Neurocomputing 316 (2018) 202–209

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Brief papers

F-score feature selection based Bayesian reconstruction of visual image from human brain activity

Wei Huang, Hongmei Yan*, Ran Liu, Lixia Zhu, Huangbin Zhang, Huafu Chen*

The Clinical Hospital of Chengdu Brain Science Institute, MOE Key Lab for Neuroinformation, University of Electronic Science and Technology of China, Chengdu 610054, PR China

ARTICLE INFO

Article history:
Received 4 September 2017
Revised 30 May 2018
Accepted 9 July 2018
Available online 8 August 2018

Keywords:
fMRI
Bayesian
F-score
Visual image reconstruction

ABSTRACT

Decoding perceptual experience from human brain activity is a big challenge in neuroscience. Recent advances in human neuroimaging have shown that it is possible to reconstruct a person's visual experience based on the retinotopy in the early visual cortex and the multivariate pattern analysis (MVPA) method using functional magnetic resonance imaging (fMRI). Previous researches reconstructed binary contrast-defined images using combination of multi-scale local image decoders in V1, V2 and V3, where contrast for local image bases was predicted from fMRI activity by sparse multinomial logistic regression (SMLR) and other models. However, the precision and efficiency of the visual image reconstruction remain insufficient. Proper feature selection is widely known to be as critical for prediction and reconstruction. Aiming at the shortcomings of existing reconstruction models, we proposed a new model of Bayesian reconstruction based on F-score feature selection (Bayes+F). The results indicate that the proposed Bayes+F model has better reconstruction accuracy and higher efficiency than the SMLR and other models, showing better robustness and noise resistant ability. It can improve the spatial correlation coefficient (Mean ± variance: 0.7078 ± 0.2104) and decrease the standard error (Mean ± variance: 0.2693 ± 0.0871) between the stimulus and the reconstructed image. Furthermore, the proposed model can reconstruct the images extremely rapid, 100 times faster than SMLR does.

1. Introduction

Vision is among the important senses of humans and other mammals. However, two basic aspects of visual science have yet to be elucidated. (1) How does our brain respond and encode the visual world? (2) Can we decode and reconstruct our visual perceptual experiences in terms of brain activity? With advancements in single-unit cellular recording, electroencephalogram (EEG), functional magnetic resonance imaging (fMRI), and other techniques, receptive fields, functional columns, and visual pathways associated with the first question have been sufficiently studied [1–3]. Nevertheless, the second question remains a major challenge in neuroscience. Multi-voxel pattern classification (MVPC) studies have revealed that brain activity patterns generated in the early visual cortex in different mental states can be used to decode different visual perceptions. Haxby et al. first identified distinct patterns of fMRI responses in the ventral temporal cortex for several object categories, including different types of small man-made objects [4]. Kamitani et al. demonstrated that fMRI activity patterns in early visual areas contain detailed orientation and motion information that can reliably predict subjective orientation perception and motion direction [5,6]. Gerven et al. showed that the orientation and rotation direction of a continuously rotating grating can be accurately decoded by using linear dynamical systems and hidden Markov models [7]. Natural images [8–12], dynamic movies [13], handwritten letters [14], faces [15], and visual imagery during sleep [16] can also be identified and decoded among numerous candidate images by using visual encoding models.

Visual identification is constrained due to limited candidate image sets or classification categories. As such, categorical constraint-free visual image reconstruction methods should be developed to decode the visual perception. Miyawaki et al. utilized retinotopy in the early visual cortex to accomplish a constraint-free reconstruction of contrast-defined arbitrary visual images from fMRI signals of the human early visual cortex [17]. They proposed sparse multinomial logistic regression (SMLR) by using multi-voxel patterns of fMRI signals and multi-scale visual representation. A stimulus state at each local element is predicted by using a decoder with multi-voxel patterns. The outputs of each local decoder are subsequently combined to reconstruct the presented image.

* Corresponding authors.
E-mail addresses: hmyan@uestc.edu.cn (H. Yan), chenhf@uestc.edu.cn (H. Chen).

https://doi.org/10.1016/j.neucom.2018.07.068
0925-2312/

However, we found that visual images reconstructed via SMLR proposed by Miyawaki et al. contained a lot of noise, and more than 10 h are needed to complete the entire reconstruction process. As a result, a very low reconstruction efficiency is obtained. Recently, Yu Zhan and Sutao Song et al. proposed a support vector machine (SVM) and a naive Bayesian classifier based on independent component analysis (NB-ICA) and improved the time consumption and spatial correlation between a stimulus and the reconstructed image to a certain degree [18,19]. Nevertheless, better reconstruction methods should be developed to explore potential practical applications. Proper feature selection is widely known to be critical for prediction and reconstruction. Selecting appropriate features and removing irrelevant or redundant features may reduce the noise of reconstructed images, decrease computational complexity, and improve reconstruction efficiency. In this study, an F-score feature selection-based Bayesian model was proposed to reconstruct visual images. The results show that the proposed model is more robust and noise resistant than SMLR and other reconstruction methods, such as SVM and RF with/without F-score feature selection. Furthermore, the proposed model can also reconstruct the images extremely rapid, 100 times faster than SMLR does.

2. Material and methods

2.1. Data sources

The dataset is collected from public data recorded by Miyawaki et al. [17]. Three types of experimental sessions were conducted to record the fMRI responses of the visual cortex: (1) conventional retinotopy mapping session, (2) random image session, and (3) figure image (geometric and alphabet shapes) session. The conventional retinotopy mapping session was used to delineate the borders between visual cortical areas, and to identify the retinotopy map on the flattened cortical surfaces. In the current study, voxels of V1, V2 and V3 were selected as features, with the dimensions of 1017, 1045 and 1237 for the three visual areas. The retinotopy mapping procedure adopted a rotating wedge and an expanding ring of flickering checkerboard. In the random image session, 440 10×10 flickering checkerboard spatially random patterns were presented to each subject in 22 blocks and 20 runs. Each stimulus block was 6 s long followed by 6 s rest period. Extra rest periods were added at the beginning (28 s) and at the end (12 s) of each run. The model training was performed based on the 440 single-trial block-averaged data. In the figure image session, five alphabet letters and five geometric shapes were shown to subjects six or eight blocks while fMRI data were recorded. Each stimulus block was 12s long followed by a 12s rest period. Block-averaged data was used as the test data. The experimental details can be found in the experimental procedures of Miyawaki's study (2008).

