Towards Understanding Climate Change Perceptions A Social Media Dataset

Katharina Prasse
University of Siegen
katharina.prasse@uni-siegen.de

Steffen Jung
Max Planck Institute for Informatics,
Saarland Informatics Campus
University of Siegen

Isaac Bravo
Technical University Munich

Stefanie Walter
Technical University Munich

Margret Keuper
University of Mannheim
Max Planck Institute for Informatics,
Saarland Informatics Campus

Abstract

Climate perceptions shared on social media are an invaluable barometer of public attention. By directing research towards this topic, we can eventually improve the effectiveness of climate change communication, increase public engagement, and enhance climate change education. We propose two real-world image datasets to promote impactful research both in the Computer Vision community and beyond. Firstly, *ClimateTV*, a dataset containing over 700,000 climate change-related images posted on Twitter and labelled on basis of the image hashtags. Secondly, *ClimateCT*, a Twitter dataset containing images with five-dimensional annotations in super-categories (i) animals, (ii) climate action, (iii) consequences, (iv) setting, and (v) type. These challenging classification datasets contain classes which are designed according to their relevance in the context of climate change. The challenging nature of the datasets is given by varying class diversities (e.g. polar bear vs. land mammal) and foci (e.g. arctic vs. snowy residential area). The analyses of our datasets using CLIP embeddings and query optimization (CoCoOp) further showcase the challenging nature of *ClimateTV* and *ClimateCT*. Both datasets are available at `http://data.climatevisions.eu/`.

1 Background & Motivation

Climate Change is indubitably one of the biggest challenges of our time. We aim to promote this highly relevant research topic to the computer vision community by providing social media data and defining a multi-label classification problem with class design that considers relevant social science literature. Moreover, we hope to contribute to interdisciplinary research with social sciences in advancing the understanding of climate change discourse.

Importance of Climate Change Perceptions. Within Social Sciences, consequences, nature, and political discourse or engagement are investigated Mooseder et al. [2023], Weaver et al. [2022], O'Neill [2020], DiFrancesco and Young [2011], Hayes and O'Neill [2021], Wang et al. [2018], Casas and Williams [2019], Rebich-Hespanha et al. [2015], Johann et al. [2023]. Moreover, climate change consequences, and animals, are frequently used categories Mooseder et al. [2023], Weaver et al.

Tackling Climate Change with Machine Learning: workshop at NeurIPS 2023.

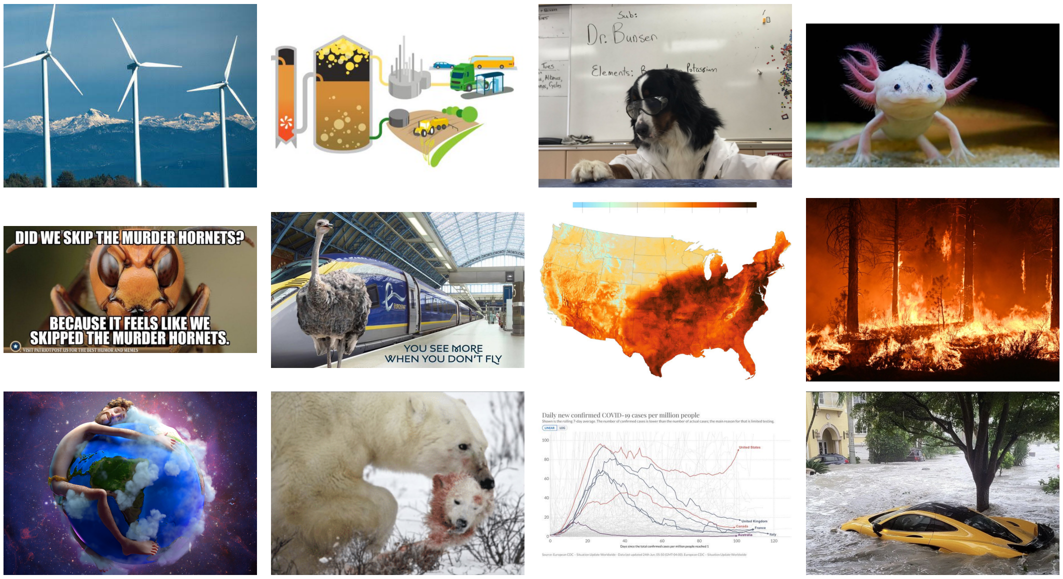

Figure 1: The collected data has a wide variety in terms of image types (photos, plots, info-graphics, drawings, etc.) and visual content. The selection criteria are identical for ClimateCT and ClimateTV, their sole difference is the types of labels assigned (manually vs. hashtag-based).

[2022], O'Neill [2020], Hayes and O'Neill [2021], besides climate change engagement Johann et al. [2023], Hayes and O'Neill [2021], Wang et al. [2018].

Several works specifically investigate the climate change perceptions using Twitter images from a social sciences perspective Mooseder et al. [2023], Rebich-Hespanha et al. [2015] and add the categories *text/quotes* and *infographic* to the list of relevant categories describing the type of an image. Additionally, Mooseder et al. [2023] starts describing the setting of images (i.e. Nature, Urban, Agriculture), which we find highly interesting, as they frequently occur in the background. We argue in favour of multiple categories per image to improve its descriptiveness (e.g. a photo (type) of a wildfire (consequence) in a forest (setting)). We thus created the *Climate Change on Twitter* dataset *ClimateCT*, which contains manual annotations in a five-dimensional label space. The employed super-categories include both descriptive labels, (i) image type, (ii) image setting, and (iii) animals, and analytical labels, (iv) consequences, and (v) climate action. Further, we created the *Climate Twitter Visuals* dataset *ClimateTV*, which contains links to a large number of images shared on Twitter and their hashtag-based labels.

Importance of Computer Vision Datasets. While computer vision research generally focuses on large-scale, curated benchmark datasets Deng et al. [2009], Krizhevsky et al. [2009], Cordts et al. [2016], Liu et al. [2015], Lin et al. [2015], Everingham et al. [2010], Chua et al. [2009], Sumbul et al. [2019], Haurum and Moeslund [2021], we want to stress the importance of evaluating models on real-life data in order to leverage the potential of computer vision for society as a whole. Recently, researchers have found that detection heavily relies on the style of the depicted item Geirhos et al. [2022]. Stylized ImageNet includes various depictions of objects in order to forgo this limitation. Our datasets offer the same advantage, as images with the same label are displayed in various styles. For example, a residential area is displayed during sunshine, rain, and snow. The authors of BigEarthNet Sumbul et al. [2019] had similar intentions when including remote sensing images from all four seasons. By allowing all kinds of image types, e.g. photos, drawings, memes, etc., we allow the differences between the content of the same class to be even larger, i.e. a dog can be photographed, drawn, or painted. One particular difference is that the classes in classical computer vision benchmark datasets are often “semantically balanced" in their class definition, e.g., different animal species are considered equally important. However, from a climate change perspective, this is not necessarily the case: Images of iconic objects or animals (e.g. polar bears) are often used in very specific ways and deserve special attention while other animals can be summarized more widely, e.g. forming semantically diverse classes such as farm animals. This semantic imbalance makes the proposed datasets challenging beyond the perspective of proposing a novel application area.

