



Technical Report No. 185

**Semi-supervised subspace  
analysis of human functional  
magnetic resonance imaging  
data**

Jacquelyn Shelton,<sup>1</sup> Matthew Blaschko,<sup>1</sup>  
Andreas Bartels<sup>2</sup>

May 2009

<sup>1</sup> Department Schölkopf, email: [jshelton;blaschko@tuebingen.mpg.de](mailto:jshelton;blaschko@tuebingen.mpg.de)

<sup>2</sup> Department Logothetis, email: [abartels@tuebingen.mpg.de](mailto:abartels@tuebingen.mpg.de)

# Semi-supervised subspace analysis of human functional magnetic resonance imaging data

*Jacquelyn Shelton, Matthew Blaschko, Andreas Bartels*

**Abstract.** Kernel Canonical Correlation Analysis is a very general technique for subspace learning that incorporates PCA and LDA as special cases. Functional magnetic resonance imaging (fMRI) acquired data is naturally amenable to these techniques as data are well aligned. fMRI data of the human brain is a particularly interesting candidate. In this study we implemented various supervised and semi-supervised versions of KCCA on human fMRI data, with regression to single- and multi-variate labels (corresponding to video content subjects viewed during the image acquisition). In each variate condition, the semi-supervised variants of KCCA performed better than the supervised variants, including a supervised variant with Laplacian regularization. We additionally analyze the weights learned by the regression in order to infer brain regions that are important to different types of visual processing.

---

## 1 Introduction/Motivation

Subspace methods have been shown to yield good performance in modeling visual data, wherein various statistical methods are applied to aligned sets of images. In this work we are interested in modeling visual data by means of variants of kernel canonical correlation analysis (KCCA), namely via (semi-supervised) Laplacian regularization (Blaschko, Lampert, Gretton, 2008). As KCCA contains principal components analysis (PCA), Fisher Linear Discriminant Analysis (LDA), and ridge regression as special cases, this framework allows us to incorporate a very rich family of regression techniques. Members of this family have been applied to face data in the past (Belhumeur, Hespanha, & Kriegman, 1996; Turk & Pentland, 1991), yielding interesting representations of the most prominent features of those faces (i.e. 'eigenfaces' and 'Fisherfaces'). When modeling visual data in this fashion, the data must meet the criterion of being reasonably well-aligned, otherwise translational biases are introduced. For this reason, images obtained via functional magnetic resonance imaging (fMRI), such as of human brains, are naturally suited to these techniques. Such fMRI data is obtained via an MRI scanner, which holds the subjects' heads immobile, thereby acquiring reasonably well-aligned data from the start. The data can be then further aligned using a suite of techniques, Statistical Parametric Mapping (SPM), developed for neuroscience image analysis. This data processing pipeline is automatic and can be assumed to be available at no additional cost, making fMRI data a perfect candidate for the application of KCCA and related techniques.

In neuroscience there has been a recent surge of interest in analyzing brain activity in more natural, complex settings, e.g. with volunteers viewing movies, in order to gain insight in brain processes and connectivity underlying more natural processing. The problem has been approached from different routes: linear regression was used to identify brain areas correlating with particular labels in the movie (Bartels, Zeki, 2004), the perceived content was inferred based on brain activity (Hasson et al., 2004), data-driven methods were used to subdivide the brain into units with distinct response profiles (Bartels, Zeki, 2005), and correlation across subjects was used to infer stimulus-driven brain processes at different timescales (Hasson et al., 2008). The current approach would allow one to broaden this repertoire to achieve exciting applications. For example, the current approach may allow one to obtain brain maps associated to a new or unknown stimulus for which no labels are available, yielding a variant of known 'brain reading' techniques. Applied across different brains, the technique may be used to extract commonalities across brains. Finally, weight vectors could e.g. be trained on particular artifacts common to fMRI, such as those induced by heart-rate, breathing or eye-movements, and then could be used to detect related artifacts in novel datasets. With this in mind, we tested the various KCCA variants on data and labels for which the associated and expected brain maps were well known from prior regression analyses (Bartels, Zeki, 2004; Bartels, Zeki, Logothetis, 2008).

