Novel Tools and Methods

Multitask fMRI Data Classification via Group-Wise Hybrid Temporal and Spatial Sparse Representations

Limei Song,1,∗ Yudan Ren,1,∗ Yuqing Hou,1 Xiaowei He,1 and Huan Liu2

https://doi.org/10.1523/ENEURO.0478-21.2022

1School of Information Science & Technology, Northwest University, Xi’an, 710127, China and 2School of Computer Science and Communication Engineering, Jiangsu University, Jiangsu, 212013, China

Abstract

Task-based functional magnetic resonance imaging (tfMRI) has been widely used to induce functional brain activities corresponding to various cognitive tasks. A relatively under-explored question is whether there exist fundamental differences in fMRI signal composition patterns that can effectively classify the task states of tfMRI data, furthermore, whether there exist key functional components in characterizing the diverse tfMRI signals. Recently, fMRI signal composition patterns of multiple tasks have been investigated via deep learning models, where relatively large populations of fMRI datasets are indispensable and the neurologic meaning of their results is elusive. Thus, the major challenges arise from the high dimensionality, low signal-to-noise ratio, interindividual variability, a small sample size of fMRI data, and the explainability of classification results. To address the above challenges, we proposed a computational framework based on group-wise hybrid temporal and spatial sparse representations (HTSSR) to identify and differentiate multitask fMRI signal composition patterns. Using relatively small cohorts of Human Connectome Project (HCP) tfMRI data as test-bed, the experimental results demonstrated that the multitask of fMRI data can be successfully classified with an average accuracy of 96.67%, where the key components in differentiating the multitask can be characterized, suggesting the effectiveness and explainability of the proposed method. Moreover, both task-related components and resting-state networks (RSNs) can be reliably detected. Therefore, our study proposed a novel framework that identifies the interpretable and discriminative fMRI composition patterns and can be potentially applied for controlling fMRI data quality and inferring biomarkers in brain disorders with small sample neuroimaging datasets.

Key words: multitask classification; task-based fMRI; group-wise; hybrid temporal and spatial sparse representations

Significance Statement

Task-based functional magnetic resonance imaging (tfMRI) is known to be able to induce functional brain activities corresponding to various cognitive tasks. However, the neuroscience mechanism of inherent functional differences that can effectively classify the multi-tfMRI data and the key functional components in composition patterns have been rarely tapped. Our proposed framework can uncover the fundamental differences in fMRI signal composition patterns and classify the multitask fMRI data with an average accuracy of 96.67%. In addition, our framework can effectively identify the key components with greater capacity in multitask classification and disclose the underlying network mechanism of these key components.
Introduction

Researchers have long been endeavoring to induce and decode functional brain activities using functional magnetic resonance imaging (fMRI) data (Poldrack et al., 2009; Jang et al., 2017; Rubin et al., 2017). To detect neural activations embedded in tfMRI data, various computational and statistical methods have been proposed over the past decades, where the general linear model (GLM) is the most prevailing method for tfMRI analysis (Friston, 2009; Mueller et al., 2013). Moreover, independent component analysis (ICA) is another effective approach to characterize functional brain networks (FBNs) based on an assumption of statistically independent relationships between components (Xu et al., 2013), while independence is an ideal hypothesis in mathematics. The study of the visual and auditory perceptual cortex shows that the activity of neurons is highly sparse (Wright et al., 2010). Based on this, a sparse representation algorithm has been applied and demonstrated its effectiveness for functional network identification (Olshausen and Field, 2004; Lv, 2013; Zhao et al., 2018).

While a large number of studies focus on characterizing the concurrent task-evoked brain regions/networks (Calhoun et al., 2001; Lv et al., 2015a; Darnai et al., 2019), the neuroscience of inherent functional differences in composition patterns of multitask fMRI signals has been rarely tapped. Investigating the differences between different task fMRI signals composition could improve better understanding for the organization of the brain’s cognitive functioning, and might contribute to disease diagnosis and classification. In addition, while different cognitive functions are induced by diverse task paradigms, which require functional interactions among different specialized brain regions/networks, another challenging issue rarely explored is whether there exist interpretable and distinctive spatial-temporal components in differentiating fMRI signals under different task designs.

As far as we know, there are several challenges in addressing the above questions. First, as whole-brain fMRI data generally consists of enormous amounts of voxels, group-wise tfMRI signals composed of multiple tasks and subjects have relatively high dimensionality, which inevitably causes overloaded computational burden. Therefore, we would need an efficient computational framework with the capacity of handling the high dimensionality of fMRI signals. Second, the variability and the noises in fMRI signals could be remarkable. Thus, it has been challenging to derive consistent activation patterns from whole-brain fMRI signals of multiple subjects with such a variety of noises, awaiting an effective computational framework (Mueller et al., 2013).

