Semi-supervised Analysis of Human fMRI Data

J. A. Shelton[1,3], M. B. Blaschko[2], C. H. Lampert[1], and A. Bartels[1,3]

[1]Max Planck Institute for Biological Cybernetics, [2]University of Oxford, [3]Universität Tübingen

Motivation

- **Analyze brain activity in natural, complex setting, to assess natural processing**
- **Problem: natural stimuli data need labels - expensive and time consuming**

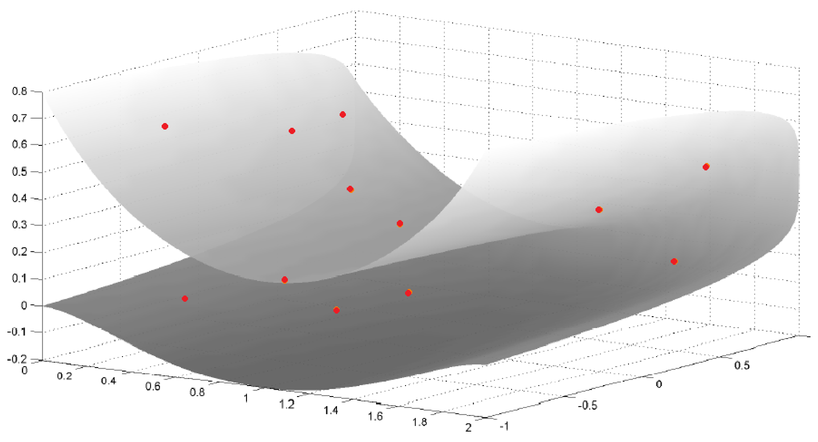

- **Goal: use unlabeled data and labeled data (semi-supervised) for dimensionality reduction and to better approximate cortical activity**

1 Methods and Materials

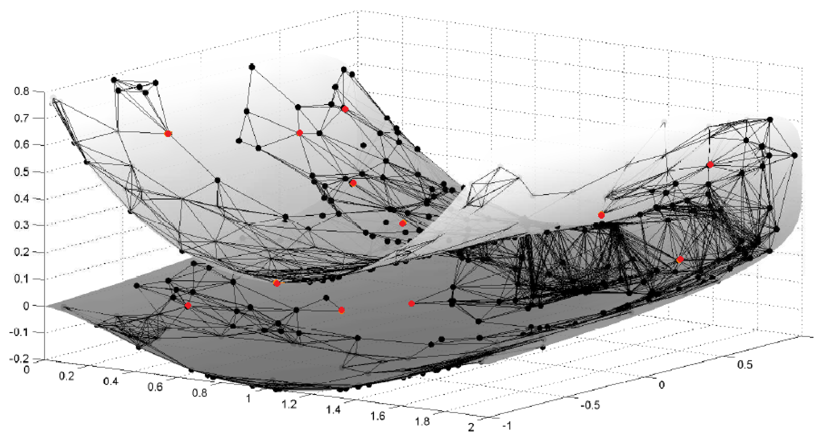
- fMRI data of one human volunteer during viewing of 2 movies.
- 350 time slices of 3-dimensional fMRI brain volumes acquired with Siemens 3T TIM scanner, separated by 3.2 s (TR), with a spatial resolution of 3x3x3 mm.
- Pre-processed according to standard procedures using the Statistical Parametric Mapping (SPM) toolbox [6].
- **Labels:** Continuous labels of one movie obtained via subjective ratings averaged across an independent set of five human observers [2]:
  - Human faces
  - Color
  - Human bodies
  - Language
  - Motion

2 Semi-Supervised Laplacian Regularization of Canonical Correlation Analysis

- Labeled fMRI data: $\{x_1, \ldots, x_n\}$.
- Corresponding labels: $\{y_1, \ldots, y_n\}$.
- Paired data (fMRI with labels): $(x_1, y_1), \ldots, (x_n, y_n)$.
- Additional unlabeled data: $x_{n+1}, \ldots, x_{n+p_x}$.
- Data matrices: $X = (x_1, \ldots, x_n)^T$, $\quad Y = (y_1, \ldots, y_n)^T$, $\hat{X} = (x_1, \ldots, x_{n+p_x})^T$.
- Graph Laplacians [3]: $\mathcal{L}_{\hat{x}} = I - D_{\hat{x}\hat{x}}^{-1/2} K_{\hat{x}\hat{x}} D_{\hat{x}\hat{x}}^{-1/2}$ for $(K_{\hat{x}\hat{x}})_{ij} = \exp\left(\frac{-\|x_i - x_j\|^2}{\sigma^2}\right)$ and diagonal $(D_{\hat{x}\hat{x}})_{ii} = \sum_{j=1}^{n+p_x} (K_{\hat{x}\hat{x}})_{ij}$.



