2018 24th International Conference on Pattern Recognition (ICPR)
Beijing, China, August 20-24, 2018

# Multi-label Semantic Decoding from Human Brain Activity

Dan Li[1,2], Changde Du[1,2], Lijie Huang[1], Zhiqiang Chen[1,2], Huiguang He[1,2,3]
[1]Research Center for Brain-inspired Intelligenceand National Laboratory of Pattern Recognition
Institute of Automation Chinese Academy of Sciences, Beijing, China,
[2]University of Chinese Academy of Sciences, Beijing, China
[3]Center for Excellence in Brain Science and Intelligence Technology,
Chinese Academy of Sciences, Beijing, China
Email: {lidan2017, duchangde2016, lijie.huang, chenzhiqiang2014, huiguang.he}@ia.ac.cn

***Abstract*—It is meaningful to decode the semantic information from functional magnetic resonance imaging (fMRI) brain signals evoked by natural images. Semantic decoding can be viewed as a classification problem. Since a natural image may contain many semantic information of different objects, the single label classification model is not appropriate to cope with semantic decoding problem, which motivates the multi-label classification model. However, most multi-label models always treat each label equally. Actually, if dataset is associated with a large number of semantic labels, it will be difficult to get an accurate prediction of semantic label when the label appears with a low frequency in this dataset. So we should increase the relative importance degree to the labels that associate with little instances. In order to improve multi-label prediction performance, in this paper, we firstly propose a multinomial label distribution to estimate the importance degree of each associated label for an instance by using conditional probability, and then establish a deep neural network (DNN) based model which contains both multinomial label distribution and label co-occurrence information to realize the multi-label classification of semantic information in fMRI brain signals. Experiments on three fMRI recording datasets demonstrate that our approach performs better than the state-of-the-art methods on semantic information prediction.**

## I. INTRODUCTION

Brain decoding, including semantic decoding [1], [2], voice and speech decoding [3], image reconstruction [4], etc, can be viewed as the study of finding the corresponding stimulus by the evoked brain responses. Semantic decoding is one of the most important parts of brain decoding, which aims to predict the semantic labels about the objects or landscapes presented in natural image from human brain activity signals measured by fMRI. Since a natural image may contain many objects like fish, water, building, people etc, brain signal might be associated with more than one semantic labels simultaneously. The single label classification model is not appropriate to cope with semantic decoding problem, which motivates the multi-label classification model.

The multi-label learning is to induce a function to predict a subset of labels for an unseen instance from a given label set [5]. During the past years, many multi-label models have been proposed [6], [7] and successfully applied in different domain [8], [9]. Geng *et al.* proposed a multi-label learning approach [10] which used an iterative label propagation (LP) technique to estimation relative importance of each label. As a result, the more times a label appears in the neighbor of an instance, the more important this label is for the instance. However, this label propagation technique may aggravate the imbalance when some labels correspond more instances than others.

Actually, if dataset is associated with a large number of semantic labels, it will be difficult to get an accurate prediction of semantic label when the label appears with a low frequency in this dataset. So we should increase the relative importance degree to the labels that associate with little instances. Besides, reasonable multinomial label distribution could help improving prediction performance as the fMRI data itself has a lot of noise in the measurement.

In this paper, we propose a DNN based model which contains both multinomial label distribution and label co-occurrence information by weight initialization to realize the multi-label classification of semantic information in fMRI brain signals. The multinomial label distribution model can increase the weight to the labels that associate with little instances in the process of semantic decoding. Considering the label co-occurrence information, we treat some of the neurons as dedicated neurons for each pattern of label co-occurrence in the final hidden layer. Then we use multinomial label distribution in weight initialization for dedicated neurons. Finally, extensive experiments demonstrate that our approach can predict semantic information from fMRI measurements more accurately.

## II. RELATED WORK

In the field of brain decoding, there are a relatively limited number of semantic decoding models have been proposed to date. Huth *et al.* decoded the semantic content of natural movies from human brain activity by using a hierarchical logistic regression (HLR) model [1]. Stansbury *et al.* used Latent Dirichlet Allocation [11] method to explore the issue that how human brain aggregate information about objects to represent scene categories [2]. However, this methods just using linear features of the fMRI data and the linear model can't learn deep representation from raw data.