All the data were obtained using a 3.0-Tesla Siemens MAGNETOM Trio A Tim scanner located at the ATR Brain Activity Imaging Center. An interleaved T2*-weighted gradient-echo echo-planar imaging (EPI) scan was performed to acquire functional images to cover the entire occipital lobe (TR, 2000 ms; TE, 30 ms; flip angle, 80°; FOV, 192 × 192 mm; voxel size, 3 × 3 × 3 mm; slice gap, 0 mm; number of slices, 30). T2-weighted turbo spin echo images were scanned to acquire high-resolution anatomical images of the same slices used for the EPI (TR, 6000 ms; TE, 57 ms; flip angle, 90°; FOV, 192 × 192 mm; voxel size, 0.75 × 0.75 × 3.0 mm). T1-weighted magnetization prepared rapid-acquisition gradient-echo (MP-RAGE) fine-structural images of the whole-head were also acquired (TR, 2250 ms; TE, 2.98 or 3.06 ms; TI, 900 ms; flip angle, 9°; field of view, 256 × 256 mm; voxel size, 1.0 × 1.0 × 1.0 mm).

The first 8s fMRI signals of each run were discarded to avoid instability of the MRI scanner. The acquired fMRI data first underwent slice-timing correction and three-dimensional motion correction by Statistical Parametric Mapping (SPM) software. Then the data were coregistered to the within session high resolution anatomical image of the same slices used for EPI and subsequently to the whole-head high resolution anatomical image. Finally, the fMRI data underwent linear trend removal within each run. Amplitude normalization relative to the mean amplitude of the first 20 s rest period in each run was performed to minimize the baseline difference across runs. The fMRI signals of each voxel were averaged within each stimulus block time after shifting the data by 4 s to compensate for hemodynamic delays. In the end, 1017 voxels in V1, 1045 voxels in V2 and 1237 voxels in V3 were selected as fMRI features for model training and reconstruction according to the retinotopy mapping session. The dataset can be found online at http://www.neuron.org/supplemental/S08966273(08)00958-6.

2.2. General procedure of image reconstruction

We presented a general framework of visual image reconstruction by using multi-voxel fMRI signal patterns and multi-scale visual representation. Previous studies showed that the visual cortex response is caused not only by corresponding visual stimulation, but also by the adjacent stimuli. Similarly, a visual image can be represented at multiple spatial scales in the visual cortex, which may serve to retain the visual sensitivity to fine-to-coarse patterns at a single visual field location [20,21]. Therefore, multiscale information from fMRI signals may achieve better reconstruction [17]. We assumed that a local image is represented by a linear combination of convolved image on different convolution kernel. The stimulus state in each local pixel can be predicted by using a decoder via multi-voxel patterns analysis with a weight set for each decoder, and the outputs of these decoders can be combined statistically to reconstruct the presented local pixel. This reconstruction involves the following main steps. (1) Random image sequences were used to present binary contrast at multiple visual field locations simultaneously and to determine the spatial linearity of neural responses corresponding to localized visual stimuli. (2) An image decoder was established to directly simulate the relationship between the stimulus and the fMRI activity at a specific time when the image was presented to the subject. The stimulus at each local pixel was then predicted by using a local decoder. The outputs of all pixels were combined together to obtain the whole predicted image. The reconstruction principle is illustrated in Fig. 1.

An F-score feature selection method was applied to select discriminant voxels, estimate their weights prior to classification, and improve the quality and efficiency of reconstruction. Instead of SMLR, a naive Bayesian classifier was used to decode the contrast of multi-scale local images. Other classification methods, including SMLR, RF, and SVM with/without F-score feature selection, were compared in terms of reconstruction accuracy and efficiency.

2.3. F-score feature selection

For a classifier that may refer to a large amount of observational variables, feature selection is necessary to select a relatively small subset of variables, reduce computation consumption, and improve algorithm performance. F-score is a simple and frequently used feature selection technique to quantify the discrimination of two sets of real numbers [22–27]. With the training vectors $x_k, k = 1, 2, \ldots, m$, the F-score of the ith feature is defined as follows if the numbers of positive and negative instances are n_+ and n_-, respectively.:

$$F_i = \frac{(\bar{x}_i^{(+)} - \bar{x}_i)^2 + (\bar{x}_i^{(-)} - \bar{x}_i)^2}{\frac{1}{n_+ - 1}\sum_{k=1}^{n_+}(x_{ki}^{(+)} - \bar{x}_i^{(+)})^2 + \frac{1}{n_- - 1}\sum_{k=1}^{n_-}(x_{ki}^{(-)} - \bar{x}_i^{(-)})^2} \quad (1)$$

