

Article

Brain Decoding of Multiple Subjects for Estimating Visual Information Based on a Probabilistic Generative Model

Takaaki Higashi ¹, Keisuke Maeda ², Takahiro Ogawa ² and Miki Haseyama ^{2,*}

¹ Graduate School of Information Science and Technology, Hokkaido University, N-14, W-9, Kita-ku, Sapporo 060-0814, Hokkaido, Japan

² Faculty of Information Science and Technology, Hokkaido University, N-14, W-9, Kita-ku, Sapporo 060-0814, Hokkaido, Japan

* Correspondence: mhaseyama@imd.ist.hokudai.ac.jp

Abstract: Brain decoding is a process of decoding human cognitive contents from brain activities. However, improving the accuracy of brain decoding remains difficult due to the unique characteristics of the brain, such as the small sample size and high dimensionality of brain activities. Therefore, this paper proposes a method that effectively uses multi-subject brain activities to improve brain decoding accuracy. Specifically, we distinguish between the shared information common to multi-subject brain activities and the individual information based on each subject's brain activities, and both types of information are used to decode human visual cognition. Both types of information are extracted as features belonging to a latent space using a probabilistic generative model. In the experiment, an publicly available dataset and five subjects were used, and the estimation accuracy was validated on the basis of a confidence score ranging from 0 to 1, and a large value indicates superiority. The proposed method achieved a confidence score of 0.867 for the best subject and an average of 0.813 for the five subjects, which was the best compared to other methods. The experimental results show that the proposed method can accurately decode visual cognition compared with other existing methods in which the shared information is not distinguished from the individual information.

Keywords: brain decoding; functional magnetic resonance imaging (fMRI); multiple subjects; visual features; generative model



Citation: Higashi, T.; Maeda, K.; Ogawa, T.; Haseyama, M. Brain Decoding of Multiple Subjects for Estimating Visual Information Based on a Probabilistic Generative Model. *Sensors* **2022**, *22*, 6148. <https://doi.org/10.3390/s22166148>

Received: 18 July 2022

Accepted: 14 August 2022

Published: 17 August 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Brain decoding estimates human cognition from brain activities and has been actively studied. There has been recent research progress in measuring human brain activities. Certain measuring methods have been used in this regard, including the implantable microelectrode array (MEA) [1] and other noninvasive measuring methods, such as near-infrared spectroscopy [2], electroencephalogram [3], functional magnetic resonance imaging (fMRI) [4–10], and magnetoencephalography (MEG) [11,12]. MEA is an invasive measurement method, and the merit of MEA is its robustness to noise during brain activity measurement. However, it necessitates the implantation of microelectrodes in a subject's body, which is its demerit, as this imposes a significant physical burden on the subject. Therefore, noninvasive methods, such as fMRI and MEG, which are less likely to directly harm a subject, are more widely used than invasive methods. fMRI in particular is frequently used to measure brain activities and can obtain brain activities with high-spatial resolutions. Compared with MEA, the demerits of fMRI are its sensitivity to noise and the large size of the measurement equipment. MEG is superior to fMRI in terms of temporal resolution and has reasonable spatial resolutions [13,14].

Several positive results have been reported by using machine learning methods to analyze brain activities from these measuring methods. For example, emotion analysis methods have been proposed from brain activities [15–17]. Some techniques have been proposed to generate an image caption and reconstruct an image using brain activities

while seeing the image [9,18–22]. In addition, image reconstruction is attempted on the basis of brain activities while imagining [20]. Researchers believe that the advancement of machine-learning-based brain decoding will reveal the human brain mechanism. The revelation of the human brain's mechanism is expected to contribute to a society in which everyone lives comfortably by realizing effective devices that use brain activities as input. For example, a brain–machine interface (BMI) aims for humans to directly operate and communicate with external machines without physical movement, which can assist the daily lives of people with handicaps [23–25].

The estimation of visual perception from fMRI data has been actively researched in the field of brain decoding to take advantage of excellent spatial resolutions [6–8,26]. fMRI data vary depending on the imaging object [6]. Previous studies [7,26] have attempted to analyze visual perception using classical methods, such as a support vector machine [27] and a Gabor wavelet filter [28]. There is a relationship between visual information extracted from a convolutional neural network (CNN) [29] and fMRI data obtained while seeing an object [8,30]. This relationship suggests that a CNN mimics the visual perception system in the human brain, and visual features extracted by a CNN are essential in estimating visual perception. Previous studies have attempted to estimate CNN-based visual features of images using fMRI data collected while subjects see the images [8,31–34]. In a previous study [8], the authors constructed a decoder that learns the relationship between each subject's fMRI data and the visual features of a seen image, and the decoder can estimate the visual features of the seen image from the fMRI data. Their method is based on linear regression. Although this decoder can estimate visual features, the estimation accuracy strongly depends on the size of the training set that consists of fMRI data. However, it is essential to lie in a closed and narrow space for a long period to measure fMRI data. Preparing a large amount of fMRI data also places a psychological and time burden on a subject.

Meanwhile, it is still challenging to estimate visual perception using a limited amount of data. Some studies [31,32] have used multi-subject fMRI data obtained when multiple subjects see the same image. These methods construct the latent space to extract the common features (shared features) from multi-subject fMRI data. These methods are based on the generative model, and they can stably train models using multiple inputs, even though the size of training data is limited. They emphasized the concept of shared features as the common information between multi-subject fMRI data. However, they have not considered individual features on which the researchers [8,35] focused as the subject-specific information in single-subject fMRI data. Each feature contains different information, and it is possible that accuracy is improved by combining shared and individual features in terms of differences in expressive ability.

In this study, we propose a novel method for estimating the visual features of a seen image from multi-subject fMRI data. To improve the estimation accuracy, we introduce the idea of focusing on multi-subject and single-subject fMRI data. We calculate the latent space based on fMRI data and extract shared and individual features using a generative model. The generative model assumes the distributions of the extracted features and can extract effective features from a limited amount of data. We can train decoders to estimate visual features using the extracted shared and individual features. In addition, it is possible to use common cognitive information from multi-subject fMRI data and subject-specific cognitive information from single-subject fMRI data.

The remainder of this paper is organized as follows. In Section 2, we explain the proposed estimation method using multi-subject fMRI data. Section 3 presents the experimental results using an fMRI dataset when multiple subjects see an image. Finally, Section 4 presents the conclusion.

2. Estimation of Visual Features of Seen Image Using Shared and Individual Features

In this section, we explain the proposed method. The overview of the training phase is illustrated in Figure 1. We constructed a probabilistic generative model (PGM) to extract

shared and individual features from multiple and single subjects. In addition, a visual decoder was constructed to estimate visual features from both extracted features. The overview of the test phase is shown in Figure 2. We extracted shared and individual features from single-subject fMRI data using the PGM and estimate the visual features of a seen image using the trained visual decoder. The trained PGM can extract shared features from only single-subject fMRI data. The training and test phases are explained in Sections 2.1 and 2.2, respectively. A list of the variables used is presented in Appendix A.

2.1. Training Phase: Construction of PGM and Visual Decoder

The training phase consisted of two steps. In the first step, we constructed the PGM to extract shared and individual features from fMRI data separately. This model can extract features robust to the noise in fMRI data. In the second step, we trained a visual decoder to estimate the visual features of a seen image. The visual decoder can transform shared and individual features into visual features using a projection matrix.

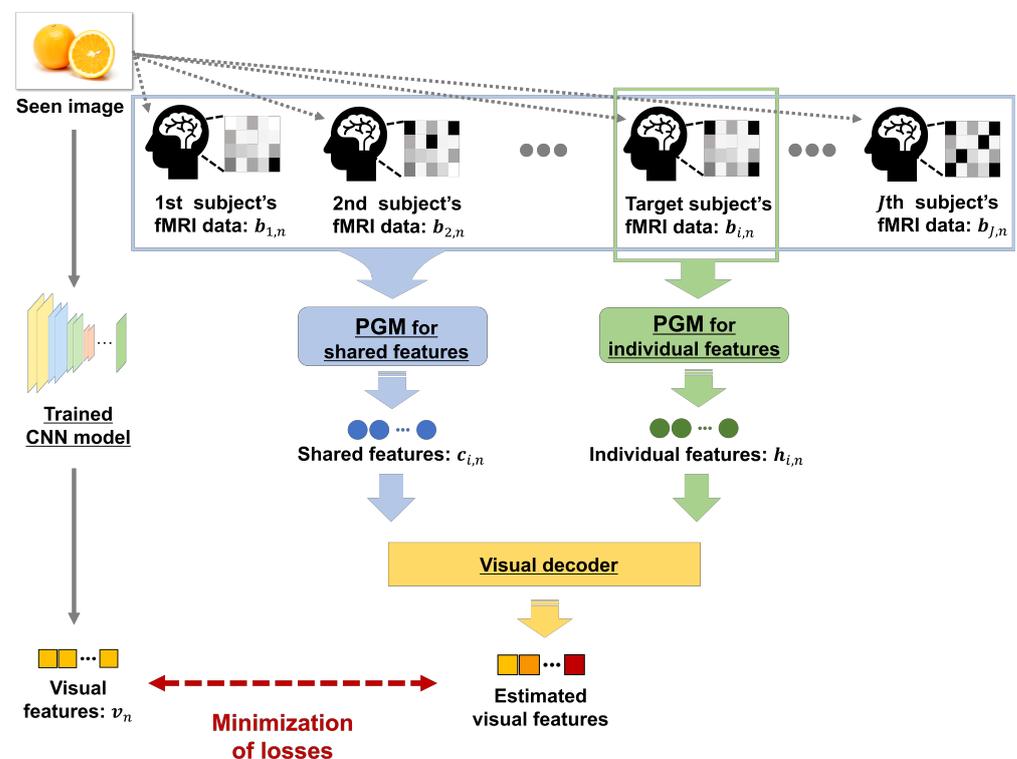


Figure 1. Overview of the training phase in the proposed method. Different PGMs are trained to extract shared and individual features of each subject. We can extract shared features from single-subject fMRI data. The PGM for individual features uses only single-subject fMRI data as a target for training. The visual decoder was trained using both extracted features from PGMs based on the minimization problem between the estimated and visual features from the trained CNN.

