Visual Information Processing Mechanism Revealed by fMRI Data[*]

Jinpeng Li, Zhaoxiang Zhang, Huiguang He[*]

[a] Research Center for Brain-inspired Intelligence
Institute of Automation, Chinese Academy of Sciences
No. 95 Zhongguancun East Road, Haidian District
Beijing, P.R. China
[b] University of Chinese Academy of Sciences (UCAS)
huiguang.he@ia.ac.cn

Abstract. The functional Magnetic Resonance Imaging (fMRI) data of both the ventral pathway and the dorsal pathway on the visual cortex in a classification task was analyzed. We found that the classification performance improved hierarchically from lower-level regions to higher-level regions in both pathways, which partly verified the visual pathway theory proposed in cognitive neuroscience. Moreover, the LO (Lateral Occipital), V3a and V3b fMRI data were good classification basis no worse than the widely-used features such as GIST, HOG and LBP. It indicated that imitating the activity patterns of visual cortex to design new feature-extraction algorithms might be favorable. Finally, the performance of V3a and V3b voxels were very close to that of LO voxels. Consequently, in the design of brain-like intelligence systems, we should consider the coordination mechanism between the two pathways rather than focusing on the ventral pathway alone. The relationship of human visual pathway and deep learning structure was also discussed tersely.

Keywords: visual cortex; ventral pathway; dorsal pathway; fMRI; classification; representation

1 Introduction

The mechanism of human visual system has long been an attractive research topic, and still under research by enthusiastic scholars. The research mainly involves biology, psychology, cognitive neuroscience and pattern recognition algorithms, thus it forms a comprehensive research field. The understanding of human visual cortex is not supposed to be merely a fundamental research issue concerning medical anatomy or biology, for the benefits it brings to the development of artificial intelligence (AI) and various engineering techniques are beyond measure. The understanding of visual system

[*] This work was supported by National Natural Science Foundation of China (61271151, 91520202) and Youth Innovation Promotion Association CAS.

will in turn instruct us to design more brain-like algorithms (e.g. new deep learning models) to accomplish pattern recognition tasks.

The visual cortex across the brain is believed to be hierarchically organized by different function-specialized regions and could be further divided into two pathways according to different functions, i.e. the ventral pathway and the dorsal pathway [1]. The former one is tightly related to object identification, while the latter one mainly deals with object localization. In 1970s, Hubel and Wiesel realized from their empirical observations that the activity mode of the neurons located in V1 resembled Gabor wavelet filters, and different neurons corresponded to different frequencies and orientations [2, 3]. Their work successfully explained the computational characters in the primary cortex V1, which was the shared "entrance" of both the ventral and dorsal pathway. Encouraged by the success of the shallow V1 model, researchers began to try deeper models under the hope of describing downstream areas and proposed the HMAX model to imitate the activity patterns of the ventral pathway [4]. The basic HMAX structure consisted of four layers. The first layer was formed by Gabor filters. The second layer performed max-pooling. The third layer extracted the output of the second layer and operated template matching, and the forth layer was another pooling layer. The multilayer HMAX model was capable of explaining the computational characters of V1 and V2, but had trouble extending to higher cortical areas such as V4 and IT (Inferior Temporal) [5]. In 1990s, researches turned to a more direct approach. The central methodology was to collect response data to various stimulus at multiple region-of-interests ROIs and used statistical fitting techniques to find model parameters that produced the observed stimulus-response relationship. However, they soon realized that multilayered networks fitted to neural data in higher areas such as V4 ended up overfitting the training data and predicting comparatively small amounts of explained variance on novel testing images [6]. Thus the features extracted from such models were not good classification basis. The reasons might include such two following aspects: (1) the data amount was not large enough to provide a precise representation of connections between regions, and (2) the process of image identification in the visual cortex could not be easily explained by the simple cascaded ventral pathway V1-V2-V4-IT. There were countless coupled and cross-pathway connections between the ventral and dorsal pathways, so the mechanism of object identification cannot be simply described by a single pathway. The contributions of the dorsal pathway should not be neglected, so the complex synergistic effects of the dorsal pathway (e.g. V3a and V3b areas) should also be included into the models. We should study the activity patterns of ROIs located along the dorsal pathway as well as ventral ones when designing new feature-extraction algorithms.

