

Objects Categorization on fMRI Data: Evidences for Feature-Map Representation of Objects in Human Brain

Sutao Song¹, Jiakai Zhang², and Yuehua Tong¹(✉)

¹ School of Education and Psychology, University of Jinan, Jinan 250022, China
sep_tongyh@ujn.edu.cn

² College of Information Science and Technology, Beijing Normal University,
Beijing 100875, China

Abstract. Brain imaging studies in humans have reported each object category was associated with different neural response pattern reflecting visual, structure or semantic attributes of visual appearance, and the representation of an object is distributed across a broader expanse of cortex rather than a specific region. These findings suggest the feature-map model of object representation. The present object categorization study provided another evidence for feature-map representation of objects. Linear Support Vector Machine (SVM) was used to analyze the functional magnetic resonance imaging (fMRI) data when subjects viewed four representative categories of objects (house, face, car and cat) to investigate the representation of different categories of objects in human brain. We designed 6 linear SVM classifiers to discriminate one category from the other one (1 vs. 1), 12 linear SVM classifiers to discriminate one category from other two categories (1 vs. 2), 3 linear SVM classifiers to discriminate two categories of objects from the other two categories (2 vs. 2). Results showed that objects with visually similar features have lower classification accuracy under all conditions, which may provide new evidences for the feature-map representation of different categories of objects in human brain.

Keywords: Feature-map representation · Support vector machine · fMRI

1 Introduction

The representation of objects in human brain is a matter of intense debate, and in the domain of neuroimaging study, three major models exist [1, 2]: category specific model, process-map model and feature-map model.

The category specific model proposes that ventral temporal cortex contains a limited number of areas that are specialized for representing specific categories of stimuli. Evidences from patients with brain lesion showed that patient with lesions in one particular brain area lost the ability to recognize facial expressions or other objects [3, 4]. One study found that a farmer with brain lesion no longer recognized his own cows [4]. The study of event-related-potentials (ERP) and magnetic encephalography (MEG) supported the face specificity in visual processing, human faces elicited a negative component peaking at about 170 ms from stimulus onset (N170 or M170) [5–7].

Functional magnetic resonance imaging (fMRI) studies described specialized areas for faces and some specific objects: the fusiform face area (FFA) for human faces, the parahippocampal place area (PPA) for scenes and the “extrastriate body area” (EBA) for visual processing of the human body [8–13].

The process-map model [1, 14–16] proposes that different areas in ventral temporal cortex are specialized for different types of perceptual processes. The studies from Gauthier et al. showed that FFA was not just specialized for faces, but for expert visual recognition of individual exemplars from any object category. For example, for bird experts, FFA shows more activity when they view the pictures of bird, and for car experts, FFA shows more activity when they view car than bird. Study also showed that the acquisition of expertise with novel objects (such as greebles, one kind of man-made object) led to increased activation in the right FFA [14].

For feature-map model, it proposes that the representations of faces and different categories of objects are widely distributed and overlapping [2, 17–21]. In the study Haxby et al. [2], fMRI data of ventral temporal cortex was recorded when subjects viewed faces, cats, five categories of man-made objects, and nonsense pictures. A correlation-based distance measure was used to predict the object categories and the prediction result indicates that the representations of faces and objects in ventral temporal cortex are widely distributed and overlapping.

The evidences for the three models came mainly from the neuroimaging study of healthy subjects or patients with brain lesion. Generally, for the analysis of the neuroimaging data, univariate method was used, such as general linear model. However, fMRI was multi-variate in nature. In recent years, multi-variate pattern analysis (MVPA) methods have been widely used in fMRI analysis [22–26]. Compared with the traditional univariate method, MVPA method takes the correlation among neurons or cortical locations into consideration and is more sensitive and informative. In this study, we further investigate the representation of objects in human brain using Support Vector Machine (SVM). As one representative MVPA method, SVM is effective in digging the information behind fMRI data. Four representative objects (house, face, car and cat) were selected as stimulus, which can be grouped in the following ways: face vs. other objects, animate vs. inanimate objects. SVM was applied to predict the label of brain states, i.e. which kind of stimulus the subject was viewing, and 6 classifiers were trained to classify one object category versus the other category (house vs. face, house vs. car, house vs. cat, face vs. car, face vs. cat, car vs. cat). To further investigate the representation of objects in human brain, 15 other classifiers were trained to cover the possible combinations of the 2-class classification problem for the four categories of objects (1 vs. 2, 2 vs. 2).

