❏ 6682

Functional magnetic resonance imaging-based brain decoding with visual semantic model

Piyawat Saengpetch, Luepol Pipanmemekaporn, Suwatchai Kamolsantiroj
Department of Computer and Information Science, King Mongkut's University of Technology North Bangkok, Thailand

Article Info	ABSTRACT
	The activity pattern of the brain has been activated to identify a person in mind. Using the function magnetic resonance imaging (fMRI) to decipher brain decoding is the most accepted method. However, the accuracy of fMRI-based brain decoder is still restricted due to limited training samples. The limitations of the brain decoder using fMRI are passed through the design features proposed for many label coding and model training to predict these characteristics for a particular label. Moreover, what kind of semantic features for deciphering the neurological activity patterns are unclear. In current work, a new calculation model for learning decoding labels that is consistent with fMRI activity responses. The approach demonstrates the proposed corresponding label's success in terms of accuracy, which is decoded from brain activity patterns and compared with conventional text-derived feature technique. Besides, experimental studies present a training model based on multi-tasking to reduce the problems of limited training data sets. Therefore, the multi-task learning model is more efficient than modern methods of calculation, and decoding features may be easily obtained.

Corresponding Author:

Piyawat Saengpetch,
Department of Computer and Information Science,
King Mongkut's University of Technology, North Bangkok, Bangkok, Thailand.
Email: s5704062910017@kmutnb.ac.th

1. INTRODUCTION

Research about understanding brain functions based on the relationship between thought and stimulation of the brain. The possibility to understand the brain's function pattern is stimulated by things such as sound, smell, light, or objects. Now there are medical analysis tools that can get data from the brain such as computed tomography (CT), electroencephalography (EEG) and magnetoencephalography (MEG) to motion brain functions. Function magnetic resonance imaging (fMRI) is a technique that allows the brain to distinguish between the states of mind. The fMRI images are generated from blood-oxygen-levels dependent (BOLD) stimulation of brain cells. Most of the research has been done on brain transcription with fMRI, such as brain-computer-interface (BCI), neural-control interface (NCI) is a direct line of communication between the brain gets wired or improvement and external devices [1], as lie detection, exposing the deliberate deception from the text of the speech. False detection by lie detector is fraud detection by evaluating the content of messages or including identifiers [2]. Brain injury or traumatic brain injury that can be measured externally [3].

The 3-dimensional image obtained from fMRI consists of a brain point called voxel (volume + pixel). A voxel shows brain activity, which indicate the areas of the brain are working. Each voxel shows a part of the brain, containing brain cells, one million cells in the mind of the subject. The brain activity is captured by identifying different actived voxels. In general, the format of decoding brain using fMRI emphasized the linear relationship between the state and the mind through statistical tests. In the literature [4]

proposed the general linear model (GLM) approach, however, this examining actived voxels approach is isolation. Moveover a multi-voxel patterns of activity (MVPA) [5] is strong technique and was proposed to detect fMRI voxels. Support vector machine (SVM) and linear discriminant analysis (LDA) algorithm is generally used for training models, a subset of the fMRI data decoding. Although the machine learning algorithm has already achieved the efficiency of MVPA, its accuracy is low due to the lack of training examples. The issues will be confined to the representation of classes applied for training, while types that do not be in training cannot be decoded. From the limitations of information for the trained model, therefore searching for and suggesting new methods. In designing features for decoding the brain that does not exist in the limited set. Pioneer literature [6] states that brain function patterns can be predicted using semantic relationships linking 25-verbs and names. Palatucci et al. [7] have investigated a 218-dimensional representation of the subjects received from 218 volunteers who were answered the questions related to the category of objects such as *"Is it cold?"* or *"Is it hot?"* or *"Can you walk on it?".* Some studies have suggested a mechanism for the automatic extraction of meaning by examining language knowledge such as Wikipedia [8] and WordNet [9]. Exiting research used the outstanding features from the text to decipher the brain, which is found to be a perfect method. In this research, we present the visual features associated with the class of objects, which increases accuracy. This method chooses the images related to the objects in the same class and different characteristics, such as image size, brightness, orientation. Therefore, we consider medium and high-level image properties in encoding many objects. We measure the performance of visual features by using color histograms and correlograms. The online image library ImageNet has included images and tagged images. We then took pictures to extract the features for a model in predicting activity in the brain of brain activity data recorded from fMRI studies based on multi-tasking learning, including in-depth learning to create features and transfer learning of deep learning such as ResNet50, VGG16. We use 150 images from ImageNet to create features. We then designed a model for brain prediction with visual semantic features and fMRI data with multi-tasking learning to predict 60 concepts. The results of this research show the model's efficiency and predictive capability when compared with the state-of-the-art. The main contributions are as follows:

− To decode the brain using images from fMRI. We explore the outstanding features from images related to the objects or images to improve the accuracy and compared with the previous research.
− To present the models that are used to decipher activities in the brain by using multiple task regression models and demonstrating the suitability of predicting properties of objects without the training set.