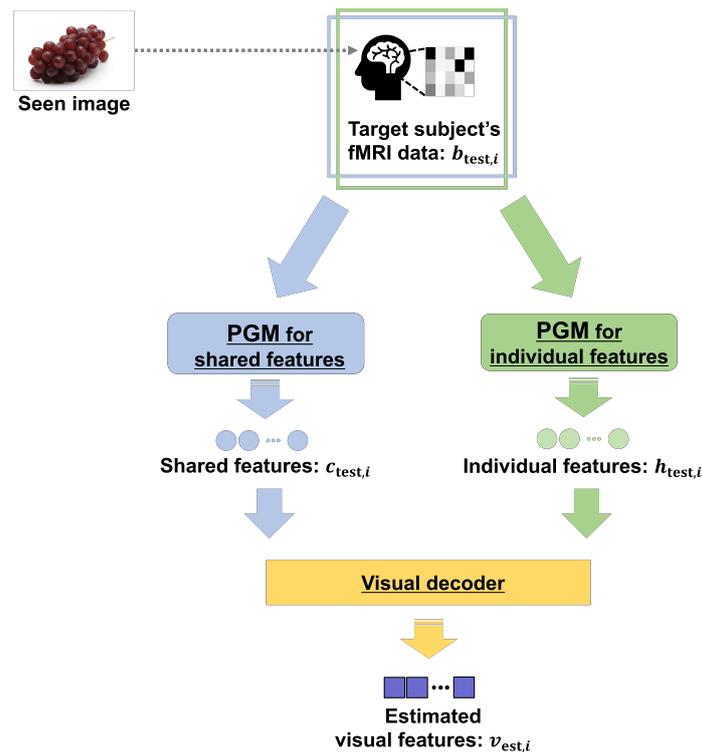


Figure 2. Overview of the test phase in the proposed method. The PGM corresponding to each feature was used to extract features from the target subject's fMRI data. The trained visual decoder can estimate visual features using shared and individual features, and this scheme realizes our approach.

2.1.1. Step 1: Construction of PGM

In step 1, we constructed the PGM for extracting of shared and individual features from fMRI data $\mathbf{B}_i = [\mathbf{b}_{i,1}, \dots, \mathbf{b}_{i,N}] \in \mathbb{R}^{d_i \times N}$ ($i = 1, \dots, J$; here, J represents the number of subjects, d_i denotes the dimension of the fMRI data in i th subject, and N represents the number of fMRI data training corresponding to each seen image). First, we describe the scheme for extracting the shared features $\mathbf{C} = [\mathbf{c}_1, \dots, \mathbf{c}_N] \in \mathbb{R}^{d_{\text{com}} \times N}$ (d_{com} being the dimensions of the shared features). An algorithm table is shown in Algorithm 1. The Gaussian distribution is introduced as a prior distribution of shared features \mathbf{C} into the following minimization problem:

$$\min_{\mathbf{P}_i, \mathbf{C}} \sum_{i=1}^J \|\mathbf{B}_i - \mathbf{P}_i \mathbf{C}\|_F^2 \quad \text{s.t. } \mathbf{P}_i^\top \mathbf{P}_i = \mathbf{I}, \quad (1)$$

where $\mathbf{P}_i \in \mathbb{R}^{d_i \times d_{\text{com}}}$ denotes the projection matrix that transforms the fMRI data \mathbf{B}_i into the shared features \mathbf{C} , and \mathbf{I} represents the identity matrix. The prior distribution of the shared features \mathbf{c}_n ($n = 1, \dots, N$) and the conditional Gaussian distribution $p(\mathbf{b}_{i,n} | \mathbf{c}_n)$ are given as follows:

$$p(\mathbf{c}_n) \sim \mathcal{N}(\mathbf{0}, \Sigma_c), \quad (2)$$

$$p(\mathbf{b}_{i,n} | \mathbf{c}_n) \sim \mathcal{N}(\mathbf{P}_i \mathbf{c}_n + \boldsymbol{\mu}_i, \rho_i^2 \mathbf{I}), \quad (3)$$

where $\Sigma_c \in \mathbb{R}^{d_{\text{com}} \times d_{\text{com}}}$ denotes the covariance matrix of the shared features \mathbf{c}_n , $\boldsymbol{\mu}_i = \frac{1}{N} \sum_{n=1}^N \mathbf{b}_{i,n} \in \mathbb{R}^{d_i}$ represents the mean of the fMRI data \mathbf{B}_i , and ρ_i^2 represents the variance of \mathbf{B}_i . The fMRI data of each subject are combined under the assumption that multiple subjects see the same image, and multi-subject fMRI data corresponding to an image $\mathbf{b}_n = [\mathbf{b}_{1,n}^\top, \dots, \mathbf{b}_{J,n}^\top]^\top$ are defined. Moreover, the marginal probability distribution $p(\mathbf{b}_n)$ and \mathbf{b}_n are represented as follows:

$$p(\mathbf{b}_n) \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{P}\boldsymbol{\Sigma}_c\mathbf{P}^\top + \boldsymbol{\Psi}), \quad (4)$$

$$\mathbf{b}_n = \mathbf{P}\mathbf{c}_n + \boldsymbol{\mu} + \boldsymbol{\epsilon}, \quad (5)$$

where $\mathbf{P} = [\mathbf{P}_1^\top, \dots, \mathbf{P}_J^\top]^\top$, $\boldsymbol{\mu} = [\boldsymbol{\mu}_1^\top, \dots, \boldsymbol{\mu}_J^\top]^\top$ and $\boldsymbol{\Psi} = \text{diag}(\rho_1^2\mathbf{I}, \dots, \rho_J^2\mathbf{I}) \in \mathbb{R}^{d_{\text{all}} \times d_{\text{all}}}$ are combined parameters ($d_{\text{all}} = \sum_{i=1}^J d_i$), and $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Psi})$ is an error term. To calculate the marginal probability distribution in Equation (4), we define the joint distribution of the shared features c_n and fMRI data $\mathbf{b}_{i,n}$ and take the logarithm. The mean and covariance matrix of $p(\mathbf{b}_n)$ can be calculated by computing the exponential part of the joint matrix distribution for the second-order and first-order terms [36].

