ESTIMATING VIEWED IMAGE CATEGORIES FROM HUMAN BRAIN ACTIVITY VIA SEMI-SUPERVISED FUZZY DISCRIMINATIVE CANONICAL CORRELATION ANALYSIS

Yusuke Akamatsu, Ryosuke Harakawa, Takahiro Ogawa and Miki Haseyama

Graduate School of Information Science and Technology, Hokkaido University
N-14, W-9, Kita-ku, Sapporo, Hokkaido, 060-0814, Japan
E-mail: {akamatsu, harakawa, ogawa}@lmd.ist.hokudai.ac.jp, miki@ist.hokudai.ac.jp

ABSTRACT

This paper presents a method to estimate viewed image categories from human brain activity via newly derived semi-supervised fuzzy discriminative canonical correlation analysis (Semi-FDCCA). The proposed method can estimate image categories from functional magnetic resonance imaging (fMRI) activity measured while subjects view images by making fMRI activity and visual features obtained from images comparable through Semi-FDCCA. To realize Semi-FDCCA, we first derive a new supervised CCA called FDCCA that can consider fuzzy class information based on image category similarities obtained from WordNet ontology. Second, we adopt SemiCCA that can utilize additional unpaired visual features in addition to pairs of fMRI activity and visual features in order to prevent overfitting to the limited pairs. Furthermore, Semi-FDCCA can be derived by combining FDCCA with SemiCCA. Experimental results show that Semi-FDCCA enables accurate estimation of viewed image categories.

***Index Terms*—** Brain-computer interface (BCI), functional magnetic resonance imaging (fMRI), canonical correlation analysis.

1. INTRODUCTION

Daily actions such as grabbing objects with hands or communicating with others are not easy for people with severe physical disabilities, like amyotrophic lateral sclerosis (ALS) and spinal cord injury. A brain-computer interface (BCI), which enables people to send commands and messages to external computers through human brain activity [1, 2], can improve the quality of life of such people.

Toward the construction of BCI, several studies have attempted to control machines by using brain activity recorded by implantable microelectrode array (MEA) [3], electroencephalography (EEG) [4], and near-infrared spectroscopy (NIRS) [5]. Since implantable MEA belongs to invasive measurement methods, which use electrodes implanted in brains, it requires heavy burdens for people. On the other hand, EEG and NIRS are non-invasive measurement methods; however, these methods have low spatial resolutions [6]. In this situation, functional magnetic resonance imaging (fMRI), which is the non-invasive measuring method that has a high spatial resolution, is the mainstay of neuroimaging in cognitive neuroscience [7]. In particular, machine learning of fMRI activity has enabled the interpretation of cognitive states including what people see [8, 9] and imagine [10]. These studies estimate image categories (*e.g.*, faces, houses, and chairs) by using fMRI activity measured when people viewed or imagined images. Earlier studies [8–10] demonstrated that the feasibility of estimating viewed image categories by multivariate statistical analysis such as linear discriminant analysis (LDA) [11] and support vector machines (SVM) [12]. These methods classified patterns of fMRI activity evoked by the visual presentation of various image categories. Subsequently collected fMRI data were classified according to the similarities between their fMRI activity and that of prior training examples. However, these methods [8–10] cannot estimate novel image categories that were not presented in the training phase since the outputs are limited to the categories used for training of the classifiers.

Recent studies [13–15] have overcome this limitation by associating fMRI activity with visual features obtained from the viewed images by using a convolutional neural network (CNN) [16], a Gabor wavelet filter [17], and an HMAX model [18]. The method [14] trained linear regression models to predict visual features from fMRI activity measured while subjects viewed images; then the trained models can predict visual features from fMRI activity measured when viewing novel image categories that were not used in the training phase. In our previous work [15], we adopted canonical correlation analysis (CCA) [19] to associate fMRI activity with visual features. We calculate linear transformations that project fMRI activity and visual features into the same latent space through CCA; then we enable direct comparison between fMRI activity measured when viewing novel image categories and visual features. However, acquisition of fMRI data requires time and body burdens for subjects. Therefore, these methods [14, 15] can obtain only a small amount of fMRI data measured when viewing limited image categories. Hence, the above previous methods have the following two problems.

Problem (i): The performance of estimating image categories may be degraded when the number of pairs of fMRI activity and visual features is limited.

Problem (ii): It is difficult to accurately estimate image categories that were not used in the training phase.

In this paper, we propose a novel method that estimates viewed image categories from fMRI activity via semi-supervised fuzzy discriminative canonical correlation analysis (Semi-FDCCA). We solve the aforementioned problems by the following approaches.

Approach (i): To solve **Problem (i)**, we introduce a supervised scheme that can consider image category information to improve the performance of estimating image categories.

Approach (ii): To solve **Problem (ii)**, we enable utilization of many additional images that were not presented when measuring fMRI activity.

In **Approach (i)**, we derive novel CCA, called fuzzy discriminative CCA (FDCCA), which incorporates image category information into the framework of CCA. FDCCA aims to calculate projections to maximize the correlation of the projected samples according to the similarities between categories calculated on the basis of the on-

This work was partly supported by JSPS KAKENHI Grant Number JP17H01744 and the MIC/SCOPE #181601001.

tology of WordNet [20]. In **Approach (ii)**, we use semi-supervised CCA (SemiCCA) [21] in order to prevent projections from overfitting to the limited categories by utilizing many additional images. Furthermore, we newly derive Semi-FDCCA by combining FDCCA with SemiCCA. By using Semi-FDCCA, we can calculate projections that enable accurate estimation of novel image categories that were not used in the training phase. Experimental results show that both **Approaches (i) and (ii)** can improve the performance of estimating novel image categories.