The remainder of this document has the following structure. The second section reviews the problem of brain transcription, along with the model of brain transcription. The third section presents the image features. The proposed model is described in Section 4. Section 5 present the experimental setup, and the results achieved and discussion in Section 6. The summary and conclusion of the presentation in section 7.

2. BRAIN TRANSCRIPTION

2.1. Understanding brain decoding

In this path, we describe the brain decoding problem by applying 3D images obtained from the fMRI dataset. The image for fMRI data $X = x_{ij}$ be $m \times n$ matrix its (i, j) th element, comprising n is dimensional of fMRI data in each m row. The function category f to predict a label. In general, the classification model can be expressed as $f(xw) = y + e$, where e is the calculated error. Then w is weight direction. The difficulty is to find w to decrease $e = y - f(xw)$ and then is formulate the way out to the optimization intricacy can be defined as:

$$w^* = \underset{w}{argmin} \frac{1}{2} \|y - Xw\|_2^2 + \lambda \Omega(w) \tag{1}$$

where $\Omega(w)$ symbolizes the part experience of the challenge and regularizes the model. In the typical case, given $n_x \leq n_y$, where y possible objects are excluded from the training set, constraining the outputs for decrypting. Because it is hard to get an fMRI image for every possible object, past research can only decipher the information taught. Therefore, this solution solves the above problems for data that has never been seen before and supports data expansion.

2.2. Representation models of brain decoding

Brain decoding is presented utilizing semantic features to obtain the label set for new input data. Regular data sets obtained from experience, meaning for describing the most likely concepts Figure 1. Brain Decoding Model. It is composed of 3 levels: 1) Voxels activation vector $X = \{x_1, x_2, ..., x_d\}$ is represented in the input layer is obtained fMRI processing; 2) Semantic feature vector is represented in intermediate layer

for describing objects $Z = \{z_1, z_2, ..., z_k\}$; and 3) Possible objects $Y = \{y_1, y_2, ..., y_n\}$ is represented in the decoding layer. The correlation between X and Z is learned by using an fMRI subset for predicting the feature values, whereas the forecast features are used for decoding the correlated object [10].
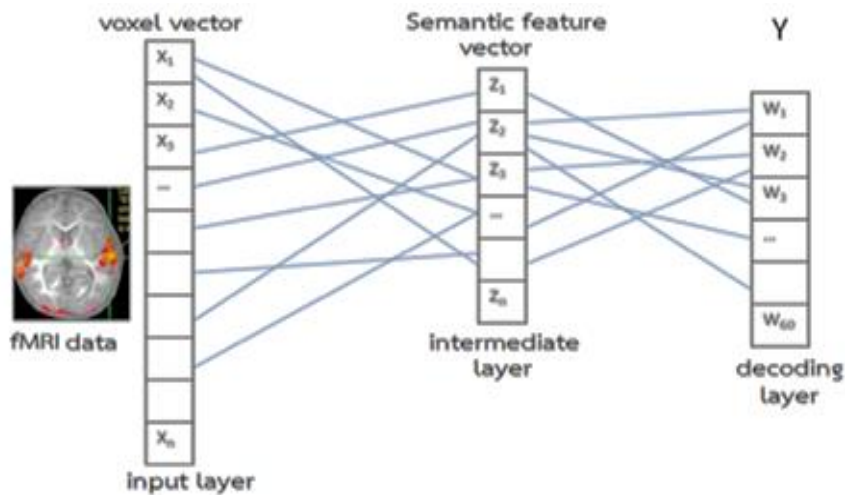

Figure 1. Brain decoding model

3. IMAGE FEATURES

Creatures tend to think out, *"Is it cold?"* or *"Is it hot?"*, or *"Can you walk on it?"*. and *"How do you feel when viewing the image?"*. It is thus the capability to identify the possible items based on standard visual features. In this part, we want to construct semantic features from images to describe activities that occur in the brain, by finding various methods in describing the images related to the concept. It was then creating a model of medicine, the type of activity in the brain.