2 Method

2.1 Subjects and fMRI Data Acquisition

The data came from one of our previous study [26]. Fourteen healthy college students participated in this study (6 males, 8 females). Subjects gave written informed consent. A 3-T Siemens scanner equipped for echo planar imaging (EPI) at the Brain Imaging

Center of Beijing Normal University was used for image acquisition. Functional images were collected with the following parameters: repeat time (TR) = 2000 ms; echo time (TE) = 30 ms; 32 slices; matrix size = 64×64 ; acquisition voxel size = $3.125 \times 3.125 \times 3.84 \text{ mm}^3$; flip angle (FA) = 90° ; field of view (FOV) = 190 ~ 200 mm. In addition, a high-resolution, three-dimensional T1-weighted structural image was acquired (TR = 2530 ms; TE = 3.39 ms; 128 slices; FA = 7° ; matrix size = 256×256 ; resolution = $1 \times 1 \times 1.33 \text{ mm}^3$).

2.2 Stimuli and Experimental Procedure

The experiment was designed in a blocked fashion. Subject participated in 8 runs and each run consisted of 4 task blocks and 5 control blocks. During each task block lasted for 24 s, 12 gray-scale images belonging to one category (houses, faces, cars or cats) were presented which were chosen randomly from 40 pictures of that particular category, and subject had to press a button with left or right thumb as long as images were repeated consecutively. Two identical images were displayed consecutively 2 times randomly during each task block. Each stimulus was presented for 500 ms followed by a 1500 ms blank screen. Control blocks were 12 s fixation in the beginning of a run and at the end of every task block (Fig. 1). Each kind of objects block was presented once during each run, and the order of them was counterbalanced in the whole session which lasted 20.8 min.



Fig. 1. The experimental procedure for one run.

2.3 Data Preprocessing

The preprocessing steps were the same as our previous study [26]. SPM2 (<http://www.fil.ion.ucl.ac.uk/spm/>) was used to finish the preprocessing job. It mainly contains 3 steps: realignment, normalization and smoothing. Subjects were preprocessed separately. In the beginning, the first 3 volumes were discarded as the initial images of each session showed some artifacts related to signal stabilization (according to the SPM2 manual). Images were realigned to the first image of the scan run and were normalized to the Montreal Neurological Institute (MNI) template. The voxel size of the normalized images was set to be $3 * 3 * 4 \text{ mm}$. At last, images were smoothed with 8 mm full-width at half maximum (FWHM) Gaussian kernel. The baseline and the low frequency components were removed by applying a regression model for each voxel [23]. The cut-off period chosen was 72 s.

2.4 Voxel Selection

Voxels that activated for any kind of object within the whole brain were selected for further analysis (family-wise error correction, $p = 0.05$) (Fig. 2).

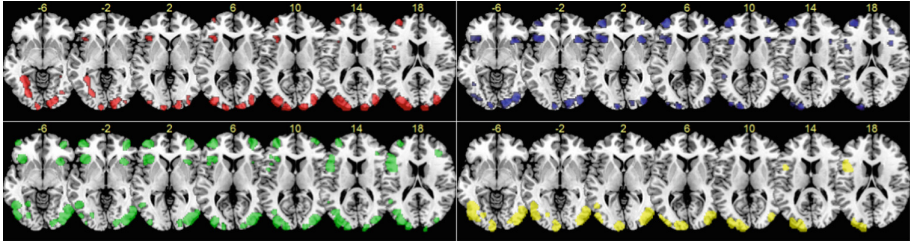


Fig. 2. The voxel selection for one representative subject (slices $-8 \sim 20$), (Red: house; Blue: Face; Green: Car; Yellow: Cat). (Color figure online)