Recently, deep neural networks have shown great superiority in learning deep representation from raw data [12]. It has also

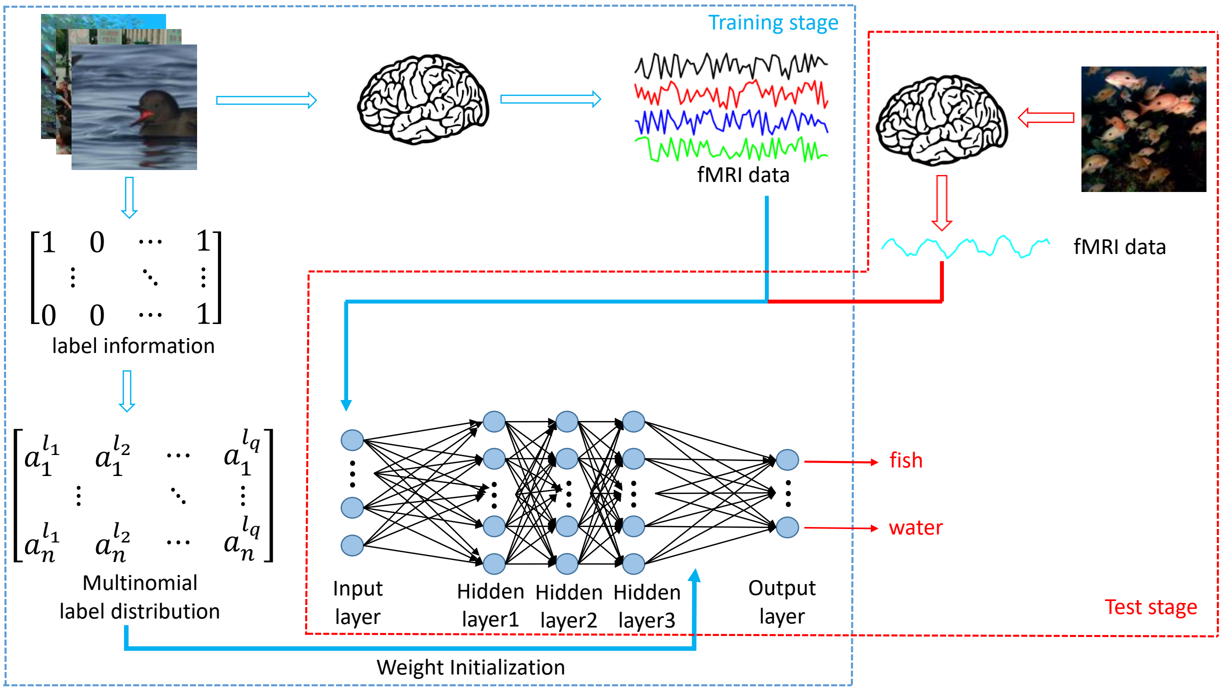


Fig. 1. Our framework for semantic decoding from human brain activity. $l_i$ $(i = 1, ..., q)$ is the label. $n$ is the number of label co-occurrence patterns $(n \leq N)$.

been successfully applied to multi-label tasks [13], [14]. But most of them treated each label equally and did not consider the relationship among instances and labels.

## III. PROPOSED METHOD

To predict the complex semantic labels from human brain activity, we firstly propose a model for calculating the multinomial distribution of labels by using conditional probability, and then use the information of this distribution for deep neural network initialization. The proposed multi-label learning framework is showed in Fig.1.