The instruments for visual cortex research differs according to diverse application fields, in which fMRI is suitable for cerebral cortex imaging. The fMRI technology measures the Blood Oxygen Level Dependent (BOLD) in the brain vessels, which is tightly related to image-understanding process. Simultaneously, fMRI is able to offer us a deep and precise insight into different ROIs at a time resolution of less than one second and a space resolution in millimeter. With the help of fMRI and distributed pattern analysis method, researchers were able to investigate where and how complex natural scene information was encoded and discriminated by the brain [7, 8].

In this paper, we analyzed fMRI data for both the ventral pathway V1-V2-V4-LO and the dorsal pathway V1-V2-V3-V3a-V3b in a natural image classification task, and the results were instructive.

2 Data and Task

In order to explore the organization and function of the visual cortex, as well as to identify which areas are involved in object recognition process, the classification accuracy (based on BOLD values of each region) is introduced as the analyzing tool and evaluation criterion. BOLD is a direct measurement of the cerebral cortex activities, so the better the image is encoded by cortex regions, the higher classification accuracy of the BOLD-based classifiers will acquire.

The data set contained the fMRI responses of 1750 natural photographs, and the stimulus included animals, buildings, food, humans, indoor scenes, manmade objects, outdoor scenes, and textures. During the experiment, the subject looked at a sequence of natural photographs displayed on a screen, and at the same time, the BOLD responses of multiple cortex regions were recorded by fMRI scanning synchronously. The experiment used flashing technique to enhance the signal-to-noise ratio of voxel responses. The fMRI responses for each image were recorded according to the stimulus design shown in Fig.1. Seven ROIs were considered, including V1, V2, V3, V3a, V3b, V4 and LO, and their overall tridimensional distribution on the occipital lobe was shown in Fig.2 (a). In Fig.2 (b), two different paths were illustrated, in which V1, V2, V4, LO belonged to the ventral pathway, and V1, V2, V3, V3a, V3b belonged to the dorsal pathway. In order to optimize the data structure, several steps for preprocessing were performed, i.e. the alignment was performed manually and the data were temporally interpolated to account for differences in slice time acquisition [9]. Peak BOLD responses to each of the 1750 images were then estimated from the preprocessed data and stored. The responses for each voxel were z-scored, so for a given voxel the units of each “response” were standard deviations from that voxel’s mean response [9]. Notice that in an fMRI map, the voxel numbers of each ROI was different according to the researchers’ selection, as is shown in Table 1.

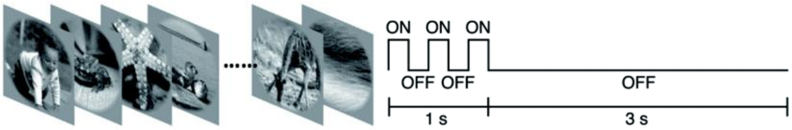

Fig. 1. Stimulus design. Every image was shown in a 1s-3s schedule. During the first 1s, the same image flashed three times (each time for 200ms) to stimulate the brain’s corresponding response patterns to the maximum, and the following 3s was grey background, then the next picture was shown.

The dataset was originally contributed by Jack Gallant et al. at UC Berkeley [9, 10]. For more detailed information about the data set, or download it for research purpose, log on to the website (https://crcns.org/data-sets/vc/vim-1/about-vim-1).