Algorithm 1 PGM for shared features c_n .

```

1: Initialize:
2:    $P_i, \rho_i, \boldsymbol{\Sigma}_c, \boldsymbol{\mu}_i, J, N, \text{MAXLOOP}, A_i, \rho_i^{\text{new}^2}, \boldsymbol{\Sigma}_c^{\text{new}}$ 
3: for  $i = 1 \dots J$  do
4:   for  $n = 1 \dots N$  do
5:      $\boldsymbol{\mu}_i \leftarrow \boldsymbol{\mu}_i + \mathbf{b}_{i,n}$ 
6:   end for
7:    $\boldsymbol{\mu}_i \leftarrow \frac{1}{N}\boldsymbol{\mu}_i$ 
8: end for
9:  $l \leftarrow 1$ 
10: while  $\text{MAXLOOP} \geq l$  do // EM algorithm for updating parameters
11:    $\text{Var}_{c|b}[c] \leftarrow \boldsymbol{\Sigma}_c - \boldsymbol{\Sigma}_c^\top \mathbf{P}^\top (\mathbf{P}\boldsymbol{\Sigma}_c\mathbf{P}^\top + \boldsymbol{\Psi})^{-1} \mathbf{P}\boldsymbol{\Sigma}_c$  // Start expectation step
12:   for  $n = 1 \dots N$  do
13:      $E_{c|b}[c_n] \leftarrow (\mathbf{P}\boldsymbol{\Sigma}_c)^\top (\mathbf{P}\boldsymbol{\Sigma}_c\mathbf{P}^\top + \boldsymbol{\Psi})^{-1} (\mathbf{b}_n - \boldsymbol{\mu})$ 
14:      $E_{c|b}[c_n c_n^\top] \leftarrow \text{Var}_{c|b}[c] + E_{c|b}[c_n] E_{c|b}[c_n]^\top$ 
15:   end for // End expectation step
16:   for  $i = 1 \dots J$  do // Start maximization step
17:     for  $n = 1 \dots N$  do
18:        $A_i \leftarrow A_i + (\mathbf{b}_{i,n} - \boldsymbol{\mu}_i) E_{c|b}[c_n]^\top$ 
19:     end for
20:      $A_i \leftarrow \frac{1}{2} A_i$ 
21:      $\mathbf{P}_i^{\text{new}} \leftarrow A_i (A_i^\top A_i)^{-1/2}$ 
22:     for  $n = 1 \dots N$  do
23:        $\rho_i^{\text{new}^2} \leftarrow \rho_i^{\text{new}^2} + \|\mathbf{b}_{i,n} - \boldsymbol{\mu}_i\|^2 - 2(\mathbf{b}_{i,n} - \boldsymbol{\mu}_i)^\top \mathbf{P}_i^{\text{new}} E_{c|b}[c_n] + \text{tr}(E_{c|b}[c_n c_n^\top])$ 
24:     end for
25:      $\rho_i^{\text{new}^2} \leftarrow \frac{1}{Nd_i} \rho_i^{\text{new}^2}$ 
26:   end for
27:   for  $n = 1 \dots N$  do
28:      $\boldsymbol{\Sigma}_c^{\text{new}} \leftarrow \boldsymbol{\Sigma}_c^{\text{new}} + E_{c|b}[c_n c_n^\top]$ 
29:   end for
30:    $\boldsymbol{\Sigma}_c^{\text{new}} \leftarrow \frac{1}{N} \boldsymbol{\Sigma}_c^{\text{new}}$ 
31:    $P_i, \rho_i^2, \boldsymbol{\Sigma}_c \leftarrow \mathbf{P}_i^{\text{new}}, \rho_i^{\text{new}^2}, \boldsymbol{\Sigma}_c^{\text{new}}$  // End maximization step
32:   Initialize:
33:      $A_i, \rho_i^{\text{new}^2}, \boldsymbol{\Sigma}_c^{\text{new}}$ 
34:    $l \leftarrow l + 1$ 
35: end while

```

We introduce the expectation maximization (EM) algorithm [37] for updating the model parameters P_i , ρ_i^2 , and $\boldsymbol{\Sigma}_c$. The posterior distribution $p(c_n|\mathbf{b}_n)$ is calculated in the expectation step of the EM algorithm. The posterior distribution $p(c_n|\mathbf{b}_n)$ follows the

Gaussian distribution, and we can analytically calculate the expected value $E_{c|b}[c_n]$ and the variance $\text{Var}_{c|b}[c]$ as follows:

$$E_{c|b}[c_n] = (\mathbf{P}\boldsymbol{\Sigma}_c)^\top (\mathbf{P}\boldsymbol{\Sigma}_c\mathbf{P}^\top + \boldsymbol{\Psi})^{-1}(\mathbf{b}_n - \boldsymbol{\mu}), \quad (6)$$

$$\text{Var}_{c|b}[c] = \boldsymbol{\Sigma}_c - \boldsymbol{\Sigma}_c^\top \mathbf{P}^\top (\mathbf{P}\boldsymbol{\Sigma}_c\mathbf{P}^\top + \boldsymbol{\Psi})^{-1} \mathbf{P}\boldsymbol{\Sigma}_c. \quad (7)$$

The expected value $E_{c|b}[c_n]$ and the variance $\text{Var}_{c|b}[c]$ are calculated using the joint distribution of the shared features c_n and fMRI data \mathbf{b}_n , similarly to the calculation in Equation (4). However, the joint distribution is defined on the basis of the marginal probability distribution $p(\mathbf{b}_n)$. The joint distribution is taken by logarithm. We calculate the expected value $E_{c|b}[c_n]$ and the variance $\text{Var}_{c|b}[c]$ using the exponential part of the joint matrix distribution for the second-order and first-order terms [36]. Furthermore, the second-order moment $E_{c|b}[c_n c_n^\top]$ is calculated as follows:

$$E_{c|b}[c_n c_n^\top] = \text{Var}_{c|b}[c] + E_{c|b}[c_n] E_{c|b}[c_n]^\top. \quad (8)$$

The parameters P_i , ρ_i^2 , and $\boldsymbol{\Sigma}_c$ are updated to maximize the expected value $R(\theta, \theta^{\text{old}})$ by $(\theta, \theta^{\text{old}} \sim \{P_i, \rho_i^2, \boldsymbol{\Sigma}_c\})$ in the maximization step of the EM algorithm. Note that θ^{old} is a fixed parameter in the expectation step. The expected value $R(\theta, \theta^{\text{old}})$ is expressed as follows:

$$\begin{aligned} R(\theta, \theta^{\text{old}}) &= \frac{1}{N} \sum_{n=1}^N \int p(c_n | \mathbf{b}_n; \theta^{\text{old}}) \log p(\mathbf{b}_n, c_n; \theta) dc_n \\ &= \frac{1}{N} \sum_{n=1}^N \int p(c_n | \mathbf{b}_n; \theta^{\text{old}}) (\log p(\mathbf{b}_n | c_n; \theta) \\ &\quad + \log p(c_n; \theta)) dc_n \\ &= \sum_{n=1}^N E_{c|b} [\log p(\mathbf{b}_n | c_n; \theta) + \log p(c_n; \theta)]. \end{aligned} \quad (9)$$

In the above equation, the transformation is performed with respect to θ . A term with only θ^{old} as parameters can be regarded as a constant and is excluded from the expectation value R in the maximization step. The expected value $R(\theta, \theta^{\text{old}})$ is calculated using partial derivatives with respect to the parameters $\{P_i, \rho_i^2, \boldsymbol{\Sigma}_c\}$ and maximized. The updated parameters P_i^{new} , $\rho_i^{\text{new}2}$ and $\boldsymbol{\Sigma}_c^{\text{new}}$ are defined as follows:

$$P_i^{\text{new}} = A_i (A_i^\top A_i)^{-1/2}, \quad (10)$$

$$A_i = \frac{1}{2} \left(\sum_{n=1}^N (\mathbf{b}_{i,n} - \boldsymbol{\mu}_i) E_{c|b}[c_n]^\top \right), \quad (11)$$

$$\begin{aligned} \rho_i^{\text{new}2} &= \frac{1}{Nd_i} \sum_{n=1}^N (\|\mathbf{b}_{i,n} - \boldsymbol{\mu}_i\|^2 - 2(\mathbf{b}_{i,n} - \boldsymbol{\mu}_i)^\top \\ &\quad \mathbf{P}_i^{\text{new}} E_{c|b}[c_n] + \text{tr}(E_{c|b}[c_n c_n^\top])), \end{aligned} \quad (12)$$

$$\boldsymbol{\Sigma}_c^{\text{new}} = \frac{1}{N} \sum_{n=1}^N E_{c|b}[c_n c_n^\top]. \quad (13)$$

Each parameter is updated using the expected value $E_{c|b}[c_n]$ and the variance $\text{Var}_{c|b}[c]$ in the expectation step of the EM algorithm. We repeatedly calculated both steps to update the model parameters P_i^{new} , $\rho_i^{\text{new}2}$ and $\boldsymbol{\Sigma}_c^{\text{new}}$. Finally, the shared features c_n can be calculated using the following equation:

$$c_n = \mathbf{P}^{-1}(\mathbf{b}_n - \boldsymbol{\mu}). \quad (14)$$

In addition, shared features of each subject can be extracted as follows:

$$\mathbf{c}_{i,n} = \mathbf{P}_i^{-1}(\mathbf{b}_{i,n} - \boldsymbol{\mu}_i), \quad (15)$$

where $\mathbf{c}_{i,n}$ denotes the shared features in the i th subject. We can extract shared features following the Gaussian distribution, and the extracted features are effective for estimating visual features due to the robustness of noise.