3.1. Hierarchical visual features

Implementing hierarchical learning in which this method has convolution neural networks (CNNs), which are commonly used to classify objects [11]. CNN has a spatial structure consisting of layers of secret units [12], including convolution, pooling, and fully connected layers. In the convolution layer, it acts as a characteristic separator from the given input image. A pooling layer process produces down-sampling forward with the spatial area. Finally, the fully connected layer operates as a classifier that predicts the product of the input picture. The combination of these layers enables us to discover a hierarchical interpretation of the input picture. In this act, we objective at the appropriate CNN to increase the accuracy of the brain transcription model by employing features from the fully connected layer. For this reason, three advanced levels of CNNs are used: VGG16, ResNet50, and Xception, to learn the vector properties of figures concepts connected with each notion.

− VGG16 [13]

The important VGG16 structure is the popular CNN model. The network also presents the highest 5 test accuracy in ImageNet with 92.7% for image classification. This model is shown utilizing layers. Convolutional layers 3x3 are packed on top of each layer, and the other two layers are connected by each node, with 4,096 connections shown in Figure 2. VGG16 construction.

Figure 2. VGG16 construction

− ResNet50 [14]

ResNet won the first prize of ILSVRC2005 for its image classification, shown in Figure 3. Skip connection in ResNet50. The ResNet50 structure consists of 50 deep conversion layers. In this network, there are a total of 16 remaining blocks, each with blocks. There are three-layer feed layers ready.
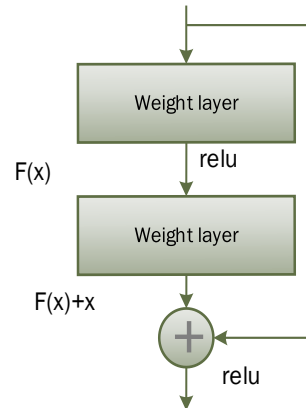

Figure 3. Skip connection in ResNet50

− Xception [15]

Xception by Google stands for the Ultimate version of Inception. With a mitigated depthwise detachable convolution, it is even better than Inception-v3 [16] (also by Google, 1st Runner Up in ILSVRC 2015) for both ImageNet ILSVRC and JFT datasets. The Xception construction has 36 convolutional layers building the feature extraction base of the network and fully-connected layers before the logistic regression layer. The 36 convolutional layers are defined into 14 sections, all of which have direct residual links around them, except for the first and last units. Figure 4 demonstrates the Xception construction.

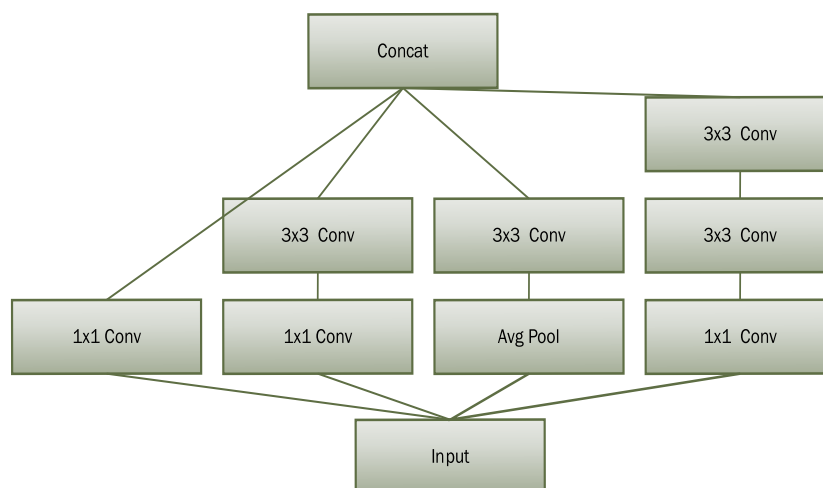Figure 4. Xception architecture

To recover the characteristic, the picture we use VGG16, ResNet50, and Xception candidly with importance trained on data sets larger as, i.e., ImageNet has combined a fresh level before production or layers softmax layer. Actions as a separate feature from the input picture. The neural network was retrained, adopting a fresh set of pictures while the earlier part was frozen.