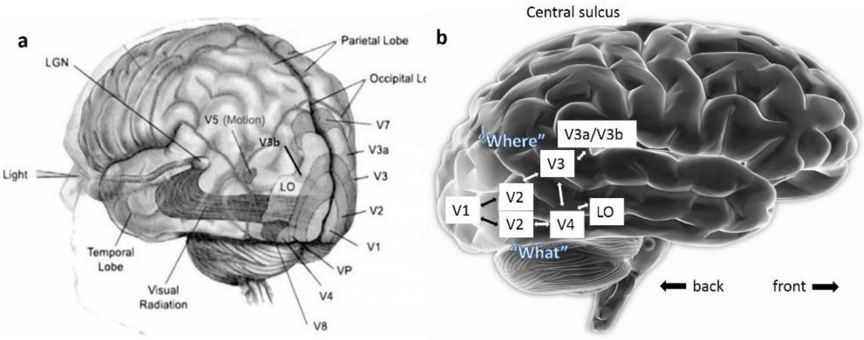

Fig. 2. The distribution and voxel number of ROIs. (a) The brief structure of human visual system[1]. The visual information is first collected by the retina and transmitted to the Lateral Geniculate Nucleus (LGN), then get into the visual cortex mainly located at the Occipital Lobe (OL) via the visual radiation. Henceforth, the brain extracts complex features in a highly-nonlinear way and begins the understanding process. (b) The two visual pathways. The dorsal pathway deals with the "where" problem and the ventral pathway deals with the "what" problem. The double sided arrows indicate that the information flow in both pathways are bidirectional rather than unidirectional, and there are connections between the two channels.

Table 1. The voxel numbers considered in seven ROIs.

ROIs	**V1**	**V2**	**V3**	**V3a**	**V3b**	**V4**	**LO**
Voxel number	1294	2036	1973	484	314	701	928

In order to perform classification task, we tagged the output labels for the 1750 fMRI maps by hand. We selected 1575 samples (90%) for supervised training and 175 samples (10%) for validation.

3 Experiments and Results

The dimensionality of fMRI signal (often more than 1000 for each region) was too high in terms of the limited sample amount (1575). So the full-connected shallow networks with backpropagation (BP) algorithm were not favorable (generally to implement a model with full-connection networks, the training cases should be at least ten times the number of total parameters of the networks [5]. Our data set apparently could not meet such strict requirement). In order to efficiently perform classification with small sample amount and high-dimensional features, we chose SVM classifiers and performed PCA before classification.

All the results in our work were obtained by three-fold cross-validation and shown in the form of mean $\pm$SD.

[1] https://quizlet.com/11094814/neuro-3-vision-2-chp-6-flash-cards/

3.1 The Rising Trend of Performance along Both Pathways

We found that there were distinguishable differences among each ROI's performance along both pathways, and there were some regular patterns or distinct trends that acted in accordance with cognitive neuroscience findings. The results were summarized in Fig. 3. In the ventral pathway V1-V2-V4-LO, the performance was 23.6%, 30.7%, 35.2% and 47.8% respectively. The dorsal pathway V1-V2-V3-V3a-V3b showed a similar trend that classification accuracy improved as the level of ROIs advanced, and the performance was 23.6%, 30.7%, 32.2%, 41.3%, and 45.9% respectively. Apparently, there was a common phenomenon in both pathways that classification performance improved significantly as visual information passed on from lower areas to higher areas. Among all the ROIs considered, LO (47.8%) played the best, followed by V3b (45.9%) at the top of the dorsal pathway in this experiment.

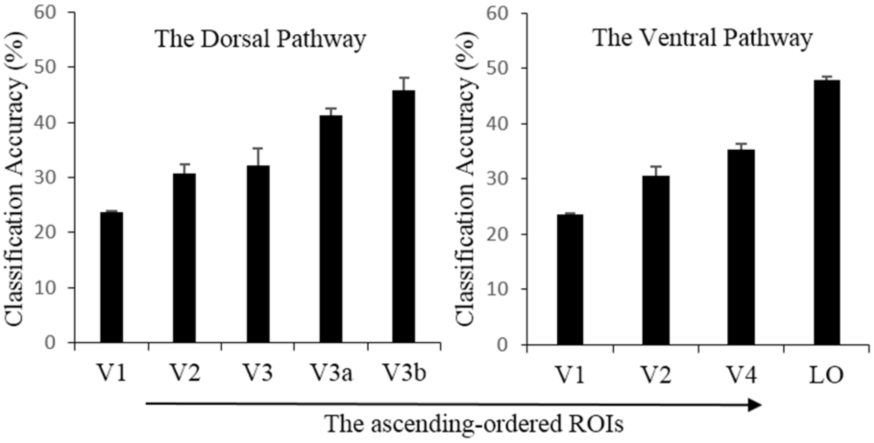

Fig. 3. Performance trend along both pathways. (a) The ventral pathway (four ROIs). (b) The dorsal pathway (five ROIs).