Similarly, we used the PGM to extract individual features $\mathbf{H}_i = [\mathbf{h}_{i,1}, \dots, \mathbf{h}_{i,N}] \in \mathbb{R}^{d_{\text{ind}} \times N}$ from the fMRI data \mathbf{B}_i (d_{ind} being the dimension of the individual features). We also introduce the Gaussian distribution as a prior distribution of the individual features \mathbf{H}_i into the following minimization problem, and the prior distribution of the individual features $\mathbf{h}_{i,n}$ is given as follows:

$$\min_{\mathbf{P}'_i, \mathbf{H}_i} \|\mathbf{B}_i - \mathbf{P}'_i \mathbf{H}_i\|_F^2 \quad \text{s.t. } \mathbf{P}'_i{}^\top \mathbf{P}'_i = \mathbf{I}, \quad (16)$$

where $\mathbf{P}'_i \in \mathbb{R}^{d_i \times d_{\text{ind}}}$ denotes the projection matrix that transforms the fMRI data \mathbf{B}_i to the individual features \mathbf{H}_i . The prior distribution of the shared features $\mathbf{h}_{i,n}$ and the conditional Gaussian distribution $p(\mathbf{b}_{i,n} | \mathbf{h}_{i,n})$ are given as follows:

$$p(\mathbf{h}_{i,n}) \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}'_{\mathbf{h}_i}), \quad (17)$$

$$p(\mathbf{b}_{i,n} | \mathbf{h}_{i,n}) \sim \mathcal{N}(\mathbf{P}'_i \mathbf{h}_{i,n} + \boldsymbol{\mu}_i, \rho_i'^2 \mathbf{I}), \quad (18)$$

where $\boldsymbol{\Sigma}'_{\mathbf{h}_i} \in \mathbb{R}^{d_{\text{ind}} \times d_{\text{ind}}}$ denotes the covariance matrix of $\mathbf{h}_{i,n}$, and $\rho_i'^2$ denotes the variance of \mathbf{B}_i . The marginal probability distributions $p(\mathbf{b}_{i,n})$ and $\mathbf{b}_{i,n}$ are represented as follows:

$$p(\mathbf{b}_{i,n}) \sim \mathcal{N}(\boldsymbol{\mu}_i, \mathbf{P}'_i \boldsymbol{\Sigma}'_{\mathbf{h}_i} \mathbf{P}'_i{}^\top + \boldsymbol{\Psi}'_i), \quad (19)$$

$$\mathbf{b}_{i,n} = \mathbf{P}'_i \mathbf{h}_{i,n} + \boldsymbol{\mu}_i + \boldsymbol{\epsilon}'_i, \quad (20)$$

where $\boldsymbol{\epsilon}'_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Psi}'_i)$ represents an error term. Note that $\boldsymbol{\Psi}'_i = \text{diag}(\rho_i'^2 \mathbf{I}) \in \mathbb{R}^{d_i \times d_i}$, and $\rho_i'^2$ denotes the variance of \mathbf{B}_i . We can extract the individual features $\mathbf{h}_{i,n}$ following the calculation steps of the shared features \mathbf{c}_n using the EM algorithm. The obvious difference between individual features and the shared features \mathbf{c}_n is that it is not necessary to combine these parameters in multiple subjects. Finally, the individual features $\mathbf{h}_{i,n}$ can be calculated using the following equation:

$$\mathbf{h}_{i,n} = \mathbf{P}'_i^{-1}(\mathbf{b}_{i,n} - \boldsymbol{\mu}_i). \quad (21)$$

2.1.2. Step 2: Construction of Visual Decoder

In step 2, we trained the visual decoder that convert shared and individual features into visual features $\mathbf{V} = [\mathbf{v}_1, \dots, \mathbf{v}_N] \in \mathbb{R}^{d_v \times N}$ (d_v being the dimensions of the visual features). We calculated projection matrices $\mathbf{P}_{\text{com},i} \in \mathbb{R}^{d_v \times d_{\text{com}}}$ and $\mathbf{P}_{\text{ind},i} \in \mathbb{R}^{d_v \times d_{\text{ind}}}$. The following minimization problem was computed with respect to the projection matrices $\mathbf{P}_{\text{com},i}$ and $\mathbf{P}_{\text{ind},i}$:

$$\begin{aligned} \min_{\mathbf{P}_{\text{com},i}, \mathbf{P}_{\text{ind},i}} & \|\mathbf{V} - (\mathbf{P}_{\text{com},i} \mathbf{C}_i + \mathbf{P}_{\text{ind},i} \mathbf{H}_i)\|_F^2 \\ & + \lambda_{\text{com},i} \|\mathbf{P}_{\text{com},i}\|_F^2 + \lambda_{\text{ind},i} \|\mathbf{P}_{\text{ind},i}\|_F^2, \end{aligned} \quad (22)$$

where $\lambda_{\text{com},i}$ and $\lambda_{\text{ind},i}$ represent regularization parameters. By repeating the partial derivatives with respect to $\mathbf{P}_{\text{com},i}$ and $\mathbf{P}_{\text{ind},i}$, we can simply obtain the following optimal projections:

$$\mathbf{P}_{\text{com},i} = (\mathbf{V} - \mathbf{P}_{\text{ind},i} \mathbf{H}_i) \mathbf{C}_i{}^\top (\mathbf{C}_i \mathbf{C}_i{}^\top + \lambda_{\text{com},i} \mathbf{I})^{-1}, \quad (23)$$

$$\mathbf{P}_{\text{ind},i} = (\mathbf{V} - \mathbf{P}_{\text{com},i} \mathbf{C}_i) \mathbf{H}_i{}^\top (\mathbf{H}_i \mathbf{H}_i{}^\top + \lambda_{\text{ind},i} \mathbf{I})^{-1}. \quad (24)$$

2.2. Test Phase: Estimation of Visual Features of Seen Image

The test phase consisted of two steps. In the first step, we extracted shared and individual features using the developed PGM. In the second step, we estimated visual features using shared and individual features using the constructed visual decoder.

2.2.1. Step 1: Extraction of Shared and Individual Features

We extracted the shared features $c_{\text{test},i}$ using the parameters of the PGM as follows:

$$c_{\text{test},i} = P_i^{-1}(\mathbf{b}_{\text{test},i} - \boldsymbol{\mu}_i), \quad (25)$$

where $\mathbf{b}_{\text{test},i} \in \mathbb{R}^{d_i}$ denotes fMRI data for the i th subject in the test phase. The shared features could be extracted as each subject's feature. Similarly, individual features $h_{\text{test},i}$ were extracted as follows:

$$h_{\text{test},i} = P_i'^{-1}(\mathbf{b}_{\text{test},i} - \boldsymbol{\mu}_i). \quad (26)$$

Both features followed a Gaussian distribution, and we can extract effective features from fMRI data.

2.2.2. Step 2: Estimation of Visual Features

Visual features were estimated using shared and individual features using the trained visual decoder as follows:

$$\mathbf{v}_{\text{est},i} = P_{\text{com},i}c_{\text{test},i} + P_{\text{ind},i}h_{\text{test},i}, \quad (27)$$

where $\mathbf{v}_{\text{est},i}$ denotes the visual features estimated using visual decoders in the i th subject. In Equation (27), visual features are estimated from shared and individual features using each projection matrix corresponding to the features.

3. Experimental Results

This section presents the experimental results of the image category estimation. In Section 3.1, the datasets used in constructing the proposed method are explained. In Section 3.2, the experimental conditions are described. In Section 3.3, the comparison methods are explained. In Section 3.4, the experimental results are presented.

3.1. Dataset

In this experiment, we used the fMRI data (approximately 4500-dimensional vectors) published in a previous study [8]. fMRI data comprise data on visual cortex activities of five subjects while observing images with measuring equipment (Siemens MAGNETOM Prisma (<https://www.siemens-healthineers.com/jp/magnetic-resonance-imaging/research-systems/magnetom-prisma> (accessed on 10 August 2022))). To obtain fMRI data in [8], four males and one female between the ages of 23 and 38 were chosen as the subjects, and the functional localizer [38–40] and the standard retinotopy [4,41] experiments were conducted to identify the visual cortex of each subject. There were 1200 seen images of 150 categories collected in ImageNet [42] (eight images per category). We used these images as pairs of fMRI data.

We performed cross-validation to examine the effectiveness of the proposed method through unbiased experiments. Due to the significant burden on the subject during brain activity acquisition, preparing several samples for the fMRI dataset is difficult. Therefore, as shown in Figure 3, we divided these 1200 pairs into 900, 150, and 150 pairs as training, validation, and test data, respectively. All categories were equally divided into the training, validation, and test data. In addition, we applied 7-fold cross-validation to 1050 pairs consisting of training and test data, respectively. If the training and validation data are interchanged, as is widely done in the machine learning field, the validity of the proposed method on small amounts of test data would be verified.

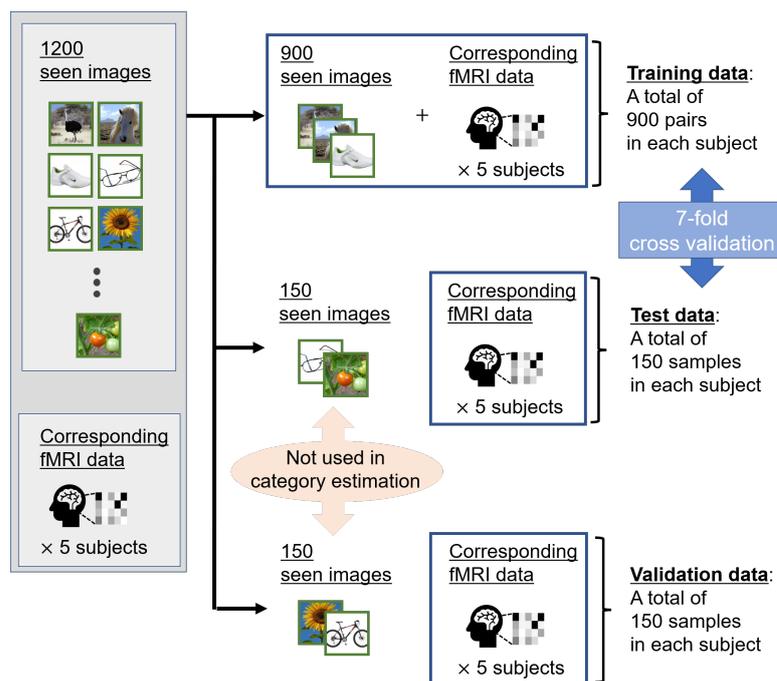


Figure 3. Overview of fMRI datasets of five subjects. We divided a total of 1200 seen images corresponding to measured fMRI data into 900, 150, and 150 images as training, test, and validation data, respectively. The validation data were fixed, and 7-fold cross-validation was applied to 1050 pairs of the training and test data. For category estimation, we used the candidate visual features averaged from other images belonging to the same seen category. Thus, the seen images were not included in the test and validation data.