3.2. Low-level visual descriptors

In creating an image representation, there are many prevalent methods. We were using low-level descriptors to capture general recognition features (such as color, surface, structure, edge). These basic features can be drawn immediately from the image and easily. In this analysis, we worked on the visual properties that are:

3.2.1. Simple color histogram (CH) [17]

Low-level feature utilizing a color histogram. Set to the distribution the number of pixels in the image for each container. We will use the effects of the color histogram to have 64 boxes in line with the part of the color saturation spectrum. We create a visual semantic model with the use of the default RGB color values and set the number of each component of the RGB color into four, which is the easiest method. (4x4x4).

3.2.2. Edge histogram (EH) [18]

Edge histogram - is a coding of the spatial distribution of the direction edge. Specifically, the images are divided into 72 boxes, each of which has several corners with a direction that is measured in 5-degree intervals. In this article, use Canny filters for edge detection and Sobel Edge Detector operators to Measure the course by the gradient of every edge point. The semantic feature has 72 dimensions.

3.2.3. Color correlogram (CORR) [19]

Correlogram color to encode spatial relationships of colors. One of the two-dimensional and three-dimensional histogram is the color of any pixel and three-dimensional spatial distances. The color correlogram is defined as:

$$r_{i,j}^{(N)} = \Pr_{P_1 \in I_{c(i)}, P_2 \in I} \left[P_2 \in I_{c(j)} \vee |P_1 - P_2| = d \right] \tag{2}$$

where $|P_1 - P_2|$ is the measure between P_1 and P_2, L is the distance intervals number, $i, j \in \{1, 2, \dots, N\}$, N is the number of boxes and $d \in \{1, 2, \dots, L\}$. We distribute the RGB value element into 36 boxes and set off the space metric to 4 odd intervals of $d = \{1, 3, 5, 7\}$. Thus, the color correlogram has a dimension of 144 (36x4).

3.2.4. Scale invariance feature transform (SIFT) [20]

SIFT is the most useful object identification algorithm in computer vision (CV) and has been used extensively. For any object in the picture, the SIFT key points on purpose can be separated into smaller parts to give a specific descriptor that defines the small image area around the mark on that object [21]. Because of the SIFT method results in the characteristic descriptor. The large size created in the picture, we use the theme of the bag of features. For each picture, first we analyze the SIFT descriptor above the local area based on the key points. We then calculated the vector's quantity on the SIFT district descriptor to create an image vocabulary using k-mean grouping. In this analysis, we created 500 groups, following in the size of the visual features being 500, for representing the image.

3.2.5. Wavelet Transform (WT) [22]

Image representation Remedies surface analysis of the image. In the surface analysis plan, the wavelet transform can be used to detect the surface of the image effectively. Wavelet transforms done on pictures that suggest repeated filtering and sub-sampling. At each level, pictures are divided into 4 sub-bands: LL, LH, HL, and HH, where L stands for low frequencies, and H means high frequencies. After that, the 2 types of wavelet transforms are the wavelet transform with the pyramid structure (PWT) and the wavelet transform with the tree structure (TWT). PWT breaks the LL band repeatedly, and TWT breaks the band. Other music to preserve relevant information that appears on the medium frequency channel. After passing the classification into parts, the vector is extracted using the mean and dispersion of each sub-district set of the strength distribution at each level. In this study, we used three levels of breakdowns to create vector properties of each object image. When performing PWT, they receive the characteristics of a 24 (3x4x2) vector. At the same time, the TWT answer in a visual semantic feature size 104 (52x2). Therefore, the total wavelength is 128.