## 2 Methods

### 2.1 Data and Experimental Setup

fMRI data of one human volunteer was acquired using a Siemens 3T TIM scanner, and consisted of 350 time slices of 3-dimensional fMRI brain volumes. Time-slices were separated by 3.2 s (TR), each with a spatial resolution of 46 slices (2.6 mm width, 0.4 mm gap) with 64x64 pixels of 3x3 mm, resulting in a spatial resolution of 3x3x3 mm. The subject watched 2 movies of 18.5 min length, one of which had labels indicating the continuous content of the movie (i.e. degree of visual contrast, or the degree to which a face was present, etc.). The imaging data were pre-processed using standard procedures using SPM running in matlab (<http://www.fil.ion.ucl.ac.uk/spm/software/spm5/>) before analysis. This included a slice-time correction to compensate for acquisition delays between slices, a spatial realignment to correct for small head-movements, a spatial normalization to the SPM standard brain space (near MNI), and spatial smoothing using a Gaussian filter of 6 mm full width at half maximum (FWHM). Subsequently, images were skull-and-eye stripped and the mean of each time-slice was set to the same value (global scaling). A temporal high-pass filter with a cut-off of 512 s was applied, as well as a low-pass filter with the temporal properties of the hemodynamic response function (hrf), in order to reduce temporal acquisition noise.

The label time-series were obtained using two separate methods, using computer frame-by-frame analysis of the movie (Bartels, Zeki, Logothetis, 2008), and using subjective ratings averaged across an independent set of five human observers (Bartels, Zeki, 2004). The computer-derived labels indicated luminance change over time (temporal contrast), visual motion energy (i.e. the fraction of temporal contrast that can be explained by motion in the movie). The human-derived labels indicated the intensity of subjectively experienced color, and the degree to which faces and human bodies were present in the movie. In prior studies, each of these labels had been shown to correlate with brain activity in particular and distinct sets of areas specialized to process the particular label in question (Bartels, Zeki, 2004; Bartels, Zeki, Logothetis, 2008).

### 2.2 Semi-supervised Kernel Canonical Correlation Analysis

We may formalize the above setting as follows. The brain volumes at each time slice may be viewed as training data in a vector space  $\{x_1, \dots, x_n\}$ , where each voxel corresponds to a dimension of the vector space. Associated with this training data are labels  $\{y_1, \dots, y_n\}$  which may be single or multi-variate. As labels require an expensive manual step, we may optionally include additional training data  $\{x_{n+1}, \dots, x_{n+p_x}\}$  that do not have corresponding values of  $y$ . We use the variables  $m_x = n + p_x$  to denote the total number of labeled and unlabeled samples in modality  $x$ . We denote the  $d \times n$  data matrix  $X = (x_1, \dots, x_n)$ , and the matrix including all data with and without correspondences  $\hat{X} = (x_1, \dots, x_n, x_{n+1}, \dots, x_{n+p_x})$ . For full generality, we symmetrize the formulation and also consider the case where there are data in  $Y$  without correspondence to  $X$ , using the analogous notation  $\hat{Y}$  and  $m_y$ . Furthermore we denote kernel matrices between the various sets of data as follows:  $\Phi_x(X)^T \Phi_x(X) = K_{xx}$ ,  $\Phi_x(\hat{X})^T \Phi_x(X) = K_{\hat{x}x}$ ,  $\Phi_x(\hat{X})^T \Phi_x(\hat{X}) = K_{\hat{x}\hat{x}}$ , etc. Kernel matrices for  $Y$  are defined analogously. We wish to optimize the following generalization of Tikhonov regularized KCCA (Blaschko, Lampert, Gretton, 2008):

$$\frac{\alpha^T K_{\hat{x}x} K_{y\hat{y}} \beta}{\sqrt{\alpha^T (K_{\hat{x}x} K_{x\hat{x}} + R_{\hat{x}}) \alpha \beta^T (K_{\hat{y}y} K_{y\hat{y}} + R_{\hat{y}}) \beta}} \quad (1)$$

where  $R_{\hat{x}} = \varepsilon_x K_{\hat{x}\hat{x}} + \frac{\gamma_x}{m_x^2} K_{\hat{x}\hat{x}} \mathcal{L}_{\hat{x}} K_{\hat{x}\hat{x}}$  and  $\mathcal{L}_{\hat{x}}$  is the empirical graph Laplacian estimated from the  $m_x$  samples of labeled and unlabeled data. The regularization term  $R$  contains both Tikhonov regularization and semi-supervised Laplacian regularization terms, controlled by parameters  $\varepsilon$  and  $\gamma$ . By setting these parameters appropriately, we may recover all of the other regression methods in the family, including LDA, and ridge regression.

In practice, we use a model selection procedure to determine  $\varepsilon$  and  $\gamma$  as specified in (Hardoon, Szedmak, Shawe-Taylor, 2004). As this procedure has not been experimentally validated in the literature, we have also performed a comparison to a model selection technique using cross-validation.

### 2.3 Results

We implemented three experimental variants of KCCA on the aforementioned data. In all cases, we used linear kernels for both  $x$  and  $y$ . This allows us to interpret the output of the model by analyzing the weights assigned to different spatially localized brain regions, which are analyzed in the discussion section (see figure 1). In the experiments with Laplacian regularization, similarities for computing the Laplacian matrix were determined using a Gaussian similarity measure where the bandwidth was set to the median distance between training data points. In

Table 1: Mean holdout correlations across the six variables in all experiments with model selection criterion of Hardoon et al. (2004).