First, 4096-dimensional visual features were extracted from VGG19 [43]. VGG19 was generally pre-trained for the 1000 categories in ImageNet, and the fully connected layers that extracted visual features were selected farther from the output. Furthermore, the principal component analysis (PCA) [44] was applied to the visual features. The visual features have high dimensions, and we used PCA to prevent overfitting. We selected the cumulative contribution ratio of PCA as 0.8 (the dimensions of visual features applied PCA d_v , being approximately 70).

3.2. Experimental Conditions

The estimated accuracy was evaluated by image category estimation. CNNs are mostly trained for image categorization, and category estimation is an appropriate evaluation metric for the representation ability of visual features extracted from a pre-trained CNN. Among the categories of seen images of the fMRI data used in this experiment, some were not used for the pre-trained CNN classifiers. Therefore, we evaluated the estimated accuracy using visual features. Figure 4 shows an overview of category estimation. The image category ranks of the estimated and candidate visual features indicate categories via VGG19 based on the correlations. We selected from 10,000 categories in the fMRI dataset and calculated averaged visual features in each category as candidate visual features. Note that of the 10,000 categories, 150 image categories were used in the fMRI dataset in the test phase. Candidate visual features were 5–10 images chosen at random from each category. The ranks of the estimated visual features and 10,000 candidate visual features were calculated and rearranged in descending order; the ground truth (GT) rank was defined as the image category rank. Finally, the confidence category score S was calculated from the image category rank G as follows:

$$S = \frac{M - G}{M - 1}, \quad (28)$$

where M represents the total number of image categories, and we set M to 10,000 in this experiment. The confidence category score S approaches 1 for better image category ranks G and 0 for worse. The confidence category scores were averaged in 150 test data points, 7-fold cross-validation sets, and five subjects. We used this metric for the experiment evaluation.

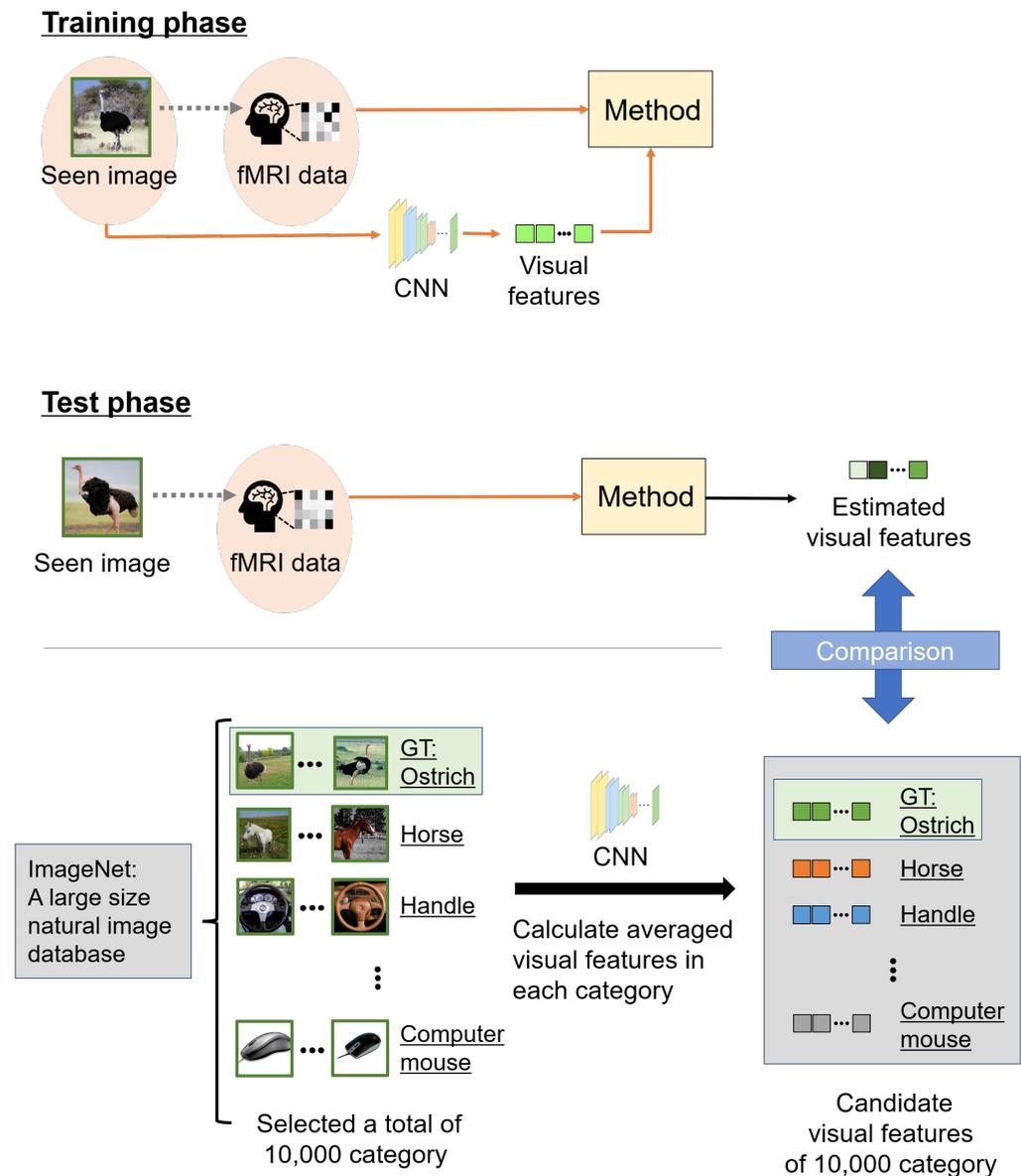


Figure 4. Overview of the scheme of category estimation. In the training phase, the relationship between a seen image and the corresponding fMRI data in each subject was learned. In the test phase, we estimated visual features using fMRI data based on the learned relationship. However, the visual features to be compared were computed from images chosen at random from ImageNet. These other images were 5–10 samples in each category, and we selected 10,000 categories from ImageNet. The 10,000 categories included 150 image categories belonging to the fMRI data in the test phase. We defined candidate visual features, averaged visual features, extracted from these other images in each category, and compared them with estimated visual features from fMRI data. Finally, we calculated the correlations between the estimated and candidate visual features, and the accuracy of the estimations was evaluated with the seen image category as the ground truth (GT).

3.3. Comparison Methods

We compared the proposed method (hereafter denoted as PM) with several comparison methods (hereafter denoted as CMs) based on the evaluation metric to validate the effectiveness of the PM. The CMs have seven patterns. Two CMs use multi-subject fMRI data and five CMs use single-subject fMRI data.

Multi-subject probabilistic generative model (MSPGM):

MSPGM is a method based on the PGM, and PM uses multi-subject fMRI data. Visual features are estimated from shared features using ridge regression [45]. We set the number of dimensions in the latent space to the same number of dimensions of the PM and searched for $\{0.1, 1, 10\}$ in the regularization parameter of ridge regression.

Multi-view Bayesian generative model for multi-subject fMRI Data (MVBGM-MS):

MVBGM-MS [46] exhibited state-of-the-art performance in the field of brain decoding for visual cognitive contents. MVBGM-MS uses multi-subject fMRI data, and the generative model estimates visual features via the latent space. MVBGM-MS uses visual features, multi-subject fMRI data, and semantic features extracted by inputting image category names into Word2vec [47] to improve accuracy. Therefore, for a fair evaluation of the PM, we used MVBGM-MS without semantic features. We set the number of dimensions in the latent space in the same manner as in the previous study [46].

Single-subject probabilistic generative model (SSPGM):

The SSPGM method is based on fMRI data and uses single-subject fMRI data. Visual features are estimated from individual features using ridge regression. We set the number of dimensions in the latent space to the same number of dimensions of the PM and searched for $\{0.1, 1, 10\}$ in the regularization parameter of the ridge regression.