Under the multi-label learning framework, we assume the training set $D = \{(\boldsymbol{x}_1, \boldsymbol{y}_1), (\boldsymbol{x}_2, \boldsymbol{y}_2)..., (\boldsymbol{x}_N, \boldsymbol{y}_N\} = \{\mathbf{X}, \mathbf{Y}\}$ denote a set of $d$ dimensional training instances $\mathbf{X} \in \mathbb{R}^{d \times N}$ and the associated labels $\mathbf{Y} \in \mathbb{R}^{q \times N}$, where $N$ and $q$ are the numbers of instances and label attributes respectively. For a training instance $\boldsymbol{x}_i$ $(i = 1, ..., N)$, its label $\boldsymbol{y}_i = (y_{1i}, ..., y_{qi})^T$ is defined as:

$$y_{ji} = \begin{cases} 1, & \textit{if } x_i \textit{ associated with label } l_j, \\ 0, & \textit{otherwise}. \end{cases} \quad (1)$$

In this paper, the instances denote the fMRI activity patterns. The task of multi-label classification is to learn a model from training set, so that the label $\widetilde{\boldsymbol{y}}$ of a test instance $\widetilde{\boldsymbol{x}}$ can be predicted accordingly.

### A. Multinomial label distribution estimation

If an instance is associated with many semantic labels, it will be difficult to get an accurate prediction of semantic label when the label appears with a low frequency in the dataset. Therefore, treating each label equally is unreasonable. We should increase the relative importance degree to the labels that associate with little instances. Inspired by this, we naturally propose a model (i.e., multinomial label distribution) to get the importance degree of each semantic label for an instance.

But the difficulty is that there is no other information to guide the computation of importance degree for each semantic label except the associated labels $\mathbf{Y} \in \{0, 1\}^{d \times N}$ in the training set. Considering the semantic label usually does not appear independently, the concept network [6] is adapted for estimating the multinomial label distribution.

Let $l_j$ and $l_k$ are two labels, $\boldsymbol{y}_{l_j}$ and $\boldsymbol{y}_{l_k}$ are the row vectors corresponding to the two labels in the matrix $\mathbf{Y}$, where $\boldsymbol{y}_{l_j} = (y^1_{l_j}, ..., y^N_{l_j})$, $y^i_{l_j} \in \{0, 1\}$. Given training set, conditional probabilities are as follows:

$$p(l_j|l_k) = \frac{\sum_{i=1}^{N} y^i_{l_k} y^i_{l_j}}{\sum\limits_{i=1}^{N} y^i_{l_k}}, \qquad p(l_k|l_j) = \frac{\sum_{i=1}^{N} y^i_{l_k} y^i_{l_j}}{\sum\limits_{i=1}^{N} y^i_{l_j}}. \quad (2)$$

In general, $p(l_j|l_k) \neq p(l_k|l_j)$. If an image is labeled with “fish”, it is usually labeled with “water”; however, if an image is labeled with “water”, the possibility that the image is also labeled with “fish” is not very high due to the fact that this image may also be related to other labels such as “boat”, “swimmer”, and “beach” [7]. If $p(l_j|l_k) > p(l_k|l_j)$, we note that the object corresponds to the label $l_k$ is more surprisal than label $l_j$. That is to say, the frequency of the label $l_k$ is less than label $l_j$.

We use $a^{l_k}_x$ to denote the importance degree of semantic label $l_k$ for fMRI activity pattern $x$. The definition of multinomial label distribution in mathematical is as follows:

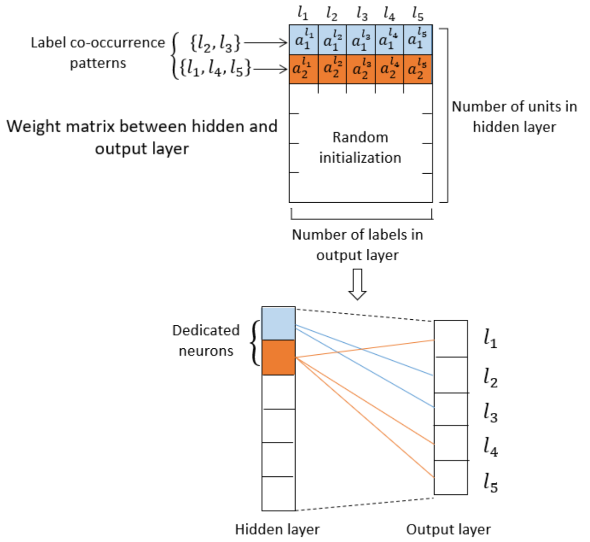


Fig. 2. Use multinomial label distribution in weight initialization for dedicated neurons.