	Motion	Temporal Contrast	Human Body	Color	Faces	Language
Exp 1	$-0.012 \pm 0.081$	$0.042 \pm 0.065$	$0.095 \pm 0.086$	$-0.075 \pm 0.069$	$0.173 \pm 0.073$	$0.172 \pm 0.070$
Exp 2	$0.065 \pm 0.066$	$0.088 \pm 0.084$	$0.274 \pm 0.093$	$-0.002 \pm 0.079$	$0.203 \pm 0.075$	$0.231 \pm 0.074$
Exp 3	$0.170 \pm 0.074$	$0.116 \pm 0.101$	$0.340 \pm 0.043$	$0.128 \pm 0.089$	$0.303 \pm 0.054$	$0.365 \pm 0.057$

Table 2: Mean holdout correlations across the six variables in all experiments with five-fold cross-validation.

	Motion	Temporal Contrast	Human Body	Color	Faces	Language
Exp 1	$-0.012 \pm 0.081$	$0.042 \pm 0.065$	$0.095 \pm 0.086$	$-0.075 \pm 0.069$	$0.173 \pm 0.073$	$0.172 \pm 0.070$
Exp 2	$0.065 \pm 0.066$	$0.088 \pm 0.084$	$0.274 \pm 0.093$	$-0.002 \pm 0.079$	$0.203 \pm 0.075$	$0.231 \pm 0.074$
Exp 3	$0.170 \pm 0.074$	$0.116 \pm 0.101$	$0.340 \pm 0.043$	$0.128 \pm 0.089$	$0.303 \pm 0.054$	$0.365 \pm 0.057$

the first of our three experiments, no Laplacian regularization was used, and model selection was performed only over the Tikhonov regularization parameter,  $\epsilon$ . In the second experiment, Laplacian regularization was employed, but only using the labeled data. In the third experiment, model selection was employed using full semi-supervised Laplacian regularization. These experiments were performed twice: once with the model selection criterion of (Hardoon, Szedmak, Shawe-Taylor, 2004) and once using five-fold cross-validation. The results of the single variate output are summarized in Tables 1 and 2.

We ran this set of experiments additionally as multi-variate experiments for three sets of variables. The first set consists of visual motion energy, human bodies, and degree of color, the second set contains motion and faces, and the final set contains motion, visual motion energy, color, and faces. The results of these multi-variate experiments are summarized in Table 3 and were derived using the model selection criterion of (Hardoon, Szedmak, Shawe-Taylor, 2004).

### 3 Discussion

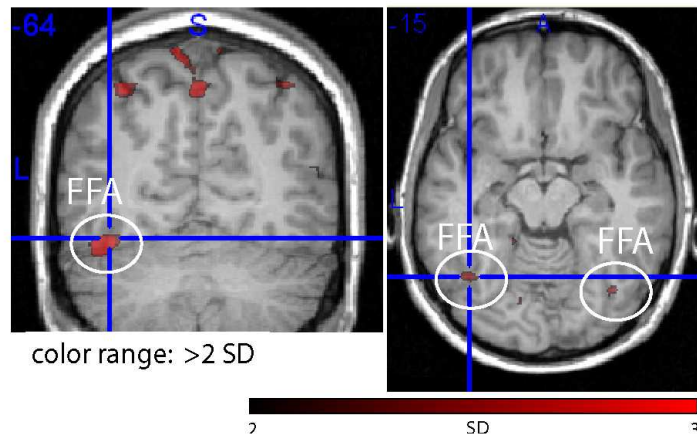
We observe several trends in Tables 1 and 2. First, our major hypotheses were confirmed: for every variate label, the performance improved with the Laplacian regularization on the labeled data, and performance was best in the semi-supervised condition. In the semi-supervised conditions (experiments 3 and 6 as shown in tables 1 and 2) the additional [unlabeled] data is sufficiently close to the  $X$  distribution to improve results significantly, thus the additional data improves the results without any information about the correspondences of the data. Second, as indicated by the similarity of the experimental results, the model selection criterion worked very well; the criterion suggested by Hardoon, Szedmak, and Shawe-Taylor (2004) performed equally as well as the five-fold cross-validation. Additionally, some variables can be better predicted than others, namely the presence of faces or human bodies in the viewing content, while some elicited relatively poorer performance in all experiments.

As we have used linear kernels in all cases, we can interpret the output of the model by analyzing the weights assigned to different spatially localized brain regions. We show results for visual stimulus consisting of *Faces* in Figure 1, *Human body* in Figure 2, *Color* in Figure 3, and *Motion* in Figure 4. In Figure 5 we show results from multivariate output consisting of *Motion* and *Faces*.