2.5 SVM Method

LibSVM (<http://www.csie.ntu.edu.tw/~cjlin/libsvm>) was used to predict the brain states. The data of first 4 runs was used to train the model, and the data of last 4 runs was used to test the model. To reduce the number of features, principle component analysis (PCA) procedure was conducted over the features and PCs accumulatively accounting for 95% of the total variance of the original data were kept for the subsequent classification. Then the attributes of training data was scaled to the range $[-1, 1]$ linearly; and the attributes of the test data was scaled using the same scaling function of the training data. To compensate the hemodynamic delays, the fMRI signals of each voxel were shifted by 4 s.

3 Results

For all the 21 combinations of two-class classification problems of the four categories of objects (1 vs. 1, 1 vs. 2, 2 vs. 2), the classification accuracies were all above the chance level (Kappa coefficients: 0.73 ± 0.13 , $M \pm SD$).

3.1 Classification Results for One vs. One Classifiers

Classification performances for discriminating one category from another category were shown in Fig. 3. In this situation, 6 classifiers were trained (house vs. face, house vs. cat, face vs. car, car vs. cat). Significant differences were found among the 6 classifiers ($F(3.2, 42) = 11.88, p < 0.001, \eta^2 = .478$). And two groups, house vs. car, face vs. cat, have the lowest classification accuracy. The lower performance in distinguishing houses from car (or face from cat) suggests that houses (or face) share more common activity with car (or cat) and therefore less dissociable.

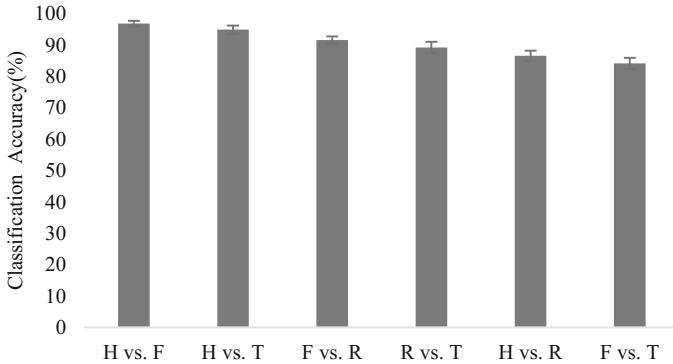


Fig. 3. Classification results for One vs. One Classifiers ($M \pm SE$) (H: house; F: Face; R: Car; T: Cat).

3.2 Classification Results for One vs. Two Classifiers

Classification performances for discriminating one category from the other two categories of objects were shown in Fig. 4. In this situation, 4 groups, and totally 12 classifiers were trained, each group contains three categories of objects (e.g., group one: face vs. house and car; house vs. face and car; car vs. house and face). Significant difference were found among the 3 classifiers for each group ($F(2, 26) = 31.22, p < 0.001, \eta^2 = .706$; $F(1.39, 18.07) = 22.88, p < 0.001, \eta^2 = .638$; $F(2, 26) = 11.18, p < 0.001, \eta^2 = .462$; $F(2, 26) = 28.80, p < 0.001, \eta^2 = .689$). When distinguishing car (or cat) from the other two categories of objects, the classifier performed worst, which suggests that car (or cat) share more common activity with the other categories of objects. Take group one for example, the classifier performed worst to discriminate car from house and face, which implies the similar spatial activity of car with house and (or) face. To look the three 2-class classifiers that involved the three objects (house vs. face, house vs. car, car vs. cat in Fig. 4 further, the two classifiers included car had lower classification accuracy. The results were similar for group 2, 3 and 4.

3.3 Classification Results for Two vs. Two Classifiers

Classification performances for discriminating two categories from the other two categories of objects were shown in Fig. 5. In this situation, 3 classifiers were trained (house and car vs. cat and face; house and cat vs. face and car; house and face vs. car and cat). Significant differences were found among the 3 classifiers ($F(2, 26) = 39.59, p < 0.001, \eta^2 = .753$). When discriminating house and car from face and cat, the classifier performed best, which suggests the dissociative spatial pattern may exist.