Sparse linear regression (SLR):

SLR [8] is a baseline method in the field of brain decoding for visual cognitive contents and directly estimated visual features from fMRI data. We estimated visual features by using voxels consisting of fMRI data with a high correlation to the features. Voxels were selected in the order of increasing correlation, and the total number of voxels to be selected was set as a hyperparameter. We searched the number of voxels for $\{50, 100, 200, 400, 500, 1000\}$.

Canonical correlation analysis (CCA):

CCA [48] is a baseline method for calculating the latent space from multi-modal features. Visual features and fMRI data are converted into features belonging to the latent space, and accuracy is evaluated in the space. We searched for $\{10, 20, 30, 40, 50, d_v\}$ in the number of dimensions in the latent space.

Bayesian CCA (BCCA):

The BCCA [49] method is an extension of CCA that adopts Bayesian learning. BCCA is a generative model. The latent space consists of visual features and fMRI data, and visual features can be estimated from fMRI data via the space. We searched for $\{10, 20, 30, 40, 50, d_v\}$ in the number of dimensions in the latent space.

Deep CCA (Deep CCA):

The Deep CCA [50] method is also an extension of CCA that adopts deep learning. Similarly to CCA, visual features and fMRI data are converted into features belonging to the latent space, and accuracy is evaluated in the space. We searched for $\{10, 20, 30, 40, 50, d_v\}$ in the number of dimensions in the latent space.

3.4. Results and Discussion

3.4.1. Estimation Performance Evaluation

Table 1 shows the accuracy of the category estimation in the PM and CMs. Note that the average scores of 150 test images were calculated according to each subject as the evaluation metric. These scores range from 0 to 1, and large values indicate superiority. In the PM, MSPGM, and SSPGM, we set d_{com} , d_{ind} , and the number of iterations in the EM algorithm to 100, 100, and 10. In addition, in the PM's visual decoder, we searched each regularization parameter $\lambda_{\text{com},i}$ and $\lambda_{\text{ind},i}$ for $\{0.1, 1, 10\}$.

Table 1. Confidence category scores were averaged for 150 test images and five subjects in the PM and all CMs (The best scores for each subject, and the averages of all subjects are shown in bold).

	Subject1	Subject2	Subject3	Subject4	Subject5	Average
Proposed Method (PM)	0.756	0.806	0.867	0.860	0.778	0.813
MSPGM	0.744	0.801	0.857	0.850	0.771	0.805
MVBGM-MS [46]	0.793	0.764	0.832	0.814	0.756	0.792
SSPGM	0.696	0.802	0.859	0.851	0.763	0.794
SLR [8]	0.772	0.734	0.817	0.809	0.711	0.769
CCA [48]	0.706	0.723	0.796	0.782	0.705	0.742
BCCA [49]	0.661	0.762	0.835	0.824	0.740	0.764
Deep CCA [50]	0.622	0.697	0.792	0.755	0.685	0.710

In Table 1, the scores, of most subjects and the averages of five subjects in the PM are superior to those in MVBGM-MS and MSPGM based on multi-subject fMRI data. The PM's superior scores indicate its effectiveness in distinguishing between shared and individual features in multi-subject fMRI data. Furthermore, MVBGM-MS is state-of-the-art, but the PM outperformed it. The scores of all subjects and the averages of the PM are superior to those of SSPGM, SLR, CCA, Deep CCA, and BCCA based on single-subject fMRI data. The SSPGM method is based on the PGM and PM, and its effectiveness in combining shared and individual features is exhibited. The effectiveness of the PM in estimating the visual features of seen images from fMRI data was confirmed compared with SLR, and based on the quantitative accuracy of the PM, it is reliable. Moreover, compared with CCA and Deep CCA, our method can derive their latent spaces successfully. In particular, the PM significantly outperformed Deep CCA, which is the only model that incorporates deep learning among the CMs. Deep CCA is also inferior to simple CCA in terms of score. Deep learning may not be compatible with fMRI data, for which only a small sample size is available. Furthermore, compared with BCCA as a generative model, we can confirm the superiority of our generative model for extracting shared and individual features.

3.4.2. Qualitative Evaluation

Figure 5 shows the qualitative evaluation of PM and MVBGM-MS, SLR, CCA, BCCA, and Deep CCA, which are methods based on other studies [8,46,48–51]. For the image categories of "shirt" and "saddle", the PM has the best confidence category scores for most subjects. These results demonstrate the effectiveness of the PGM as a feature extractor and the idea of using shared and individual features.

Figure 6 shows the qualitative evaluation of the PM, MSPGM, and SSPGM based on our PGM. For the image category of "hand calculator", PM has the best confidence category scores for most subjects. However, for the image category of "obelisk", MSPGM has the best confidence category scores for most subjects. These results indicate that although the PGM can extract valid shared features, there exists a possibility that it cannot extract individual features. Due to its characteristics, the PGM is superior in extracting shared features from multi-subject fMRI data. For the image categories "spectacles" and "camera

tripod”, all methods did not achieve sufficient confidence category scores compared with the quantitative evaluation in most subjects in Table 1. These images contained multiple objects, and a subject’s gaze may not be focused on a single object during fMRI data acquisition. In addition, for the category of “camera tripod”, a part of a human face also appears in the image, which may have affected the subjects’ cognition. Category estimation may still be a difficult task when seeing images containing multiple objects or objects not related to the image categories.

		
	Image category: Shirt	Image category: Saddle
Subject 1	PM : 0.989 MVBGM-MS : 0.943 SLR : 0.984 CCA : 0.908 BCCA : 0.737 DeepCCA : 0.715	PM : 0.976 MVBGM-MS : 0.709 SLR : 0.762 CCA : 0.651 BCCA : 0.774 DeepCCA : 0.570
Subject 2	PM : 0.974 MVBGM-MS : 0.819 SLR : 0.527 CCA : 0.943 BCCA : 0.833 DeepCCA : 0.965	PM : 0.958 MVBGM-MS : 0.840 SLR : 0.889 CCA : 0.636 BCCA : 0.923 DeepCCA : 0.655
Subject 3	PM : 0.965 MVBGM-MS : 0.522 SLR : 0.655 CCA : 0.819 BCCA : 0.790 DeepCCA : 0.499	PM : 0.987 MVBGM-MS : 0.870 SLR : 0.917 CCA : 0.936 BCCA : 0.805 DeepCCA : 0.912
Subject 4	PM : 0.993 MVBGM-MS : 0.875 SLR : 0.915 CCA : 0.925 BCCA : 0.966 DeepCCA : 0.826	PM : 0.808 MVBGM-MS : 0.546 SLR : 0.787 CCA : 0.589 BCCA : 0.483 DeepCCA : 0.528
Subject 5	PM : 0.993 MVBGM-MS : 0.857 SLR : 0.203 CCA : 0.988 BCCA : 0.772 DeepCCA : 0.815	PM : 0.813 MVBGM-MS : 0.882 SLR : 0.393 CCA : 0.910 BCCA : 0.644 DeepCCA : 0.934

Figure 5. Examples of the category estimation results of PM and five CMs (MVBGM-MS, SLR, CCA, BCCA, and Deep CCA) in the quantitative evaluation. The confidence category scores range from 0 to 1, and the best scores for each subject are shown in bold.

				
	Image category: <u>Hand calculator</u>	Image category: <u>Obelisk</u>	Image category: <u>Spectacles</u>	Image category: <u>Camera tripod</u>
Subject 1	PM : 0.657 <u>MSPGM : 0.870</u> SSPGM : 0.389	PM : 0.998 <u>MSPGM : 1.000</u> SSPGM : 0.996	PM : 0.613 <u>MSPGM : 0.729</u> SSPGM : 0.504	PM : 0.621 <u>MSPGM : 0.690</u> SSPGM : 0.542
Subject 2	<u>PM : 0.728</u> MSPGM : 0.411 SSPGM : 0.626	PM : 0.987 <u>MSPGM : 0.988</u> SSPGM : 0.983	PM : 0.300 MSPGM : 0.320 <u>SSPGM : 0.416</u>	<u>PM : 0.004</u> <u>MSPGM : 0.004</u> SSPGM : 0.001
Subject 3	<u>PM : 0.997</u> MSPGM : 0.996 SSPGM : 0.995	PM : 0.932 <u>MSPGM : 0.953</u> SSPGM : 0.913	<u>PM : 0.722</u> MSPGM : 0.614 SSPGM : 0.554	PM : 0.453 MSPGM : 0.450 <u>SSPGM : 0.496</u>
Subject 4	<u>PM : 0.940</u> MSPGM : 0.905 SSPGM : 0.867	<u>PM : 0.989</u> MSPGM : 0.981 SSPGM : 0.983	PM : 0.432 <u>MSPGM : 0.723</u> SSPGM : 0.632	PM : 0.089 MSPGM : 0.205 <u>SSPGM : 0.268</u>
Subject 5	<u>PM : 1.000</u> <u>MSPGM : 1.000</u> SSPGM : 0.999	PM : 0.921 <u>MSPGM : 0.932</u> SSPGM : 0.853	<u>PM : 0.468</u> MSPGM : 0.385 SSPGM : 0.385	<u>PM : 0.807</u> MSPGM : 0.491 SSPGM : 0.491

Figure 6. Examples of the category estimation results of PM and two CMs (MSPGM and SSPGM) based on PGM. The confidence category scores range from 0 to 1, and the best scores for each subject are shown in bold.