2 Datasets

We propose *ClimateTV* which contains more than 700,000 image links and the corresponding hashtags, posted on Twitter in the context of climate change. We define a mapping from hashtags to our labels. Additionally, we have created the *ClimateCT* image link dataset, which contains five manually assigned labels per image and sufficient instances to conduct 16-shot learning and validation. Both datasets are available at `http://data.climatevisions.eu/`. All licensing concerns are discussed in the supplementary material.

2.1 Data Collection

All tweeted images must contain either *#climatechange*, *climate change* or *climatechange*.

Manually labeled ClimateCT contains links to Twitter images from January 1, 2019, to December 31, 2022. In total, it contains 1,038 images with five labels. The data collection period was chosen to validate the choice of super-categories and categories. The top ten most popular images in terms of absolute numbers of likes and re-tweets were chosen for each month. In order to have a sufficient number of instances for each class, additional images were manually downloaded from Twitter to forgo any model bias. The number of images differs between categories, but we ensure sufficient data for 16-shot learning and subsequent testing, hence at least 21 images per class.

Hashtag-based labeled ClimateTV includes links to images tweeted in the period of January 1, 2019, to December 31, 2019. This dataset contains more than 700,000 images and has no overlap with the *ClimateCT* dataset. Climate change-relevant class labels and hashtags have been embedded using SONAR Duquenne et al. [2023] and matched based on their cosine similarity. We assign the class label as quasi-labels if its cosine similarity to the image hashtag is above the threshold of 0.9.

2.2 Super-Categories

We propose a five-dimensional labelling scheme arguing that a high-dimensional topic such as climate change requires also high-dimensional labels. The labels are organized into five super-categories, i.e.

(i) Animals: Pets, Farm animals, Polar bear, Land mammal, Sea mammal, Fish, Amphibian / Reptile, Insects, Birds, Other animals

(ii) Climate action: Protest, Politics, Sustainable energy, Fossil energy, Other climate action

(iii) Consequences: Biodiversity loss, Covid / Health, Drought, Floods, Wildfires, Other extreme weather, Melting ice, Sea level rise, Rising temperature, Human rights, Economic consequences, Other consequences

(iv) Setting: Residential / Commercial, Industrial, Agricultural, Indoor space, Arctic / Antarctic, Ocean / coastal, Desert, Forest / Jungle, Other nature, Outer space, Other setting

(v) Type: Photo, Photo Collage, Illustration, Meme, Data Visualization, Screenshot, Infographic, Poster / Event invitation, Other type

The super-categories are selected in accordance with relevant climate change literature Rebich-Hespanha et al. [2015], Mooseder et al. [2023] with the goal of avoiding overlapping categories. While animals contains various general animal categories, the polar bear, the most popular animal in climate change images is given its own category Born [2019]. The super-category settings describes *What kind of physical space is shown in the image*, ranging from various urban to natural settings. This distinction allows for a more fine-grained classification, as photos depicting animals in a city and animals in the wild send distinctly different messages. Further, the super-category type describes the format of the images and contains the categories proposed by Mooseder et al. and Rebich-Hespanha Mooseder et al. [2023], Rebich-Hespanha et al. [2015], i.e. infographic, photograph, screenshot, etc. Moreover, we use the climate consequences proposed by Mooseder et al. [2023]. Lastly, the super-category climate action describes whether the image contains climate change-related actions.

Annotation process. For each super-category, all images were annotated individually, thus each image was seen five times. For each image, at least two annotators assigned a label, and conflicting labels were resolved during discussion. All labels were explained to the five annotators and it was ensured that for a single image, all labels were given by the same set of annotators.

3 Analysis and Baseline Results

With this section, we aim to both give an insight into the structure of our datasets and to exhibit its challenging nature. We use CLIP Radford et al. [2021] to make zero-shot predictions for each label category within the *ClimateCT* dataset. The accuracies are compared between the manually designed CLIP queries Radford et al. [2021] and their feature-engineered query counterparts learned using Conditional Context Optimization [CoCoOp] Zhou et al. [2022]. CoCoOp Query optimization Zhou et al. [2022] is trained in a 16-shot learning setting and evaluated on both the manually-labelled *ClimateCT* and the hashtag-based labelled *ClimateTV*. We report the results for the super-category **(iii) consequences**. Further results are reported in the supplementary material.

The classes and corresponding keywords for the super-category consequences are shown in Table 1. The majority of classes comprise a single query, except for classes covid / general health and other extreme weather, where two queries were needed to describe the class. Queries similarities are highest between the general classes, i.e. economic consequences, other extreme weather, and climate change consequences. Melting ice is the most distinct keyword. The full query similarity matrix is given in the supplementary material. The model accuracies for the super-category consequences are reported in Table 3. For *ClimateCT*, the overall accuracy significantly increases when the queries are optimized. While the majority of classes greatly increase in accuracy, economic consequences and human rights increase by a smaller amount, possibly due to the abstract nature of these classes. For the strongest CLIP class wildfires, the class accuracy decreases by more than 10%.

Class	**Keywords**
Biodiversity loss	Biodiversity loss
Covid / general health	Covid, health
Drought	Drought
Floods	Floods
Wildfires	Wildfires
Other extreme weather	Extreme weather events, Hurricane
Melting Ice	Melting Ice
Sea level rise	Sea level rise
Rising temperature	Rising temperature
Human rights	Human rights
Economic consequences	Economic consequences
Other Consequences	Climate Change Consequences

Table 1: Classes and keywords used for the super-category consequences. The standard CLIP query "A photo of a {*Keyword*}" is used.

Dataset **Model**	**CT** **CLIP** Acc.	**CT** **CoCoOp** Acc.	**TV** **CoCoOp** Acc.
Total	40.51	71.12	33.06
Biodiversity loss	50.00	77.16	36.76
Covid / general health	0.00	56.68	37.20
Drought	29.17	78.98	23.80
Floods	66.67	86.64	34.68
Wildfires	95.24	83.72	41.78
Other extreme weather	55.33	87.50	17.38
Melting ice	66.67	80.00	61.64
Sea level rise	44.74	66.64	26.78
Rising temperature	64.00	98.00	27.64
Human rights	75.00	88.76	43.14
Economic consequences	50.00	51.44	27.02
Other consequences	15.25	n/a	n/a

Table 2: The super-category consequences evaluated using CLIP and CoCoOp on both datasets.

When evaluating the performant CoCoOp Zhou et al. [2022] queries on the *ClimateTV* dataset, the accuracy decreases due to the high diversity of this dataset. The most visually distinct classes, melting ice and wildfires have the highest class accuracies, while abstract classes, such as other extreme weather have the lowest accuracy scores. Examples of *ClimateTV* images are included in the supplementary material.