The rest of this paper is organized as follows. In Section 2, CCA and some of its extensions are reviewed briefly. In Section 3, a method for estimating viewed image categories via Semi-FDCCA is presented. In Section 4, we present the experimental results. Conclusions of this paper are given in Section 5.

2. PRELIMINARY

2.1. Canonical Correlation Analysis

First, we explain canonical correlation analysis (CCA) [19]. Given a pair of matrices $\boldsymbol{X} = [\boldsymbol{x}_1, \cdots, \boldsymbol{x}_N] \in \mathbb{R}^{D_x \times N}$ and $\boldsymbol{Y} = [\boldsymbol{y}_1, \cdots, \boldsymbol{y}_N] \in \mathbb{R}^{D_y \times N}$, where each sample is a vector with dimensions D_x and D_y, and N is the number of samples. We assume that $\boldsymbol{X}$ and $\boldsymbol{Y}$ are both centered. The aim of CCA is to find a pair of projection vectors $\boldsymbol{w}_x \in \mathbb{R}^{D_x}$ and $\boldsymbol{w}_y \in \mathbb{R}^{D_y}$ to maximize the correlation between $\boldsymbol{w}_x^{\mathrm{T}}\boldsymbol{X}$ and $\boldsymbol{w}_y^{\mathrm{T}}\boldsymbol{Y}$ as follows:

$$\begin{aligned} &\max_{\boldsymbol{w}_x, \boldsymbol{w}_y} \frac{\boldsymbol{w}_x^{\mathrm{T}}\boldsymbol{X}\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{w}_y}{\sqrt{\boldsymbol{w}_x^{\mathrm{T}}\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{w}_x}\sqrt{\boldsymbol{w}_y^{\mathrm{T}}\boldsymbol{Y}\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{w}_y}} \\ &\text{s.t.} \quad \boldsymbol{w}_x^{\mathrm{T}}\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{w}_x = \boldsymbol{w}_y^{\mathrm{T}}\boldsymbol{Y}\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{w}_y = 1. \end{aligned} \tag{1}$$

The solution of formula (1) can be obtained by solving the following generalized eigenvalue problem:

$$\begin{bmatrix} \boldsymbol{0} & \boldsymbol{X}\boldsymbol{Y}^{\mathrm{T}} \\ \boldsymbol{Y}\boldsymbol{X}^{\mathrm{T}} & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{w}_x \\ \boldsymbol{w}_y \end{bmatrix} = \lambda \begin{bmatrix} \boldsymbol{X}\boldsymbol{X}^{\mathrm{T}} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Y}\boldsymbol{Y}^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} \boldsymbol{w}_x \\ \boldsymbol{w}_y \end{bmatrix}, \tag{2}$$

where λ is the eigenvalue.

2.2. Discriminative Canonical Correlation Analysis

Next, we explain discriminative canonical correlation analysis (DCCA) [22]. DCCA is a supervised CCA that utilizes the class information. DCCA aims to find projection vectors $\boldsymbol{w}_x \in \mathbb{R}^{D_x}$ and $\boldsymbol{w}_y \in \mathbb{R}^{D_y}$ to ensure that the within-class correlation between $\boldsymbol{w}_x^{\mathrm{T}}\boldsymbol{x}$ and $\boldsymbol{w}_y^{\mathrm{T}}\boldsymbol{y}$ is maximized, while the between-class correlation between $\boldsymbol{w}_x^{\mathrm{T}}\boldsymbol{x}$ and $\boldsymbol{w}_y^{\mathrm{T}}\boldsymbol{y}$ is minimized as follows:

$$\begin{aligned} &\max_{\boldsymbol{w}_x, \boldsymbol{w}_y} \frac{\boldsymbol{w}_x^{\mathrm{T}}\boldsymbol{C}_w\boldsymbol{w}_y - \eta \cdot \boldsymbol{w}_x^{\mathrm{T}}\boldsymbol{C}_b\boldsymbol{w}_y}{\sqrt{\boldsymbol{w}_x^{\mathrm{T}}\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{w}_x}\sqrt{\boldsymbol{w}_y^{\mathrm{T}}\boldsymbol{Y}\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{w}_y}} \\ &\text{s.t.} \quad \boldsymbol{w}_x^{\mathrm{T}}\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{w}_x = \boldsymbol{w}_y^{\mathrm{T}}\boldsymbol{Y}\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{w}_y = 1, \end{aligned} \tag{3}$$

where η is a parameter, and $\boldsymbol{C}_w$ and $\boldsymbol{C}_b$ denote covariance matrices of the within-class samples and the between-class samples, respectively. $\boldsymbol{C}_w$ and $\boldsymbol{C}_b$ are defined as

$$\boldsymbol{C}_w = \sum_{i=1}^{C}\sum_{k=1}^{N_i}\sum_{l=1}^{N_i} \boldsymbol{x}_k^{(i)}\boldsymbol{y}_l^{(i)\mathrm{T}}, \boldsymbol{C}_b = \sum_{i=1}^{C}\sum_{j=1, j\neq i}^{C}\sum_{k=1}^{N_i}\sum_{l=1}^{N_j} \boldsymbol{x}_k^{(i)}\boldsymbol{y}_l^{(j)\mathrm{T}},$$

