2016 23rd International Conference on Pattern Recognition (ICPR)
Cancún Center, Cancún, México, December 4-8, 2016

Sparse-coded Cross-domain Adaptation from the Visual to the Brain Domain

Pouya Ghaemmaghami*, Moin Nabi*, Yan Yan*, and Nicu Sebe*
*Department of Information Engineering and Computer Science, University of Trento, Italy

***Abstract*—Brain decoding (i.e., retrieving information from brain signals by employing machine learning algorithms) has recently received considerable attention across many communities. In a typical brain decoding paradigm, different types of stimuli are shown to the participant of the neuroimaging experiment, while his/her concurrent brain activity is captured using neuroimaging techniques. Then a machine learning algorithm is employed to categorize the measured brain signal into the target stimuli classes. Accurate prediction of the stimulus category by the algorithm is considered a positive evidence of the hypothesis of the existence of stimulus-related information in brain data. However, most of the brain decoding studies suffer from the constraint of having few and noisy samples. In order to overcome this limitation, in this paper, an adaptation paradigm is employed in order to transfer knowledge from visual domain to brain domain. We experimentally show that such adaptation procedure leads to improved results for the object recognition task in the brain domain, outperforming significantly the results achieved by the brain features alone. This is the first study in the direction of transferring knowledge by adapting representations learned on visual domain to the brain modality. We believe this paper opens up avenues for exploiting large-scale visual datasets to achieve performance gain in brain decoding.**

I. Introduction

For many year, reading someone's mind has been the domain of science fiction. Recently however, after all new discoveries about the brain, "Mind Reading" has become the province of science [1]. In fact, a challenging goal in neuroscience is decoding mental contents from brain activities. Recent progress in neuroimaging suggests the possibility of brain decoding [2]. This has received considerable attention in Brain Computer interfacing (BCI) and rehabilitation communities particularly due to its potential for helping disabled and paralyzed people [3], [4].

Brain decoding can be generally formulated as the classification of stimuli into a set of pre-defined categories. In a typical brain decoding paradigm, different categories of stimuli are presented to experimental subjects, while their brain signals are recorded simultaneously using various neuroimaging methods. Then a machine learning approach is employed to categorize the measured signal into the target stimuli classes. Among various neuroimging techniques for recording brain activity, the most widely used methods for noninvasive brain recording in humans are Functional Magnetic Resonance Imaging (fMRI), Magnetoencephalography (MEG) and Electroencephalography (EEG). Once brain signals are recorded, the aforementioned decoding systems can be applied to the measured signal. However, due to the low signal-to-noise ratio and non-stationarity nature of the signals, the performance of the decoding system is often not very accurate. Besides, the neuroimaging datasets suffer from few samples due to the cost of recording brain signals and subject's fatigue. This small number of samples drastically decreases the performance of machine learning algorithms.

In machine learning literature, researchers tackle this problem by employing the transfer learning paradigm. In this paradigm, shared knowledge can be transfered from a large set of samples from a source domain to a target domain with fewer samples. In such cases, the performance in the target domain strictly relies on the performance in the source domain and the similarity between the two domains. These methods aim at finding representations such that the domain divergence and consequently the modeling error on the target domain would be minimized. Transfer learning can truly be beneficial in cases where collecting data is extremely expensive or even impossible [5]. This situation arises often in brain studies.

The problem with generic transfer learning algorithms is that they have been shown to be highly sensitive to the discrepancy across source and target domains. Nevertheless, recent progress in Deep Neural Nets (DNN) provides the transfer learning community the opportunity to learn generic representations which are capable of capturing the semantics, hence they can be transfered across domains [6], [7] and modalities [8].

Due to the transferability power of such representations specifically in an object recognition task, in this study we investigate the possibility of transferring them for the same object recognition task using brain signals. Prior works in brain studies have shown that there is a region in the human brain called the "Ventral Temporal Cortex" (VTC) containing information about colour, object categories, concepts and semantics [9]. Inspired by this, and because of the importance of VTC in visual perception and object recognition, in this paper, we address the specific problem of transferring knowledge learned in ImageNet [10] to the brain domain. We hypothesize that such adaptation can be done successfully yielding increased performance of the machine learning algorithm on the target domain (brain datasets).