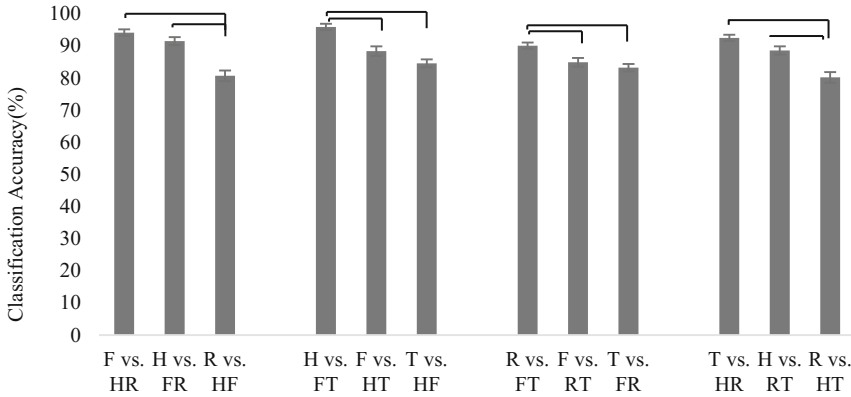


Fig. 4. Classification results for One vs. Two Classifiers ($M \pm SE$) (H: house; F: Face; R:Car; T: Cat).

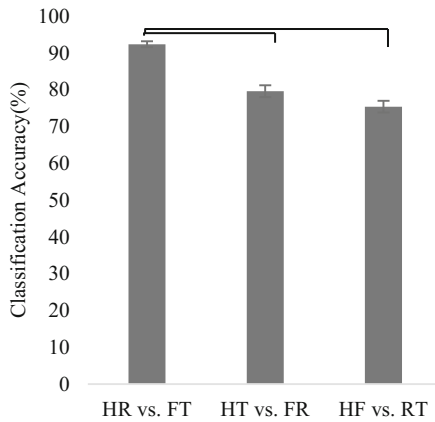


Fig. 5. Classification results for Two vs. Two Classifiers ($M \pm SE$) (H: house; F: Face; R: Car; T: Cat).

3.4 Classification Results for Regions Maximally Responsive to One Category of Objects

The classification accuracies for One vs. One and Two vs. Two classifiers were also provided when the voxels that responded maximally to one category were chosen as features (Fig. 6). Again, significant differences were found (all $ps < .001$). And similar patterns were observed across voxel selection schemes.