**Definition.** *Multinomial label distribution*
*The $a_x^{l_k}$ satisfies the following constraints:*

- *non-negativity and normalization: $a_x^{l_k} \in [0, 1]$, $\sum_{l_k} a_x^{l_k} = 1$.*
- *If there is no relationship between the label $l_k$ and the instance x, then $a_x^{l_k} = 0$.*
- *The non-zero probability $a_x^{l_k}$ should be different among all possible labels associated with a fMRI pattern x.*

Assume the associated labels of an instance $x$ are $l_i$, $l_j$, $l_t$, $l_k$, the surprisal value for label $l_k$ is defined as:

$$s_x^{l_k} = \sum_{m=l_i,l_j,l_t} p(m|l_k). \tag{3}$$

The higher the value of $s_x^{l_k}$ is, the more important the label $l_k$ is. Based on the definition, the importance degree for label $l_k$ with instance $x$ is estimated as:

$$a_x^{l_k} = \frac{exp(s_x^{l_k})}{\sum_{m=l_i,l_j,l_t,l_k} exp(s_x^m)}. \tag{4}$$

This multinomial label distribution can reflect the relationship among semantic label and fMRI activity pattern, and increase the weight of the labels that associate with little instances. People always automatically pay more attention to some unusual objects. We wonder if the proposed importance degree distribution model which learned from dataset shares similarity with human attention. So in the later experiments, we use the salience object detection method to validate the relationship among human attention and our proposed multinomial label distribution model.

### *B. Multi-label Deep network*

Every instance corresponds to a label co-occurrence pattern, and each label co-occurrence pattern corresponds to a label distribution. The distribution reflects the importance degree for the labels in a label co-occurrence pattern. However, label co-occurrence itself also contains information.

Kurata *et al*. proposed a neural network initialization method to active the pattern of label co-occurrence [15]. This method mainly aimed to solve multi-label text classification problem. They used the frequency of the label co-occurrence patterns in training data to initialize the weight. In this weight initialization process, they also treated each label equally in a label co-occurrence pattern.

Based on the above method, we embed the multinomial label distribution information in a three layer DNN according to the weight initialization.

If semantic labels of images are same, we assume they have the same label distribution and label co-occurrence pattern. In the final hidden layer, we treat some of the neurons as dedicated neurons for each pattern of label co-occurrence. The number of dedicated neurons is equal to the number of label co-occurrence patterns in the training data. We use multinomial label distribution in weight initialization for dedicated neurons. Fig.2 shows the key idea of our method.

For simplicity in the following explanation, we assume that the training dataset has a finite label set $\{l_1, l_2, l_3, l_4, l_5\}$. There are two label co-occurrence $\{l_2, l_3\}$ and $\{l_1, l_4, l_5\}$. For each pattern of label co-occurrence, we initialize a weight matrix row as its label distribution, i.e., the value of each element in the row is the importance degree of the label in its pattern of label co-occurrence. $a_i^{l_j}$ denotes the importance degree of semantic label $l_t$ in the pattern $i$. According to the previous definition, we can get $a_1^{l_1} = a_1^{l_4} = a_1^{l_5} = a_2^{l_2} = a_2^{l_3} = 0$. So these dedicated neurons are initialized to connect the corresponding co-occuring labels with stronger weights than others. And non-zero probability increases the weight of the important labels further. Therefore, this three layer DNN structure contains both label co-occurrence information and multinomial label distribution information.

The remaining rows that are not associated with the label co-occurrence pattern are randomly initialized, which plays a role as a balance factor in traditional machine learning methods. All the connection weights in the DNN including the connection weights between the dedicated neurons and all labels are updated through back-propagation. Training cost function for the DNN is selected by binary cross entropy.

## IV. Experiments

To validate the effectiveness of the proposed method, we perform experiments on three human subjects' fMRI recording datasets. The experiments was divided into two parts, firstly, we use the salience object detection method to validate the relationship among human attention and our proposed multinomial label distribution model, and the second is the performance of semantic decoding using fMRI recoding data.