where $\boldsymbol{x}_j^{(i)}$ and $\boldsymbol{y}_j^{(i)}$ denote jth samples in ith class, N_i is the number of samples belonging to ith class, and C is the number of classes. The solution of formula (3) can be obtained by solving the following generalized eigenvalue problem:

$$\begin{bmatrix} \boldsymbol{0} & \bar{\boldsymbol{C}}_{xy} \\ (\bar{\boldsymbol{C}}_{xy})^{\mathrm{T}} & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{w}_x \\ \boldsymbol{w}_y \end{bmatrix} = \lambda \begin{bmatrix} \boldsymbol{X}\boldsymbol{X}^{\mathrm{T}} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Y}\boldsymbol{Y}^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} \boldsymbol{w}_x \\ \boldsymbol{w}_y \end{bmatrix}, \tag{4}$$

where λ is the eigenvalue, and $\bar{\boldsymbol{C}}_{xy} = \boldsymbol{C}_w - \eta \cdot \boldsymbol{C}_b$.

2.3. Semi-supervised Canonical Correlation Analysis

Furthermore, we explain semi-supervised canonical correlation analysis (SemiCCA) [21]. In addition to a pair $\boldsymbol{X} \in \mathbb{R}^{D_x \times N}$ and $\boldsymbol{Y} \in \mathbb{R}^{D_y \times N}$, SemiCCA utilizes unpaired samples $\boldsymbol{Y}_U \in \mathbb{R}^{D_y \times M}$, where M is the number of unpaired samples. SemiCCA combines CCA and principal component analysis (PCA) [23]. In SemiCCA, CCA utilizes only paired samples, while PCA also utilizes unpaired samples to reveal the global structure. The solution of SemiCCA can be obtained by solving the following generalized eigenvalue problem:

$$\overline{\boldsymbol{B}} \begin{bmatrix} \boldsymbol{w}_x \\ \boldsymbol{w}_y \end{bmatrix} = \lambda \underline{\boldsymbol{B}} \begin{bmatrix} \boldsymbol{w}_x \\ \boldsymbol{w}_y \end{bmatrix}, \tag{5}$$

where

$$\overline{\boldsymbol{B}} = \beta \begin{bmatrix} \boldsymbol{0} & \boldsymbol{C}_{P_{xy}} \\ (\boldsymbol{C}_{P_{xy}})^{\mathrm{T}} & \boldsymbol{0} \end{bmatrix} + (1-\beta) \begin{bmatrix} \boldsymbol{C}_{P_{xx}} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{C}_{yy} \end{bmatrix}, \tag{6}$$

$$\underline{\boldsymbol{B}} = \beta \begin{bmatrix} \boldsymbol{C}_{P_{xx}} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{C}_{P_{yy}} \end{bmatrix} + (1-\beta) \begin{bmatrix} \boldsymbol{I}_{D_x} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_{D_y} \end{bmatrix}. \tag{7}$$

$\boldsymbol{C}_{P_{xy}}$, $\boldsymbol{C}_{P_{xx}}$, and $\boldsymbol{C}_{P_{yy}}$ are covariance matrices of paired samples $\boldsymbol{C}_{P_{xy}} = \boldsymbol{X}\boldsymbol{Y}^{\mathrm{T}}/N$, $\boldsymbol{C}_{P_{xx}} = \boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}/N$, $\boldsymbol{C}_{P_{yy}} = \boldsymbol{Y}\boldsymbol{Y}^{\mathrm{T}}/N$, and $\boldsymbol{C}_{yy}$ is covariance matrices of paired and unpaired samples $\boldsymbol{C}_{yy} = (\boldsymbol{Y}\boldsymbol{Y}^{\mathrm{T}} + \boldsymbol{Y}_U\boldsymbol{Y}_U^{\mathrm{T}})/(N+M)$. In Eqs. (6) and (7), β is a parameter, which controls the trade-off between CCA and PCA. Namely, when $\beta = 1$, Eq. (5) is reduced to the CCA eigenvalue problem shown in Eq. (2), while when $\beta = 0$, Eq. (5) is reduced to the PCA eigenvalue problem.

3. ESTIMATION OF VIEWED IMAGE CATEGORIES FROM FMRI ACTIVITY VIA SEMI-FDCCA

Our method consists of the training phase (**3.1**) and test phase (**3.2**). In the training phase, we calculate linear transformations that project fMRI activity and visual features into the same latent space via a novel canonical correlation analysis, Semi-FDCCA. In the test phase, we estimate viewed image categories from fMRI activity.