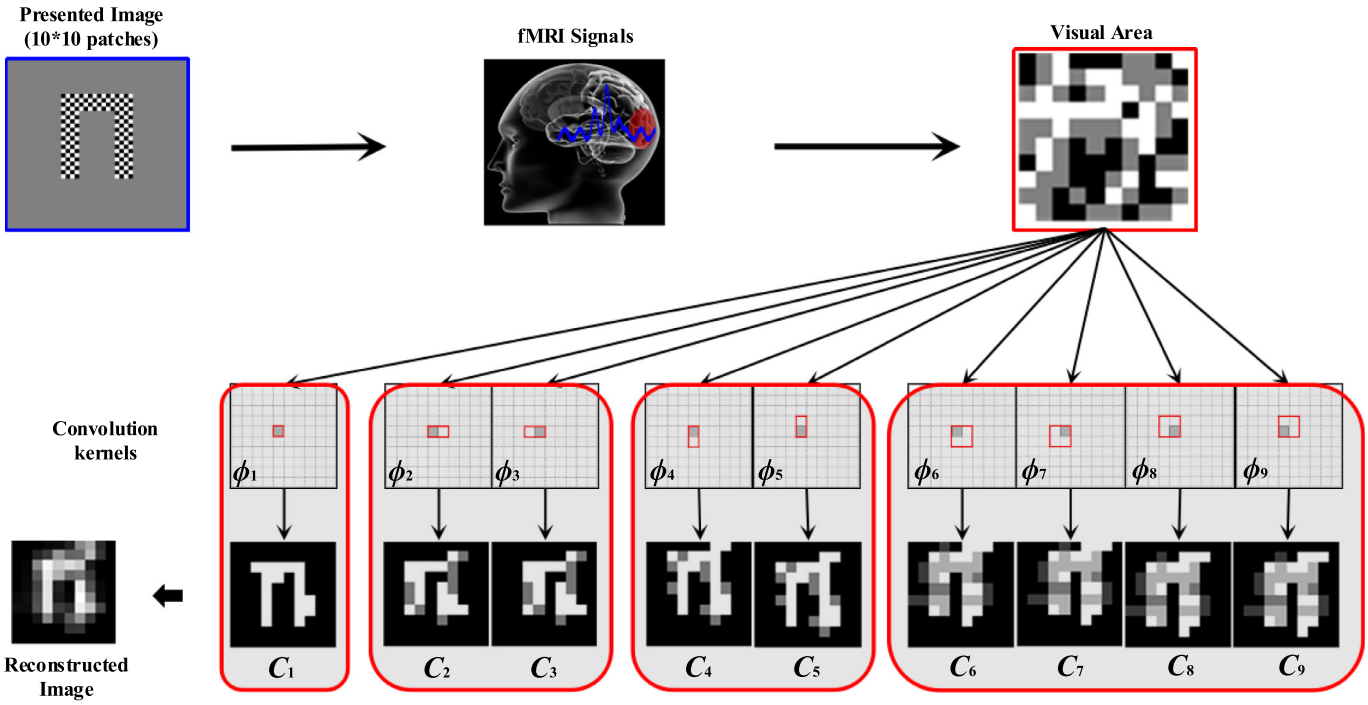


Fig. 1. Principle of visual image reconstruction. The fMRI signals were measured when subjects viewed the Chess board visual stimuli (10×10 patches). Image decoders (Bayes+F or SMLR and so on) were established to predict the contrast of image bases on different convolution kernels. The predicted value of a specific pixel was calculated by summing up the class labels of all the "local image bases".

where $\bar{x}_i$, $\bar{x}_i^{(+)}$ and $\bar{x}_i^{(-)}$ are the averages of the ith feature of the whole, positive, and negative data sets, respectively; $x_{k,i}^{(+)}$ is the ith feature of the kth positive instance; and $x_{k,i}^{(-)}$ is the ith feature of the kth negative instance. The numerator indicates the discrimination between positive and negative sets, and the denominator denotes the value within each of the two sets. A high F-score shows that a feature is highly discriminative.

Considering more than two sets of real numbers, Xie et al. proposed an improved F-score method [28]. Given the training vectors $x_k, k = 1, 2, \ldots, m$, and the number of data sets $l(l \geq 2)$, the F-score of the ith feature is defined as follows if the number of the jth data set is $n_j, j = 1, 2, \ldots, l$.:

$$F_i = \frac{\sum_{j=1}^{l} (\bar{x}_i^{(j)} - \bar{x}_i)^2}{\sum_{j=1}^{l} \frac{1}{n_j - 1} \sum_{k=1}^{n_j} (\bar{x}_{ki}^{(j)} - \bar{x}_i^{(j)})^2} \tag{2}$$

where $\bar{x}_i$ and $\bar{x}_i^{(j)}$ are the averages of the ith feature of the whole data set and the jth data set, respectively; $x_{k,i}^{(j)}$ is the ith feature of the kth instance in the jth data set. The numerator indicates the discrimination between data sets, and the denominator denotes the value within each of the data set. A high F-score shows that the feature is highly discriminative. In this study, the improved multiple classification F-score method was applied to select features during the training of the multi-scale local decoder.

2.4. Convolution processing of random images

Four convolution kernels (see Supplementary Material 1) are used to process each random image to produce nine convolved random images (Fig. 2). After the random image is processed by the convolution kernel 1, the resulting image is exactly the same as the original one, namely, there is no change before and after the convolution kernel 1. But after the random image is processed by kernel 2 or 3, the resulting image is one column or one row less than the original size, and the mean contrast value of each local image pixel becomes 0, 0.5 or 1. Similarly, the resulting image is one column and one row less than the original size after being processed by the convolution kernel 4, and the mean contrast value of the local pixel becomes 0, 0.25, 0.5, 0.75 or 1.

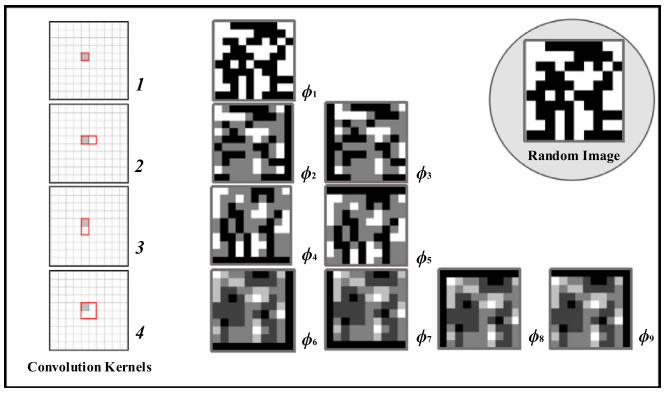


Fig. 2. Random image for convolution operation. The left represents four different types of convolution kernels. The remaining images represent nine convolved images.

In order to keep the convolved image consistent with the original image size, zero-padding is applied for each of the kernel convolved images. The nine convolved images after zero-padding are illustrated as $\phi_1 \sim \phi_9$ in Fig. 2.

2.5. Visual image reconstruction model

Visual image reconstruction is to create a model to predict the contrast of image bases on different spatial scales for each pixel location. For the stimulation of the visual image (440 random images or 80 figure images), each image can obtain the average activation value of each voxel in early visual area (V1, V2, V3) from the fMRI.

Each random image can obtain nine convolved images by convolution processing (see Supplementary Material 2). These convolved images and the average activation value of each voxel in these visual areas are used as the training data set. The average activation value of each voxel in V1, V2, V3 of geometric shapes and alphabet letters images are used as the test data (see Supplementary Material 2).

The following is the data processing steps of the visual image reconstruction model in the visual area:

(1) For convolved image ϕ_1, the pixel values of 440 random images in the first position is used as the training label, which is a vector of size 440×1. The corresponding average activation value of each voxel in V1 were used as training data, which is a matrix of size 440×1017 (V1 contains 1017 voxels). The training data and label were combined into training samples, and the Bayesian classifier was used to determine the classification under the ten-fold cross validation. The predictive label (x_1) of the training data can be obtained after the Bayesian classifier. For the other eight convolved images $(\phi_2, \phi_3, \ldots, \phi_9)$, repeat the above steps so all the nine predictive labels $(x_1, x_2, \ldots, x_9)$ of the training data can be obtained. The nine predictive labels $(x_1, x_2, \ldots, x_9)$ of the training data are combined with training label (y) to establish multiple linear regression models $(y = \omega_1 \times x_1 + \omega_2 \times x_2 + \cdots + \omega_9 \times x_9 + \varepsilon)$, where ε is the residual. The training label (y) is the actual pixel value of the first position in the random image. These association coefficients $(\omega_1, \omega_2, \ldots, \omega_9)$ are estimated using the least squares method.