### *A. Experimental setup*

*1) **Data description**:* We conduct experiments on three fMRI datasets obtained from three subjects [4], and the datasets are available at the website[1]. The three fMRI datasets

[1]https://crcns.org/data-sets/vc/vim-2

TABLE I
THE DETAIL OF THE THREE DATASETS USED IN OUR EXPERIMENTS

| *Datasets* | *#Instances* | *#Feature* | *#ROI* | *#Training* |
|---|---|---|---|---|
| Subject 1 | 7740 | 734 | V4 | 7200 |
| Subject 2 | 7740 | 1032 | V4 | 7200 |
| Subject 3 | 7740 | 753 | V4 | 7200 |

correspond to the same image stimuli. We use fMRI data from visual area V4 for analysis. The number of labels in training dataset and test dataset are respectively 113 and 84, and each instance has 4 labels on average. The label annotation tool is available at the website[2]. Table I summarizes the detail characteristics of the fMRI datasets.

2) ***Comparing algorithms**:* In this paper, we compare the performance of our method with four relevant algorithms:

- **DNN+Label propagation**: Label propagation algorithm estimates the importance degree of each associated label for an instance by iterative techniques [10]. To make a fair comparison with our method, we apply the label propagation algorithm to calculate the label distribution matrix which is used between the final hidden layer and output layer in the DNN structure to initialize the weights. And then use this DNN for multi-label classification prediction.
- **DNN+** $\sqrt{f} \times UB$: This algorithm assumes that a specific pattern of label co-occurrence appears in the training data $f$ times, and used $\sqrt{f} \times UB$ for weight initialization [15]. $UB$ is the upper bound which determined by the number of units in the final hidden layer and output layer [16].
- **CA2E**: CA2E is a multi-label classification method that joint feature and label embedding by a deep latent space [14], which is composed of 2 layer of fully connected neural network and both single fully connected layer structures.

3) ***Evaluation metrics**:* To evaluate the performance of the semantic decoding, we use the hamming loss, microF1 and top-5 accuracies as the evaluation metrics [17].

## B. Experimental Results

1) ***Multinomial label distribution and the Saliency detection**:* We have built a model which assumes that the importance degree of each semantic label is different. We wonder if the proposed model which learned from dataset shares similarity with human attention. Saliency detection methods can simulate human visual attention via detecting saliency in images and videos. Therefore we use a spectral approach [18] to detect salient object in 6 images as showed in Fig.3. These images are the stimulations corresponding to the training data (fMRI), which contain different kinds of objects. Table II shows the result of the label distribution calculated by our model correspond to the image in Fig.3. As it can be seen, the importance degree distribution is in line with the result of salient object detection, for example, the reef is in the saliency ares in image A but sea is not, the importance degree for reef and sea calculated by our distribution model is 0.42 and 0.067 respectively. This suggests that the proposed multinomial label distribution model on this dataset can reflect human attention and the model is reasonable.

[2]https://imagga.com/

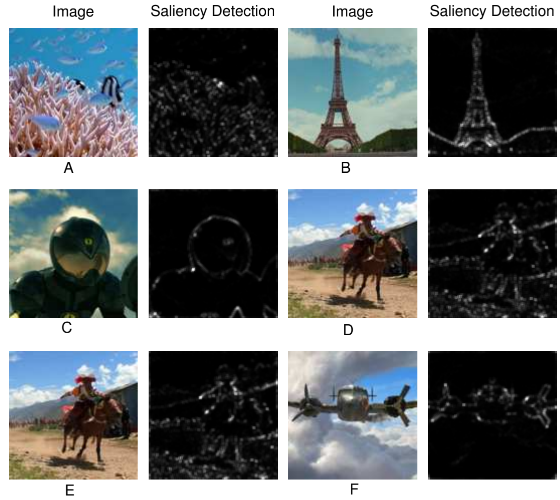


Fig. 3. Images and Saliency detection results.