3.1. Training Phase

First, we obtain fMRI data measured when viewing images; however, the presence of many irrelevant fMRI features for estimation could lead to poor performance. To solve this problem, we use sparse logistic regression (SLR) [24], which automatically selects relevant fMRI features. SLR is often used in fMRI activity analysis [14, 25, 26]. By selecting fMRI features via SLR, we obtain fMRI activity features $\boldsymbol{X} = [\boldsymbol{x}_1, \cdots, \boldsymbol{x}_N] \in \mathbb{R}^{D_x \times N}$, where each sample is a vector with dimension D_x, and N is the number of samples. Also, we extract visual features from images presented when measuring fMRI activity. We used visual features obtained from the third pooling layer of Inception-v3 architecture [27], which was trained with images in ImageNet [28] to classify image categories. Due to its high discriminant power, features obtained from Inception-v3 are one of the most commonly used visual features [29, 30]. We reduce dimensions of visual features by PCA to prevent Semi-FDCCA from overfitting; then we obtain visual features $\boldsymbol{Y} = [\boldsymbol{y}_1, \cdots, \boldsymbol{y}_N] \in \mathbb{R}^{D_y \times N}$, where each sample is a vector with dimension D_y. We assume that $\boldsymbol{X}$ and $\boldsymbol{Y}$ are both centered. Here, we find a pair of projection vectors $\boldsymbol{w}_x \in \mathbb{R}^{D_x}$ and $\boldsymbol{w}_y \in \mathbb{R}^{D_y}$ via FDCCA. FDCCA is derived from the idea of DCCA (*see* **2.2**), and FDCCA aims to calculate projections to maximize the correlation of the projected samples according to the similarities between categories as follows:

$$\max_{\boldsymbol{w}_x, \boldsymbol{w}_y} \frac{\boldsymbol{w}_x^{\mathrm{T}} \boldsymbol{C}_h \boldsymbol{w}_y - \alpha \cdot \boldsymbol{w}_x^{\mathrm{T}} \boldsymbol{C}_l \boldsymbol{w}_y}{\sqrt{\boldsymbol{w}_x^{\mathrm{T}} \boldsymbol{X}\boldsymbol{X}^{\mathrm{T}} \boldsymbol{w}_x}\sqrt{\boldsymbol{w}_y^{\mathrm{T}} \boldsymbol{Y}\boldsymbol{Y}^{\mathrm{T}} \boldsymbol{w}_y}} \quad \text{s.t.} \quad \boldsymbol{w}_x^{\mathrm{T}} \boldsymbol{X}\boldsymbol{X}^{\mathrm{T}} \boldsymbol{w}_x = \boldsymbol{w}_y^{\mathrm{T}} \boldsymbol{Y}\boldsymbol{Y}^{\mathrm{T}} \boldsymbol{w}_y = 1, \tag{8}$$

where α is a parameter, and $\boldsymbol{C}_h$ and $\boldsymbol{C}_l$ denote covariance matrices of similar samples and dissimilar samples, respectively. For calculating $\boldsymbol{C}_h$ and $\boldsymbol{C}_l$, we define

$$\boldsymbol{X} = [\boldsymbol{x}_1^{(1)}, \cdots, \boldsymbol{x}_{N_1}^{(1)}, \cdots\cdots, \boldsymbol{x}_1^{(C)}, \cdots, \boldsymbol{x}_{N_C}^{(C)}] \in \mathbb{R}^{D_x \times N},$$

$$\boldsymbol{Y} = [\boldsymbol{y}_1^{(1)}, \cdots, \boldsymbol{y}_{N_1}^{(1)}, \cdots\cdots, \boldsymbol{y}_1^{(C)}, \cdots, \boldsymbol{y}_{N_C}^{(C)}] \in \mathbb{R}^{D_y \times N},$$

$$\boldsymbol{e}_{N_i} = [\underbrace{0 \cdots 0}_{\sum_{t=1}^{i-1} N_t}, \underbrace{1 \cdots 1}_{N_i}, \underbrace{0 \cdots 0}_{N - \sum_{t=1}^{i} N_t}]^{\mathrm{T}} \in \mathbb{R}^N,$$

$$\boldsymbol{e}_{N_j} = [\underbrace{0 \cdots 0}_{\sum_{t=1}^{j-1} N_t}, \underbrace{1 \cdots 1}_{N_j}, \underbrace{0 \cdots 0}_{N - \sum_{t=1}^{j} N_t}]^{\mathrm{T}} \in \mathbb{R}^N,$$

$$\boldsymbol{S} = \begin{bmatrix} \boldsymbol{S}_{11} & \boldsymbol{S}_{12} & \cdots & \boldsymbol{S}_{1C} \\ \boldsymbol{S}_{21} & \boldsymbol{S}_{22} & \cdots & \boldsymbol{S}_{2C} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{S}_{C1} & \boldsymbol{S}_{C2} & \cdots & \boldsymbol{S}_{CC} \end{bmatrix} \in \mathbb{R}^{N \times N},$$

$$\boldsymbol{S}_{ij} = \begin{bmatrix} S(i,j) & \cdots & S(i,j) \\ \vdots & \ddots & \vdots \\ S(i,j) & \cdots & S(i,j) \end{bmatrix} \in \mathbb{R}^{N_i \times N_j},$$

where $\boldsymbol{x}_t^{(i)}$ and $\boldsymbol{y}_t^{(i)}$ denote tth samples in ith category, N_i is the number of samples in ith category, and C is the number of categories. $S(i,j)$ represents the similarity between ith category and jth category. For the similarities between categories, we used path similarity [31], which denotes the similarity between two categories based on the shortest path that connects the categories on the ontology of WordNet [20]. Then $\boldsymbol{C}_h$ and $\boldsymbol{C}_l$ are defined as