(2) For training data and training labels, the F-value of each voxel is calculated by the F-score feature selection algorithm. We arrange the F values of all the voxels in ascending order and select the first ten values. The average activation voxel values corresponding to the first ten F scores are chosen as the filtered features (see Supplementary Material 5). The filtered training data is a matrix of size 440×10. Likewise, the corresponding filtered features of 80 figure image stimuli are calculated as the test data, which is a matrix of size 80×10. For the convolved image ϕ_1, the F-scored pixel values of 440 random images in the first position are trained with a naive Bayesian classifier. Then, the trained model is used to predict the label of the test data. For the other eight convolved images $(\phi_2, \phi_3, \ldots, \phi_9)$, repeat the above steps and then obtain eight predictive labels $(C_2, \ldots, C_9)$ of size 80×1 of the test data.

(3) The nine predictive labels $(C_1, C_2, \ldots, C_9)$ of the testing sample and their association coefficients $(\omega_1, \ldots, \omega_9)$ are summed up to obtain the final predicted pixel value $(\sum_{i=1}^{9} \omega_i \times C_i)$ of the first position.

Above steps are repeated for each position of the test images (100 locations), then the corresponding reconstructed images can be obtained. Schematic illustration of the decoding steps is shown as Fig. 3. Detailed process is given in the supplementary material 3.

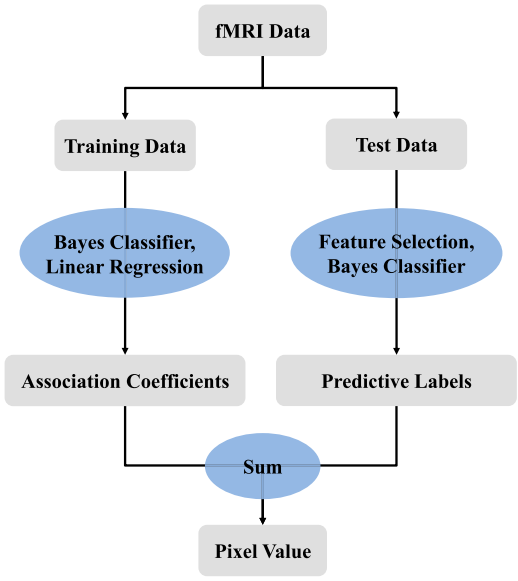


Fig. 3. Schematic illustration of the contrast-defined visual stimuli decoding steps using F-score feature selection based Bayesian classifier.

3. Results

3.1. Reconstruction accuracy

The reconstructing stimulus included geometric shapes and alphabet shapes. We reconstructed the images with the activities of the voxels in V1. In this part, the F-score combined Bayesian method, three representative classification methods [SMLR, random forest (RF), and SVM], and two F-score combined methods (RF+F and SVM+F) were applied to compare their performances and reconstruction efficiencies. The reconstructed images obtained by the six algorithms are shown seperately in Fig. 4. The presented stimulus and their corresponding final reconstructed images obtained via the six methods are compared in Fig. 5. Although geometric and alphabet shapes are not used to train the reconstruction model, the six algorithms can effectively reconstruct the simple contrast stimulus images. Fig. 4(d) reveals the reconstructed results of SMLR algorithm proposed by Miyawaki et al. The reconstructed images display the essential features of the original geometric and alphabet shapes (Figs. 4 and5). The images reconstructed by the F-score-combined methods (Figs. 4(a)–(c), bottom 3 rows in Fig. 5) showed less noise and higher target-background contrast than those without F-score methods (Fig. 4(d)–(f) and upper 3 rows in Fig. 5). The quality and noise resistance of the proposed F-score-combined Bayesian model are excellent. Therefore, the F-score feature selection can be applied to reduce the noise of the reconstructed images by selecting task-related voxels and disregarding irrelevant voxels.

Spatial correlations between the stimulus and reconstructed images are objective indices used to evaluate reconstruction accuracy [17,19]. Fig. 6 shows the average standard error and spatial correlation coefficient between the represented and reconstructed images respectively. F-score-combined methods yield lower standard error and higher correlation coefficient than the methods without F-score. The proposed F-score-combined Bayesian model exhibits the highest correlation coefficient and the lowest standard error. The main contribution lies in that F-score feature selection method filters out the noise which is irrelevant to the local image reconstruction. For SMLR, the average standard error and spatial correlation between the presented and reconstructed images are 0.4059 ± 0.0613 and 0.6335 ± 0.1540, respectively. By comparison, the standard error and average spatial correlation of the F-score combined Bayesian model are 0.2693 ± 0.0871 and 0.7078 ± 0.2104, respectively. Hence, F-score feature selection contributes a lot to the reconstruction quality improvement.

3.2. Reconstruction time

Our brain can rapidly recognize and process natural scenes [29]. Reconstruction time should also be considered as an important