4 Discussion

With *ClimateCT* we offer the Computer Vision community the opportunity to evaluate and improve models and methods on high-impact data with high-quality labels. Additionally, we offer the hashtag-based labelled *ClimateTV* for evaluation on a large scale. Simultaneously we advance the research into understanding public attention to the topic of climate change, by providing novel datasets and fostering research in the automatic content description. Insights can be employed to reduce public misinformation and to better steer public engagement in this context.

5 Acknowledgements

This work is partially supported by the BMBF project 16DKWN027b Climate Visions and DFG research unit 5336 Learning2Sense. All experiments were run on the computational resources of the University of Siegen and the Max Planck Institute for Informatics. We thank Shashank Agnihotri for his guidance.

References

D. Born. Bearing witness? polar bears as icons for climate change communication in national geographic. *Environmental Communication*, 13(5):649–663, 2019.

A. Casas and N. W. Williams. Images that matter: Online protests and the mobilizing role of pictures. *Political Research Quarterly*, 72(2):360–375, 2019. ISSN 10659129. URL `http://www.jstor.org/stable/45276914`.

T.-S. Chua, J. Tang, R. Hong, H. Li, Z. Luo, and Y. Zheng. Nus-wide: a real-world web image database from national university of singapore. In *Proceedings of the ACM International Conference on Image and Video Retrieval*, page 1–9, Santorini, Fira Greece, Jul 2009. ACM. ISBN 978-1-60558-480-5. doi: 10.1145/1646396.1646452. URL `https://dl.acm.org/doi/10.1145/1646396.1646452`.

M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele. The cityscapes dataset for semantic urban scene understanding. In *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

X. Corp. Twitter developer agreement. `https://developer.twitter.com/en/developer-terms/agreement`, a. Accessed: 2023-08-26.

X. Corp. Twitter terms of service. `https://twitter.com/en/tos#intlTerms`, b. Accessed: 2023-08-26.

J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.

D. A. DiFrancesco and N. Young. Seeing climate change: the visual construction of global warming in canadian national print media. *cultural geographies*, 18(4):517–536, 2011. doi: 10.1177/1474474010382072. URL `https://doi.org/10.1177/1474474010382072`.

P.-A. Duquenne, H. Schwenk, and B. Sagot. Sonar: Sentence-level multimodal and language-agnostic representations. *arXiv preprint arXiv:2308.11466*, 2023.

M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The pascal visual object classes (voc) challenge. *International Journal of Computer Vision*, 88(2):303–338, June 2010.

T. Fenzl and P. Mayring. Qcamap: eine interaktive webapplikation für qualitative inhaltsanalyse. 2017.

R. Geirhos, P. Rubisch, C. Michaelis, M. Bethge, F. A. Wichmann, and W. Brendel. Imagenet-trained cnns are biased towards texture; increasing shape bias improves accuracy and robustness. *arXiv preprint arXiv:1811.12231*, Nov 2022. URL `http://arxiv.org/abs/1811.12231`. arXiv:1811.12231 [cs, q-bio, stat].

J. B. Haurum and T. B. Moeslund. Sewer-ml: A multi-label sewer defect classification dataset and benchmark. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13456–13467, 2021. doi: 10.48550/arXiv.2103.10895. URL `http://arxiv.org/abs/2103.10895`.

S. Hayes and S. O'Neill. The greta effect: Visualising climate protest in uk media and the getty images collections. *Global Environmental Change*, 71:102392, 2021.

M. Johann, L. Höhnle, and J. Dombrowski. Fridays for future and mondays for memes: How climate crisis memes mobilize social media users. *Media and Communication*, 11(3), 2023.

W. Kay, J. Carreira, K. Simonyan, B. Zhang, C. Hillier, S. Vijayanarasimhan, F. Viola, T. Green, T. Back, P. Natsev, M. Suleyman, and A. Zisserman. The kinetics human action video dataset. *arXiv preprint arXiv:1705.06950*, May 2017. URL `http://arxiv.org/abs/1705.06950`. arXiv:1705.06950 [cs].

K. Krippendorff. *Content analysis: An introduction to its methodology*. Sage publications, 2018.

A. Krizhevsky, G. Hinton, et al. Learning multiple layers of features from tiny images. 2009.

R. J. Light. Measures of response agreement for qualitative data: some generalizations and alternatives. *Psychological bulletin*, 76(5):365, 1971.

T.-Y. Lin, M. Maire, S. Belongie, L. Bourdev, R. Girshick, J. Hays, P. Perona, D. Ramanan, C. L. Zitnick, and P. Dollár. Microsoft coco: Common objects in context. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13*. arXiv, Feb 2015. doi: 10.48550/arXiv.1405.0312. URL `http://arxiv.org/abs/1405.0312`. arXiv:1405.0312 [cs].

Z. Liu, P. Luo, X. Wang, and X. Tang. Deep learning face attributes in the wild. In *Proceedings of International Conference on Computer Vision (ICCV)*, December 2015.

A. Mooseder, C. Brantner, R. Zamith, and J. Pfeffer. (social) media logics and visualizing climate change: 10 years of# climatechange images on twitter. *Social Media+ Society*, 9(1):20563051231164310, 2023.

S. O'Neill. More than meets the eye: A longitudinal analysis of climate change imagery in the print media. *Climatic Change*, 163(1):9–26, 2020.

A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, G. Krueger, and I. Sutskever. Learning transferable visual models from natural language supervision, 2021.

S. Rebich-Hespanha, R. E. Rice, D. R. Montello, S. Retzloff, S. Tien, and J. P. Hespanha. Image themes and frames in us print news stories about climate change. *Environmental Communication*, 9(4):491–519, 2015.

L. Smaira, J. Carreira, E. Noland, E. Clancy, A. Wu, and A. Zisserman. A short note on the kinetics-700-2020 human action dataset. *arXiv preprint arXiv:2010.10864*, Oct 2020. URL `http://arxiv.org/abs/2010.10864`. arXiv:2010.10864 [cs].

G. Sumbul, M. Charfuelan, B. Demir, and V. Markl. Bigearthnet: A large-scale benchmark archive for remote sensing image understanding. In *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*, page 5901–5904, Jul 2019. doi: 10.1109/IGARSS.2019.8900532. URL `http://arxiv.org/abs/1902.06148`. arXiv:1902.06148 [cs].

S. Wang, A. Corner, D. Chapman, and E. Markowitz. Public engagement with climate imagery in a changing digital landscape. *Wiley Interdisciplinary Reviews: Climate Change*, 9(2):e509, 2018.

I. Weaver, N. Westwood, T. Coan, S. O'Neill, and H. T. Williams. Sponsored messaging about climate change on facebook: Actors, content, frames. *arXiv preprint arXiv:2211.13965*, 2022.

K. Zhou, J. Yang, C. C. Loy, and Z. Liu. Conditional prompt learning for vision-language models. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022.