TABLE II
IMAGES AND CORRESPONDING LABEL DISTRIBUTION

| Images | Label distribution |
|---|---|
| A | reef: 0.42 coral: 0.34 fish: 0.12 sea: 0.067 |
| B | tower: 0.39 tourism: 0.27 city: 0.1934 sky: 0.047 |
| C | face: 0.501 sky: 0.499 |
| D | horse: 0.7 grass: 0.12 people: 0.1 sky: 0.054 |
| E | airplane: 0.51 fly: 0.26 sky: 0.16 |
| F | building: 0.27 sky: 0.23 water: 0.21 plant: 0.15 people: 0.14 |

TABLE III
INVESTIGATION NEURAL NETWORK AFTER BACK-PROPAGATION

| average-kl | weight-dedicated (mean,variance) | weight-all (mean,variance) |
|---|---|---|
| 0.01 | (0.2142, 0.0024) | (-0.003, 0.0032) |

2) ***Performance of semantic decoding**:* We use the same DNN structure in all experiments to guarantee the fairness. As mentioned above, we combine the label co-occurrence information with the multinomial label distribution information for the weight initialization. Now we investigate the situation of the two information after back-propagation. Since each label co-occurrence pattern corresponds to a label distribution, we use the Kullback-Leibler (KL) divergence to measured the similarity of the label distributions before and after back propagation, and average all KL values. Then we calculate the mean of connection weights between the dedicated neurons and corresponding co-occurring labels and compared with the mean of all connections in this weight matrix. As is shown in Table III, the network retains a lot of information for the