To summarize, the main contribution of this study is as follows: We are the first to introduce the idea of cross-modal domain adaptation in brain studies. Our proposed method overcomes the limitation of brain datasets (i.e., few noisy samples) by transferring knowledge from the image modality

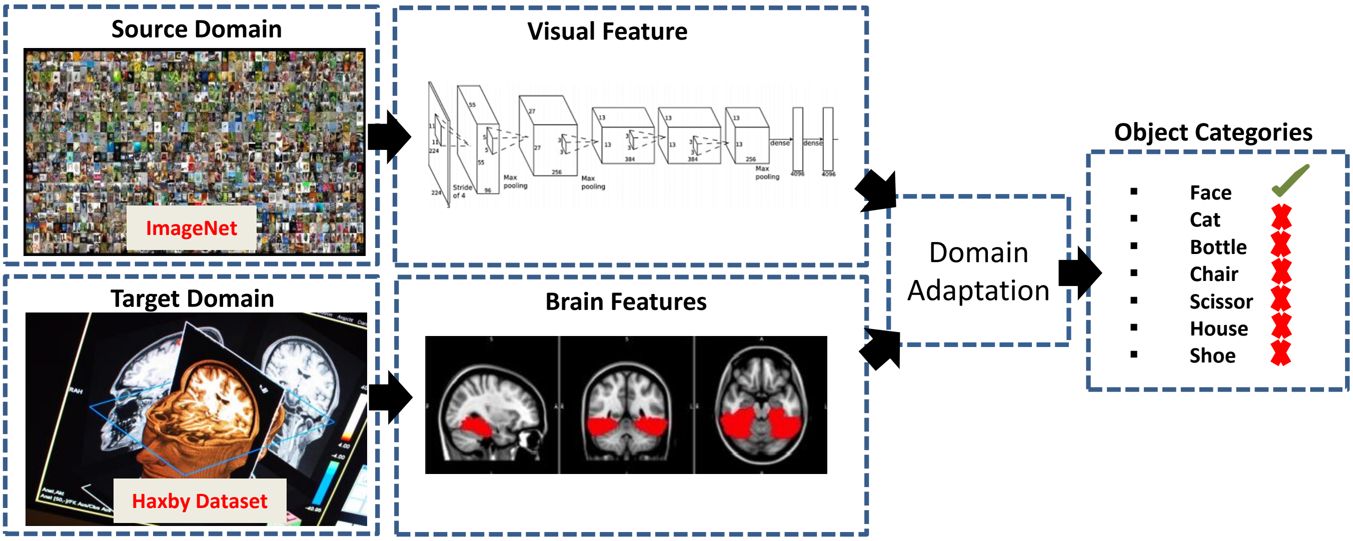


Fig. 1: Domain Adaptation Pipeline.

resulting in better performance as compared to the use of brain features alone. The proposed framework is a generic one and can be easily applied to other brain datasets (i.e., other neuroimaging modalities). This study can open a new door for scientists to investigate brain signals from a new perspective.

II. Related Works

Domain adaptation (also called transfer learning) aims at learning a good model from a source data distribution which can also perform well on a different (but related) target data distribution. There are several approaches to transfer learning [5]. *Instance transfer* [11] involves feeding a few labeled target samples along with many source samples to the classifier, assuming that certain parts of the source data are still useful for learning in the target domain. *Feature-based transfer* [12] involves minimizing the differences between source and target by representing both in a common feature space. In *parameter-transfer* [13], shared parameters or priors between the source and target models are exploited. In *knowledge transfer* [14], the source domain classifier is adapted to the target employing adaptive classifier.

Convolutional Neural Networks have recently resurfaced as a powerful tool for learning from big data (*e.g.*, ImageNet [10] with $\sim$1M images), providing models with excellent representational capacities. These models have been trained via backpropagation through several layers of convolutional filters [15]. It has been shown that such models are not only able to achieve state-of-the-art performance for the visual recognition task, but the learned representation can be readily applied to other relevant tasks [6]. These models perform extremely well in domains with large amounts of training data. With limited training data, however, they will likely dramatically over-fit the training data. Attracted by their amazing capability to produce a generic semantic representation, in this paper, we investigate transferring the learned deep representation from a large set of samples of visual domain to a small set of samples from the brain domain.

There has been a large body of works on representation transfer across domains belonging to the same modality. The representation transfer aims at encoding the knowledge used to transfer across domains into a learned representation by minimizing the domain discrepancy and the classification error. This problem is known as "common feature learning" in the field of multitask learning [16]. Recently, Tzeng, et al. [7] proposed a deep architecture which is simultaneously optimized for domain divergence and uses a soft label distribution matching loss. All these lines of work focused on the problem of domain adaptation within the same modality. In this work we, however, tackle the more difficult problem of domain adaptation across different modalities. This cross-model adaptation problem has received much less attention. While a few methods have been proposed for the text/image [17] and depth/image [8] adaptation, as far as we know, we are the first showing that deep-net-based cross-model adaptation can be used for brain signals.