4. MULTITASK VISUAL SPACE LEARNING

We introduced multi-task lasso learning (MTL) [23, 24] in which many learning tasks are improved at the same time. This method produces accurate predictions and is highly effective. In classification, MTL aims to increase the efficiency of multiple classification tasks by co-operating learning. One example is the spam filter, which can be classified as a complex classification mission. The reason for this is that learning the native fMRI representation may not have enough impact on marking multiple classes due to the training representation that Limited and challenging to find data sets. The aim is to learn the standard fMRI descriptions related to properties that are separated from many classes. The reason for this is that learning the original fMRI representation may not be sufficient for many states' characterization due to the small and limited training examples. A significant impact on learning is the finding of representatives of linear and nonlinear transformations of fMRI image data, as illustrated by mathematical equations. Consider $T = \{T_i\}_{i=1}^K$ is a set of k separate tasks where a task T_i directs on learning model for estimate the ith value. Training dataset is $D_i = \{x_j^i, y_j^i\}_{j=1}^{n_i}$, where $x_j^i \in R^d$ is the jth training fMRI image made of d voxels in T_i and y_j^i is its ith feature value. The training fMRI data matrix is X^i for T_i, i.e. $X^i = (x_1^i, \dots, x_{n_i}^i)$. For the task T_i, we learn $\hat{y}^i = \sum_{p=1}^d \beta_p^i X_p^i + \varepsilon^i$ where $\hat{y}^i, \beta_p^i, \varepsilon^i \in R^{n_i}$ Moreover, p is the number of voxels. We presume that the data is regulated so the constant terms can be dropped, i.e. X_p^i and $\hat{y}^i$ have average 0 and $\|X_p^i\|_2 = 1$ where $\|.\|_2$ is the L2 Euclidean norm. Let $\beta_p = (\beta_p^1, \dots, \beta_p^K)$ be the vector of all coefficients for the pth voxel across various tasks. To achieve a compact and discriminative representation, the multi-task Lasso is formulated as the answer to the optimization problem.

$$\min_{\beta} \left\{ \frac{1}{2} \sum_{i=1}^K \left\| Y^i - \sum_{p=1}^d \beta_p^i X_p^i \right\|_2^2 + \lambda \sum_{p=1}^d \|\beta_p\|_\infty \right\} \tag{3}$$

Where $\|\beta_p\|_\infty = max_k |\beta_p^i|$ is the sup-norm in the Euclidean space. It additionally has the impression of "grouping" the parts in β_p such that they can obtain zeros concurrently. After teaching K models, we still have to build a decision rule to choose the most important class to be related to a given fMRI image. For a given fMRI image X, K predicted feature values are obtained by using β coefficients. Lastly, Pearson's correlation coefficient [25] are used to measure the association between X and a target class C.

$$(X, C) = \frac{\sum_{i=1}^K (\Omega_i(X) - \acute{x})(c_i - \acute{c})}{\sqrt{\sum_{i=1}^K (X_i - \acute{x})^2} \times \sqrt{\sum_{i=1}^K (c_i - \acute{c})^2}} \tag{4}$$

where mapping function is $\Omega_i(X)$ that reconstructs X to the ith feature value. The result with the highest correlation fMRI concept.

5. EXPERIMENT SETUP

5.1. Datasets

In this research, we use fMRI 3D data from Carnegie Mellon University, collected from 9 right-handed adult volunteers, consisting of fMRI data that allows volunteers to view 60 line drawings and nouns, six times each. Images (6X60 = 360) 1 fMRI image will be converted to a voxel vector with approximately 20000 voxels. Then decrease the size of the vector to 500 voxels using the searchlight method [24]. In the technique [6] and [7], fMRI data are decreased to 500-features.

5.2. Text-based c features

The proposed model performing is compared with the text-based semantic features state-of-the-art models, as described follows:
− Verb25 [6] provides the notion in the form of the noun and describing the co-occurrence vector between nouns and verbs of 25 names such as "*run", "push", "eat"* and many others. These common verbs are often definite nouns in English sentences and designed by the structuralists. In an investigation, Verb-25 is a practical effect of this dataset.
− Human218 [7] holds the notion in the form of 218 attributes. The pattern vector is obtained by 218 questions, such as *"Is it cold?", "Is it hot?", "Can you walk on it?" e*tc. The linguists created these questions and collected the answers from cloud sourcing and computed the average answers correlated to each notion.

5.3. Image-based features

In this research, we have selected visual properties from pictures for the design of experiments by using properties from images from lower to upper levels. Color histogram (CH), color correlogram (CORR), edge histogram (EH), Wavelet Transform (WT), BoW + SIFT are on NUS-WIDE [26] database research. The NUS-WIDE dataset holds 269,648 pictures compiled from the Flickr online photograph database to adapt to the fMRI image concept and Use the following characteristics from the CNNs, namely VGG16, ResNet50, and Xception, and use the properties from all the images to create a feature vector to describe the concept.