$$\begin{aligned} \boldsymbol{C}_h &= \sum_{i=1}^{C}\sum_{j=1}^{C} \overline{S(i,j)} \sum_{k=1}^{N_i}\sum_{l=1}^{N_j} \boldsymbol{x}_k^{(i)} \boldsymbol{y}_l^{(j)\mathrm{T}} \\ &= \sum_{i=1}^{C}\sum_{j=1}^{C} \overline{S(i,j)} (\boldsymbol{X}\boldsymbol{e}_{N_i})(\boldsymbol{Y}\boldsymbol{e}_{N_j})^{\mathrm{T}} \\ &= \boldsymbol{X}\overline{\boldsymbol{S}}\boldsymbol{Y}^{\mathbf{T}}, \end{aligned}$$

$$\begin{aligned} \boldsymbol{C}_l &= \sum_{i=1}^{C}\sum_{j=1}^{C} \underline{S(i,j)} \sum_{k=1}^{N_i}\sum_{l=1}^{N_j} \boldsymbol{x}_k^{(i)} \boldsymbol{y}_l^{(j)\mathrm{T}} \\ &= \sum_{i=1}^{C}\sum_{j=1}^{C} \underline{S(i,j)} (\boldsymbol{X}\boldsymbol{e}_{N_i})(\boldsymbol{Y}\boldsymbol{e}_{N_j})^{\mathrm{T}} \\ &= \boldsymbol{X}\underline{\boldsymbol{S}}\boldsymbol{Y}^{\mathbf{T}}, \end{aligned}$$

where

$$\overline{S(i,j)} = \begin{cases} S(i,j) & (S(i,j) \geq \theta_a) \\ 0 & (\text{otherwise}) \end{cases},$$

$$\underline{S(i,j)} = \begin{cases} S(i,j) & (S(i,j) \leq \theta_b) \\ 0 & (\text{otherwise}) \end{cases},$$

$$\overline{\boldsymbol{S}} = \begin{bmatrix} \overline{\boldsymbol{S}_{11}} & \overline{\boldsymbol{S}_{12}} & \cdots & \overline{\boldsymbol{S}_{1C}} \\ \overline{\boldsymbol{S}_{21}} & \overline{\boldsymbol{S}_{22}} & \cdots & \overline{\boldsymbol{S}_{2C}} \\ \vdots & \vdots & \ddots & \vdots \\ \overline{\boldsymbol{S}_{C1}} & \overline{\boldsymbol{S}_{C2}} & \cdots & \overline{\boldsymbol{S}_{CC}} \end{bmatrix}, \ \underline{\boldsymbol{S}} = \begin{bmatrix} \underline{\boldsymbol{S}_{11}} & \underline{\boldsymbol{S}_{12}} & \cdots & \underline{\boldsymbol{S}_{1C}} \\ \underline{\boldsymbol{S}_{21}} & \underline{\boldsymbol{S}_{22}} & \cdots & \underline{\boldsymbol{S}_{2C}} \\ \vdots & \vdots & \ddots & \vdots \\ \underline{\boldsymbol{S}_{C1}} & \underline{\boldsymbol{S}_{C2}} & \cdots & \underline{\boldsymbol{S}_{CC}} \end{bmatrix},$$

$$\overline{\boldsymbol{S}_{ij}} = \begin{bmatrix} \overline{S(i,j)} & \cdots & \overline{S(i,j)} \\ \vdots & \ddots & \vdots \\ \overline{S(i,j)} & \cdots & \overline{S(i,j)} \end{bmatrix}, \ \underline{\boldsymbol{S}_{ij}} = \begin{bmatrix} \underline{S(i,j)} & \cdots & \underline{S(i,j)} \\ \vdots & \ddots & \vdots \\ \underline{S(i,j)} & \cdots & \underline{S(i,j)} \end{bmatrix}.$$

Here, θ_a and θ_b are thresholds that distinguish similar samples from dissimilar samples. The solution of formula (8) can be obtained by solving the following generalized eigenvalue problem:

$$\begin{bmatrix} \mathbf{0} & \hat{\boldsymbol{C}}_{xy} \\ (\hat{\boldsymbol{C}}_{xy})^{\mathrm{T}} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{w}_x \\ \boldsymbol{w}_y \end{bmatrix} = \lambda \begin{bmatrix} \boldsymbol{X}\boldsymbol{X}^{\mathrm{T}} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{Y}\boldsymbol{Y}^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} \boldsymbol{w}_x \\ \boldsymbol{w}_y \end{bmatrix}, \tag{9}$$

where λ is the eigenvalue, and $\hat{\boldsymbol{C}}_{xy} = \boldsymbol{X}\overline{\boldsymbol{S}}\boldsymbol{Y}^{\mathbf{T}} - \alpha \cdot \boldsymbol{X}\underline{\boldsymbol{S}}\boldsymbol{Y}^{\mathbf{T}}$. By considering the similarity between categories when calculating the projections $\boldsymbol{w_x}$ and $\boldsymbol{w_y}$, FDCCA obtains projections that can discriminate categories more precisely than DCCA.