4. Conclusions and Future Work

In this article, we proposed a method for estimating visual information from multi-subject fMRI data obtained while subjects observed images. The PM estimated visual features using shared features in multi-subject fMRI data and individual features in single-subject fMRI data. The PGMs were constructed with respect to each feature from fMRI data and used as effective feature extractors. In addition, we constructed the visual decoder using the shared and individual features to estimate visual features. The experimental results verified the effectiveness of the proposed approaches. Although fMRI data tend to contain measured noises and large individual differences compared with other biological activities, such as an eye gaze, this experiment confirmed the effectiveness of combining multi-subject fMRI data. These findings validated the use of machine learning for biological activity analysis with time and physical factor constraints.

Apart from the increase in the sample size due to the expansion of the fMRI dataset, using modalities other than fMRI data may provide a hint as to how to improve the accuracy. In particular, introducing other information that represents an image, such as image captions, is expected to improve the results. For example, some theories suggest estimating visual features directly from fMRI data and caption features using an image captioning model (caption features) or constructing a latent space combining fMRI data and visual and caption features to improve expressive ability. The human brain contains regions specialized for object recognition related to an image category and regions related to lower-order information, such as object color and shape [41]. Although visual features extracted using a CNN contain information specific to image category classification, they may not contain sufficient information, such as an image color and shape. Therefore, the introduction of image captions that can represent the colors and shapes in images is considered an effective method for extracting information related to images in fMRI data obtained while subjects see the images.

Author Contributions: Conceptualization, T.H., K.M., T.O. and M.H.; methodology, T.H., K.M., T.O. and M.H.; software, T.H.; validation, T.H., K.M., T.O. and M.H.; data curation, T.H.; writing—original draft preparation, T.H.; writing—review and editing, K.M., T.O. and M.H.; visualization, T.H.; funding acquisition, T.O. and M.H. All authors have read and agreed to the published version of the manuscript.

Funding: This work was partly supported by JSPS KAKENHI, grant number JP21H03456, and AMED, grant number JP21zf0127004 .

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Publicly available dataset was analyzed in this study. This data can be found here: <https://github.com/KamitaniLab/GenericObjectDecoding> (accessed on 10 August 2022).

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1 presents a list of the variables used in Section 2.

Table A1. List of variables used for the proposed method in the training and test phases corresponds to Sections 2.1 and 2.2.

Section 2.1.1: Construction of PGM

$b_{i,n}$	fMRI data corresponding to n th image in i th subject
B_i	fMRI data in i th subject ($B_i = [b_{i,1}, \dots, b_{i,N}]$)
c_n	Shared features corresponding to n th image
C	Shared features ($C = [c_1, \dots, c_N]$)
P_i	Projection matrix that transforms fMRI data in i th subject into shared features
I	Identity matrix
Σ_c	Covariance matrix of shared features
μ_i	Mean of fMRI data in i th subject
ρ_i^2	Variance of fMRI data in i th subject
b_n	Concatenated fMRI data corresponding to n th image for total J subjects
μ	Concatenated mean μ_i for total J subjects
P	Concatenated projection matrix P_i for total J subjects
ϵ	Error term of shared features
Ψ	Joint covariance
E	Expected value of expectation maximization (EM) algorithm
Var	Variance of EM algorithm
R	Expected value in maximization step of EM algorithm
P_i^{new}	Updated projection matrix that transforms fMRI data in i th subject
$\rho_i^{\text{new}2}$	Updated variance of fMRI data in i th subject
Σ_c^{new}	Updated covariance matrix of shared features
$c_{i,n}$	Estimated shared features corresponding to n th image in i th subject
J	Number of subjects
N	Number of seen images
n	Index of seen images ($n = 1, \dots, N$)
i	Index of subjects ($i = 1, \dots, J$)
d_{com}	Dimensions of shared features
d_i	Dimensions of fMRI data in i th subject

Table A1. Cont.

d_{all}	Sum of dimensions for total J subjects
θ	Updated PGM parameters ($\theta \sim \{P_i^{\text{new}}, \rho_i^{\text{new}^2}, \Sigma_c^{\text{new}}\}$)
θ^{old}	PGM parameters before update ($\theta^{\text{old}} \sim \{P_i, \rho_i^2, \Sigma_c\}$)
$h_{i,n}$	Individual features corresponding to n th image in i th subject
H_i	Individual features in i th subject ($H_i = [h_{i,1}, \dots, h_{i,N}]$)
P_i'	Projection matrix that transforms fMRI data in i th subject into individual features
$\rho_i'^2$	Variance of fMRI data in i th subject
Σ_{h_i}'	Covariance matrix of individual features
Ψ_i'	Joint covariance in i th subject
e_i'	Error term of individual features in i th subject
d_{ind}	Dimensions of individual features
Section 2.1.2: Construction of visual decoder	
v_n	Visual features of n th image
V	Visual features ($V = [v_1, \dots, v_N]$)
C_i	Estimated shared features in i th subject ($C_i = [c_{i,1}, \dots, c_{i,N}]$)
$P_{\text{com},i}$	Projection matrix that transforms shared features into visual features in i th subject
$P_{\text{ind},i}$	Projection matrix that transforms individual features into visual features in i th subject
$\lambda_{\text{com},i}$	Regularization parameter corresponding to shared features in i th subject
$\lambda_{\text{ind},i}$	Regularization parameter corresponding to individual features in i th subject
d_v	Dimensions of visual features
Section 2.2.1: Extraction of shared and individual features	
$b_{\text{test},i}$	fMRI data in i th subject
$c_{\text{test},i}$	Shared features in i th subject
$h_{\text{test},i}$	Individual features in i th subject
Section 2.2.2: Estimation of visual features	
$v_{\text{est},i}$	Estimated visual features by visual decoder in i th subject

References

- Ponce, C.R.; Xiao, W.; Schade, P.F.; Hartmann, T.S.; Kreiman, G.; Livingstone, M.S. Evolving images for visual neurons using a deep generative network reveals coding principles and neuronal preferences. *Cell* **2019**, *177*, 999–1009. [\[CrossRef\]](#) [\[PubMed\]](#)
- Fazli, S.; Mehnert, J.; Steinbrink, J.; Curio, G.; Villringer, A.; Müller, K.R.; Blankertz, B. Enhanced performance by a hybrid NIRS–EEG brain computer interface. *Neuroimage* **2012**, *59*, 519–529. [\[CrossRef\]](#) [\[PubMed\]](#)
- Müller, K.R.; Tangermann, M.; Dornhege, G.; Krauledat, M.; Curio, G.; Blankertz, B. Machine learning for real-time single-trial EEG-analysis: From brain–computer interfacing to mental state monitoring. *J. Neurosci. Methods* **2008**, *167*, 82–90. [\[CrossRef\]](#)
- Engel, S.A.; Rumelhart, D.E.; Wandell, B.A.; Lee, A.T.; Glover, G.H.; Chichilnisky, E.J.; Shadlen, M.N. fMRI of human visual cortex. *Nature* **1994**, *369*, 525. [\[CrossRef\]](#) [\[PubMed\]](#)
- Engel, S.A.; Glover, G.H.; Wandell, B.A. Retinotopic organization in human visual cortex and the spatial precision of functional MRI. *Cereb. Cortex* **1997**, *7*, 181–192. [\[CrossRef\]](#) [\[PubMed\]](#)
- Haxby, J.V.; Gobbini, M.I.; Furey, M.L.; Ishai, A.; Schouten, J.L.; Pietrini, P. Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* **2001**, *293*, 2425–2430. [\[CrossRef\]](#)
- Kay, K.N.; Naselaris, T.; Prenger, R.J.; Gallant, J.L. Identifying natural images from human brain activity. *Nature* **2008**, *452*, 352–355. [\[CrossRef\]](#)
- Horikawa, T.; Kamitani, Y. Generic decoding of seen and imagined objects using hierarchical visual features. *Nat. Commun.* **2017**, *8*, 15037. [\[CrossRef\]](#)
- Shen, G.; Dwivedi, K.; Majima, K.; Horikawa, T.; Kamitani, Y. End-to-end deep image reconstruction from human brain activity. *Front. Comput. Neurosci.* **2019**, *13*, 21. [\[CrossRef\]](#)
- Han, K.; Wen, H.; Shi, J.; Lu, K.H.; Zhang, Y.; Fu, D.; Liu, Z. Variational autoencoder: An unsupervised model for encoding and decoding fMRI activity in visual cortex. *NeuroImage* **2019**, *198*, 125–136. [\[CrossRef\]](#)
- Baillet, S. Magnetoencephalography for brain electrophysiology and imaging. *Nat. Neurosci.* **2017**, *20*, 327–339. [\[CrossRef\]](#) [\[PubMed\]](#)