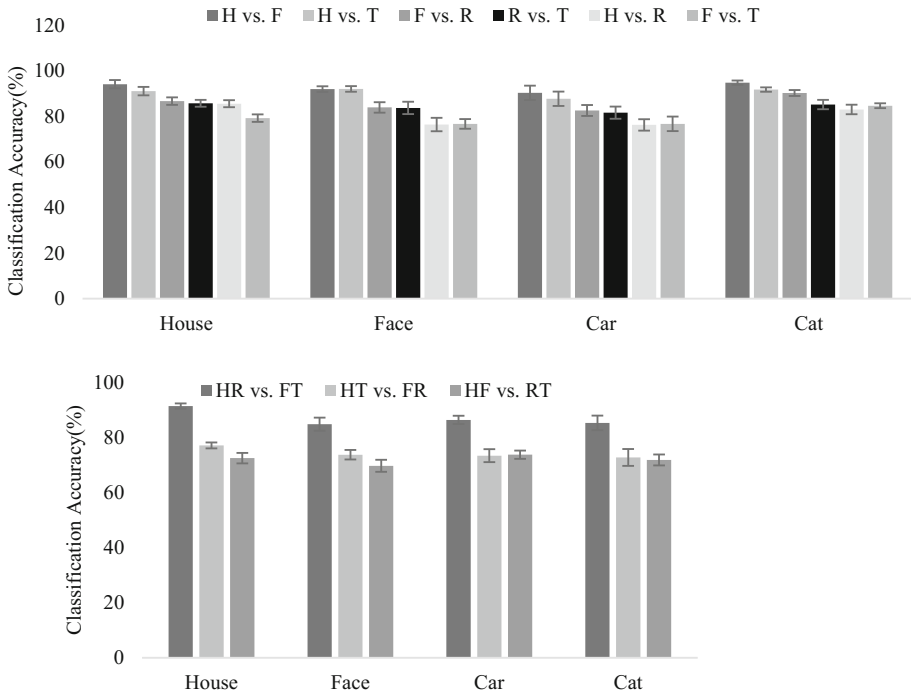


Fig. 6. Classification results for One vs. One (up) and Two vs. Two (below) classifiers when the voxels that responded maximally to one category (i.e. House, face, car or cat) were chosen as features ($M \pm SE$) (H: house; F: Face; R: Car; T: Cat).

4 Discussion and Conclusions

In this study, one MVPA method, SVM was used to analyze the fMRI data when subjects viewed faces and other objects. We investigated the possibility to classify the brain states by various groups. This study selected four representative objects to study the representation of objects in human brain in large scale (i.e. the scale of fMRI technology). Totally 21 classifiers were trained to cover most of the possible combinations of the four objects (the 1 vs. 3 classifiers were not included, as they provide no useful information about the representation way of objects in human brain). Results showed that objects with visually similar features have lower classification accuracy under all conditions, which may provide new evidence for the feature-map representation of different category of objects in human brain.

The current analysis applied linear SVM to predict the categories of objects that the subject viewed. SVM finds a linear combination of features which characterize or separate two or more classes of objects or events. Thus, the higher the classification accuracy is, the less in common the spatial activities are, and vice versa. As one multi-variate analysis method, SVM is powerful in digging information behind fMRI data. However, the use of multi-variate analysis method in fMRI study when subjects

viewed the pictures of faces and objects was not new. Haxby et al. applied correlation based method (it is the first time multi-variate method was used to analyze fMRI data) to classify the brain states evoked by face, cat and five other man-made objects (houses, shoes, scissors, bottles and chairs), and the result supported the feature-map model [2]. Different from Haxby's study, we chose four objects (house, face, car and cat), which can be further classified as animate (face and cat) vs. inanimate objects (house and car). Besides, intuitively, face and cat contain information about face processing (such as features related with eyes, mouth and ears), and house and car contain information related with scene processing. The result of Fig. 3 shows that it is most difficult to classify the brain activities elicited by the following two groups, face vs. cat and house vs. car, which is more likely to support the feature-map model. The similar visual features are represented adjacent spatially in brain, and the brain activity patterns recorded by fMRI are adjacent or overlapped, as the patterns of voxel activities corresponding to each category on the whole brain shown in Fig. 2. Thus, the classification accuracy for linear SVM is low. Besides, when the voxels that responded maximally to one category of objects were chosen as features, similar patterns of classification accuracies were observed as that shown in Figs. 2 and 5, and the accuracies were all above the chance level, indicating the overlapped representations of faces and objects. If the definition of feature is not clear, when we grouped any two categories of objects as one class, the classification result (Fig. 5) shows that the classifier performed best when discriminating house and car from face and cat, while the classifier performed worst when discriminating house and face from car and cat. This result indicates house and car share more features, face and cat share more features in common, and thus have similar brain activity pattern. In other situations, results also showed that objects with visually similar features achieved lower classification accuracy (Fig. 4), which further supports the feature-map representation of different category of objects in human brain.

In conclusion, MVPA methods and fMRI technology provide new way to under the representation of different categories of objects in human brain. The current study shows new evidence for feature-map representation of objects.

Acknowledgments. This research was financially supported by Young Scientist Fund of National Natural Science Foundation of China (NSFC) (31300924), NSFC General program (61375116), the Fund of University of JiNan (XKY1508, XKY1408).