At last, we note that our work is different from the transfer learning methodologies used in [18]–[20], which mainly focus on learning a common feature space across subjects. A key difference compared to these methods is that they transfer knowledge only within the brain modality whereas our setting provides the means to inject semantics inherited from the visual domain into the brain modality.

III. Method

Sparse coding was shown to be able to find succinct representations of stimuli from the brain [21]. In this section, we describe the details of our domain adaptation sparse coding method. Figure 1 illustrates the overview of our adaptation paradigm.

The source task (i.e., the image domain) consists of data samples denoted by $\mathbf{X_s} = \{\mathbf{x_s^1}, \mathbf{x_s^2}, ..., \mathbf{x_s^{n_s}}\} \in I\!R^{n_s \times d}$, where $\mathbf{x_s^i} \in I\!R^d$ is a d-dimensional feature vector and n_s is the number of samples in the source task. The target task is defined as the brain fMRI data. Similarly, the target task consists of data samples denoted by $\mathbf{X_t} = \{\mathbf{x_t^1}, \mathbf{x_t^2}, ..., \mathbf{x_t^{n_t}}\} \in I\!R^{n_t \times d}$, where $\mathbf{x_t^i} \in I\!R^d$ is a d-dimensional feature vector and n_t is the number of samples in the target task.

To better adapt useful knowledge from the source domain to the target domain, we are going to learn a shared subspace across the two domains, obtained by an orthonormal projection $\mathbf{W} \in I\!R^{d \times b}$, where b is the dimensionality of the subspace. In this learned subspace, the data distributions between the source domain and the target domain should be similar to each other. The benefits of this strategy is that we can improve the coding quality of the target task by transferring knowledge from the source task. This can be realized through the following optimization problem:

$$\begin{aligned} \min_{\mathbf{C_s},\mathbf{D_s},\mathbf{C_t},\mathbf{D_t},\mathbf{W},\mathbf{D}} & \|\mathbf{X_s} - \mathbf{C_s D_s}\|_F^2 + \lambda_1 \|\mathbf{C_s}\|_1 \\ & + \|\mathbf{X_t} - \mathbf{C_t D_t}\|_F^2 + \lambda_2 \|\mathbf{C_t}\|_1 \\ & + \lambda_3 \|\mathbf{X_s W} - \mathbf{C_s D}\|_F^2 + \lambda_4 \|\mathbf{X_t W} - \mathbf{C_t D}\|_F^2 \\ s.t. \quad & \begin{cases} \mathbf{W^T W} = \mathbf{I} \\ (\mathbf{D_s})_{\mathbf{j}\cdot} (\mathbf{D_s})'_{\mathbf{j}\cdot} \leq 1, \quad \forall j = 1, ..., l \\ (\mathbf{D_t})_{\mathbf{j}\cdot} (\mathbf{D_t})'_{\mathbf{j}\cdot} \leq 1, \quad \forall j = 1, ..., l \\ \mathbf{D_{j\cdot}} \mathbf{D}'_{\mathbf{j}\cdot} \leq 1, \quad \forall j = 1, ..., l \end{cases} \end{aligned} \tag{1}$$

where $\mathbf{D_s}, \mathbf{D_t} \in I\!R^{l \times d}$ are overcomplete dictionaries ($l > d$) with l prototypes of the source and target task; $(\mathbf{D_s})_{\mathbf{j}\cdot}$ and $(\mathbf{D_t})_{\mathbf{j}\cdot}$ in the constraints denote the j-th row of $\mathbf{D_s}$ and $\mathbf{D_t}$, respectively; $\mathbf{C_s} \in I\!R^{n_s \times l}$ and $\mathbf{C_t} \in I\!R^{n_t \times l}$ correspond to the sparse representation coefficients of $\mathbf{X_s}$ and $\mathbf{X_t}$, respectively. In the last two terms of Eqn.(1), $\mathbf{X_s}$ and $\mathbf{X_t}$ are projected by $\mathbf{W}$ into the subspace to explore the relationship between the source and the target tasks. $\mathbf{D} \in I\!R^{l \times b}$ is the dictionary learned in the shared subspace between the source and the target tasks. $\mathbf{D_{j\cdot}}$ in the constraints denotes the j-th row of $\mathbf{D}$. $\mathbf{I}$ is the identity matrix. $(\cdot)'$ denotes the transpose operator. $\lambda' s$ are the regularization parameters. The first constraint guarantees the learned $\mathbf{W}$ to be orthonormal, and the other constraints prevent the learned dictionary to be arbitrarily large. In our objective function, we learn dictionaries $\mathbf{D_s}$, $\mathbf{D_t}$ for the source and the target task respectively and one shared dictionary $\mathbf{D}$ between the source and the target tasks.