6.   RESULTS AND DISCUSSION

This section is a test to compare the efficiency of using the properties obtained from text using nine fMRI images of each volunteer, 360 images from 60 images divided by concept, for use in teaching 300 images and 60 images for testing to find the effectiveness of models and features extracted with different methods. Table 1 shows predictability of the concept seen in the LR model accuracy (%) from 9 fMRI volunteers and the proposed algorithms (i.e., WT, EDH, ResNet50, and Xception). It can be observed that the proposed visual feature significantly enhances the performance of brain decoding. Moreover, the best performance models are obtained by using extracted features based on VGG16, ResNet50, and Xception. The multi-task learning (MTL) method is also compared with a linear regression (LR) method. As shown in Table 1 and Table 2 all the MTL models significantly outperformed the LR models. Thus, the given results emphasize the highlight of the MTL approach for improving the generalization performance. The leave-two-out cross-validation technique [6, 7] have been employed for evaluating the effectiveness of the visual feature for decoding novel concepts (unseen).

Table 1. Predictability of the concept seen in the LR model Accuracy (%) from 9 fMRI volunteers

Model	volunteer									Avg.	#Feature
	1	2	3	4	5	6	7	8	9		
Verbs 25	68.99	62.45	64.85	63.41	54.74	58.57	55.26	55.05	59.98	60.37	25
Human 218	62.14	58.43	54.90	58.45	56.30	53.30	56.52	48.59	49.53	55.35	218
CH	59.92	60.03	59.26	56.51	50.08	50.46	54.65	51.06	51.11	54.79	64
CORR	61.47	59.40	53.19	54.92	49.76	50.07	54.58	58.53	54.85	55.20	144
WT	61.23	57.38	55.77	54.78	50.41	52.46	51.61	53.05	56.58	54.81	128
EDH	60.27	55.53	56.33	55.64	53.40	51.25	51.95	53.29	51.05	54.30	73
BoW (SIFT)	70.52	63.03	61.58	61.55	49.10	53.82	54.61	52.61	55.81	60.07	500
VGG 16	65.21	60.79	59.29	62.31	54.60	56.51	53.80	54.13	51.91	57.62	4,096
ResNet 50	77.15	69.27	67.92	71.09	60.60	62.80	63.45	54.29	59.63	65.13	2,048
Xception	56.76	62.67	55.39	56.12	53.22	54.49	53.50	52.28	54.56	55.33	2,048

We tested the model on all the fMRI images of 58 concepts for the training set, and then fMRI images of 2 concepts for the testing set. The comparison results of all the predicting 2 unseen concepts model have been illustrated in Table 2. As recorded in Table 2, Both VGG16 and ResNet50 significantly bettered Verb25 and Human218, whereas the ResNet50 used 106 voxels of the whole brain showed the highest correctness. Figure 5 demonstrates that the MTL model exceeds the LR model for decoding V1 and V2 overall the semantic features. As a result, the high-level visual feature is better for identifying brain activity patterns in terms of accuracy.

Table 2. Predictability of the concept seen in the MTL model Accuracy (%) from 9 fMRI volunteers

| Model	volunteers									Avg.	#Feature	Selected Voxels
	1	2	3	4	5	6	7	8	9			
Verbs 25	70.16	67.08	67.68	66.35	59.03	56.56	58.23	56.65	57.14	62.10	25	328
Human 218	61.38	60.82	57.94	59.94	57.59	55.22	58.13	49.66	49.46	56.68	218	444
CH	64.79	63.47	60.87	58.45	53.11	52.25	58.95	50.78	51.01	57.08	64	147
CORR	63.38	63.12	58.64	58.73	51.99	52.07	58.50	55.35	57.34	57.68	144	233
WT	59.36	61.79	59.02	61.23	52.32	53.85	57.61	53.85	53.42	56.94	128	127
EDH	67.04	57.25	58.47	59.55	54.68	51.32	54.89	54.16	53.05	56.71	73	134
BoW(SIFT)	79.50	71.20	70.78	69.41	53.87	57.59	59.47	55.85	55.59	63.70	500	241
VGG 16	79.57	70.04	70.71	71.86	61.58	65.47	64.59	54.13	58.20	**66.24**	4,096	485
ResNet 50	79.18	72.36	71.12	73.37	62.52	64.73	63.90	55.30	58.43	**66.77**	2,048	482
Xception	82.72	87.67	79.29	86.36	77.27	87.01	78.78	82.32	78.57	**82.22**	2,048	435

(a)