Supplementary Material

A Datasets

A.1 Licence

ClimateTV and *ClimateCT* consist of a list of links to images that users have publicly shared on Twitter. Users grant Twitter "a worldwide, non-exclusive, royalty-free license (with the right to sublicense) to use, copy, reproduce, process, adapt, modify, publish, transmit, display and distribute their content in any and all media or distribution methods now known or later developed."Corp. [b]. Content here is defined as the tweet including any audio, image, or video contained in it. Users explicitly allow Twitter to redistribute their content to third parties. According to the Developer Agreement, researchers are allowed to use the data for non-illegal research excluding surveillance, monitoring sensitive events, or profiling Corp. [a]. In providing the links to the dataset, we accommodate users' changes in privacy settings following the example of the Kinect Datasets Kay et al. [2017], Smaira et al. [2020]. For maximum transparency, the image links can be found on our github (published upon acceptance).

A.2 Annotation process

The image annotation made use of the open-source platform QCAMAP (`https://www.qcamap.org/`) Fenzl and Mayring [2017]. Annotation guidelines were provided to all annotators to explain the process and give details regarding the various class labels. These guidelines are supplied as part of the supplementary material. In total there were three rounds of annotation: The first confirmed the choice of super-categories and class labels. The second round extended the number of annotated images for less frequent classes. Any missing images in order to have a sufficient amount of training and testing data for 16-shot learning were annotated in round three.

The quality of the annotation guideline was confirmed by the high level of agreement between the annotators, as Table 3 indicates.

Category	Kappa	Krippendorff	Agreement
Animals	88.7	88.7	92.3
Climate action	85.8	85.8	97.2
Consequences	84.1	84.0	89.2
Setting	87.5	87.5	88.9
Type	88.7	88.7	92.3

Table 3: Annotation metrics for each of the higher-level categories.

We measured the overlap between class labels assigned by different annotators using social science metrics Light [1971], Krippendorff [2018]. We hope that comparable levels of accuracy can be achieved by computer vision models.

B ClimateCT: Analysis and Baseline Results

We use CLIP Radford et al. [2021] to make zero-shot predictions for both *ClimateTV* and *ClimateCT*. Further, we perform query optimization using CoCoOp Zhou et al. [2022] in a 16-shot learning setting and evaluate its effect on classification accuracy. We report results for the other categories here. Table 4 gives an overview of accuracy per category.

Super-Category	Accuracy incl. None cat.	Accuracy excl. None cat.
Animals	79.00	64.68
Climate action	21.77	46.95
Consequences	28.90	40.51
Setting	34.10	26.04
Type	51.64	n/a

Table 4: CLIP classification accuracy varies between super-categories, reported with (left) and without *None*-category (right). Super-categories with a distinct *None*-category benefit from having it, while general ones decrease the overall model performance.

B.1 Animals

The classes, keywords, and queries used for the super-category *animals* are shown in Table 5. animals are generally expected to be easily detected, however, since our data contains various types of images, their correct classification is not trivial. For example, besides the wild polar bear, also images containing a polar bear costume are labeled as polar bear. Furthermore, the varying granularity and specificity of classes is a novelty that we provide.

#	Class	Keywords	Query
1	No animals	No animals	A photo of no animals.
2	Pets	Pets	A photo of animals of type pet.
3	Farm animals	farm animals	A photo of animals of type farm animal.
4	Polar bear	Polar bear	A photo of animals of type polar bear.
5	Land mammals excl. classes 2-4	Land mammal	A photo of animals of type land mammal.
6	Sea mammals	Sea mammal	A photo of animals of type sea mammal.
7	Fish excl. class 2	Fish	A photo of animals of type fish.
8	Amphibian or Reptile excl. class 2	Amphibian, Reptile	A photo of animals of type amphibian. A photo of animals of type reptile.
9	Insects excl. class 2	Insects	A photo of animals of type insects.
10	Birds excl. class 2	Birds	A photo of animals of type bird.
11	animals excl. animal classes 2 - 10	Animals	A photo of animals.

Table 5: Classes, keywords, and queries used for the category Animals.

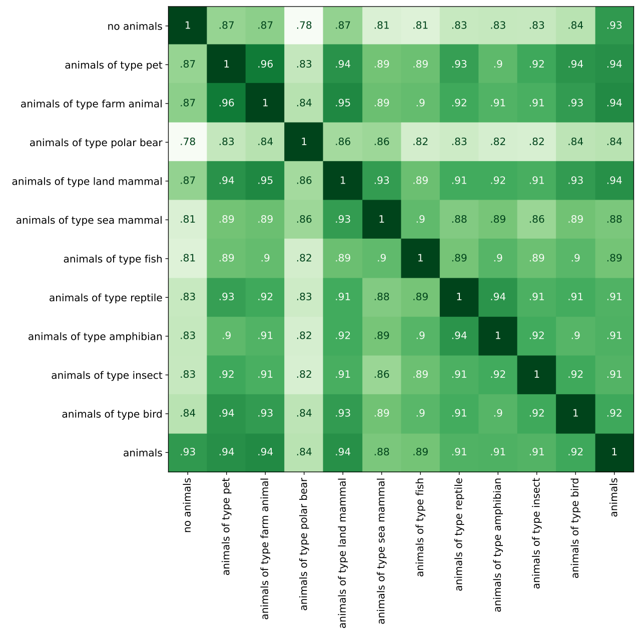

Figure 2: Query cosine similarity matrix for all queries regarding animals.

The query similarity matrix shows that the polar bear query is most distinct from the other queries in Figure 2. pets and farm animals have very similar embeddings, even though they can have very different semantics. The confusion matrix shows less confusion between the two classes than expected from the similarity of their queries. Most outstanding is the fact that polar bears are more frequently misclassified than expected. Figure 3 shows that the "None"-class is most often predicted, possibly due to the various forms of representations of the animals in *ClimateCT*.

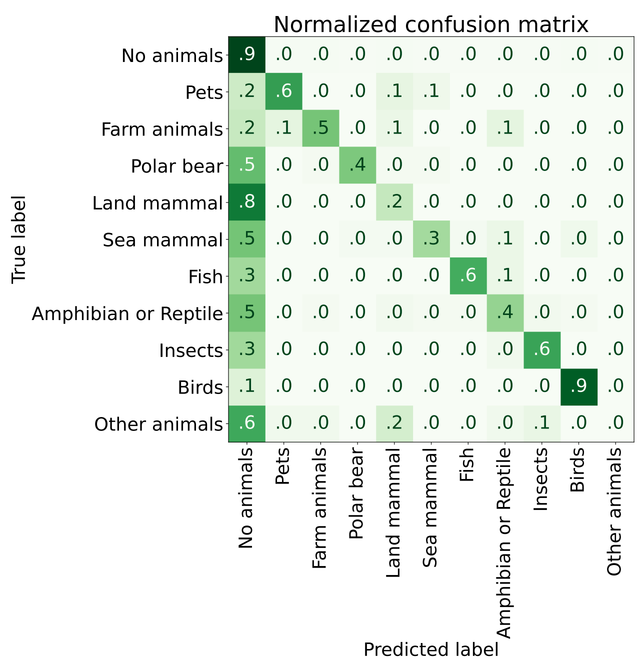

Figure 3: Confusion matrix for animals including all classes.