Optimization: To solve the proposed objective problem of Eqn.(1), we adopt the alternating minimization algorithm to optimize it with respect to $\mathbf{D}$, $\mathbf{D_s}$, $\mathbf{C_s}$, $\mathbf{D_t}$, $\mathbf{C_t}$ and $\mathbf{W}$ respectively in five steps as follows:

Step1: Fixing $\mathbf{D_s}$, $\mathbf{C_s}$, $\mathbf{W}$, $\mathbf{D_t}$, $\mathbf{C_t}$, Optimize $\mathbf{D}$. If we stack $\mathbf{X} = [\mathbf{X_s}; \mathbf{X_t}]$, $\mathbf{C} = [\mathbf{C_s}; \mathbf{C_t}]$, Eqn.(1) is equivalent to:

$$\begin{aligned} & \min_{\mathbf{D}} \|\mathbf{XW} - \mathbf{CD}\|_F^2 \\ & s.t. \quad \mathbf{D_{j\cdot}} \mathbf{D_{j\cdot}^T} \leq 1, \quad \forall j = 1, ..., l \end{aligned} \tag{2}$$

This is equivalent to the dictionary update stage in the traditional dictionary learning algorithm. We adopt the dictionary update strategy of Algorithm 2 in [22] to efficiently solve it.

Step2: Fixing $\mathbf{D}$, $\mathbf{C_s}/\mathbf{C_t}$, $\mathbf{W}$, Optimize $\mathbf{D_s}/\mathbf{D_t}$. This is the same as Step 1 which is equivalent to the dictionary update stage in the traditional dictionary learning for k tasks. We adopt the dictionary update strategy of Algorithm 2 in [22] to efficiently solve it.

Step3: Fixing $\mathbf{D_s}/\mathbf{D_t}$, $\mathbf{W}$, $\mathbf{D}$, Optimize $\mathbf{C_s}/\mathbf{C_t}$. Eqn.(1) is equivalent to:

$$\begin{aligned} \min_{\mathbf{C_s},\mathbf{C_t}} & \|\mathbf{X_s} - \mathbf{C_s D_s}\|_F^2 + \lambda_1 \|\mathbf{C_s}\|_1 \\ & + \|\mathbf{X_t} - \mathbf{C_t D_t}\|_F^2 + \lambda_2 \|\mathbf{C_t}\|_1 \\ & + \lambda_3 \|\mathbf{X_s W} - \mathbf{C_s D}\|_F^2 + \lambda_4 \|\mathbf{X_t W} - \mathbf{C_t D}\|_F^2 \end{aligned} \tag{3}$$

This formulation can be decoupled into $(n_s + n_t)$ distinct problems. We adopt the Fast Iterative Shrinkage-Thresholding Algorithm (FISTA) [23] to solve the problem.

Step4: Fixing $\mathbf{D_s}$, $\mathbf{C_s}$, $\mathbf{D}$, $\mathbf{D_t}$, $\mathbf{C_t}$, Optimize $\mathbf{W}$. If we stack $\mathbf{X} = [\mathbf{X_s}; \mathbf{X_t}]$, $\mathbf{C} = [\mathbf{C_s}; \mathbf{C_t}]$, Eqn.(1) is equivalent to:

$$\begin{aligned} & \min_{\mathbf{W}} \|\mathbf{XW} - \mathbf{CD}\|_F^2 \\ & s.t. \quad \mathbf{W^T W} = \mathbf{I} \end{aligned} \tag{4}$$

Substituting $\mathbf{D} = (\mathbf{C^T C})^{-1} \mathbf{C^T X W}$ back into the above function, we achieve

$$\begin{aligned} & \min_{\mathbf{W}} \left\| (\mathbf{I} - \mathbf{C}(\mathbf{C^T C})^{-1} \mathbf{C^T}) \mathbf{XW} \right\|_F^2 \\ & = \min_{\mathbf{W}} tr(\mathbf{W^T X^T} (\mathbf{I} - \mathbf{C}(\mathbf{C^T C})^{-1} \mathbf{C^T}) \mathbf{XW}) \\ & s.t. \quad \mathbf{W^T W} = \mathbf{I} \end{aligned} \tag{5}$$

The optimal $\mathbf{W}$ is composed of eigenvectors of the matrix $\mathbf{X^T} (\mathbf{I} - \mathbf{C}(\mathbf{C^T C})^{-1} \mathbf{C^T}) \mathbf{X}$ corresponding to the s smallest eigenvalues.

We summarize our algorithm for solving Eqn.(1) as Algorithm 1.