(b)
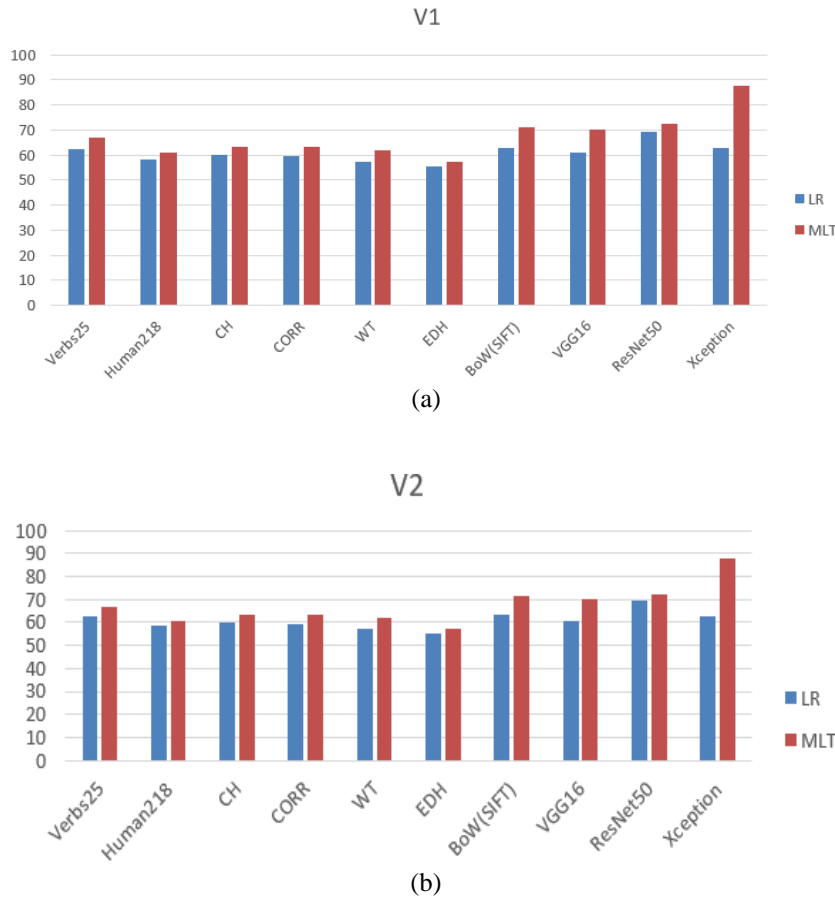
Figure 5. Performance comparison of MTL model and LR model on V1 and V2

7. CONCLUSION

In this paper, we have proposed an exotic fMRI brain transcription model using image-to-image features. The visual characteristics of many applicants are examined from low to high-level features that focus on the fMRI decoding model precisely. Also, the study presented is a compact and discriminating pattern learning-the relationship between the voxel activation model and the image properties to decipher the concept. The results show the operational success of our model and the advantage over state-of-the-art brain decoding with fMRI.

REFERENCES

[1] R. Sitaram, et al., "fMRI brain-computer interfaces," *IEEE Signal Process. Mag.,* vol. 25, no. 1, pp. 95-106, 2008.
[2] Y.-F. Lai, M.-Y. Chen, and H.-S. Chiang, "Constructing the lie detection system with fuzzy reasoning approach," *Granul. Comput.,* vol. 3, no. 2, pp. 169-176, 2017.
[3] J. K. Huttunen et al., "Detection of Hyperexcitability by Functional Magnetic Resonance Imaging after Experimental Traumatic Brain Injury," *Journal of neurotrauma,* vol. 35, no. 22, pp. 2708-2717, 2018.
[4] K. J. Worsley et al., "A General Statistical Analysis for fMRI Data," *Neuroimage,* vol. 15, no. 1, pp. 1-15, 2002.
[5] A. Mahmoudi, et al., "Multivoxel pattern analysis for fMRI data: A review," *Comput. Math. Methods Med.,* vol. 2012, 2012.
[6] T. M. Mitchell et al., "Predicting human brain activity associated with the meanings of nouns.," *science 320.5880,* vol. 320, no. 5880, pp. 1191-1195, 2008.
[7] M. Palatucci, et al., "Zero-Shot Learning with Semantic Output Codes," *Neural Inf. Process. Syst.,* pp. 1-9, 2009.
[8] L. Pipanmaekaporn, et al., "Designing Semantic Feature Spaces for Brain-Reading," *Esann,* pp. 22-24, 2015.
[9] A. B. Jelodar, et al., "WordNet Based Features for Predicting Brain Activity associated with meanings of nouns," *NAACL HLT 2010 First Work. Comput. Neurolinguistics,* pp. 18-26, 2010.
[10] P. Saengpetch, L. Pipanmemekaporn, and S. Kamolsantiroj, "Visual representation model for fMRI-based brain decoding," *ACM Int. Conf. Proceeding Ser.,* pp. 58-63, 2019.
[11] K. Kavukcuoglu, et al., "Learning convolutional feature hierarchies for visual recognition," *Adv. Neural Inf. Process. Syst. 23 24th Annu. Conf. Neural Inf. Process. Syst. 2010,* no. 1, pp. 1-9, 2010.