3.2 LO fMRI Data Contains Substantial Information of Images

The visual tasks (usually classification) have long been a difficult challenge to modern computer science. Numerous algorithms aimed at representing images were previously proposed [11, 12]. They were designed quite statistical and mathematical for computer calculation, but far from imitating the way the brain worked. Unsurprisingly, if our goal was simply classification accuracy, the opinions diverged as to whether more biological detailed models would ultimately be needed [13].

In order to show the superiority of the brain over traditional feature-extraction methods in image representation, we extracted 512-D GIST features (unlike SIFT who aimed at giving pictures local and regional descriptions, GIST aimed at offering global and overall features), 576-D HOG features and 256-D LBP features for each of the 1750 natural photographs, and designed SVM classifiers accordingly. The results were summarized in Fig. 4. It turned out to be that in this task, LBP (48.4%) was slightly better than LO fMRI (47.8%), LO fMRI played better than GIST (46.8%), while HOG (42.1%) played the worst.

Therefore, if we designed deep models that could eventually simulated the response activities of voxels located at LO or V3b (and even-higher areas) to imitate the way the brain-extracted features, the classification performance might be better than classifiers designed on the basis of traditional features. Introducing the prior knowledge of human visual pathways into the designation of computer vision systems to form more bionic visual models for various tasks was commendable. The brain is undoubtedly more effective than human-assigned feature-extraction approaches.

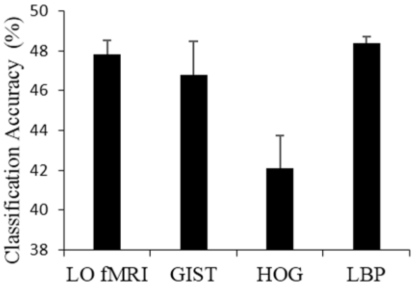

Fig. 4. Performance comparison of LO fMRI with GIST, HOG, and LBP.

3.3 The Dorsal Pathway Contributes to Object Identification

The relatively high performance of LO was natural, because modern neuroscience had found numerous evidences of the specific function of ventral pathway in object recognition. However, we found that the performance of V3a and V3b (41.3% and 45.9% respectively) were not far from LO. The results indicated that the dorsal pathway (including V3, V3a and V3b at least) also contributed to object identification.

We also performed canonical correlation analysis (CCA) [12] to seek for the correlationship of V3a and V3b with other ROIs. The results were shown in Fig.5. CCA algorithm linearly mapped two sets of variables to new spaces respectively, and then maximized the correlationship of two sets of mapped data. Therefore, CCA was an appropriate method to evaluate the linear correlationship of two given variables. Here voxel activities of different regions were considered as variable sets, and the linear unoriented correlationship (the strongest relationship) of V3a and V3b with other regions was thus excavated. The introduction of CCA aimed at analyzing unoriented functional connectivity of ROIs rather than the oriented effective connectivity, and was often done by electroencephalograph.

Simultaneously, the recent findings in deep learning also confirmed our speculation. In 2012, Krizhevsky et al. built the famous convolutional neural network and won the ImageNet competition [14]. It marked the beginning of the dominance of deep neural networks in computer vision. In the past four years, error rates had dropped further, roughly matching (or even exceeded) human performance in the domain of visual object classification [13]. In order to give the high performance a physiological explanation and improve recent deep model structure, researchers tried to compare the state-of-art deep neural networks with the visual pathway to see how much they match in architecture. For example, Michael Eickenberg et al. extracted the outputs of all layers after rectified linear units (ReLU) of OverFeat (2013). They used L2 penalized linear regression to fit a predictive model to each voxel of the measured brain activity after spatial smoothing and subsampling. They found that the outputs of some layers e.g. the fourth or fifth convolutional layer were able to predict the activities of V3a and V3b voxels at relatively high accuracy. It implied that there were some internal relationships between network layers and the two ROIs. Consequently, if the deep networks were confirmed to be "brain-like" (some scholars are working on the interesting topic, such as Nikolaus [13], Cambridge and DiCarlo [5], MIT), then we might come to the conclusion that V3a and V3b played an important role in object recognition. Moreover, the role of V3a and V3b in the real visual pathway might be similar with the corresponding layers in the deep network.