Finally, the classification algorithm can be applied to $\mathbf{C_t}$ with corresponding labels to train classification models to be used in the target domain.

IV. Experiments and Results

In this section, we first introduce the details of the employed datasets, then explain the features extraction method and classification scenario and finally present the experiments in detail and discuss the results.

Algorithm 1: Domain adaptation method.

```
Input:
    Data sample matrix $\mathbf{X}$; Subspace dimensionality $b$, Dictionary size $l$, Regularization parameters $\lambda_s$.
Output:
    Optimized $\mathbf{W} \in I\!R^{d \times b}$, $\mathbf{C} \in I\!R^{n \times l}$, $\mathbf{D_s} \in I\!R^{l \times d}$, $\mathbf{D_t} \in I\!R^{l \times d}$, $\mathbf{D} \in I\!R^{l \times b}$.
Initialize $\mathbf{W}$ using any orthonormal matrix;
Initialize $\mathbf{C}$ with $l_2$ normalized columns;
repeat
    Compute $\mathbf{D}, \mathbf{D_s}, \mathbf{D_t}$ using Algorithm 2 in [22];
    Adopting FISTA [23] to solve $\mathbf{C}$;
    Compute $\mathbf{W}$ by eigen decomposition of $\mathbf{X^T}(\mathbf{I} - \mathbf{C}(\mathbf{C'C})^{-1}\mathbf{C'})\mathbf{X}$;
until Convergence;
```

A. Databases and Features

Brain Domain (Target): In this work, as our target domain, we used a well-known dataset (i.e., Haxby dataset) introduced in a study on face and object representation in human ventral temporal cortex [24]. This dataset consists of the fMRI data of 6 subjects in which each subject had undergone 12 sessions (runs). In each run, the subjects passively viewed greyscale images of eight object categories (faces, houses, cats, bottles, scissors, shoes, chairs, and nonsense patterns)[1], grouped in 24s time blocks separated by rest periods. Each image was shown for 500ms and was followed by a 1500ms inter-stimulus interval.

In this work, we use brain features (voxels) within the Ventral Temporal Cortex (VTC) in order to select the relevant features for the object recognition task. VTC is the area in the brain where high-level visual regions reside and it is involved in visual perception and recognition [9]. In order to obtain VTC voxels, we used the *Atlas-based* approach employed in [25][2]. We refer to these features as "Brain-Features".

Visual Domain (Source): We use ImageNet images [10] selected from the synsets corresponding to the seven object categories of: faces, houses, cats, bottles, chairs, shoes and scissors.[3] The number of images for each category of interest is more than 1000 images. For each sample, we extract the output of the *fc7* layer of the pre-trained AlexNet model [15] using the standard CNN Caffe toolbox [27].

B. Classification Scenario

Following [28], a linear SVM is employed to classify the features into the set of categories. The classifier is trained and tested on the data for each subject separately (within-subject analysis). This is repeated for all six subjects. The evaluation has been done in a leave-one-run-out fashion.

C. Experiments

We first study the effectiveness of the adaptation method explained in section III. We, then, evaluate the parameter sensitivity of our proposed method.

1) Experiment 1: To evaluate the effectiveness of adaptation, we employed the classification scenario explained above, only replacing the Brain-Features with the Adapted-Brain-Features. Adapted-Brain-Features are computed using the adaption method explained in section III.

Table I summarizes the results of this experiment. The average accuracy using the Adapted-Brain-Features is significantly superior compared to the average accuracy obtained by Brain-Features ($p - value < 0.005$). This difference suggests the impact of transferring knowledge from the visual modality to the brain modality. Regardless of the big differences in these modalities, the semantic representations learned in ImageNet are transferred successfully to brain features. Besides, our result shows improvement in 5 out of 6 subjects.

TABLE I: Seven-Class Classification Accuracy (average accuracy over all runs for each subject)

Subjects	**Brain-Features [28]**	**Adapted-Brain-Features**
Subject 1	0.64 ± 0.10	**0.78** ± 0.09
Subject 2	0.59 ± 0.07	**0.65** ± 0.11
Subject 3	0.43 ± 0.08	**0.47** ± 0.07
Subject 4	0.49 ± 0.10	**0.57** ± 0.08
Subject 5	0.66 ± 0.08	**0.73** ± 0.05
Subject 6	**0.65** ± 0.10	0.63 ± 0.06
Average	0.58 ± 0.13	**0.64** ± 0.13

To allow the category-wise analysis, the confusion matrices for object classification are illustrated in Figure 2a and 2b. In both cases, "face" and "house" categories are predicted with higher confidence compared to the other categories. In 5 out of 7 categories, the classification using Adapted-Brain-Features outperforms the Brain-Features. The "House" category performs similarly in both feature spaces (Brain-Features and Adapted-Brain-Features). The "Face" category, however, is predicated better using Brain-Features. This is probably due to the importance of Fusiform Face Area (FFA) for face recognition [29], [30] and the effect of this area might be lost after adaptation.