[12] C. Sample et al., "A general modeling framework for describing spatially structured population dynamics," *Ecology and Evolution,* vol. 8, no. 1, pp. 493-508, 2018.

[13] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556,* 2014.

[14] S. Wu, S. Zhong, and Y. Liu, "Deep residual learning for image steganalysis," *Multimed. Tools Appl.,* pp. 1-17, 2017.

[15] F. Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions," *Proceedings of the IEEE conference on computer vision and pattern recognition,* pp. 1251-1258, 2017.

[16] C. Szegedy, et al., "Rethinking the Inception Architecture for Computer Vision," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.,* vol. 2016, pp. 2818-2826, 2015.

[17] L. Shapiro and G. Stockman, "Computer Vision," *Prentice Hall,* vol. 21, no. 3. 2001.

[18] N. Prajapati, A. K. Nandanwar, and G. S. Prajapati, "edge histogram descriptor, geometric moment and sobel edge detector combined features based object recognition and retrieval system," *Int. J. Comput. Sci. Inf. Technol.,* vol. 7, no. 1, pp. 407-412, 2016.

[19] J. Huang, et al., "Image indexing using color correlograms," *Proceedings of IEEE computer society conference on Computer Vision and Pattern Recognition,* pp. 762-768, 1997.

[20] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints Abstract by Matthijs Dorst," *International journal of computer vision,* vol. 60, no. 2, pp. 91-110, 2004

[21] J. Zhang, et al., "Local features and kernels for classification of texture and object categories: A comprehensive study," *International journal of computer vision,* vol. 73, no. 2, pp. 213-238, 2007.

[22] B. S. Manjunath, "Texture features for browsing and retrieval of image data," *IEEE Trans. Pattern Anal. Mach. Intell.,* vol. 18, no. 8, pp. 837-842, 1996.

[23] X. Chen, J. He, R. Lawrence, and J. G. Carbonell, "Adaptive multi-task sparse learning with an application to fMRI study," *Proc. 12th SIAM Int. Conf. Data Mining, SDM 2012,* pp. 212-223, 2012.

[24] Y. Zhang and Q. Yang, "A Survey on Multi-Task Learning," *arXiv preprint arXiv:1707.08114,* 2017.

[25] J. Lee Rodgers and W. Alan Nice Wander, "Thirteen ways to look at the correlation coefficient," *The American Statistician,* vol. 42, no. 1, pp. 59-66, 1988.

[26] T.-S. Chua, et al., "NUS-WIDE: A Real-World Web Image Database from National University of Singapore," *Proceedings of the ACM international conference on image and video retrieval,* 2009.

BIOGRAPHIES OF AUTHORS

Piyawat Saengpetch, he received B.S. degree from the Department of Computer Science at Songkhla Rajabhat University, Thailand, in 1994, MSc from the Prince of Songkhla University, Thailand in 2005, and the Ph.D. degree in 2014 from Department of Computer and Information Science at King Mongkut's University of Technology North Bangkok, Bangkok, Thailand. Since 2010, he has worked in the Computer Science Program at Suratthani Rajabhat University, Thailand. His current research in ML and AI.

Luepol Pipanmaekapornis is currently a lecturer at the Department of Computer and Information Science, King Mongkut's University of North Bangkok, Thailand. He holds bachelor's and master's degrees in computer science. He also earned his doctoral degree in computer science from Queensland University of Technology, Australia, graduating in 2013. His current research interests involve information retrieval, web mining, and data mining.

Suwatchai Kamolsantiroj is currently a lecturer at the Department of Computer and Information Science, King Mongkut's University of North Bangkok, Thailand. He holds master's degrees in computer science. He also earned his doctoral degree in computer engineering from Kasetsart University, Thailand, graduating in 2013. His current research interests involve fuzzy logic, neural network, and artificial intelligence.