When removing the “None"-class, the confusion decreases. Solely insects are frequently misclassified as amphibian or reptile, and all other classes have a high true-positive rate as shown in Figure 4.

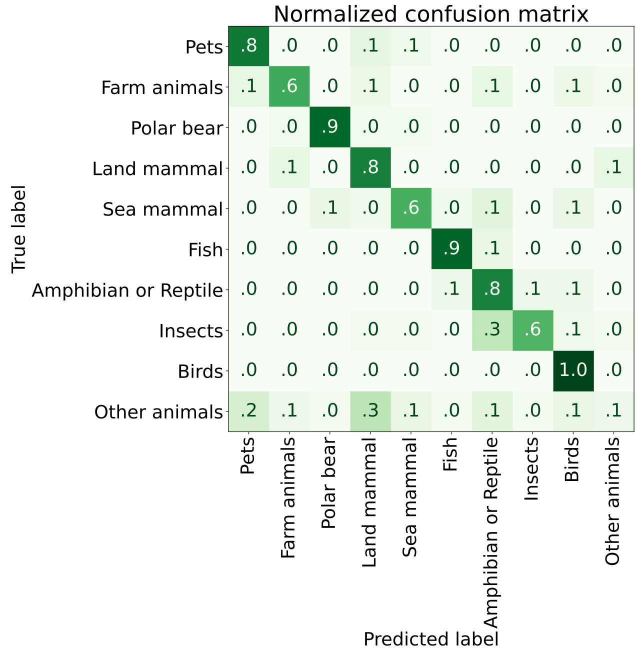

Figure 4: Confusion matrix of animals for all classes except for no animals.

B.2 Climate Action

The classes, keywords, and query for the super-category climate action are displayed in Table 6. This category, along with setting, and consequences, is one of the analytical super-categories, and is thus expected to be more difficult than a descriptive category.

#	Class	Keywords	Query
1	No climate action	No climate action	A photo of no climate action
2	Protests	Protests	A photo of protests.
3	Politics	Politics	A photo of politics.
4	Sustainable energy	Sustainable / wind / solar energy, hydropower, biogas	A photo of sustainable energy. A photo of wind energy. A photo of solar energy. A photo of hydropower. A photo of biogas.
5	Fossil Energy	Fossil / carbon energy, natural gas, oil, fossil fuel	A photo of fossil energy. A photo of carbon energy. A photo of natural gas. A photo of oil. A photo of fossil fuel.
6	Other excl. classes 2-5	Climate action	A photo of climate action.

Table 6: Classes, keywords, and queries used for the category Climate action.

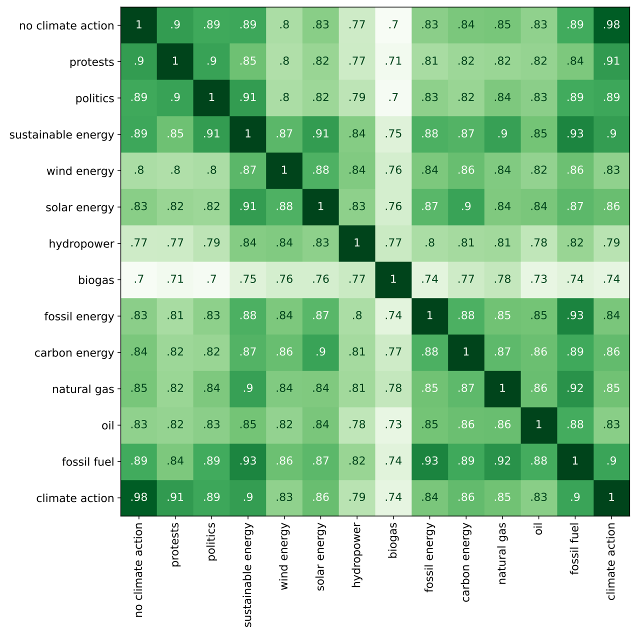

Figure 5: Query cosine similarity matrix for all queries regarding climate action.

The query similarity matrix indicates that the term *Biogas* has the most distinct query overall. Due to the mention of “gas" in it, it is most similar to *natural gas*. For all other queries related to energy, the similarity is given independent of the sustainability of the energy source. This is true for all sustainable sources of energy except for *hydropower*, which is more similar to the other formats of sustainable energy. fossil energy sources are generally more or equally similar to other fossil energy sources than sustainable energy sources. Politics and protest are also very similar to each other, as shown in Figure 5.

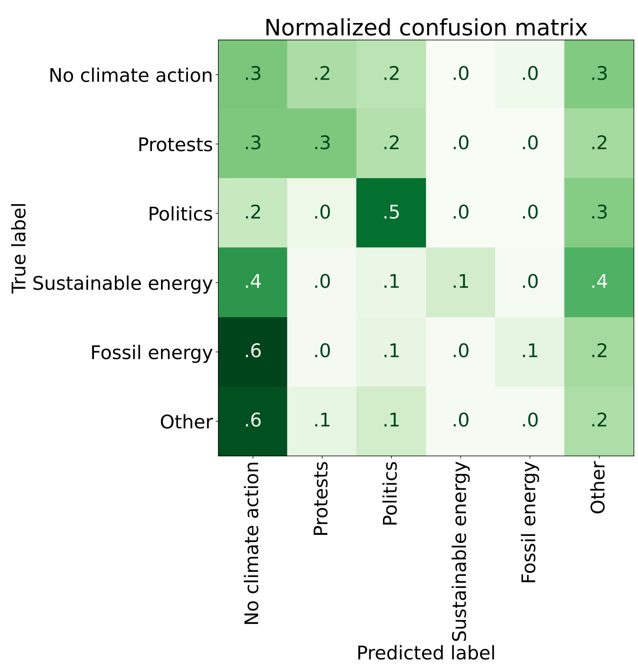

Figure 6: Confusion matrix of climate action including all classes shows frequent predictions of the “None"-class.

Especially the classes sustainable energy and fossil energy are very hard to classify correctly, as shown in Figure 7. While sustainable energy is most often classified as other, fossil energy is most often classified as sustainable energy. The classes politics and protest have better true-positive rates and are both mostly confused with other climate action.