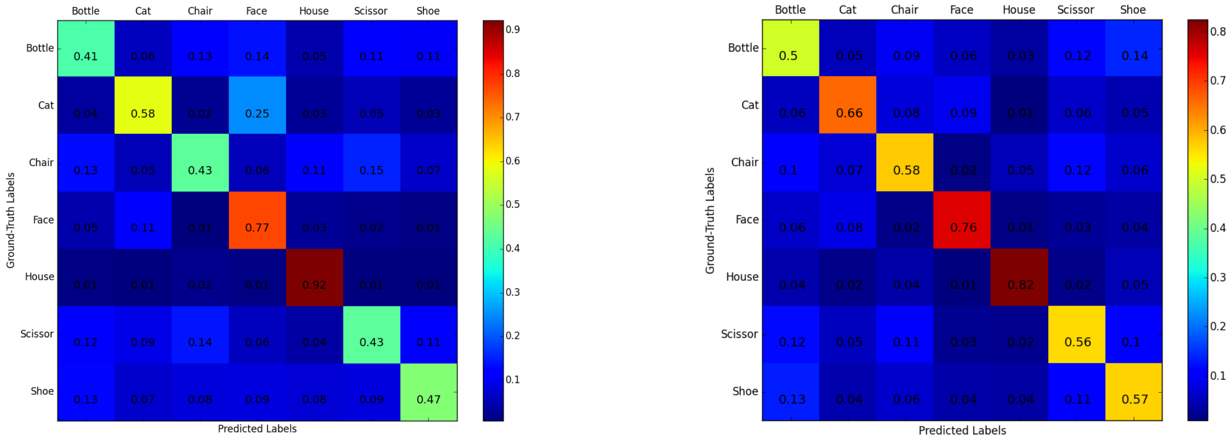
2) Experiment 2: Since the atlases we used in this study are probabilistic atlases, we investigate the effect of selecting different thresholds (probability) on our classification results. We set the threshold value in the range of $0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8$. In more details, for each threshold we discarded the voxels that their probability on the atlas are below the threshold value. Figure 3 demonstrates the classification results of each subject using different thresholds.

We also calculate the average results of all subjects over all runs for each threshold. Table II compares the results of such analysis. The results show the importance of the brain region (i.e., VTC) we used for feature selection.

[1] We discard "nonsense patterns" category in all our experiments.

[2] The atlases that we employed in this study are "Harvard-Oxford cortical and subcortical structural atlases" [26].

[3] These categories are the same categories used in the Haxby's fMRI dataset (excluding the non-sense pattern images).



(a) Brain-Features

(b) Adapted-Brain-Features

Fig. 2: Normalized Confusion Matrices. (a) Brain Features (before adaptation). (b) proposed method (Adapted-Brain-Features).

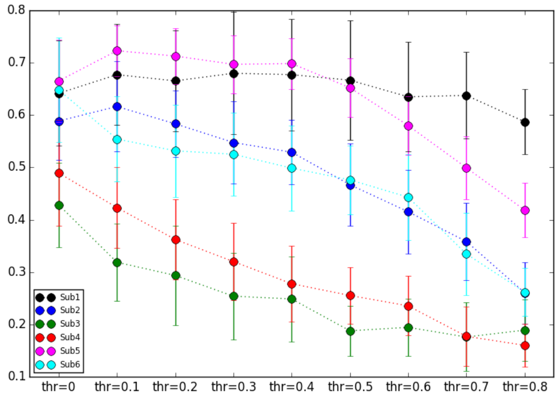


Fig. 3: Classification accuracy using different values of Atlas-threshold.

TABLE II: Average accuracy of all subject over all folds for each threshold)

Threshold	Accuracy
thr = 0.0	**0.58 ± 0.13**
thr = 0.1	0.55 ± 0.16
thr = 0.2	0.52 ± 0.17
thr = 0.3	0.50 ± 0.19
thr = 0.4	0.48 ± 0.19
thr = 0.5	0.45 ± 0.19
thr = 0.6	0.41 ± 0.18
thr = 0.7	0.36 ± 0.18
thr = 0.8	0.31 ± 0.16

3) Parameter sensitivity and convergence study: We set the regularization parameters in the range of $0.001, 0.01, 0.1, 1, 10, 100, 1000$. We present the parameter sensitivity of the proposed method in Fig.4. We fix $\lambda_3 = 0.1$, $\lambda_4 = 1$ (the values giving the best results in our experiments) and analyze the regularization parameters λ_1, λ_2 in Fig.4 (left). Meanwhile, we fix $\lambda_1 = 1$, $\lambda_2 = 1$ (the values giving the best results in our experiments) and analyze the regularization parameters λ_3, λ_4 in Fig.4 (middle). We observe that λ_3 and λ_4 are more sensitive compared with λ_1 and λ_2, which demonstrates the importance of the way to obtain the adapted features through the projection in the lower dimensional space.