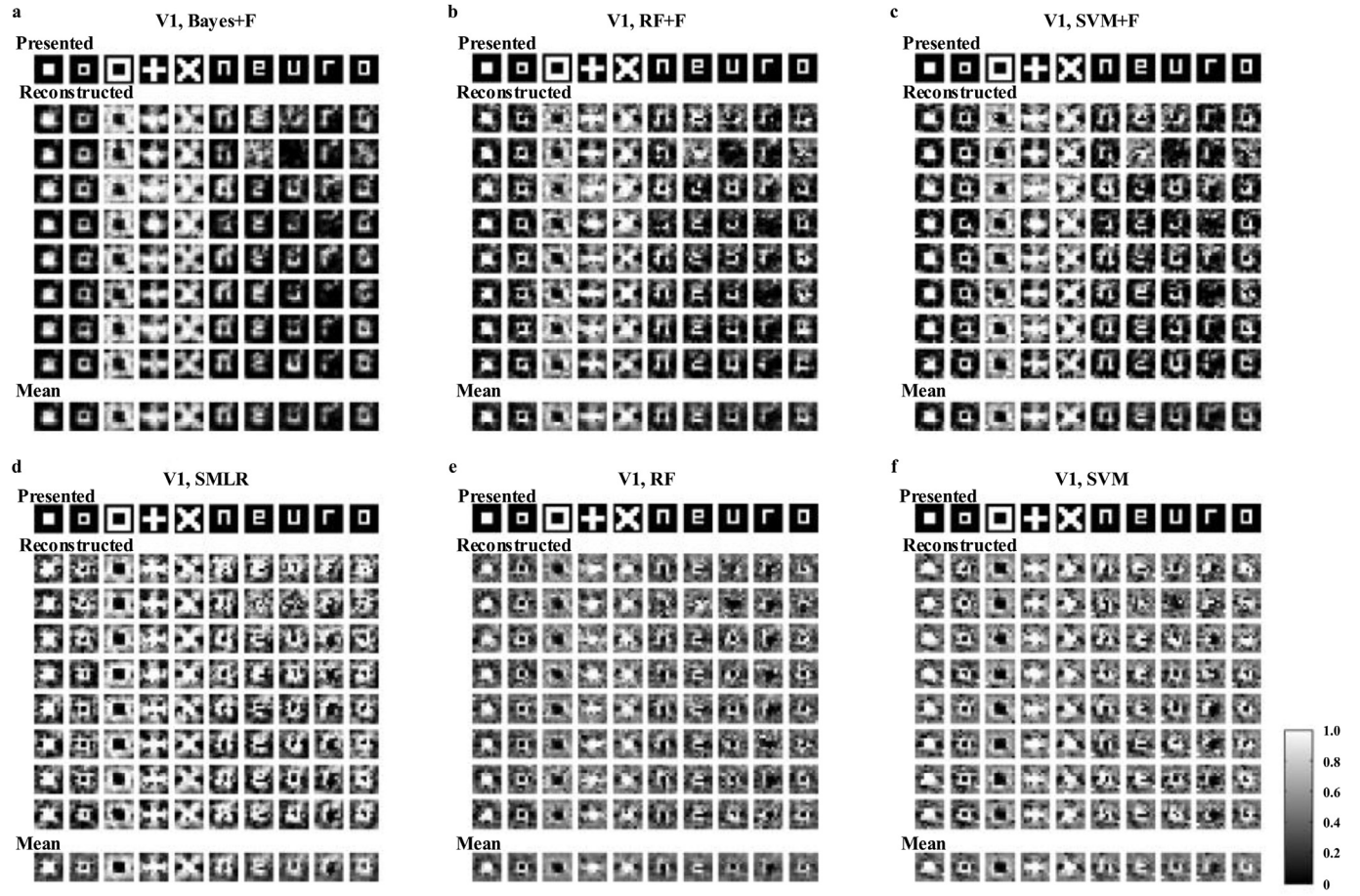


Fig. 4. Visual reconstruction results obtained by six different algorithms in v1 area. (a) The results obtained by the F-score combined naive Bayesian classifier. (b) The results obtained by the F-score combined random forest classifier. (c) The results obtained by the F-score combined support vector machine classifier. (d) Results obtained by sparse multinomial logistic regression classifiers. (e) Results obtained by a random forest classifier. (f) Results obtained a support vector machine classifier. For each panel, the first row represents the images which were presented to subjects during fMRI data recording. The 2nd–9th rows show the reconstructed results repeated 8 times. The bottom row represents the averaged reconstructed images.

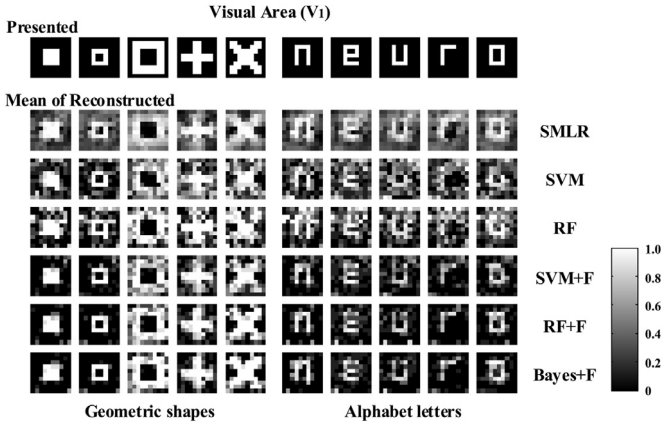


Fig. 5. The averaged reconstructed images in v1. The first row represents the images which were presented to subjects during fMRI data recording. The bottom rows represent the corresponding averaged reconstructed images using the six methods. After joining the F-score feature selection algorithm (SVM+F, RF+F, Bayes+F), the reconstruction results were closer to the real stimulation image, which showed that feature selection is necessary.

index to evaluate reconstruction algorithms. Fig. 7 shows the comparison of the reconstruction time of the six algorithms. The configuration of the computer used is shown in Supplementary Materials 4. SMLR requires the longest reconstruction time of approximately more than 14 h. By contrast, the F-score-combined Bayesian model needs about 8 min. Therefore, the F-score combined Bayesian model runs almost one hundred times faster than the SMLR method does. Although the SVM+F calculation time is slightly lower than SVM, and the RF+F calculation time is slightly lower than that of RF, but the F-score feature selection algorithm does not fundamentally reduce the computation time. In terms of the reduction in reconstruction time, the most essential factor is the low Bayesian computational complexity.

3.3. Reconstruction effects from different visual cortex

To compare the contribution of different visual areas when contrast-defined visual information is decoded, we also reconstructed the stimulus with the activities of the voxels in V2 and V3 independently by using the F-score combined Bayesian model. The results are shown in Fig. 8. For each panel in Fig. 8a, the first row represents the presented images to subjects during fMRI data recording. The 2nd–9th rows show the reconstructed result of the same image from one trial fMRI data. The bottom row represents the averaged images of all reconstructed images of the same image. The results show that reconstruction quality decreases gradually from V1 to V2 and V3. The correlation coefficients (Fig. 8b) of V1, V2, and V3 are respectively: 0.7078 ± 0.2104, 0.4614 ± 0.1644, and 0.2284 ± 0.1542, respectively. The standard error (Fig. 8c) of V1, V2, and V3 are 0.2693 ± 0.0871,

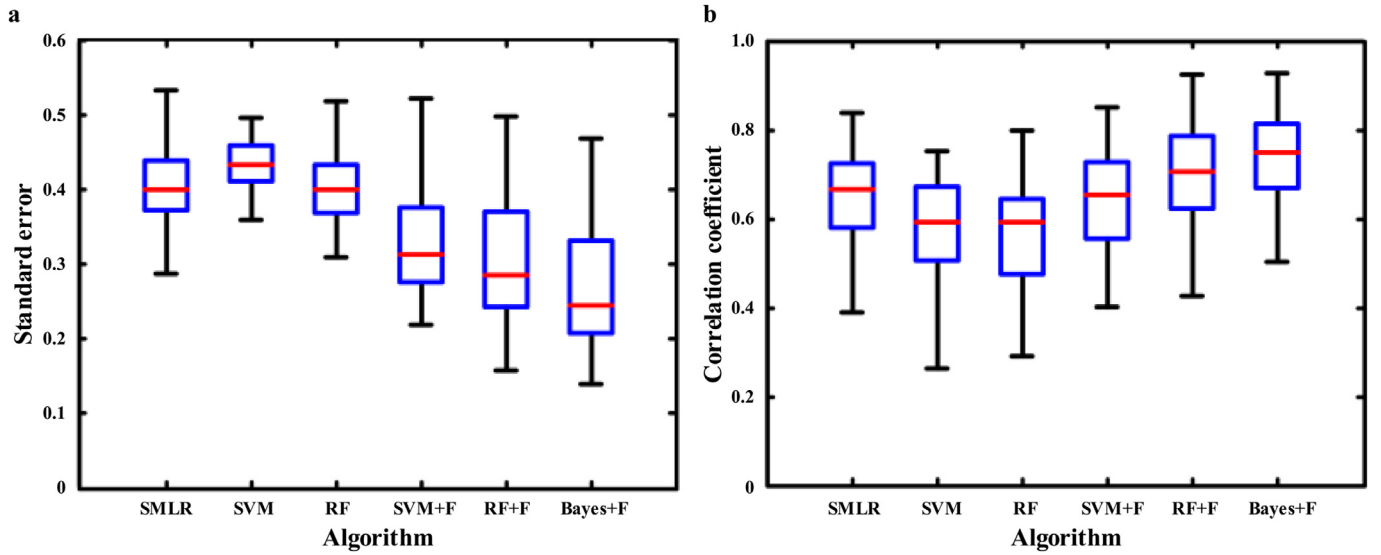


Fig. 6. Reconstruction performance of different algorithms in visual V1 area. (a) Reconstruction standard error of the six algorithms. (b) Reconstruction correlation coefficient of the six algorithms.