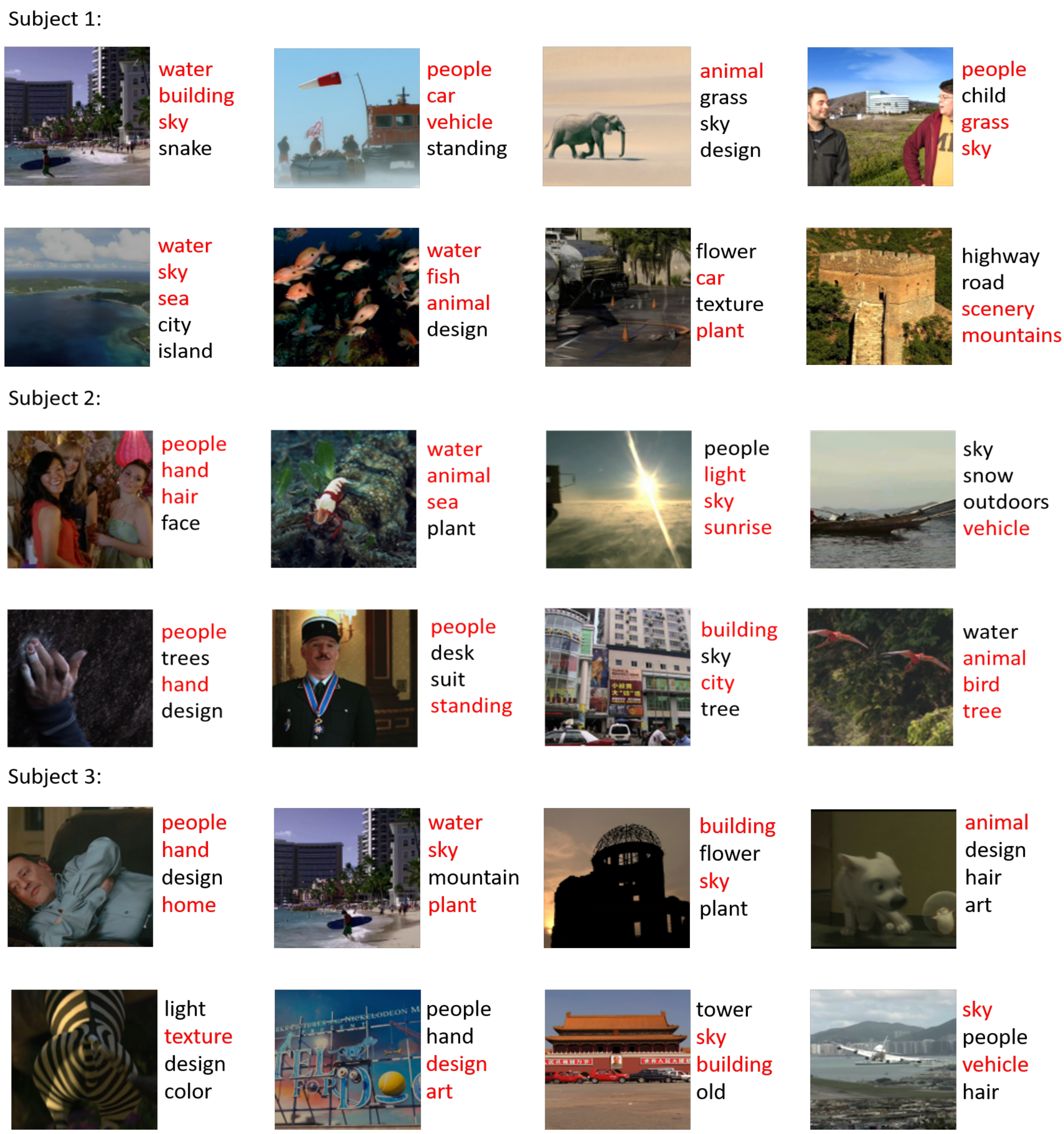


Fig. 4. The results of semantic decoding conducted by our method. The red label indicates the correct prediction.

label-label and instance-label after back-propagation.

The semantic decoding results of our method are showed in Fig 4, where the image is the stimulus, and the right side of each image is the predicted semantic label according to the corresponding fMRI data. The red label indicates the correct prediction. Many general categories, such as people, water, sky are decoded accurately, which suggests that these categories are represented by specific pattern of activity in the brain. In contrast, some subtle categories are decoded poorly, even though we can accurately decode the hyponyms of these poorly decoded general categories. For example, elephant, dog, and airplane are poorly decoded, but its hyponyms, such as animal and vehicle are decoded accurately. Considering the recall rate for each category, we compare our method with the CA2E method [14] and DNN (i.e., a three layer DNN that does not contain any prior information ) on fMRI data subject 1. As is shown in Fig.5, the number of categories predicted by our method is much larger than the CA2E method [14]. Table IV reports the detailed results of each algorithms on the three datasets. H.l, O.f1, top-5 denote hamming loss, microF1 and top-5 accuracies respectively. For each evaluation criteria, ↑ means the larger the better, ↓ means the smaller the better.

## V. Conclusion

In this paper, we propose a multi-label learning model to decode the semantic information from fMRI data. We design a novel multinomial label distribution model and obtain the label distribution for every instance. Simultaneously, we

TABLE IV
PREDICTIVE PERFORMANCE OF EACH COMPARING ALGORITHM ON THREE DATASETS.

| Datasets | Subject1 | | | Subject2 | | | Subject3 | | |
|---|---|---|---|---|---|---|---|---|---|
| | H.l ↓ | O.f1 ↑ | top-5 ↑ | H.l ↓ | O.f1 ↑ | top-5 ↑ | H.l ↓ | O.f1 ↑ | top-5 ↑ |
| DNN | 0.053 | 0.19 | 0.55 | 0.049 | 0.24 | **0.62** | 0.053 | 0.18 | 0.52 |
| DNN+Label propagation [10] | 0.054 | 0.17 | 0.52 | 0.050 | 0.23 | 0.59 | 0.054 | 0.16 | 0.48 |
| DNN+ $\sqrt{f} \times UB$ [15] | 0.053 | 0.18 | 0.54 | 0.050 | 0.22 | 0.51 | 0.053 | 0.18 | 0.51 |
| Our method | **0.052** | **0.21** | **0.56** | **0.048** | **0.26** | 0.61 | **0.052** | **0.20** | **0.56** |

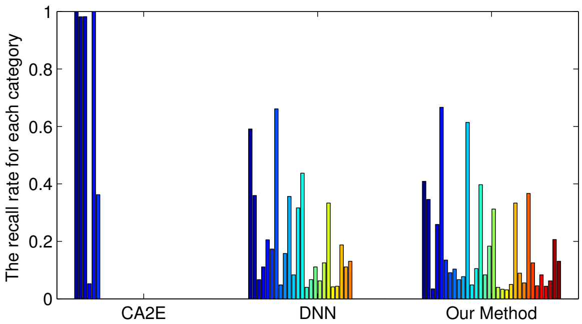


Fig. 5. Comparing CA2E method with our method on the recall rate for each category, we find that the CA2E method only predicts six categories, and distributes the same labels for a large number of different test instances. So the recall rates of the four categories are very high. The number of categories predicted by our method and DNN is 32 and 24. And the average recall rate is 0.18 and 0.15 respectively.

also consider the label co-occurrence information and then uniquely joint label-instance and label-label information in a unified DNN model. Extensive experiments clearly validate our method is effective for semantic decoding.