Furthermore, we combine FDCCA with SemiCCA (*see* **2.3**) to prevent projections $\boldsymbol{w_x}$ and $\boldsymbol{w_y}$ from overfitting to the limited categories used in the training phase. For this purpose, in addition to a pair $\boldsymbol{X}$ and $\boldsymbol{Y}$, we extract unpaired visual features $\boldsymbol{Y}_U = [\boldsymbol{y}_{U,1}, \cdots, \boldsymbol{y}_{U,M}] \in \mathbb{R}^{D_y \times M}$ (M being the number of unpaired images) from images that were not used when measuring fMRI activity. We assume that the dimension of $\boldsymbol{Y}_U$ is reduced by PCA, and $\boldsymbol{Y}_U$ is centered. We rearrange the eigenvalue problem in Eq. (9) via SemiCCA as follows:

$$\overline{\boldsymbol{B}} \begin{bmatrix} \boldsymbol{w}_x \\ \boldsymbol{w}_y \end{bmatrix} = \lambda \underline{\boldsymbol{B}} \begin{bmatrix} \boldsymbol{w}_x \\ \boldsymbol{w}_y \end{bmatrix}, \tag{10}$$

where

$$\overline{\boldsymbol{B}} = \beta \begin{bmatrix} \mathbf{0} & \boldsymbol{C}_{P_{xy}} \\ (\boldsymbol{C}_{P_{xy}})^{\mathrm{T}} & \mathbf{0} \end{bmatrix} + (1-\beta) \begin{bmatrix} \boldsymbol{C}_{P_{xx}} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{C}_{yy} \end{bmatrix},$$

$$\underline{\boldsymbol{B}} = \beta \begin{bmatrix} \boldsymbol{C}_{P_{xx}} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{C}_{P_{yy}} \end{bmatrix} + (1-\beta) \begin{bmatrix} \boldsymbol{I}_{D_x} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I}_{D_y} \end{bmatrix}.$$

$\boldsymbol{C}_{P_{xy}}$, $\boldsymbol{C}_{P_{xx}}$, and $\boldsymbol{C}_{P_{yy}}$ are covariance matrices of paired samples $\boldsymbol{C}_{P_{xy}} = \hat{\boldsymbol{C}}_{xy}/N$, $\boldsymbol{C}_{P_{xx}} = \boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}/N$, $\boldsymbol{C}_{P_{yy}} = \boldsymbol{Y}\boldsymbol{Y}^{\mathrm{T}}/N$, and $\boldsymbol{C}_{yy}$ is covariance matrices of all visual features $\boldsymbol{C}_{yy} = (\boldsymbol{Y}\boldsymbol{Y}^{\mathrm{T}} + \boldsymbol{Y}_U\boldsymbol{Y}_U^{\mathrm{T}})/(N+M)$. By solving the eigenvalue problem in Eq. (10), we obtain $\boldsymbol{W_x} = [\boldsymbol{w}_{x1}, \cdots, \boldsymbol{w}_{xD}]$ and $\boldsymbol{W_y} = [\boldsymbol{w}_{y1}, \cdots, \boldsymbol{w}_{yD}]$ including projection vectors corresponding to the D largest eigenvalues $\boldsymbol{\lambda} = [\lambda_1, \cdots, \lambda_D]$, where D satisfies $D \leq \min(D_x, D_y)$ and $D \leq C$. By using the newly derived Semi-FDCCA, we calculate projections that can **(i)** consider fuzzy class information, *i.e.*, the similarities between categories and **(ii)** improve the generalization capability by utilizing unpaired visual features.

3.2. Test Phase

We estimate image categories from fMRI activity measured when subjects viewed test images, *i.e.*, unobserved images in the training phase. By using fMRI features selected in the training phase via SLR, we obtain an fMRI activity feature $\boldsymbol{x}_{\mathrm{test}} \in \mathbb{R}^{D_x}$ from fMRI activity measured when subjects viewed a test image. We also extract visual features $\boldsymbol{y}_{\mathrm{test},i} \in \mathbb{R}^{D_y}$ ($i = 1, 2, \cdots, C_{\mathrm{test}}$; C_{test} being the number of test image categories) obtained by averaging the visual features belonging to ith test image category. We assume that

Table 1. Average ranks of correctly estimated image categories for 50 test images. It was confirmed that the differences between ranks of correctly estimated image categories for all test images via Ours (Semi-FDCCA) and those via Refs. [14], [15], DCCA [22], and FDCCA was statistically significant with $P < 0.01$ by Welch's t-test [32]. In addition, it was significant with $P < 0.3$ between Ours and SemiCCA [21].

	Ref. [14]	Ref. [15] (CCA [19])	DCCA [22]	FDCCA	SemiCCA [21]	**Ours (Semi-FDCCA)**
Subject 1	15.86	17.52	17.38	16.94	15.94	**14.40**
Subject 2	15.78	16.82	13.62	13.22	13.16	**11.68**
Subject 3	13.60	17.24	13.88	13.24	10.46	**8.84**
Subject 4	14.58	13.34	12.42	12.46	10.76	**9.90**
Subject 5	14.90	13.04	13.08	12.40	**9.44**	10.20
Mean $\pm$ SD	14.94 ± 0.93	15.59 ± 2.21	14.08 ± 1.93	13.65 ± 1.88	11.95 ± 2.61	$\mathbf{11.00 \pm 2.15}$

the dimension of $\boldsymbol{y}_{\text{test},i}$ is reduced by PCA, also $\boldsymbol{x}_{\text{test}}$ and $\boldsymbol{y}_{\text{test},i}$ are both centered by using the means of $\boldsymbol{X}$ and $\boldsymbol{Y}$, respectively. Next, we calculate new features $\boldsymbol{u}$ and $\boldsymbol{v}_i$ on the same latent space by using projections as follows:

$$\boldsymbol{u} = \boldsymbol{\Lambda} \boldsymbol{W}_x^{\mathrm{T}} \boldsymbol{x}_{\text{test}} \in \mathbb{R}^D, \quad \boldsymbol{v}_i = \boldsymbol{W}_y^{\mathrm{T}} \boldsymbol{y}_{\text{test},i} \in \mathbb{R}^D,$$

where $\boldsymbol{\Lambda}$ denotes diagonal matrix whose diagonal elements are eigenvalues $\boldsymbol{\lambda} = [\lambda_1, \cdots, \lambda_D]$. We enable accurate comparison between $\boldsymbol{u}$ and $\boldsymbol{v}_i$ by using Semi-FDCCA. Finally, we calculate correlation coefficient between $\boldsymbol{u}$ and $\boldsymbol{v}_i$. By ranking categories in the descending order of the correlation coefficient, estimation of image categories from fMRI activity becomes feasible.