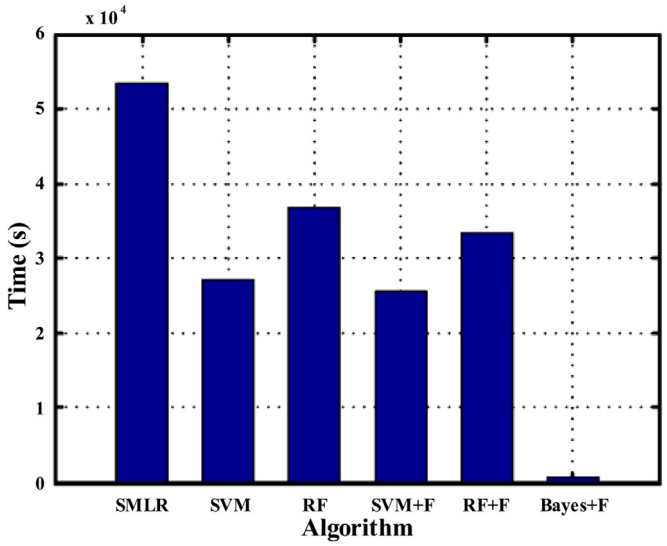


Fig. 7. Reconstruction time of the six algorithms.

0.3905 ± 0.0493, and 0.4520 ± 0.0398, respectively. Higher visual areas show significantly smaller correlation coefficient (Fig. 8b) and larger errors (Fig. 8c) than V1 (ANOVA, Bonferroni-corrected $p < 0.05$ for multiple comparisons). The results show that V1 has a better topological mapping relationship with the visual image. Although the quality of the reconstructed images becomes poorer as the visual area increases for V2 or V3, it does not necessarily mean that the higher visual area does not characterize the visual image well. The reason is that the higher visual area means more abstract image coding which results in weaker topological mapping relationship. We combined the voxels of V1, V2 and V3 together, and found that the reconstruction performance was slightly worse than that obtained by using features from only V1 (Supplementary Materials 5). The average standard error and spatial correlation between the presented and reconstructed images are 0.3182 ± 0.0786 and 0.6737 ± 0.1654, respectively. The addition of voxels in V2 and V3 as features brings noise instead.

4. Discussion

In this study, an F-score feature selection combined Bayesian model was proposed to improve the reconstruction performance on the basis of Miyawaki's framework of constraint-free visual image reconstruction. The result showed that the proposed model can be used to reconstruct the images with the highest quality (highest correlation coefficient and lowest standard error) and fastest speed among the six models. F-score method is a basic and simple technique used to determine the distinction between multi classes with real values [22–24]. F-score feature selection provides the advantages of selecting task-related voxels as input features and effectively removing irrelevant voxels. As a result, the noise of reconstructed images is greatly reduced. In this paper, the first ten voxels were selected as input features because of the following reasons: (1) we compared the reconstruction performances under different number of features, and found that the correlation coefficient and standard error are pretty stable and good when feature number is around 10. But with the increasing of voxel numbers, the reconstruction accuracy decreased and noise increased, shown as supplementary material 6; (2) an increase of the number of features cannot enhance the reconstruction quality obviously but prolong the reconstruction time.

In this study, visual images were reconstructed with five additional models (SVM and RF with/without F-score, Bayes+F) except SMLR. Thus, the reconstruction model is diverse and specific. Although lots of pattern recognition algorithms can be used to construct the model of visual image reconstruction, performance and efficiency may be different among various algorithms. Compared with other machine learning algorithms, Bayesian algorithm applies a priori knowledge to calculate posterior probability and has a simpler computing complexity, so it costs less computation time. Therefore, the proposed F-score combined Bayesian reconstruction model is more accurate and less time consuming than other models. The proposed model also yields a higher spatial correlation than NB-ICA [19] and 3-pixel image-based SVM [18]. Their correlation coefficients are 0.41 ± 0.13 and 0.6934 ± 0.1165, respectively.

Reconstruction quality continuously decreases from V1 to V2 and V3. This is because V1 has the most direct retinotopy mapping from the visual field to the cortical voxels. When the image information flow passes through the visual pathway, it will be processed in a nonlinear way [17]. Higher visual areas have a less