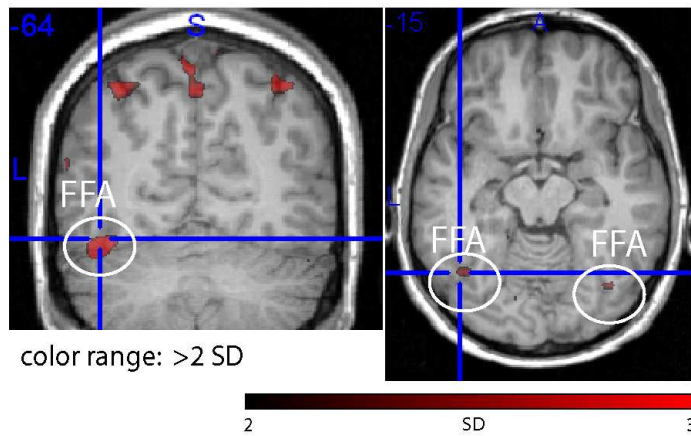
Figure 1 shows slices taken through the anatomical image of one subject, with weight maps obtained from the different analyses of its functional data superimposed in red. We show examples of three labels, for NNN analysis methods. The maps corresponding well to the known functional anatomy, and to activations obtained in the previous regression studies of free-movie-viewing data (Bartels, Zeki, 2004). Faces obtained high weights

Table 3: Mean holdout correlations across the 3 multi-variate sets in all experiments with model selection criterion of Hardoon et al. (2004).

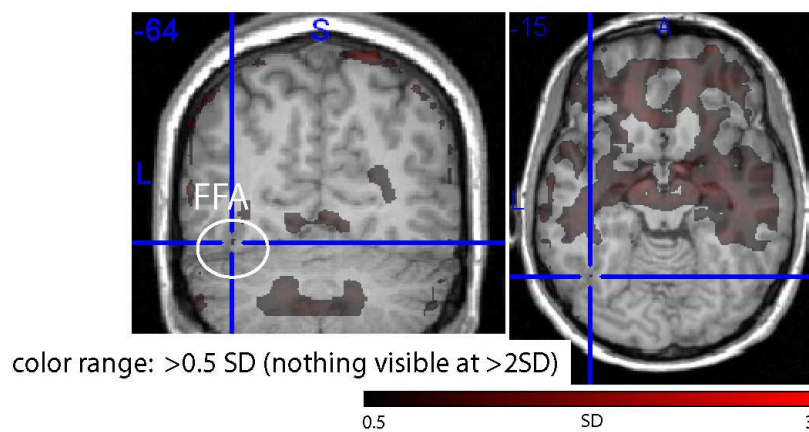
	Visual motion energy, Body, Color	Motion, Faces	Motion, Visual motion energy, Color, Faces
Experiment 1	$0.1596 \pm 0.0807$	$-0.0827 \pm 0.0460$	$0.1167 \pm 0.0785$
Experiment 2	$0.1873 \pm 0.0879$	$0.0602 \pm 0.0908$	$0.1498 \pm 0.0827$
Experiment 3	$0.2844 \pm 0.0716$	$0.1898 \pm 0.0636$	$0.2528 \pm 0.0579$



(a) Semi-supervised Laplacian regularized solution.

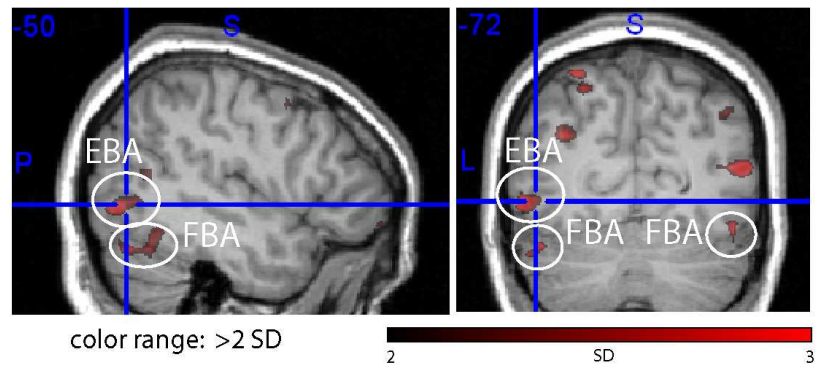


(b) Laplacian regularized solution.