4. EXPERIMENTAL RESULTS

4.1. Dataset

We used public fMRI dataset provided in the previous study [14]. This dataset contains fMRI activity measured from five subjects while viewing images collected from ImageNet [28]. Categories corresponding to the ontology of WordNet [20] are attached to each image. In the experiment in the previous study [14], a total of 1,200 images from 150 categories (eight images from each category) were presented to each subject in the training phase, and a total of 50 images from 50 categories (one image from each category) were presented to each subject in the test phase[1]. Note that test image categories were not used in the training phase. We calculated test visual features $\boldsymbol{y}_{\text{test},i}$ by using the average of visual features obtained from 100 images that belong to ith test image category on ImageNet. Furthermore, we utilized 10,000 images from Tiny ImageNet dataset[2] as unpaired images for calculating projections via Semi-FDCCA.

4.2. Experimental Conditions

In this experiment, we empirically used 150-dimensional fMRI activity features selected via SLR for each subject. We also used 203-dimensional visual features after reducing dimensions via PCA so that the cumulative contribution ratio becomes more than 80%. We compared our proposed Semi-FDCCA with the following five methods: Ref. [14] (linear regression), Ref. [15] (CCA [19]), DCCA [22], FDCCA, and SemiCCA [21]. We set $\alpha \in \{0, 0.01, \cdots, 1\}$ and $\beta \in \{0, 0.01, \cdots, 1\}$ in FDCCA, SemiCCA, and Semi-FDCCA. Also, we empirically set thresholds $\theta_a = 0.5$ and $\theta_b = 0.1$ in FDCCA and Semi-FDCCA.

4.3. Results and Discussion

Figure 1 shows examples of rankings of estimated image categories by our proposed Semi-FDCCA, and Table 1 shows average ranks of correctly estimated image categories for 50 test images. Note that we determined α and β as values that achieved the highest performance in the validation using the training dataset. First, by comparing the results by FDCCA with Ref. [14] and Ref. [15] (CCA), we can confirm the effectiveness of **Approach (i)**: the introduction of a supervised scheme that can consider image category information. Moreover, we can confirm the effectiveness of considering the similarities between categories by comparing the results by FDCCA with DCCA. Second, from the results by SemiCCA, Ref. [14] and Ref. [15] (CCA), we can confirm the effectiveness of **Approach (ii)**: the improvement of generalization capability by utilizing many additional images. Furthermore, we can see that the collaborative use of **Approaches (i) and (ii)** improves the estimating performance since our proposed Semi-FDCCA outperforms the other five methods. The above experimental results have shown that our proposed Semi-FDCCA enables accurate estimation of novel image categories that were not used in the training phase.

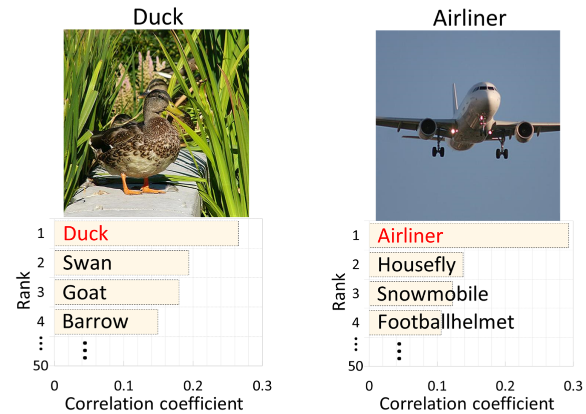

Fig. 1. Rankings of estimated image categories for test images ('Duck' and 'Airliner'). The horizontal line shows correlation coefficient between $\boldsymbol{u}$ and $\boldsymbol{v}_i$. The vertical one shows ranks of estimated categories, and correct categories are indicated in red.

5. CONCLUSIONS

In this paper, we proposed a novel method for estimating viewed image categories from fMRI activity via Semi-FDCCA. We combined FDCCA that can consider fuzzy class information, $i.e.$, category similarities with SemiCCA that can utilize many additional visual features; then Semi-FDCCA was newly derived. Consequently, experimental results showed that our proposed Semi-FDCCA can accurately estimate novel image categories that were not used in the training phase.

[1] Since the previous study [14] measured fMRI activity for each test image 35 times, we used fMRI activity averaged across all trials.

[2] https://tiny-imagenet.herokuapp.com/

6. REFERENCES

[1] M. A. Nicolelis, "Actions from thoughts," *Nature*, vol. 409, no. 6818, pp. 403–407, 2001.

[2] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, et al., "Brain–computer interfaces for communication and control," *Clinical neurophysiology*, vol. 113, no. 6, pp. 767–791, 2002.