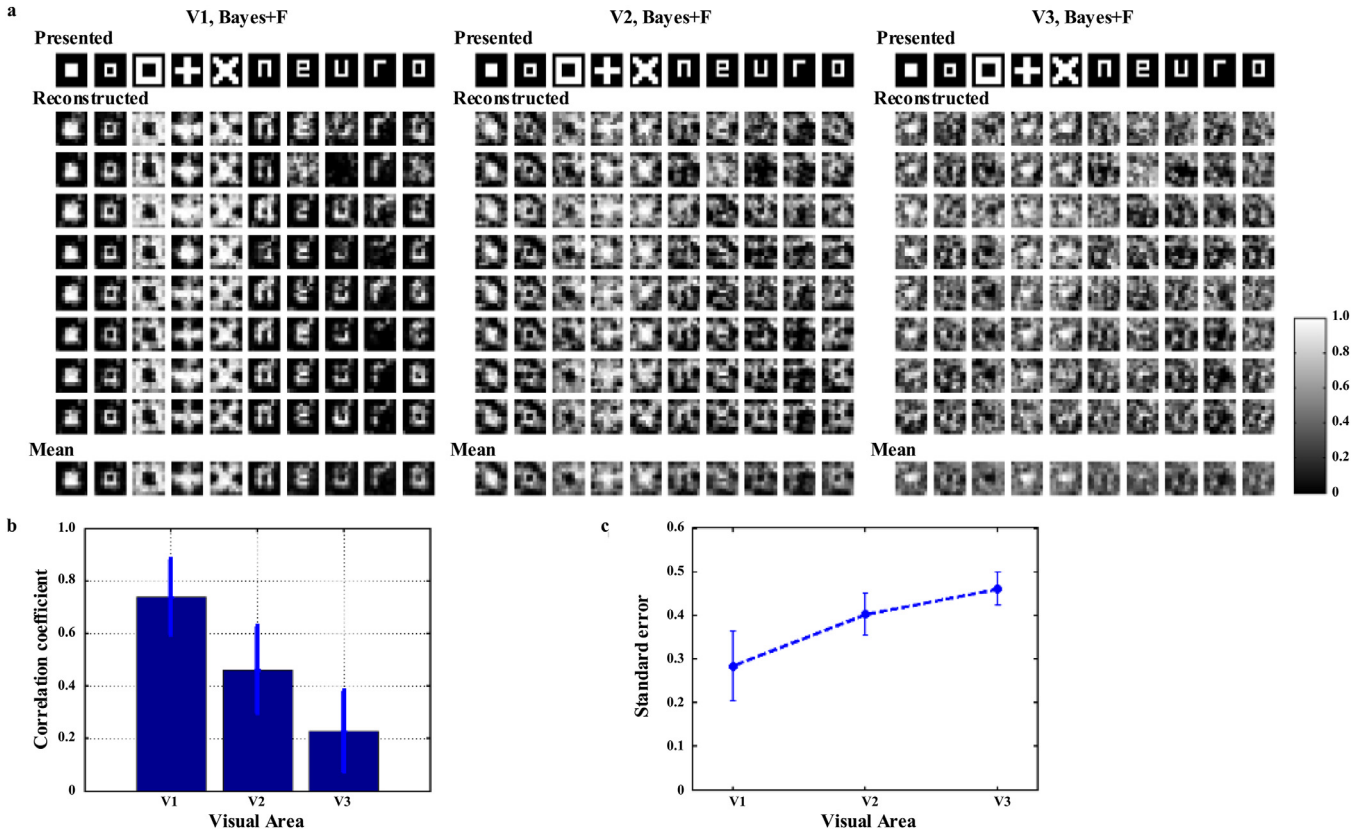


Fig. 8. Comparison of reconstructed images in different visual areas under the Bayes+F algorithm. (a) From left to right panels, the reconstructed images in V1, V2 and V3, respetively. (b) The correlation coefficient between the reconstructed and the stimulus images in V1, V2 and V3. With the increase of the visual cortex, the correlation coefficient decreased. (c) The standard errors of the reconstructed images in V1, V2 and V3.

direct retinotopy mapping from the visual field to the cortical voxels than V1, so they may have a more abstract representation of image. Meanwhile, the reconstruction model proposed in this paper is based on the topology mapping relationship model, therefore, the reconstruction effect decreases from V1 to V2 and V3. Fig. 8 reveals the comparison of the reconstructed images in individual V1, V2, and V3 by using the F-score combined Bayesian algorithm. The result is consistent with that in previous studies [17,19]. Theoretical studies have suggested that V1 uses a sparse code to represent natural scenes efficiently [30]. This preference to V1 voxels may be attributed to the reconstruction principle of retinotopic organization. The center of the reconstructed image is more accurate and clearer than the surrounding areas, therefore, fixation and visual attention may also contribute to the reconstruction quality [17]. In future research, images can be reconstructed with more fixation points to improve the average accuracy of reconstructed images in the surrounding areas. However, the phenomenon that the voxels in V1 is more effective in reconstruction does not indicate that higher visual areas do not contribute to the image reconstruction. Higher visual areas are involved in processing information of objects, faces, scenes, and other complex images [31,32], so the higher visual area voxels may play important roles in reconstructing the complex natural scenes.

5. Conclusion

In this paper, aiming at the shortcomings of existing reconstruction models, we proposed an F-score feature selection-combined Bayesian model that can be efficiently used to reconstruct visual images from human brain activities. The contribution of the F-score feature selection algorithm is mainly to reduce the background noise of the reconstructed image and improve the quality of the reconstructed image. The contribution of the Bayesian classifier algorithm is to reduce the time required for image reconstruction and improve the efficiency of reconstructing the image. The result indicates that the proposed Bayes+F model has better reconstruction accuracy and higher reconstruction efficiency than the SMLR and other models, showing better robustness and noise resistant ability. Besides, the proposed model can reconstruct the images extremely rapid, 100 times faster than SMLR does. This study provides a promising support for brain computer interface in decoding the brain activities into human perception.

Acknowledgment

This work was supported by National High Technology Development Program (863) of China (2015AA020505), 973 project (2013CB329401), and the National Natural Science Foundation of China (61573080, 91420105, 61773094). The authors declared that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.neucom.2018.07.068.

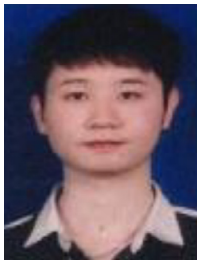
References

[1] J.S. Lund, A. Angelucci, P.C. Bressloff, Anatomical substrates for functional columns in macaque monkey primary visual cortex, Cerebral Cortex 13 (1) (2003) 15–24.