References

1. Gauthier, I.: What constrains the organization of the ventral temporal cortex? *Trends Cogn. Sci.* **4**(1), 1–2 (2000)
2. Haxby, J., et al.: Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* **293**(5539), 2425–2430 (2001)
3. Hecaen, H., Angelergues, R.: Agnosia for faces (prosopagnosia). *Arch. Neurol.* **7**(2), 92 (1962)
4. Assal, G., Favre, C., Anderes, J.: Nonrecognition of familiar animals by a farmer. Zooagnosia or prosopagnosia for animals. *Revue Neurologique* **140**(10), 580 (1984)

5. Carmel, D., Bentin, S.: Domain specificity versus expertise: factors influencing distinct processing of faces. *Cognition* **83**(1), 1–29 (2002)
6. Rossion, B., et al.: The N170 occipito-temporal component is delayed and enhanced to inverted faces but not to inverted objects: An electrophysiological account of face-specific processes in the human brain. *NeuroReport* **11**(1), 69 (2000)
7. Liu, J., et al.: The selectivity of the occipitotemporal M170 for faces. *NeuroReport* **11**(02), 337 (2000)
8. Kanwisher, N., McDermott, J., Chun, M.: The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J. Neurosci.* **17**(11), 4302 (1997)
9. McCarthy, G., et al.: Face-specific processing in the human fusiform gyrus. *J. Cogn. Neurosci.* **9**(5), 605–610 (1997)
10. Kanwisher, N.: Domain specificity in face perception. *Nat. Neurosci.* **3**, 759–763 (2000)
11. Fodor, J.A.: *The Modularity of Mind*. MIT, Cambridge (1981)
12. Downing, P.E., et al.: A cortical area selective for visual processing of the human body. *Science* **293**(5539), 2470 (2001)
13. Epstein, R., Kanwisher, N.: A cortical representation of the local visual environment. *Nature* **392**(6676), 598–601 (1998)
14. Tarr, I.G.M.J., et al.: Activation of the middle fusiform ‘face area’ increases with expertise in recognizing novel objects. *Nat. Neurosci.* **2**(6), 569 (1999)
15. Gauthier, I., et al.: Expertise for cars and birds recruits brain areas involved in face recognition. *Nat. Neurosci.* **3**, 191–197 (2000)
16. Tarr, M.J., Gauthier, I.: FFA: a flexible fusiform area for subordinate-level visual processing automatized by expertise. *Nat. Neurosci.* **3**, 764–770 (2000)
17. Ishai, A., et al.: Distributed representation of objects in the human ventral visual pathway. *Nat. Acad. Sci.* **96**, 9379–9384 (1999)
18. Ishai, A., et al.: The representation of objects in the human occipital and temporal cortex. *J. Cogn. Neurosci.* **12**(Supplement 2), 35–51 (2000)
19. Haxby, J.V., Hoffman, E.A., Gobbini, M.I.: The distributed human neural system for face perception. *Trends Cogn. Sci.* **4**(6), 428–432 (2000)
20. Chao, L.L., Haxby, J.V., Martin, A.: Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects. *Nat. Neurosci.* **2**, 913–919 (1999)
21. Tanaka, K.: Inferotemporal cortex and object vision. *Ann. Rev. Neurosci.* **19**(1), 109–139 (1996)
22. Zhang, H., et al.: Face-selective regions differ in their ability to classify facial expressions. *NeuroImage* **130**, 77–90 (2016)
23. Wegrzyn, M., et al.: Investigating the brain basis of facial expression perception using multi-voxel pattern analysis. *Cortex* **69**, 131–140 (2015)
24. Kragel, P.A., Labar, K.S.: Multivariate neural biomarkers of emotional states are categorically distinct. *Soc. Cogn. Affect. Neurosci.* **10**(11), 1437–1448 (2015)
25. Cowen, A.S., Chun, M.M., Kuhl, B.A.: Neural portraits of perception: Reconstructing face images from evoked brain activity. *NeuroImage* **94**(1), 12–22 (2014)
26. Song, S., et al.: Comparative study of SVM methods combined with voxel selection for object category classification on fMRI data. *PLoS ONE* **6**(2), e17191 (2011)