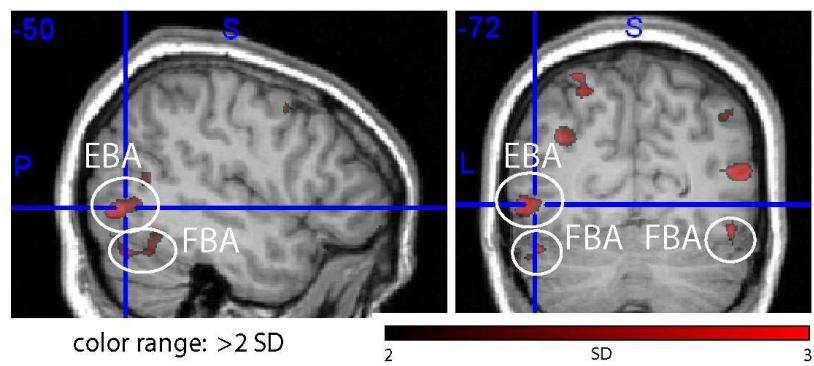


(c) KCCA without Laplacian regularization.

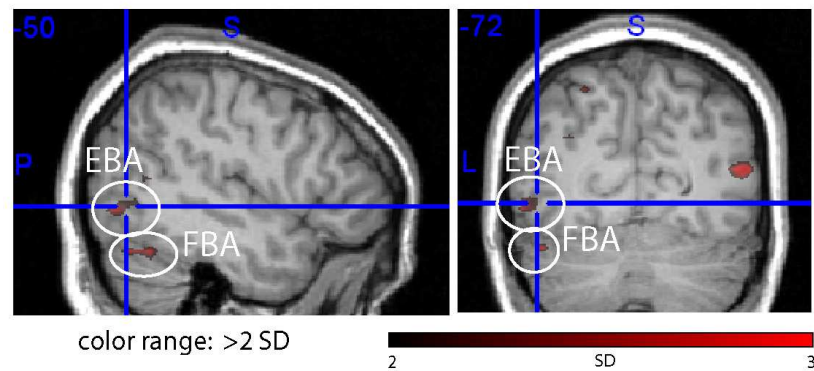
Figure 1: Faces: activation in the cortical region responsive to the visual perception of faces, the fusiform face area (FFA). Weight vectors are plotted over an anatomical image of the volunteers brain. Note that the semi-supervised Laplacian regularization led to the most specific and most significant weights in FFA.



(a) Semi-supervised Laplacian regularized solution.

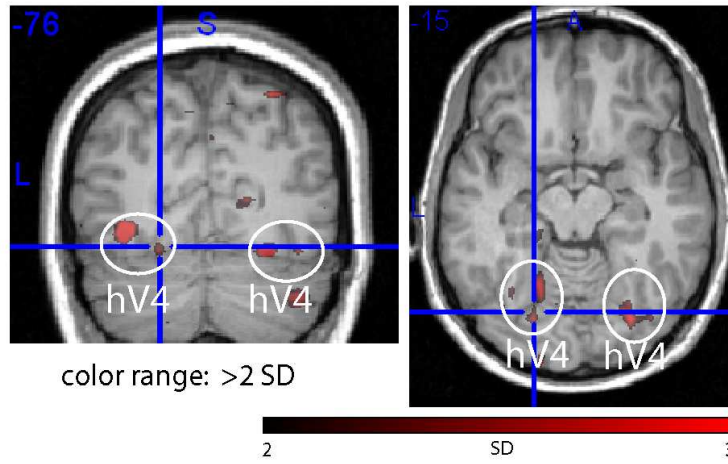


(b) Laplacian regularized solution.

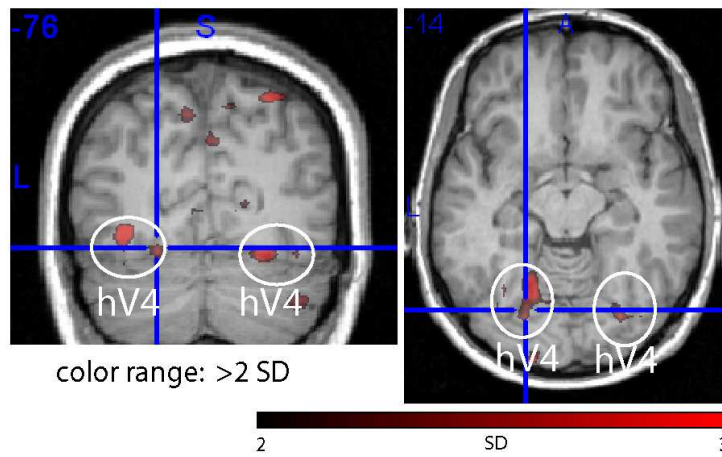


(c) KCCA without Laplacian regularization.