We also analyze the convergence of our algorithm as shown in Fig.4 (right). We observe that our algorithm converges very fast (5 iterations).

V. Conclusions

In this paper, we proposed an adaptation framework in order to transfer the semantic representations learned on the visual domain to the brain domain. We showed that despite the big difference between these two modalities, the adaptation procedure led to improved results for the object classification task, outperforming the baseline method on the fMRI dataset. This is the first study in the direction of transferring object category knowledge from big visual datasets to the brain modality. We believe such domain adaptation approaches can improve the performance of the brain decoding algorithms.

However, the proposed method presents some limitations too. As mentioned in section IV-A, we performed our analysis on the Haxby dataset which is the only publicly available fMRI dataset for “object recognition”. Although our results show the increased performance on almost all subjects after adaptation, in order to confirm the efficacy of transferring knowledge from the visual domain to the brain modality, we need to apply this adaptation paradigm on other neuroimaging datasets on the same task. Unfortunately, in brain studies, the number of publicly available datasets is limited and consequently we did not find another fMRI dataset on the same object recognition task. For this, as our future plan, we will be exploring such adaptation procedure on other tasks (e.g. Action Recognition) by employing other neuroimaging

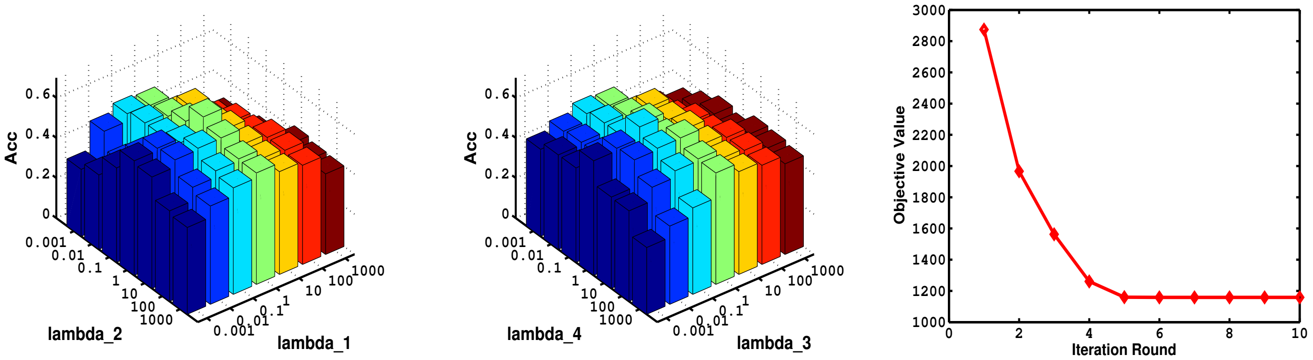


Fig. 4: (left) Sensitivity study of parameters on λ_1, λ_2 with fixed λ_3, λ_4. (middle) Sensitivity study of parameters on λ_3, λ_4 with fixed λ_1, λ_2. (right) Convergence of our algorithm with adapted features.

modalities (e.g. MEG and EEG).

References

[1] E. Klarreich, "Reading brains," *Communications of the ACM*, vol. 57, no. 3, pp. 12–14, 2014.

[2] M. Chen, J. Han, X. Hu, X. Jiang, L. Guo, and T. Liu, "Survey of encoding and decoding of visual stimulus via fmri: an image analysis perspective," *Brain imaging and behavior*, vol. 8, no. 1, pp. 7–23, 2014.

[3] N. Birbaumer, "Breaking the silence: brain–computer interfaces (BCI) for communication and motor control," *Psychophysiology*, vol. 43, no. 6, pp. 517–532, 2006.

[4] J. R. Wolpaw, D. J. McFarland, and T. M. Vaughan, "Brain-computer interface research at the wadsworth center," *IEEE Transactions on Rehabilitation Engineering*, vol. 8, no. 2, pp. 222–226, 2000.

[5] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, 2010.

[6] J. Donahue, Y. Jia, O. Vinyals, J. Hoffman, N. Zhang, E. Tzeng, and T. Darrell, "DeCAF: A deep convolutional activation feature for generic visual recognition," in *ICML*, 2014.