12. Zariief, C.N.; Hussein, W. Decoding the human brain activity and predicting the visual stimuli from magnetoencephalography (meg) recordings. In Proceedings of the 2019 International Conference on Intelligent Medicine and Image Processing, Bali, Indonesia, 19–22 April 2019; pp. 35–42.
13. Liljeström, M.; Hulten, A.; Parkkonen, L.; Salmelin, R. Comparing MEG and fMRI views to naming actions and objects. *Hum. Brain Mapp.* **2009**, *30*, 1845–1856. [[CrossRef](#)] [[PubMed](#)]
14. Taylor, M.; Donner, E.; Pang, E. fMRI and MEG in the study of typical and atypical cognitive development. *Neurophysiol. Clin./Clin. Neurophysiol.* **2012**, *42*, 19–25.
15. Zheng, W.L.; Santana, R.; Lu, B.L. Comparison of classification methods for EEG-based emotion recognition. In Proceedings of the World Congress on Medical Physics and Biomedical Engineering, Toronto, ON, Canada, 7–12 June 2015; pp. 1184–1187.
16. Zheng, W.L.; Zhu, J.Y.; Lu, B.L. Identifying stable patterns over time for emotion recognition from EEG. *IEEE Trans. Affect. Comput.* **2017**, *10*, 417–429. [[CrossRef](#)]
17. Hemanth, D.J.; Anitha, J. Brain signal based human emotion analysis by circular back propagation and Deep Kohonen Neural Networks. *Comput. Electr. Eng.* **2018**, *68*, 170–180. [[CrossRef](#)]
18. Matsuo, E.; Kobayashi, I.; Nishimoto, S.; Nishida, S.; Asoh, H. Describing semantic representations of brain activity evoked by visual stimuli. In Proceedings of the 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Miyazaki, Japan, 7–10 October 2018; pp. 576–583.
19. Takada, S.; Togo, R.; Ogawa, T.; Haseyama, M. Generation of viewed image captions from human brain activity via unsupervised text latent space. In Proceedings of the 2020 IEEE International Conference on Image Processing (ICIP), Abu Dhabi, United Arab Emirates, 25–28 October 2020; pp. 2521–2525.
20. Shen, G.; Horikawa, T.; Majima, K.; Kamitani, Y. Deep image reconstruction from human brain activity. *PLoS Comput. Biol.* **2019**, *15*, e1006633. [[CrossRef](#)]
21. Belyi, R.; Gaziv, G.; Hoogi, A.; Strappini, F.; Golan, T.; Irani, M. From voxels to pixels and back: Self-supervision in natural-image reconstruction from fMRI. In Proceedings of the Advances in Neural Information Processing Systems 32, Annual Conference on Neural Information Processing Systems 2019, Vancouver, BC, Canada, 8–14 December 2019.
22. Ren, Z.; Li, J.; Xue, X.; Li, X.; Yang, F.; Jiao, Z.; Gao, X. Reconstructing seen image from brain activity by visually-guided cognitive representation and adversarial learning. *NeuroImage* **2021**, *228*, 117602.
23. Nicolelis, M.A. Actions from thoughts. *Nature* **2001**, *409*, 403–407. [[CrossRef](#)]
24. Lebedev, M.A.; Nicolelis, M.A. Brain-machine interfaces: past, present and future. *Trends Neurosci.* **2006**, *29*, 536–546. [[CrossRef](#)]
25. Patil, P.G.; Turner, D.A. The development of brain-machine interface neuroprosthetic devices. *Neurotherapeutics* **2008**, *5*, 137–146. [[CrossRef](#)]
26. Cox, D.D.; Savoy, R.L. Functional magnetic resonance imaging (fMRI) ‘brain reading’: Detecting and classifying distributed patterns of fMRI activity in human visual cortex. *Neuroimage* **2003**, *19*, 261–270. [[CrossRef](#)]
27. Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297. [[CrossRef](#)]
28. Daugman, J.G. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *JOSA A* **1985**, *2*, 1160–1169. [[CrossRef](#)] [[PubMed](#)]
29. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. In Proceedings of the Advances in Neural Information Processing Systems 25: 26th Annual Conference on Neural Information Processing Systems 2012, Lake Tahoe, NV, USA, 3–6 December 2012; pp. 1097–1105.
30. Nonaka, S.; Majima, K.; Aoki, S.C.; Kamitani, Y. Brain Hierarchy Score: Which Deep Neural Networks Are Hierarchically Brain-Like? *IScience* **2020**, *24*, 103013. [[CrossRef](#)] [[PubMed](#)]
31. Akamatsu, Y.; Harakawa, R.; Ogawa, T.; Haseyama, M. Brain decoding of viewed image categories via semi-supervised multi-view Bayesian generative model. *IEEE Trans. Signal Process.* **2020**, *68*, 5769–5781. [[CrossRef](#)]
32. Chen, P.H.C.; Chen, J.; Yeshurun, Y.; Hasson, U.; Haxby, J.; Ramadge, P.J. A reduced-dimension fMRI shared response model. In Proceedings of the Advances in Neural Information Processing Systems, Montreal, QC, Canada, 7–12 December 2015; pp. 460–468.
33. Higashi, T.; Maeda, K.; Ogawa, T.; Haseyama, M. Estimation of viewed images using individual and shared brain responses. In Proceedings of the IEEE 9th Global Conference on Consumer Electronics, Kobe, Japan, 13–16 October 2020; pp. 716–717.
34. Higashi, T.; Maeda, K.; Ogawa, T.; Haseyama, M. Estimation of Visual Features of Viewed Image From Individual and Shared Brain Information Based on FMRI Data Using Probabilistic Generative Model. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, Toronto, ON, Canada, 6–11 June 2021; pp. 1335–1339.
35. Papadimitriou, A.; Passalis, N.; Tefas, A. Visual representation decoding from human brain activity using machine learning: A baseline study. *Pattern Recognit. Lett.* **2019**, *128*, 38–44. [[CrossRef](#)]
36. Bishop, C.M.; Nasrabadi, N.M. *Pattern Recognition and Machine Learning*; Springer: Berlin/Heidelberg, Germany, 2006; Volume 4.
37. Dempster, A.P.; Laird, N.M.; Rubin, D.B. Maximum likelihood from incomplete data via the EM algorithm. *J. R. Stat. Soc. Ser. B (Methodol.)* **1977**, *39*, 1–22.
38. Kourtzi, Z.; Kanwisher, N. Cortical regions involved in perceiving object shape. *J. Neurosci.* **2000**, *20*, 3310–3318. [[CrossRef](#)]
39. Kanwisher, N.; McDermott, J.; Chun, M.M. The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J. Neurosci.* **1997**, *17*, 4302–4311. [[CrossRef](#)]
40. Epstein, R.; Kanwisher, N. A cortical representation of the local visual environment. *Nature* **1998**, *392*, 598–601. [[CrossRef](#)]

41. Sereno, M.I.; Dale, A.; Reppas, J.; Kwong, K.; Belliveau, J.; Brady, T.; Rosen, B.; Tootell, R. Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science* **1995**, *268*, 889–893. [[CrossRef](#)] [[PubMed](#)]
42. Deng, J.; Dong, W.; Socher, R.; Li, L.J.; Li, K.; Fei-Fei, L. ImageNet: A large-scale hierarchical image database. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 248–255.
43. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
44. Wold, S.; Esbensen, K.; Geladi, P. Principal component analysis. *Chemom. Intell. Lab. Syst.* **1987**, *2*, 37–52. [[CrossRef](#)]
45. Hoerl, A.E.; Kennard, R.W. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* **1970**, *12*, 55–67. [[CrossRef](#)]
46. Akamatsu, Y.; Harakawa, R.; Ogawa, T.; Haseyama, M. Multi-View bayesian generative model for multi-Subject fMRI data on brain decoding of viewed image categories. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, Barcelona, Spain, 4–8 May 2020; pp. 1215–1219.
47. Mikolov, T.; Sutskever, I.; Chen, K.; Corrado, G.S.; Dean, J. Distributed representations of words and phrases and their compositionality. In Proceedings of the Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems 2013, Lake Tahoe, NV, USA, 5–10 December 2013.
48. Hotelling, H. Relations between two sets of variates. In *Breakthroughs in Statistics*; Springer: Berlin/Heidelberg, Germany, 1992; pp. 162–190.
49. Fujiwara, Y.; Miyawaki, Y.; Kamitani, Y. Modular encoding and decoding models derived from Bayesian canonical correlation analysis. *Neural Comput.* **2013**, *25*, 979–1005. [[CrossRef](#)]
50. Andrew, G.; Arora, R.; Bilmes, J.; Livescu, K. Deep canonical correlation analysis. In Proceedings of the International Conference on Machine Learning, Miami, FL, USA, 4–7 December 2013; pp. 1247–1255.
51. Ek, C.H.; Torr, P.H.; Lawrence, N.D. Gaussian process latent variable models for human pose estimation. In Proceedings of the International Workshop on Machine Learning for Multi-modal Interaction, Brno, Czech Republic, 28–30 June 2007; Springer: Berlin/Heidelberg, Germany, 2007; pp. 132–143.