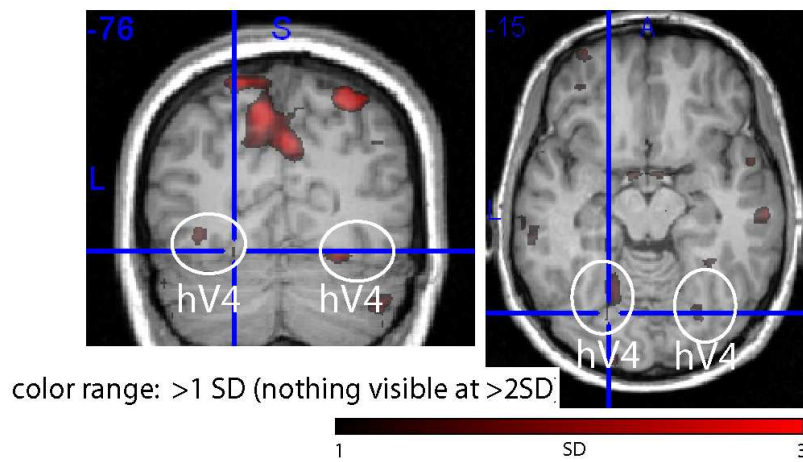
Figure 2: Human Body: activation in the cortical region responsive to the visual perception of human bodies, in the extrastriate body area (EBA) and in the fusiform body area (FBA). Same observation as in Figure 1.



(a) Semi-supervised Laplacian regularized solution.

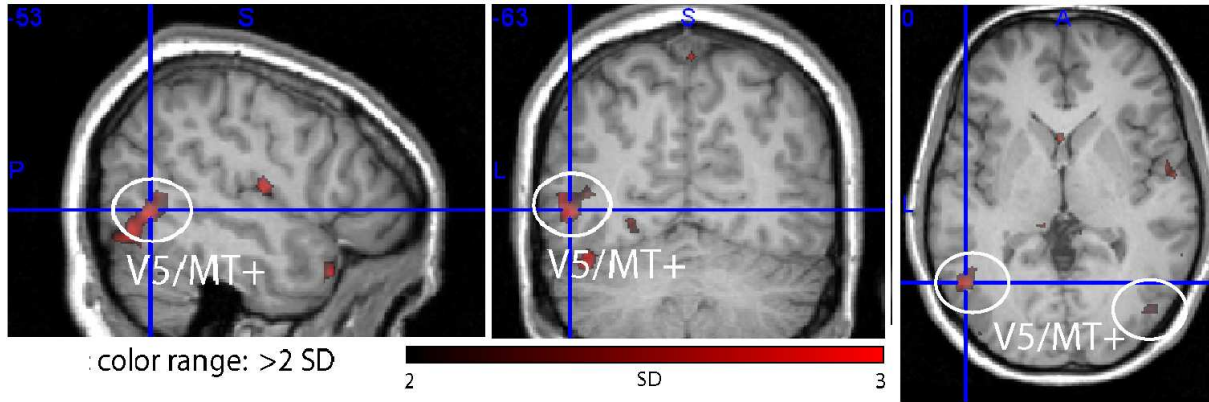


(b) Laplacian regularized solution.

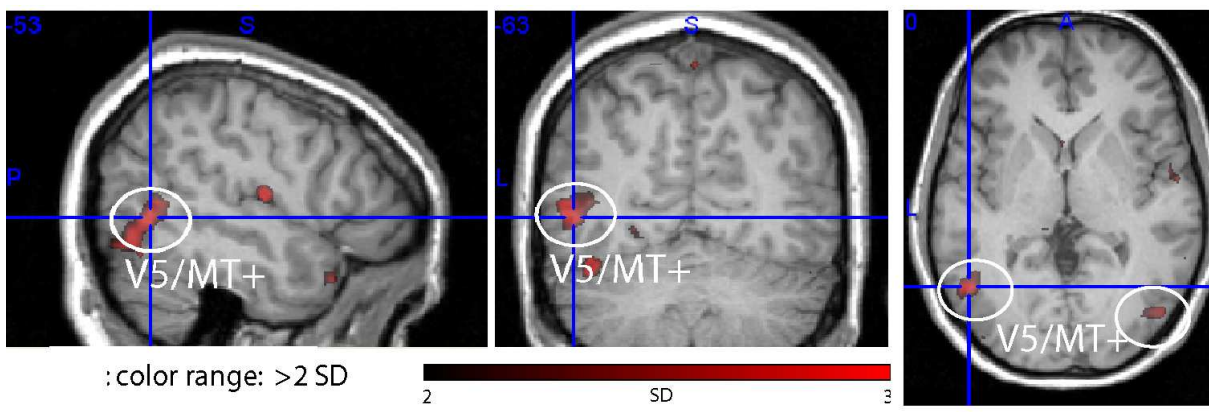


(c) KCCA without Laplacian regularization.