[7] E. Tzeng, J. Hoffman, T. Darrell, and K. Saenko, "Simultaneous deep transfer across domains and tasks," in *ICCV*, 2015.

[8] S. Gupta, J. Hoffman, and J. Malik, "Cross modal distillation for supervision transfer," *CVPR*, 2016.

[9] K. Grill-Spector and K. S. Weiner, "The functional architecture of the ventral temporal cortex and its role in categorization," *Nature Reviews Neuroscience*, vol. 15, no. 8, pp. 536–548, 2014.

[10] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. Berg, and L. Fei-Fei, "Imagenet large scale visual recognition challenge," *IJCV*, vol. 115, no. 3, pp. 211–252, 2015.

[11] W. Dai, Q. Yang, and Y. Yu, "Boosting for transfer learning," in *ICML*, 2007.

[12] H. Daume, "Frustratingly easy domain adaptation," in *ACL*, 2007.

[13] E. Bonilla, K. Chai, and C. Williams, "Multi-task gaussian process prediction," in *NIPS*, 2008.

[14] J. Yang, R. Yan, and A. G. Hauptmann, "Cross-domain video concept detection using adaptive SVMs," in *ACM Multimedia*, 2007.

[15] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *NIPS*, 2012.

[16] A. Argyriou, T. Evgeniou, and M. Pontil, "Convex multi-task feature learning," *Machine Learning*, vol. 73, no. 3, pp. 243–272, 2008.

[17] N. Srivastava and R. Salakhutdinov, "Multimodal learning with deep boltzmann machines," *Journal of Machine Learning Research*, vol. 15, pp. 2949–2980, 2014.

[18] H. Morioka, A. Kanemura, J.-i. Hirayama, M. Shikauchi, T. Ogawa, S. Ikeda, M. Kawanabe, and S. Ishii, "Learning a common dictionary for subject-transfer decoding with resting calibration," *NeuroImage*, vol. 111, pp. 167–178, 2015.

[19] V. Jayaram, M. Alamgir, Y. Altun, B. Schölkopf, and M. Grosse-Wentrup, "Transfer learning in brain-computer interfaces," *IEEE Computational Intelligence Magazine*, vol. 11, no. 1, pp. 20–31, 2016.

[20] S. Koyamada, Y. Shikauchi, K. Nakae, M. Koyama, and S. Ishii, "Deep learning of fmri big data: a novel approach to subject-transfer decoding," *arXiv:1502.00093*, 2015.

[21] B. A. Olshausen *et al.*, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature*, vol. 381, no. 6583, pp. 607–609, 1996.

[22] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online dictionary learning for sparse coding," in *ICML*, 2009.

[23] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 183–202, 2009.

[24] J. V. Haxby, M. I. Gobbini, M. L. Furey, A. Ishai, J. L. Schouten, and P. Pietrini, "Distributed and overlapping representations of faces and objects in ventral temporal cortex," *Science*, vol. 293, no. 5539, pp. 2425–2430, 2001.

[25] C. Chu, A.-L. Hsu, K.-H. Chou, P. Bandettini, C. Lin, A. D. N. Initiative *et al.*, "Does feature selection improve classification accuracy? impact of sample size and feature selection on classification using anatomical magnetic resonance images," *Neuroimage*, vol. 60, no. 1, pp. 59–70, 2012.

[26] R. S. Desikan, F. Ségonne, B. Fischl, B. T. Quinn, B. C. Dickerson, D. Blacker, R. L. Buckner, A. M. Dale, R. P. Maguire, B. T. Hyman, M. S. Albert, and R. J. Killiany, "An automated labeling system for subdividing the human cerebral cortex on mri scans into gyral based regions of interest," *Neuroimage*, vol. 31, no. 3, pp. 968–980, 2006.

[27] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, "Caffe: Convolutional architecture for fast feature embedding," in *ACM Multimedia*, 2014.

[28] M. N. Hebart, K. Görgen, and J.-D. Haynes, "The decoding toolbox (tdt): a versatile software package for multivariate analyses of functional imaging data," *Frontiers in Neuroinformatics*, vol. 8, p. 88, 2015.

[29] N. Kanwisher, J. McDermott, and M. M. Chun, "The fusiform face area: a module in human extrastriate cortex specialized for face perception," *The Journal of Neuroscience*, vol. 17, no. 11, pp. 4302–4311, 1997.

[30] N. Kanwisher and G. Yovel, "The fusiform face area: a cortical region specialized for the perception of faces," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 361, no. 1476, pp. 2109–2128, 2006.