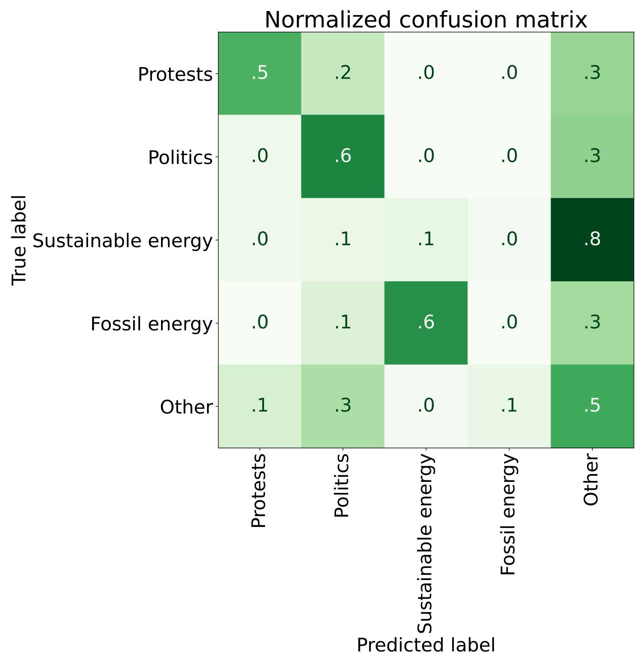

Figure 7: Confusion matrix of climate action for all classes except for no climate action can best classify politics.

B.3 Setting

The class labels, keywords, and queries used for this super-category are displayed in Table 7. This super-category was rather difficult to classify using zero-shot prediction with CLIP embeddings, given the nature of the pre-training. Image captions generally focus on the foreground rather than the background, except for the cases when the background represents a contrast to the foreground, i.e. *"My field is flooded"* or *"Our neighbour's stable is on fire"*.

#	Class	Keywords
1	No setting	–
2	Residential/commercial area (Outdoor/Outside)	Residential/commercial area
3	Industrial area	Industrial area
4	Agricultural/rural area	Agricultural/rural area
5	Indoor space excl. classes 3-4	Indoor space, room
6	Arctic or Antarctica	Arctic, Antarctica
7	Ocean, coastal	Ocean, coastal
8	Desert	Desert
9	Forest, jungle	Forest, jungle
10	Nature excl. classes 4,6-9	Nature
11	Outer space	Outer space
12	Other setting excl. classes 2-11	Area

Table 7: Classes and keywords used for the category Setting. QUERY for obtaining photos is "A photo of a {*Keyword*}".

While some categories have a high level of accuracy, e.g. residential/commercial, outer space, other categories have very low levels of accuracy, e.g. agricultural, other nature, or other space, as Table 8 indicates. Said categories are arguably highly diverse, as agricultural includes both factory farming and biological farming for both animals and crops. The diversity of other nature and other space is naturally given by their definition.

Category (# images)	Accuracy incl. None cat.	Accuracy excl. None cat.
Overall (1,038)	34.10	26.04
No setting (193)	20.00	n/a
Residential / Commercial area (142)	82.61	90.00
Industrial area (38)	75.76	51.02
Agricultural /rural area (34)	28.87	26.32
Indoor space (224)	70.73	21.25
Arctic / Antarctica (36)	19.66	12.89
Ocean / coastal (76)	53.62	54.84
Desert (35)	45.10	34.21
Forest / Jungle (56)	35.71	100.00
Other nature (78)	35.21	19.63
Outer space (22)	65.38	60.71
Other setting (104)	14.52	13.98

Table 8: The super-category Setting has a low overall accuracy. residential/commercial's and outer space's accuracies stand out, which can be explained by their distinct nature. The accuracy of forest / jungle experiences the largest increase when the no setting class is removed.

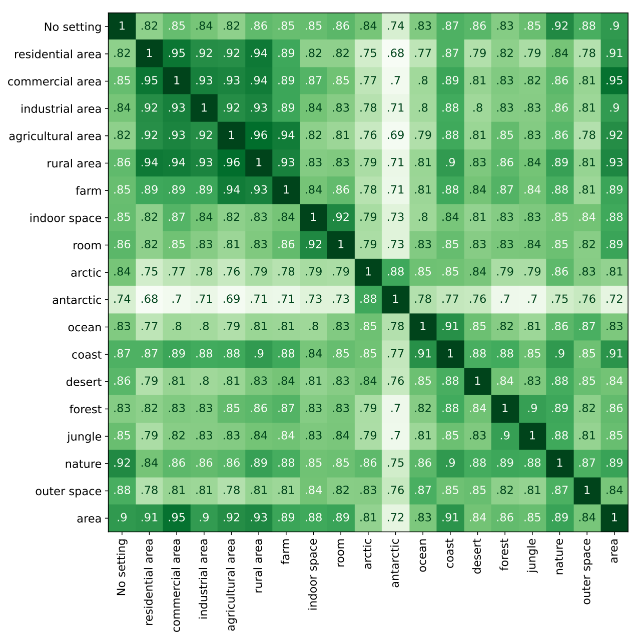

Figure 8: Query cosine similarity matrix for all queries regarding setting indicates high similarity between the human-influenced settings e.g. *residential*, *commercial*, *industrial*, etc.

When comparing CLIP queries' similarities, displayed in Figure 8, it becomes apparent that *residential area*, *commercial area*, *industrial area*, *agricultural area*, *rural area*, and *farm* have very similar embeddings. While the cosine similarities within each category residential/commercial (*residential area*, *commercial area*), agricultural/rural (*agricultural area*,*rural area*, *farm*), and indoor space (*indoor space*, *room*) are large, the similarity to other classes are not much smaller. Moreover, *Antarctica* has the most distinct query, being closest to *Arctic*. In this super-category, the no setting is rather distinct from other queries. We report the confusion matrices (incl./ excl. None), which give further insights into the deterioration of the accuracy score when the 'None' category is removed.

This super-category is very challenging due to its visual similarities. Snow, which can typically be found in the Arctic or Antarctic, might as well be present in an image depicting a residential area. A large challenge for computer vision models is to detect which part of the environment is distinct for each category and which visual features are independent of them.

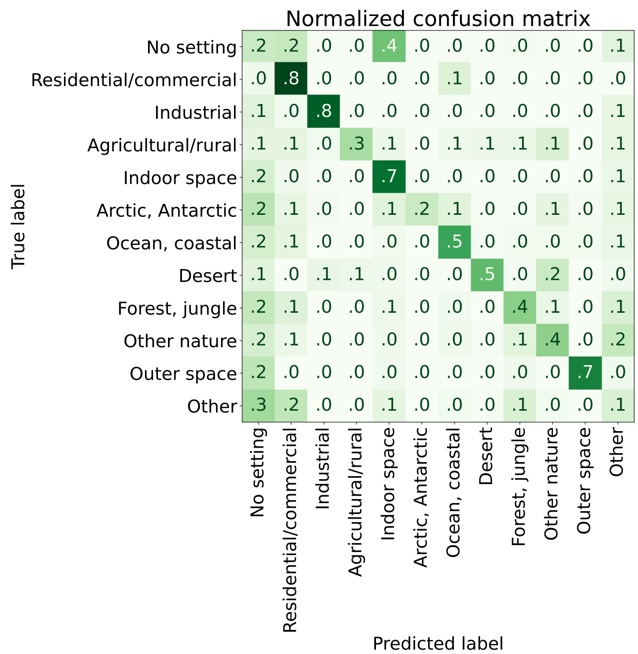

Figure 9: Confusion matrix of setting including all categories displays most confusion in the no setting class.