Although building the one-to-one correspondence between deep network layers and visual pathway regions is not accessible now (the existing deep models can only roughly imitate the visual system), yet deep networks are still regarded as best models of human visual system till today.

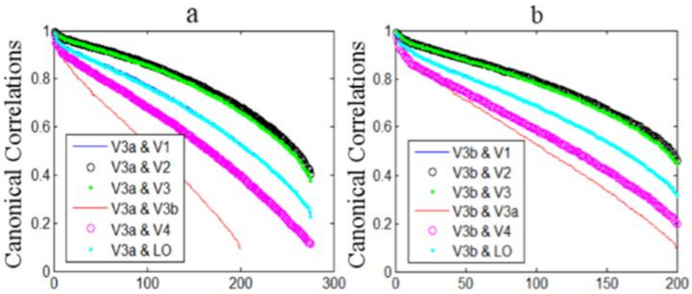

Fig. 5. The correlationship of V3a and V3b with other regions measured by CCA. (a) The canonical correlation of V3a and other regions. There was a relatively strong correlationship between V3a and V2, as well as with V3. (b) The same analysis was performed on V3b, and the result demonstrated that relatively strong correlationship existed between V3b and V2, as well as with V3. Both the figures show that the activity patterns of V3a and V3b were tightly related to V2 and V3 voxels. It should be mentioned that V4 and LO also had relatively good correlationship with the two regions.

4 Discussion

The results demonstrates that the accuracy increases along the path V1-V2-V4-LO, which is the main part of the 'ventral pathway'. The ventral path mainly solves the problem of object recognition, so the outcome is not surprising. The increasing trend of accuracy corresponds to the fact that as we track the information stream in the human visual system, the representations of the stimulus grow more and more abstract and global for comprehension. Along the entire visual path, the higher functional areas assemble the information delivered by lower ones to form more comprehensive and integrated representations. Notice that V1 is the "entrance" of the ventral pathway, and the whole information of any given image is "stored" in V1, so the representations of this region are intuitively expected to perform the best. However, we find that the following regions all perform better in classification task, which indicates that the visual information is deeply hidden in V1 with high nonlinearity, so the SVM classifiers are unable to excavate the essence of the data. But that is exactly why the ventral visual pathway exists. The V1 information is further transmitted in a highly nonlinear way among the cascaded cortex regions. In this process, the nonlinearity is decoded gradually, making the representations change from wide and shallow to deep and narrow [5]. The mechanism of vision can be described as a nonlinear data miming (DM) process.

The results also demonstrates that the prediction accuracy of LO based classifiers rivals the traditional-feature-based classifiers. Image classification is difficult because it's hard to excavate the deep statistical essence (features) of the data. The features should possess enough distinguishing ability between different samples, but represent similar samples as close as possible. Our work shows brain cortex regions have such characteristics no less than traditional features do. Therefore, we can design new brain-like feature-extraction methods to simulate the activity patterns of visual cortex regions (especially higher regions).

The dorsal pathway also shows its contribution to object recognition process, although it was traditionally believed to be tightly related to localization problems and not effective in object recognition. In fact, there are many complex connections between ventral neurons and dorsal neurons, and the contributions of dorsal regions in recognition tasks should not be neglected.