## ACKNOWLEDGMENT

This work was supported by National Natural Science Foundation of China (91520202, 81701785), Youth Innovation Promotion Association CAS, the CAS Scientific Research Equipment Development Project (YJKYYQ20170050) and the Beijing AI Project.

## REFERENCES

[1] A. G. Huth, T. Lee, S. Nishimoto, N. Y. Bilenko, A. T. Vu, and J. L. Gallant, "Decoding the semantic content of natural movies from human brain activity," *Frontiers in systems neuroscience*, vol. 10, 2016.

[2] D. E. Stansbury, T. Naselaris, and J. L. Gallant, "Natural scene statistics account for the representation of scene categories in human visual cortex," *Neuron*, vol. 79, no. 5, pp. 1025–1034, 2013.

[3] E. Formisano, F. De Martino, M. Bonte, and R. Goebel, ""who" is saying "what"? brain-based decoding of human voice and speech," *Science*, vol. 322, no. 5903, pp. 970–973, 2008.

[4] S. Nishimoto, A. T. Vu, T. Naselaris, Y. Benjamini, B. Yu, and J. L. Gallant, "Reconstructing visual experiences from brain activity evoked by natural movies," *Current Biology*, vol. 21, no. 19, pp. 1641–1646, 2011.

[5] Z. H. Zhou and M. L. Zhang, "Multi-label learning." 2017, to be published.

[6] W. Zhang, X. Xue, J. Fan, X. Huang, B. Wu, and M. Liu, "Multi-kernel multi-label learning with max-margin concept network," in *IJCAI*, vol. 22, no. 1, 2011, pp. 1615–1620.

[7] X. Li, X. Zhao, Z. Zhang, F. Wu, Y. Zhuang, J. Wang, and X. Li, "Joint multilabel classification with community-aware label graph learning," *IEEE Transactions on Image Processing*, vol. 25, no. 1, pp. 484–493, 2016.

[8] B. Wu, S. Lyu, B. G. Hu, and Q. Ji, "Multi-label learning with missing labels for image annotation and facial action unit recognition," *Pattern Recognition*, vol. 48, no. 7, pp. 2279–2289, 2015.

[9] J. Liu, W. C. Chang, Y. Wu, and Y. Yang, "Deep learning for extreme multi-label text classification," in *SIGIR*. ACM, 2017, pp. 115–124.

[10] Y. Li, M. Zhang, and X. Geng, "Leveraging implicit relative labeling-importance information for effective multi-label learning," in *ICDM*, 2015, pp. 251–260.

[11] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *Journal of machine Learning research*, vol. 3, no. Jan, pp. 993–1022, 2003.

[12] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[13] J. Wang, Y. Yang, J. Mao, Z. Huang, C. Huang, and W. Xu, "Cnn-rnn: A unified framework for multi-label image classification," in *CVPR*, 2016, pp. 2285–2294.

[14] C. K. Yeh, W. C. Wu, W. J. Ko, and Y. C. F. Wang, "Learning deep latent space for multi-label classification." in *AAAI*, 2017, pp. 2838–2844.

[15] G. Kurata, B. Xiang, and B. Zhou, "Improved neural network-based multi-label classification with better initialization leveraging label co-occurrence." in *HLT-NAACL*, 2016, pp. 521–526.

[16] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, 2010, pp. 249–256.

[17] X. Kong, M. K. Ng, and Z.-H. Zhou, "Transductive multilabel learning via label set propagation," *IEEE Transactions on Knowledge and Data Engineering*, vol. 25, no. 3, pp. 704–719, 2013.

[18] X. Hou and L. Zhang, "Saliency detection: A spectral residual approach," in *CVPR*. IEEE, 2007, pp. 1–8.