(a) Labeled data



(b) Labeled and unlabeled data

Semi-Supervised CCA [4]

- **Solve (e.g. as generalized eigenproblem):**

$$\max_{w_x, w_y} \frac{w_x^T C_{xy} w_y}{\sqrt{(w_x^T (C_{xx} + R_{\hat{x}}) w_x)(w_y^T (C_{yy} + R_y) w_y)}} \qquad (1)$$

with regularizers

$$R_{\hat{x}} = \epsilon_x I + \gamma_x \hat{X} \mathcal{L}_{\hat{x}} \hat{X}^T \text{ and } R_y = \epsilon_y I + \gamma_y Y \mathcal{L}_y Y^T \qquad (2)$$

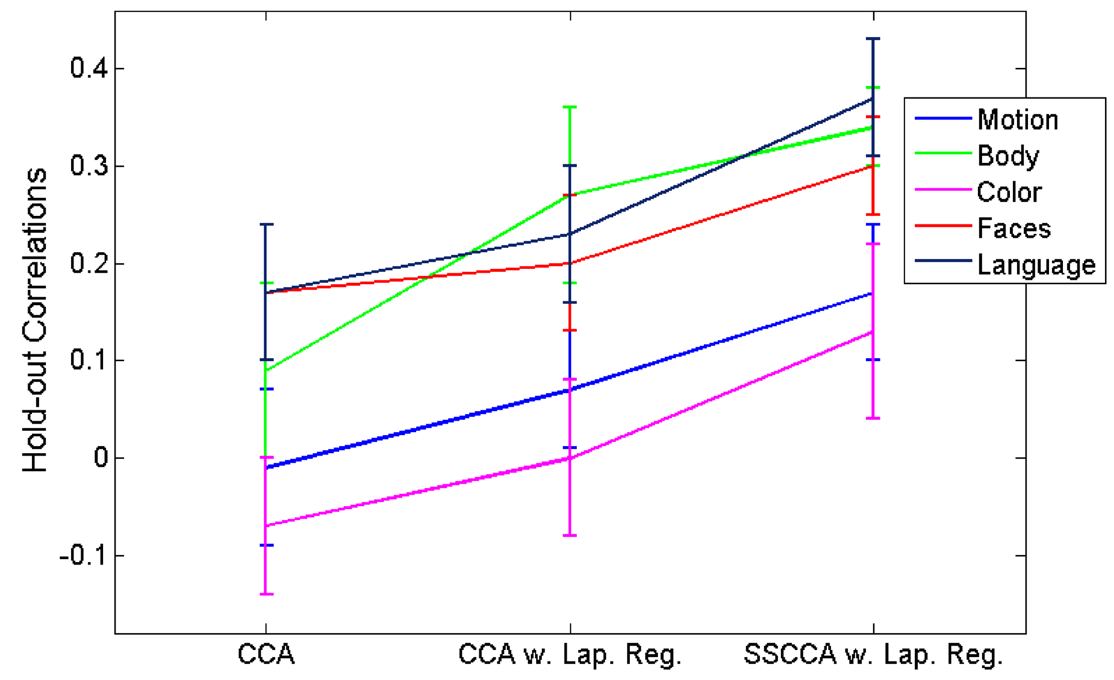
- **Finds projections that are smooth with respect to manifold structures of $\hat{X}, Y$ instead of ambient spaces.**

3 Results

Experiments:

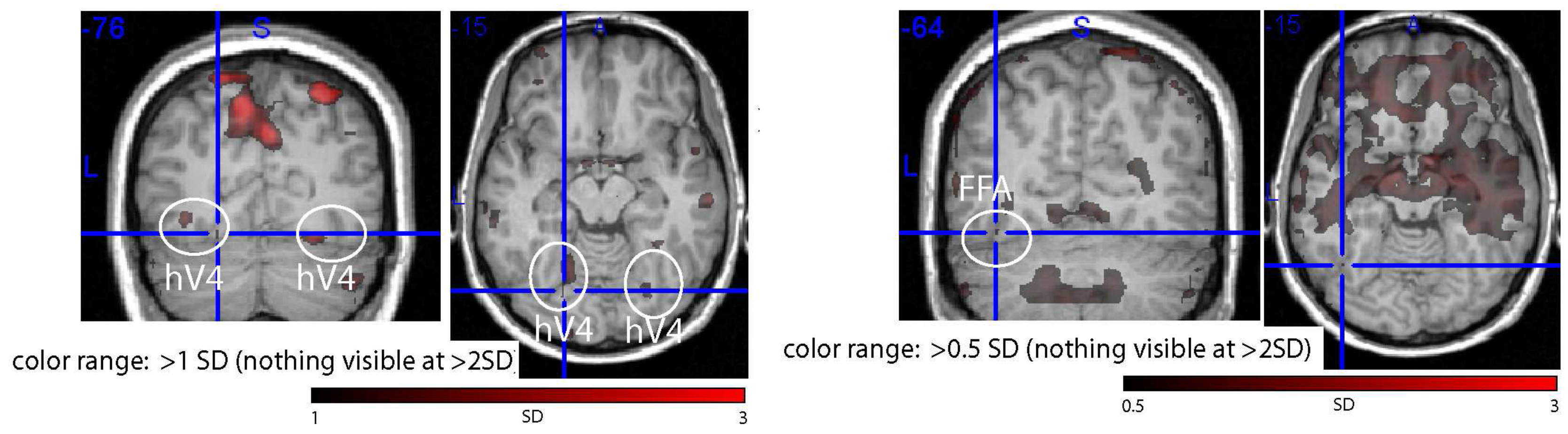
- CCA with only Tikhonov regularization - labeled data only
- CCA with Tikhonov and Laplacian regularization - labeled data only
- Semi-supervised CCA with Tikhonov and Laplacian regularization - labeled and unlabeled data

Mean holdout correlations from five-fold cross validation across [each of the five] variables in all experiments.

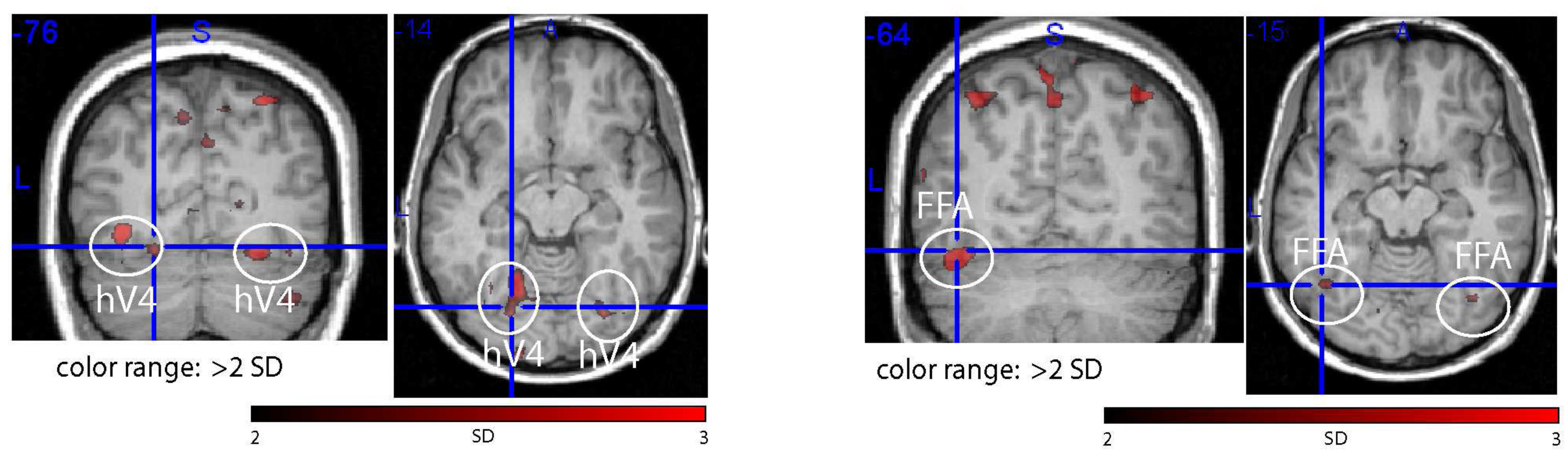


Visualization of learned weight vectors (w_x) for color and face stimuli, following [2].