The relationship between visual pathway and deep neural networks is confusing but interesting. They are similar in the hierarchically-connected structure (some scholars even matched up the layers and regions), and the basic element of artificial neural network imitates the real nerve cell, and the convolutional networks even simulates the local receptive field character. They are different because the brain is a deep and complex recurrent neural network [13], which could not be fully described by the current feed-forward deep models. Moreover, it is physiologically unlikely that the visual cortex learns exactly by BP algorithm, because true biological postnatal learning in humans may use a large amount of unsupervised data. However, the deep networks are still regarded as the best models of the brain and have achieved great success in various fields, such as speech recognition and machine translation [13]. Our results verified that there are "information pyramids" in our visual system, including the ventral pathway as well as the dorsal pathway. There's a commonly addressed question that, why our visual system (and the deep networks) are hierarchically organized? Previous studies have shown that three-layer shallow BP network can approximate continuous functions with arbitrary precision by adding a sufficient number of hidden units and suitably setting the weights [15], but why a "multi-layer pyramid" structure is needed? The reason depth matters is that deep models can represent many complex functions more concisely [13], because they are endowed with more powerful nonlinear feature-extraction ability.

Almost all researches that try to bind deep learning and visual pathway together are limited to analyzing how much they match, but until today, there's no effective way to improve deep learning structure by the foreknowledge of the visual cortex (e.g. redesign convolutional filters for each layer in accordance with corresponding cortex regions and even weight updating algorithm). This direction deserves much further research. Based on classification accuracy and previous evidences in the OverFeat network, we find that V3a and V3b located at the dorsal pathway are also involved in object recognition process. V3a and V3b have their own important status in visual information encoding, consequently, if we want to redesign each layer (or layers) by different ROIs' activity characteristics, not only the regions of the ventral pathway should be included, but also the ROIs of the dorsal pathway should be considered.

References

1. Kruger N, Janssen P, Kalkan S, et al. Deep hierarchies in the primate visual cortex: What can we learn for computer vision?[J]. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2013, 35(8): 1847-1871.
2. Hubel D H, Wiesel T N. Receptive fields of single neurones in the cat's striate cortex[J]. The Journal of physiology, 1959, 148(3): 574-591.
3. De Valois K K, De Valois R L, Yund E W. Responses of striate cortex cells to grating and checkerboard patterns[J]. The Journal of Physiology, 1979, 291: 483.
4. Riesenhuber M, Poggio T. Hierarchical models of object recognition in cortex[J]. Nature neuroscience, 1999, 2(11): 1019-1025.
5. Yamins D L, Dicarlo J J. Using goal-driven deep learning models to understand sensory cortex.[J]. Nature Neuroscience, 2016, 19(3):356-365.
6. Gallant J L, Connor C E, Rakshit S, et al. Neural responses to polar, hyperbolic, and Cartesian gratings in area V4 of the macaque monkey.[J]. Journal of Neurophysiology, 1996, 76(4):2718-2739.
7. Schmah T, Hinton G E, Zemel R S, et al. Generative versus discriminative training of RBMs for classification of fMRI images[C]// Advances in Neural Information Processing Systems 21, Proceedings of the Twenty-Second Annual Conference on Neural Information Processing Systems, Vancouver, British Columbia, Canada, December 8-11, 2008. 2008:1409-1416.
8. Walther D B, Caddigan E F L. Natural scene categories revealed in distributed patterns of activity in the human brain.[J]. Journal of Neuroscience the Official Journal of the Society for Neuroscience, 2009, 29(34):10573-10581.
9. Kay K N, Thomas N, Prenger R J, et al. Identifying natural images from human brain activity.[J]. Nature, 2008, 452(7185):352-355.
10. Kay K N, Naselaris T, Gallant J L. fMRI of human visual areas in response to natural images[J]. CRCNS. org, 2011.
11. X. Wang, T. X. Han, S. Yan. An HOG-LBP human detector with partial occlusion handling[J]. Proceedings, 2009, 30(2):32-39.
12. Cruz-Mota J, Bogdanova I, Paquier B, et al. Scale Invariant Feature Transform on the Sphere: Theory and Applications[J]. International Journal of Computer Vision, 2012, 98(2):217-241.

13. Kriegeskorte N. Deep Neural Networks: A New Framework for Modeling Biological Vision and Brain Information Processing[J]. Annual Review of Vision Science, 2015, 1: 417-446.
14. Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[C]//Advances in neural information processing systems. 2012: 1097-1105.
15. Schäfer A M, Zimmermann H G. Recurrent neural networks are universal approximators[M]//Artificial Neural Networks–ICANN 2006. Springer Berlin Heidelberg, 2006: 632-640.