The complete confusion matrix, shown in Figure 9, indicates that all classes are mainly confused with the no setting or other setting. Both classes also have a low accuracy, as they themselves are often confused with the other classes. While no setting is mainly confused with indoor space, other setting is mainly confused with no setting and residentialcommercial. It is interesting to note that even though the queries for residential/commercial and industrial are very similar, they are rarely confused with each other. Comparing the two confusion matrices in Figure 9 and Figure 10, the category gorest, jungle is strongly increasing its accuracy. Residential/commercial also increases its accuracy and becomes a very common category for model predictions. It appears that many image embeddings are most similar to the text embeddings of this category. Without the 'None'-class, the accuracies deteriorate, as shown in Figure 10.

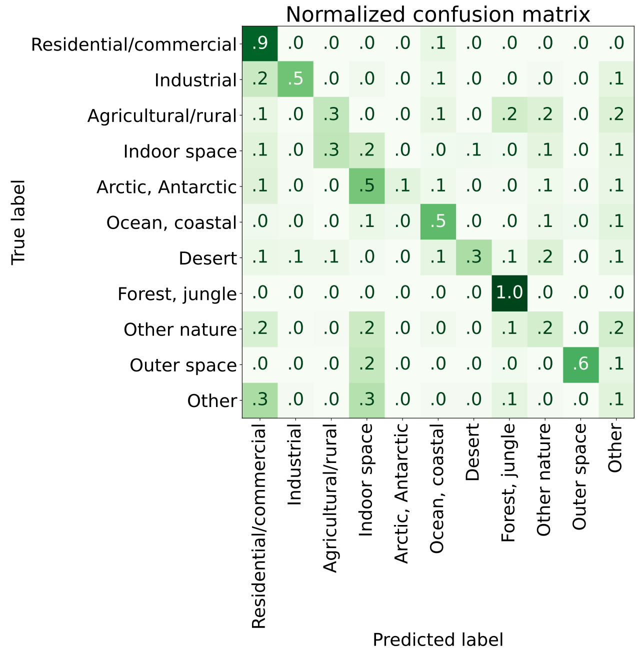

Figure 10: Confusion matrix of setting for all classes except for no setting shows as deterioration of model performance when the 'None'-class is removed.

B.4 Type

Table 9 provides an overview of which classes, keywords, and CLIP queries are used for the following analysis. For this super-category, there is no "None"-class as every image is of a type.

#	Class	Keyword(s)	Query
1	Posters/Event invitations	Posters/Event invitations	A poster of climate change. A event invitation of climate change.
2	Meme excl. class 1	Meme	A meme of climate change.
3	Infographic excl. class 2	Infographic	A infographic of climate change.
4	Data Visualization excl. classes 2-3	Data Visualization	A data visualization of climate change.
5	Illustration excl. classes 1-4	Illustration, drawing, cartoon	A illustration of climate change. A drawing of climate change. A cartoon of climate change.
6	Screenshot excl. classes 1-5	Screenshot	A screenshot of climate change.
7	Single Photo excl. classes 1-6	Photo	A single photo of climate change.
8	Two or more Photos excl. classes 1-7	Photo Collage	A photo collage of climate change.
9	Other	–	A climate change.

Table 9: Classes, keywords, and queries used for the category Type.

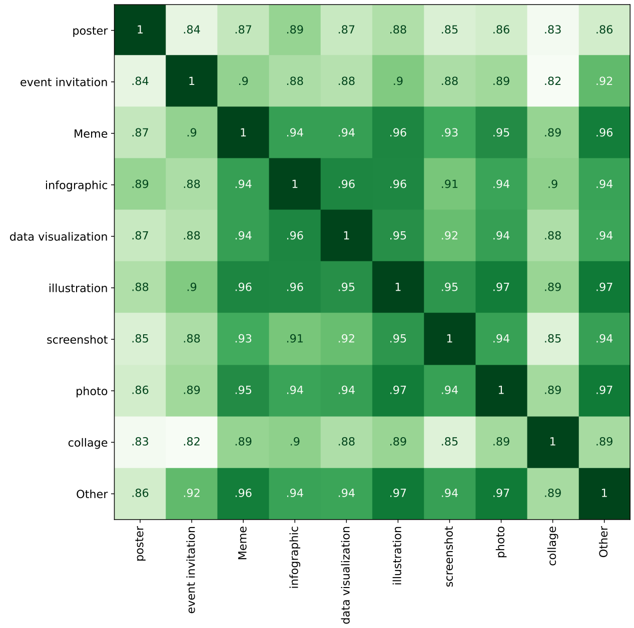

Figure 11: Query cosine similarity matrix for all queries regarding type indicates that infographics are hard to distinguish from other classes.

The query similarity matrix in Figure 11 indicates that the queries regarding type share many similarities. Infographic, data visualization, and illustration are very similar, which can be explained by the common elements they share. The classes posters, event invitation, and collage are most distinct from other classes. Even though photos and collages both consist of photos, their embeddings differ. The other type class is, as expected quite similar to the other classes.

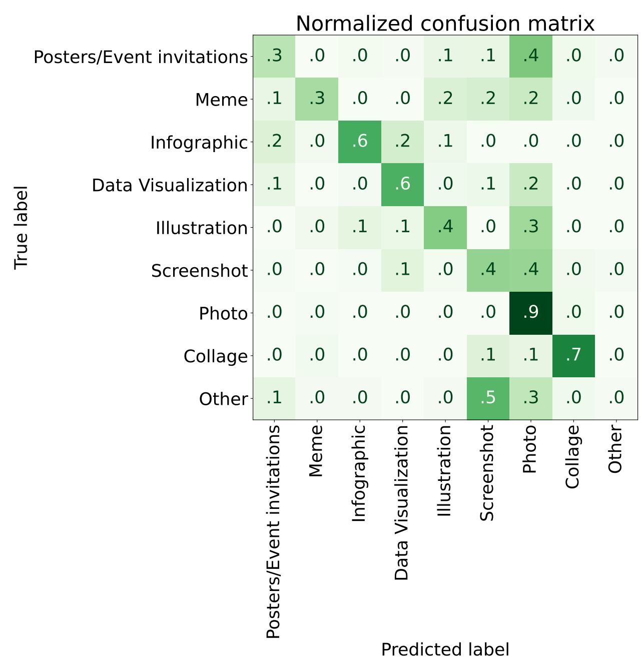

Figure 12: Confusion matrix of type including all classes shows the high true positive rate for images of type photo.

The confusion matrix indicates that photos can be detected really well, shown in Figure 12. Simultaneously, photo is the class, with which the most wrongly classified images are confused. The class posters/event invitations is even more often wrongly classified as photo than it is correctly classified. screenshot is classified equally often as photo and its true class.

C *ClimateTV*: Sample Images

In Figure 13 we share example images of the *ClimateTV* dataset which exemplify the challenging nature of this dataset.