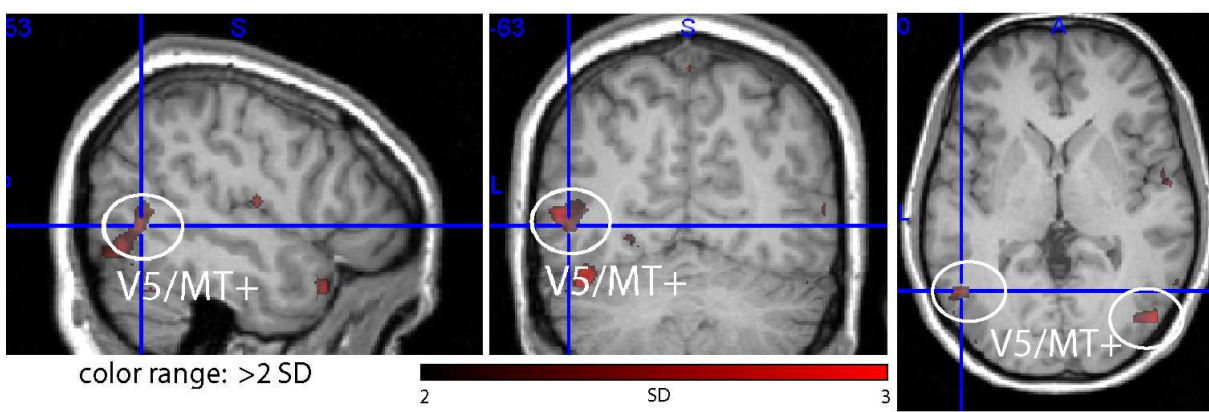
Figure 3: Color: activation in the color responsive cortex (human visual area 4, hV4). Same observation as in Figure 1.



(a) Semi-supervised Laplacian regularized solution.



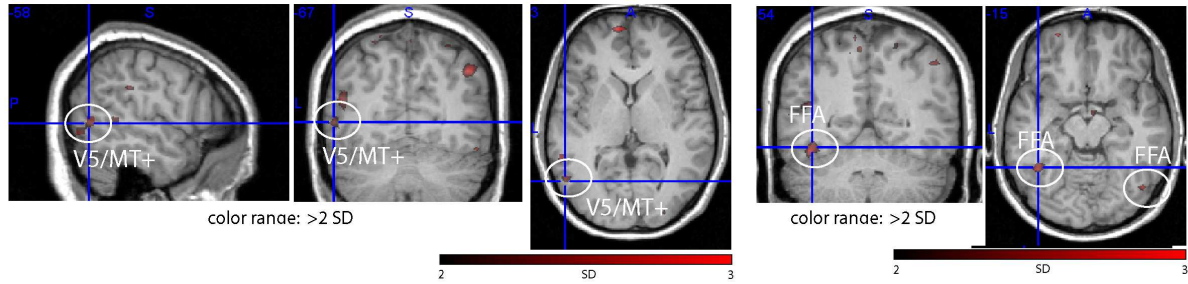
(b) Laplacian regularized solution.



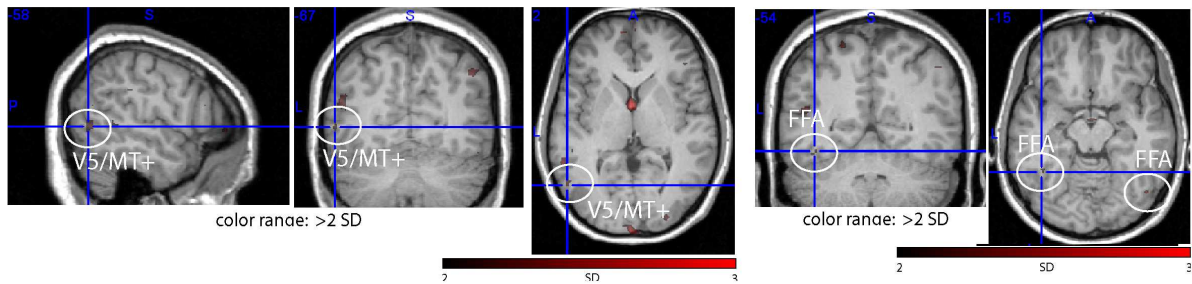
(c) KCCA without Laplacian regularization.

Figure 4: Motion: activation in the visual motion complex, area V5+/MT+. Same observation as in Figure 1.

