverse-order rollback.

Our second generation system, Netgram++, provides systematic reliability guardrails with these considerations. Our core idea is leveraging the fact that most industry-grade net- work management systems already separate logical network representations from the physical networks. The logical network data is stored in a source-of-truth database, such as FBNet at Meta and Malt at Google. All network management tasks require read and write to the network data in some form. Thus, we can abstract the network management workflows as changes to the network data and apply various database techniques to implement customized features tackling each reliability requirement.

For safety, we propose a programming model with the network object abstraction and a set of APIs. A Netgram++ program constructs network objects by scoping a number of devices, and all operations to them are through the APIs. From the programming model, the runtime system can automatically generate queries to FBNet, enforce constraints on the network, monitor task progress, and handle failures. For consistency, we build an object tree from network objects based on their dependencies and apply multi-granularity locking on the object tree to enable workflow-level transactions. For efficiency, we perform hierarchical lock scheduling on the object tree to maximize task parallelization and minimize execution time. For resilience, we identify limitations of reverse-order rollbacks and devise rollback plans with pattern matching on semantics of management operations.

Our contributions in this project are as follows. (1) We are the first to share experience of a production-grade network workflow management system and present comprehensive workflow measurements that can be used for future research in this area. (2) We demonstrate how database techniques can be customized to workflow systems to enhance reliability, and give detailed explanation of an example design with a programming model and runtime system for locking, scheduling, and failure handling. (3) We conduct extensive evaluation of the improved system with both simulation and production case studies: the programming model reduces LoC of workflow programs by 95.4% to 96.4%, effective locking and scheduling reduces task execution time by 90.1%, and conflicts between multiple workflows are resolved successfully. (4) We offer the first open-sourced task traces and simulator of net-work workflow management systems, which give academic researchers access to first-hand production data and an evaluation tool to study an industrial problem.

## 34.7 Academic Activities

### 34.7.1 Conference and Workshop Positions

#### Membership in program and organization committees

*Yiting Xia:*

- *37th ACM Special Interest Group on Data Communication, SIGCOMM 2023*, New York City, USA, September 2023 (PC member),
- *20th USENIX Symposium on Networked Systems Design and Implementation, NSDI 2023*, Boston, USA, April 2023 (PC member),
- *22nd ACM Workshop on Hot Topics in Networks, HotNets 2023*, Location TBD, November 2023 (PC member),
- *The 2022 Internet Measurement Conference, IMC 2022*, Nice, France, October 2022 (PC member),

- *6th Asia-Pacific Workshop on Networking, APNet 2022*, Fuzhou, China, July 2022 (PC member),
- *The 2021 ACM SIGCOMM Symposium on SDN Research, SOSR 2021*, Virtual Conference, October 2021 (PC member),
- *29th IEEE International Conference on Network Protocols, ICNP 2021*, Virtual Conference, November 2021 (PC member).

#### Membership in steering and other committees

*Yiting Xia:*

- *12th Networking Networking Women Professional Development Workshop, N2Women 2022*, Amsterdam, Netherlands, August 2022 (Co-chair).

### 34.7.2 Invited Talks and Tutorials

*Yiting Xia:*

- *Towards High-Performance and Power-Efficient Network Infrastructure for Cloud Computing*, 8th Winter Seminar Series at Sharif University of Technology, April 2023,
- *Towards High-Performance and Power-Efficient Network Infrastructure for Cloud Computing*, Peking University, March 2023,
- *Towards High-Performance and Power-Efficient Network Infrastructure for Cloud Computing*, Tsinghua University, March 2023,
- *Towards Deployable Optical Data Center Networks with Unified Routing*, MPI-SWS Lightning Tutorials Series, July 2022,
- *Towards Deployable Optical Data Center Networks with Unified Routing*, Huawei InnovWave HCIO2022, June 2022,
- *A Social Network under Social Distancing – Experience and Insights during COVID-19 on Risk-Driven Backbone Management*, Saarland Informatics Campus Lecture Series, February 2021,
- *A Social Network under Social Distancing – Experience and Insights during COVID-19 on Risk-Driven Backbone Management*, MIT Computer Science & Artificial Intelligence Laboratory, October 2020.

## 34.8 Teaching Activities

### Summer Semester 2021

- Distributed Systems (Co-lecturer, Yiting Xia)

### Winter Semester 2021

- Advance Topics on Data Networks (Co-lecturer, Yiting Xia and Jialong Li)

### Summer Semester 2022

– Data Networks (Co-lecturer, Yiting Xia and Jialong Li)

### Winter Semester 2021

– Advance Topics on Data Networks (Co-lecturer, Yiting Xia and Jialong Li)

### Summer Semester 2023

– Distributed Systems (Co-lecturer, Yiting Xia)

### Master and Bachelor Theses

– Zhengqing Liu: Hop-On Hop-Off System Implementation via Programmable Switches, Master's thesis, 2022 (Supervisor: Yiting Xia, Jialong Li)

## 34.9 Dissertations, Habilitations, Awards

### 34.9.1 Awards

– Yiting Xia: *N2Women Rising Star Award 2021*

## 34.10 Publications

### Conference articles

[1] S. S. Ahuja, V. Gupta, V. Dangui, S. Bali, A. Gopalan, H. Zhong, P. Lapukhov, Y. Xia, and Y. Zhang. Capacity-efficient and uncertainty-resilient backbone network planning with hose. In F. Kuipers and M. Caesar, eds., *SIGCOMM '21*, Virtual Event, USA, 2021, pp. 547–559. ACM.

[2] J. Li, Y. Lei, F. De Marchi, R. Joshi, B. Chandrasekaran, and Y. Xia. Hop-On Hop-Off routing: A fast tour across the optical data center network for latency-sensitive flows. In *APNet '22, 6th Asia-Pacific Workshop on Networking*, 2022.

[3] R. Pan, Y. Lei, J. Li, Z. Xie, B. Yuan, and Y. Xia. Efficient flow scheduling in distributed deep learning training with echelon formation. In *HotNets '22, 21st ACM Workshop on Hot Topics in Networks*, Austin, TX, USA, 2022, pp. 93–100. ACM.

[4] Y. Xia, Y. Zhang, Z. Zhong, G. Yan, C. Lim, S. S. Ahuja, S. Bali, A. Nikolaidis, K. Ghobadi, and M. Ghobadi. A social network under social distancing: Risk-driven backbone management during COVID-19 and beyond. In *Proceedings of the 18th USENIX Symposium on Networked Systems Design and Implementation*, Virtual Event, 2021, pp. 217–231. USENIX Association.

[5] Z. Zhong, M. Ghobadi, A. Khaddaj, J. Leach, Y. Xia, and Y. Zhang. ARROW: Restoration-aware traffic engineering. In F. Kuipers and M. Caesar, eds., *SIGCOMM '21*, Virtual Event, USA, 2021, pp. 560–579. ACM.

# Part IV

# Index

# Index