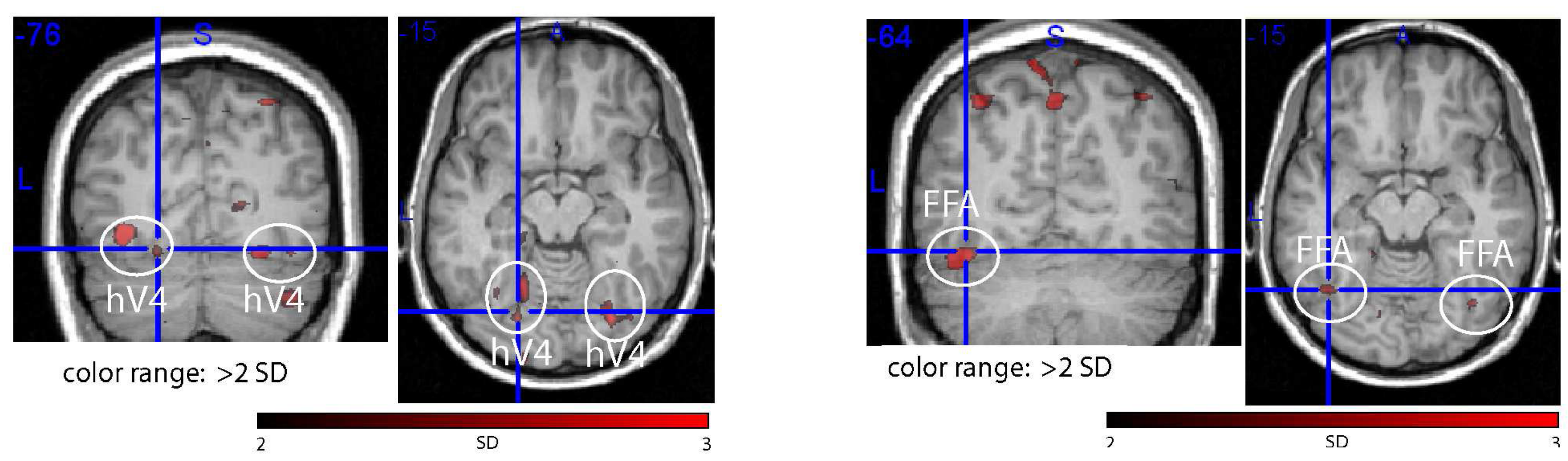
Color | Faces



(a) CCA, Tikhonov regularization



(b) CCA, Tikhonov and Laplacian regularization



(c) Semi-supervised CCA, Tikhonov and Laplacian regularization

4 Conclusions

- Semi-supervised Laplacian regularization framework consisently improves performance of dimensionality reduction [7, 8, 9]
- Weights learned by (semi-supervised) CCA identify expected regions of cortical activity [2]
- Semi-supervised learning allows the integration of unlabeled data in supervised learning to improve results
- Current and future work [9]:
  - Across-subjects comparisons
  - Unlabeled data acquired during resting state

References

[1] Bartels, A., Zeki, S., and Logothetis, N. K. (2008). *Natural vision reveals regional specialization to local motion and to contrast-invariant*, global flow in the human brain. Cereb Cortex 18:705-717.

[2] Bartels, A., Zeki, S. (2004). *The chronoarchitecture of the human brain - natural viewing conditions reveal a time-based anatomy of the brain.* NeuroImage 22:419-433.

[3] Belkin, M., Niyogi, P., Sindhwani, V.: *Manifold Regularization: A Geometric Framework for Learning from Labeled and Unlabeled Examples.* JMLR (2006)

[4] Blaschko, M.B., Lampert, C.H., Gretton, A. (2008). *Semi-supervised Laplacian Regularization of Kernel Canonical Correlation Analysis.* ECML

[5] Hardoon, D. R., S. Szedmak and J. Shawe-Taylor. (2004). *"Canonical Correlation Analysis: An Overview with Application to Learning Methods,"* Neural Computation, 16, (12), 2639-2664.

[6] Friston, K., Ashburner, J., Kiebel, S., Nichols, T., Penny, W. (Eds.) *Statistical Parametric Mapping: The Analysis of Functional Brain Images*, Academic Press (2007)

[7] Shelton, J., Blaschko, M., and Bartels, A. (2008). *Semi-supervised subspace analysis of human functional magnetic resonance imaging data*, Max Planck Institute Tech Report, (185) (05 2009)

[8] Blaschko, M., Shelton, J., Bartels, A., Lampert, C., H., and Gretton, A. (submitted) *Semi-supervised Kernel Canonical Correlation Analysis with Application to human fMRI.* Neurocomputing.

[9] Blaschko, M., Shelton, J., and Bartels, A. (submitted) *Augmenting Feature-driven fMRI Analyses: Semi-supervised learning and resting state activity.* NIPS.