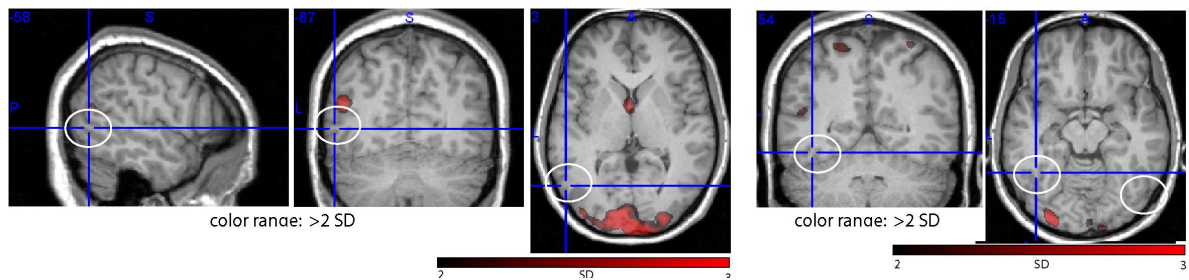




(a) Semi-supervised Laplacian regularized solution.



(b) Laplacian regularized solution.



(c) KCCA without Laplacian regularization.

Figure 5: Multivariate - *Motion* and *Faces*: activations in the visual motion complex, area V5+/MT+ (left), and activation in the cortical region responsive to the visual perception of faces, the fusiform face area (FFA) (right). Same observation as in Figure 1.

in the fusiform cortex (fusiform face area, FFA); Human Bodies dorso-lateral and ventral parts within the lateral occipital cortex (extrastriate body area (EBA) and fusiform body area (FBA)); Color obtained high weights in the medial fusiform cortex where human V4 is located. The spatial layout of the weights thus corresponds well to the previous literature, and indicates that some of the analyses applied here yield results that are neuroscientifically meaningful and that can identify distinct cortical regions involved in the distinct tasks. Analysis 3 worked well in that the thresholded weight maps (greater than 2 standard deviations) show relatively well-defined activity of the regions previously shown to be involved with the features. For other analyses, e.g. experiment. 1, we had to reduce the threshold to 0.5 or 1 (faces and color, respectively) to obtain activity in the areas in question, and the maps show additional, unspecific activity as well.

## 4 Conclusions

In this work, we used a semi-supervised Laplacian regularized generalization of KCCA which has several important regression techniques as special cases. The various experimental designs we tested improved successively (as shown by the correlations of the holdout sets for each variant), both from the supervised variant with no regularization parameter, to the supervised variant with Laplacian regularization, and further improved in the semi-supervised variant by the addition of unlabeled data. Additionally, the analysis of the weights learned by, particularly the semi-supervised experiment and the supervised Laplacian regularized experiment, display the same brain activation patterns shown in previous neuroscientific studies. This provides promise that with semi-supervised methods, one can simply add unlabeled data of a similar variety to a (significantly smaller) labeled data set, and achieve similar results, as the approximation of the labeled data distribution is thereby strengthened. These results additionally lay the groundwork for exciting neuroscience applications, such as removal of learned artifacts from fMRI data, eliminating the need for expensive labeling of natural stimuli shown during fMRI image acquisition, as well as potential for discovering brain activity patterns associated to new or unknown stimuli. In our future work we intend to include resting state data, in order to investigate if this data too comes from a distribution similar enough to the X distribution to further improve results. Given this is the more common type of data acquired and simpler to acquire, this would be a useful application for neuroscientists.

## References

- [1] Bartels, A., and Zeki, S. (2003). Functional brain mapping during free viewing of natural scenes. *Human Brain Mapping*, vol. 21: 75–83.
- [2] Bartels, A., and Zeki, S. (2004). The chronoarchitecture of the human brain - natural viewing conditions reveal a time-based anatomy of the brain. *NeuroImage*, vol. 22: 419–433.
- [3] 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*, vol. 18: 705–717.
- [4] Blaschko, M. B., C. H. Lampert and A. Gretton. (2008). Semi-Supervised Laplacian Regularization of Kernel Canonical Correlation Analysis. *Machine Learning and Knowledge Discovery in Databases, ECML PKDD 2008*, 133–145.
- [5] Belhumeur, P., Hespanha, J., and Kriegman, D.. (1996). “Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection,” *Proc. Fourth European Conf. on Computer Vision*, vol. 1: 45–58.
- [6] Haroon, D. R., Szedmak, S., and Shawe-Taylor, J. (2004). “Canonical Correlation Analysis: An Overview with Application to Learning Methods,” *Neural Computation*, vol. 16(12): 2639–2664.
- [7] Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., and Malach, R. (2004). Intersubject synchronization of cortical activity during natural vision. *Science*, vol. 303: 1634–1640.
- [8] Hasson, U., Yang, E., Vallines, I., Heeger, DJ., and Rubin, N. (2008). A hierarchy of temporal receptive windows in human cortex. *J. Neuroscience*, vol. 28: 2539–2550.
- [9] Turk and Pentland, A. (1991). “Face recognition using eigenfaces”. *Proc. IEEE Conference on Computer Vision and Pattern Recognition*: 586–591.