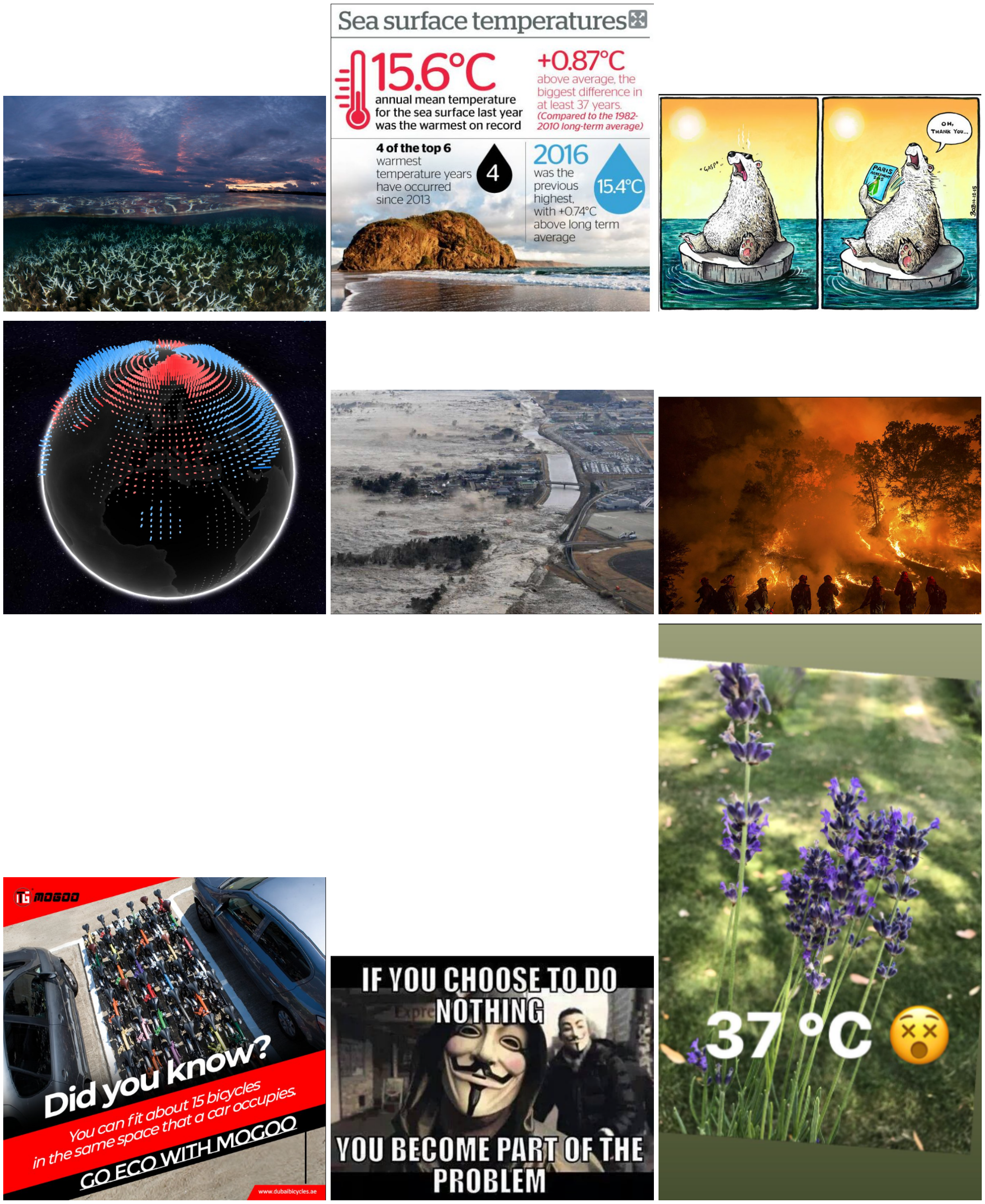

Figure 13: *ClimateTV* is a highly diverse dataset with hashtag-based labels.

D Annotation Guidelines

Software link: https://www.qcamap.org

Time frame
2019

Case selection
Download of all tweets with images & "climate change", "climatechange", "#climatechange"

General rules

- Class labels should be assigned as specific as possible.
- If no specific class label applies, assign the next broader category.
- When more than one class label applies to the image, apply to more suitable one e.g., an image of a burning forest with the text “save the animals” should be labelled as “Biodiversity loss”.

Type

What kind of image is it? Rule: When multiple types of images are present, code the predominant one.

Class labels:

- Single Photo: Photograph, can be modified e.g., include text
- Photo Collage: more than one photo
- Illustration: e.g., drawing, painting, cartoon
- Data visualization
- Infographic: Collection of imagery, data visualizations like pie charts and bar graphs, and minimal text that gives an easy-to-understand overview of a topic, not news articles
- Screenshot: e.g., news, scientific article, tweet
- Meme: photo/illustration + text, memes can be present online in multiple variations. They are playfully adapted to various contexts to generate virality and participation.
- Poster / event invitation
- Other: all images with no clear association to the aforementioned classes.

Setting

What kind of physical space is shown in the image?
Rule: When multiple settings are present, code the predominant one.

Class labels:
- Ocean, coastal
- Desert
- Forest, jungle
- Arctic and Antarctic: snow and ice in a natural setting.
- Other Nature: e.g., lake, mountainous, savannah, tundra, taiga, etc.
- Agricultural/rural
- Residential/commercial: includes governmental buildings, airports.
- Industrial area
- Indoor space: not industrial or agricultural
- Outer space: earth from space
- Other setting
- No setting: all images with no clear association to the aforementioned classes.

Animals

Does the image represent an animal?

Class labels:
- Pets, e.g., cats, dogs
- Farming animals, e.g., cows, horses
- Land Mammal: except Polar bears, pets, or farm animals.
- Polar bear
- Sea mammals
- Fish
- Insects
- Birds
- Amphibian, Reptile
- Other animals: excluding above classes.
- No animals: all images with no clear association to the aforementioned classes.

Consequences

Which consequences of climate change does the image represent?

Consequences refers to the effects that climate change is having on people and the environment. They usually refer to things that are happening in the present and/or will happen in the future.

Class labels:
- Floods
- Drought
- Wildfires
- Extreme heat
- Other extreme weather events: e.g., hurricanes, tornadoes
- Melting Ice
- Sea level rise
- Human rights: e.g., war, migration, famine
- Economic consequences
- Biodiversity loss
- Covid & general health
- Other consequences
- No consequences: all images with no clear association to the aforementioned classes.

Climate Action

Does the image depict climate change related actions or engagement?

Class labels:
- Politics: any political events or conferences e.g., COP
- Protests: any protest e.g., Friday for Future
- Sustainable energy: all forms of sustainable energy, e.g., biogas, wind energy, solar panels, etc.
- Fossil energy: all forms of fossil energy, e.g., natural gas, coal, oil, etc.
- Other climate action: all forms of activities that relate to climate change, e.g., recycling.
- No climate action: all images with no clear association to the aforementioned classes.