[3] L. R. Hochberg, D. Bacher, B. Jarosiewicz, N. Y. Masse, et al., "Reach and grasp by people with tetraplegia using a neurally controlled robotic arm," *Nature*, vol. 485, no. 7398, pp. 372–377, 2012.

[4] K. Müller, M. Tangermann, G. Dornhege, M. Krauledat, et al., "Machine learning for real-time single-trial EEG-analysis: from brain–computer interfacing to mental state monitoring," *Journal of neuroscience methods*, vol. 167, no. 1, pp. 82–90, 2008.

[5] S. Fazli, J. Mehnert, J. Steinbrink, G. Curio, et al., "Enhanced performance by a hybrid NIRS–EEG brain computer interface," *Neuroimage*, vol. 59, no. 1, pp. 519–529, 2012.

[6] B. K. Min, M. J. Marzelli, and S. S. Yoo, "Neuroimaging-based approaches in the brain–computer interface," *Trends in biotechnology*, vol. 28, no. 11, pp. 552–560, 2010.

[7] N. K. Logothetis, "What we can do and what we cannot do with fMRI," *Nature*, vol. 453, no. 7197, pp. 869–878, 2008.

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

[9] D. D. Cox and R. L. Savoy, "Functional magnetic resonance imaging (fMRI) " brain reading ": detecting and classifying distributed patterns of fMRI activity in human visual cortex," *Neuroimage*, vol. 19, no. 2, pp. 261–270, 2003.

[10] L. Reddy, N. Tsuchiya, and T. Serre, "Reading the mind's eye: decoding category information during mental imagery," *Neuroimage*, vol. 50, no. 2, pp. 818–825, 2010.

[11] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Annals of eugenics*, vol. 7, no. 2, pp. 179–188, 1936.

[12] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.

[13] K. N. Kay, T. Naselaris, R. J. Prenger, and J. L. Gallant, "Identifying natural images from human brain activity," *Nature*, vol. 452, no. 7185, pp. 352–355, 2008.

[14] T. Horikawa and Y. Kamitani, "Generic decoding of seen and imagined objects using hierarchical visual features," *Nature communications*, vol. 8, no. 15037, pp. 1–15, 2017.

[15] Y. Akamatsu, R. Harakawa, T. Ogawa, and M. Haseyama, "Estimation of viewed image categories via CCA using human brain activity," in *Proc. IEEE Global Conf. Consumer Electronics (GCCE)*, 2018, pp. 171–172.

[16] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Advances in Neural Information Processing Systems (NIPS)*, 2012, pp. 1097–1105.

[17] J. P. Jones and L. A. Palmer, "An evaluation of the two-dimensional gabor filter model of simple receptive fields in cat striate cortex," *Journal of neurophysiology*, vol. 58, no. 6, pp. 1233–1258, 1987.

[18] M. Riesenhuber and T. Poggio, "Hierarchical models of object recognition in cortex," *Nature neuroscience*, vol. 2, no. 11, pp. 1019–1025, 1999.

[19] H. Hotelling, "Relations between two sets of variates," *Biometrika*, vol. 28, no. 3/4, pp. 321–377, 1936.

[20] G. Miller, *WordNet: An electronic lexical database*, MIT press, 1998.

[21] A. Kimura, M. Sugiyama, T. Nakano, H. Kameoka, et al., "SemiCCA: Efficient semi-supervised learning of canonical correlations," *Information and Media Technologies*, vol. 8, no. 2, pp. 311–318, 2013.

[22] T. Sun, S. Chen, J. Yang, and P. Shi, "A novel method of combined feature extraction for recognition," in *Proc. IEEE Int. Conf. Data Mining (ICDM)*, 2008, pp. 1043–1048.

[23] S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," *Chemometrics and intelligent laboratory systems*, vol. 2, no. 1-3, pp. 37–52, 1987.

[24] O. Yamashita, M. Sato, T. Yoshioka, F. Tong, et al., "Sparse estimation automatically selects voxels relevant for the decoding of fMRI activity patterns," *NeuroImage*, vol. 42, no. 4, pp. 1414–1429, 2008.

[25] G. Ganesh, E. Burdet, M. Haruno, and M. Kawato, "Sparse linear regression for reconstructing muscle activity from human cortical fMRI," *Neuroimage*, vol. 42, no. 4, pp. 1463–1472, 2008.

[26] Y. Miyawaki, H. Uchida, O. Yamashita, M. Sato, et al., "Visual image reconstruction from human brain activity using a combination of multiscale local image decoders," *Neuron*, vol. 60, no. 5, pp. 915–929, 2008.

[27] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, et al., "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 2818–2826.

[28] J. Deng, W. Dong, R. Socher, L. Li, et al., "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2009, pp. 248–255.

[29] S. Hershey, S. Chaudhuri, D. P. Ellis, J. F. Gemmeke, et al., "CNN architectures for large-scale audio classification," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, 2017, pp. 131–135.

[30] X. Xia, C. Xu, and B. Nan, "Inception-v3 for flower classification," in *Proc. IEEE Int. Conf. Image, Vision and Computing (ICIVC)*, 2017, pp. 783–787.

[31] T. Pedersen, S. Patwardhan, and J. Michelizzi, "Wordnet::similarity: Measuring the relatedness of concepts," in *Proc. Demonstration Papers at HLT-NAACL*, 2004, pp. 38–41.

[32] B. L. Welch, "The generalization of 'student's' problem when several different population variances are involved," *Biometrika*, vol. 34, no. 1/2, pp. 28–35, 1947.