[2] K. Ohki, S. Chung, Y.H. Ch'ng, K. Prakash, R.C. Reid, Functional imaging with cellular resolution reveals precise micro-architecture in visual cortex, Nature 433 (7026) (2005) 597.
[3] M. Siegel, T.H. Donner, R. Oostenveld, P. Fries, A.K. Engel, Neuronal synchronization along the dorsal visual pathway reflects the focus of spatial attention, Neuron 60 (4) (2008) 709–719.
[4] J.V. Haxby, M.I. Gobbini, M.L. Furey, A. Ishai, J.L. Schouten, P. Pietrini, Distributed and overlapping representations of faces and objects in ventral temporal cortex, Science 293 (5539) (2001) 2425–2430.
[5] Y. Kamitani, F. Tong, Decoding the visual and subjective contents of the human brain, Nat. Neurosci. 8 (5) (2005) 679–685.
[6] Y. Kamitani, F. Tong, Decoding seen and attended motion directions from activity in the human visual cortex, Curr. Biol. 16 (11) (2006) 1096–1102.
[7] M.A.J. van Gerven, P. Kok, F.P. de Lange, T. Heskes, Dynamic decoding of ongoing perception, NeuroImage 57 (3) (2011) 950–957.
[8] S. Song, Z. Zhan, Z. Long, J. Zhang, L. Yao, Comparative study of SVM methods combined with voxel selection for object category classification on fMRI data, PloS One 6 (2) (2011) e17191.
[9] R.M. Cichy, J. Heinzle, J.-D. Haynes, Imagery and perception share cortical representations of content and location, Cerebral Cortex 22 (2) (2011) 372–380.
[10] S.-H. Lee, D.J. Kravitz, C.I. Baker, Disentangling visual imagery and perception of real-world objects, Neuroimage 59 (4) (2012) 4064–4073.
[11] A.M. Albers, P. Kok, I. Toni, H.C. Dijkerman, F.P. de Lange, Shared representations for working memory and mental imagery in early visual cortex, Curr. Biol. 23 (15) (2013) 1427–1431.
[12] T. Naselaris, C.A. Olman, D.E. Stansbury, K. Ugurbil, J.L. Gallant, A voxel-wise encoding model for early visual areas decodes mental images of remembered scenes, Neuroimage 105 (2015) 215–228.
[13] S. Nishimoto, A.T. Vu, T. Naselaris, Y. Benjamini, B. Yu, J.L. Gallant, Reconstructing visual experiences from brain activity evoked by natural movies, Curr. Biol. 21 (19) (2011) 1641–1646.
[14] S. Schoenmakers, M. Barth, T. Heskes, M. van Gerven, Linear reconstruction of perceived images from human brain activity, NeuroImage 83 (2013) 951–961.
[15] A.S. Cowen, M.M. Chun, B.A. Kuhl, Neural portraits of perception: reconstructing face images from evoked brain activity, Neuroimage 94 (2014) 12–22.
[16] T. Horikawa, M. Tamaki, Y. Miyawaki, Y. Kamitani, Neural decoding of visual imagery during sleep, Science 340 (6132) (2013) 639–642.
[17] Y. Miyawaki, H. Uchida, O. Yamashita, M.-a. Sato, Y. Morito, H.C. Tanabe, N. Sadato, Y. Kamitani, Visual image reconstruction from human brain activity using a combination of multiscale local image decoders, Neuron 60 (5) (2008) 915–929.
[18] Y. Zhan, J. Zhang, S. Song, L. Yao, Visual image reconstruction from fMRI activation using multi-scale support vector machine decoders, in: International Conference on Human-Computer Interaction, Springer, 2013, pp. 491–497.
[19] S. Song, X. Ma, Y. Zhan, Z. Zhan, L. Yao, J. Zhang, Bayesian reconstruction of multiscale local contrast images from brain activity, J. Neurosci. Methods 220 (1) (2013) 39–45.
[20] F.W. Campbell, J.G. Robson, Application of fourier analysis to the visibility of gratings., J. Physiol. 197 (3) (1968) 551.
[21] R.L. De Valois, D.G. Albrecht, L.G. Thorell, Spatial frequency selectivity of cells in macaque visual cortex., Vis. Res. 22 (5) (1982) 545–559.
[22] Y.-W. Chen, C.-J. Lin, Combining SVMs with various feature selection strategies, Feat. Extrac. (2006) 315–324.
[23] M.F. Akay, D. Akgöl, J. George, Support vector machines combined with feature selection for prediction of maximal oxygen uptake, in: International Symposium on Engineering, Artificial Intelligence and Applications. Kyrenia, North Cyprus, pp. 3–4.
[24] K. Polat, S. Güneş, A new feature selection method on classification of medical datasets: kernel F-score feature selection, Expert Syst. Appl. 36 (7) (2009) 10367–10373.
[25] F.-L. Chen, F.-C. Li, Combination of feature selection approaches with SVM in credit scoring, Expert Syst. Appl. 37 (7) (2010) 4902–4909.
[26] Y. Liu, G. Wang, H. Chen, H. Dong, X. Zhu, S. Wang, An improved particle swarm optimization for feature selection, J. Bionic Eng. 8 (2) (2011) 191–200.
[27] F. Liu, W. Guo, J.-P. Fouche, Y. Wang, W. Wang, J. Ding, L. Zeng, C. Qiu, Q. Gong, W. Zhang, H. Chen, Multivariate classification of social anxiety disorder using whole brain functional connectivity, Brain Struct. Funct. 220 (1) (2015) 101–115.
[28] J. Xie, C. Wang, Using support vector machines with a novel hybrid feature selection method for diagnosis of erythemato-squamous diseases, Expert Syst. Appl. 38 (5) (2011) 5809–5815.
[29] R. Vanrullen, S.J. Thorpe, The time course of visual processing: from early perception to decision-making, J. Cogn. Neurosci. 13 (4) (2001) 454–461.
[30] W.E. Vinje, J.L. Gallant, Sparse coding and decorrelation in primary visual cortex during natural vision, Science 287 (5456) (2000) 1273–1276.
[31] G. Golarai, S. Hong, B.W. Haas, A.M. Galaburda, D.L. Mills, U. Bellugi, K. Grill-Spector, A.L. Reiss, The fusiform face area is enlarged in williams syndrome, J. Neurosci. 30 (19) (2010) 6700–6712.
[32] K.S. Scherf, M. Behrmann, K. Humphreys, B. Luna, Visual category-selectivity for faces, places and objects emerges along different developmental trajectories, Dev. Sci. 10 (4) (2007).

Wei Huang received the M.S. degree in biomedical engineering from University of Electronic Science and Technology of China in 2017. He is currently working toward the Ph.D. degree in biomedical engineering from University of Electronic Science and Technology of China. His research interests include visual cognitive decoding, deep learning, computer vision and functional magnetic resonance imaging.

Hongmei Yan received her M.S. and Ph.D. degrees from Chongqing University in 2000, 2003 respectively. She is now a professor in the Clinical Hospital of Chengdu Brain Science Institute, MOE Key Lab for Neuroinformation, University of Electronic Science and Technology of China. Her research interests include visual cognition, eye movements, visual attention, and saliency detection.



Ran Liu received the B.E. degree in biomedical engineering from Sichuan University in 2015. He is currently working toward the M.S. degree in physical biology from University of Electronic Science and Technology of China. His research interests include big data mining in medicine, data analysis and artificial intelligence.

Lixia Zhu received the M.S. degree in biomedical engineering from University of Electronic Science and Technology of China in 2017. Her research interests include frequency specificity, emotion regulation, cognitive science and functional magnetic resonance imaging.

Huangbin Zhang received the B.S. degree in information and computer science from Gannan Normal University in 2014. He is currently working toward the M.S. degree in biomedical engineering from University of Electronic Science and Technology of China. His research interests include deep learning, computer vision and functional magnetic resonance imaging.

Huafu Chen Prof. received the Ph.D. degree in Biomedical Engineering from University of Electronic Science and Technology of China in 2004. He is currently works in the Clinical Hospital of Chengdu Brain Science Institute, MOE Key Lab for Neuroinformation, University of Electronic Science and Technology of China. His research interests include brain imaging and pattern recognition method, functional magnetic resonance imaging, brain connectivity and mental illness brain imaging analysis.