With the successful application of sparse representation and deep learning algorithms in fMRI signals classification studies (Zhang et al., 2016; Parhi, 2019; Wang et al., 2020), there are still some problems. For example, a previous study identified the difference between tfMRI and resting-state fMRI signals via sparse representation method, but it ignored the fundamental differences among different types of task paradigms and lacked the further investigation of the key component of classification (Zhang et al., 2016). In addition, this study derived representative spatial and temporal characteristics via subject-level sparse representation framework, where the correspondence among subjects was confirmed through manual inspection, resulting in time-consuming and laborious work. Recently, while deep learning algorithms are considered to be promising approaches to decode multitask fMRI data, there are still limitations of current studies, including the demand for a large training sample size of fMRI data, which is hard to collect for clinical populations (Litjens et al., 2017; Wen et al., 2018), manual or experimental setting of numerous hyperparameters that are time-consuming and suboptimal (Ching et al., 2018), and the moderate explainability of deep learning models (Guo et al., 2016). For example, Wang et al. (2020) reported that 1034 subjects were used to classify fMRI signals of seven tasks, where these deep learning models are hard to apply to clinical data because of limited sample size, and the neurologic meaning of their findings is elusive. Considering the above challenges and pitfalls of existing research, it is desirable for an appropriate framework that can classify multitask fMRI signals and characterize the key components which play a key role in classification with a small sample case.

In this study, we proposed a group-wise two-stage framework based on hybrid temporal and spatial sparse representations (HTSSR) to identify the intrinsic differences in tfMRI signal composition patterns. Our results demonstrated that both temporal and spatial features could be obtained group-wisely by analyzing only a small proportion (10%) of whole brain fMRI signals. Seven HCP tasks can be classified simultaneously with an average accuracy of 96.67%. Moreover, our framework cannot only effectively identify the key components that can well characterize and differentiate multitask signals, but also imply the underlying neuroscience implications of these components, offering an effective methodology for basic neuroscience and clinical research.

Materials and Methods

Overview

The overall framework (Fig. 1) consists of three stages: (1) data preprocessing and preparation; (2) training stage...
on training set; (3) classification and pos-hoc analyses stage on testing set. In data preprocessing and preparation, for each subject, the whole-brain fMRI data of seven different tasks were extracted and then spatially concatenated to a matrix $S^1$ (Fig. 1, top panel). The whole dataset of 60 subjects was randomly divided into training set and testing set, that is, we randomly selected $p$ subjects from all the subjects as training set and set the rest subjects as testing set. Then, all the training subjects’ signal matrices were spatially concatenated to one large matrix $S^1$ for training model (Fig. 1a1). In temporal sparse representation (TSR) training stage, the online dictionary learning algorithm was used to factorize the large matrix $S^1$ of training set into groupwise time-series dictionaries $D^1$ and the corresponding loading coefficients $a^1$ (Fig. 1b1). Afterwards, in the spatial sparse representation (SSR) training stage, the groupwise time-series dictionaries $D^1$ and the corresponding loading coefficients $a^1$ were fed into the spatial sparse representation (SSR) training to derive the groupwise spatial dictionary $D^2$ and loading coefficient $a^2$ (Fig. 1c1). After the training stage, we then conducted the classification and pos-hoc analyses on testing set. Specifically, based on the group-wise time-series dictionaries $D^1$ and $D^2$ and the trained model derived from training stage, the loading coefficient $a^1$ test...
and $x^2$ test for testing set were obtained for classification analysis (Fig. 1b2-c2). Afterwards, the loading coefficient $x^2$ of training set was adopted to train a support vector machine (SVM) for classification, and the loading coefficient $x^2_{\text{test}}$ was fed into the trained SVM model to obtain the label for the testing set and assess the classification performance of proposed method. Note that all the parameters used in the testing set were learned and trained by the training set. In general, we conducted the experiment above 10 times for validation. For each time, we randomly selected $p$ subjects from the whole datasets as the training set and set the rest for testing, where 10% of voxels of whole brain for each subject were randomly selected for model training or testing, resulting in that the subjects in the training set and testing set were different for each experiment. Furthermore, the common dictionary $D^1$ contains intrinsic functional patterns, and its atoms could estimate spatial maps. By analyzing the active components in the loading coefficient $x^2$, the most discriminative atoms in $x^2$ can be selected as the key components in classification features (Fig. 1d). The temporal features and representative functional networks can be obtained during TSR and SSR, respectively (Fig. 1e).

Data acquisition and preprocessing

In this study, we adopted the seven task fMRI data in Human Connectome Project (HCP) Q1 release (Barch et al., 2013). The acquisition parameters of tfMRI data are as follows: $90 \times 10^3$ matrix, 220 mm Field-of-view (FOV), 72 slices, The repetition time (TR) = 0.72s, The echo time (TE) = 33.1 ms, flip angle = $52^\circ$, bandwidth (BW) = 2290 Hz/Px, in-plane FOV = 208 $\times$ 180 mm, 2.0 mm isotropic voxels.

fMRI preprocessing pipelines included motion correction, spatial smoothing, temporal prewhitening, slice timing correction, global drift removal, and nonlinear registration into 2-mm MNI152 space using FSL FNIRT (http://www.fmrib.ox.ac.uk/fsl/feat5/index.html). Then, we used the standard MNI152 template as the mask to extract each individual tfMRI data, resulting in group-wise spatial correspondence of all the subjects. In this work, 60 subjects in the released dataset were used. More details about data acquisition and preprocessing are referred to (Barch et al., 2013).

The numbers of time points for each task are: emotion (176 frames), motor (284 frames), gambling (253 frames), language (316 frames), relational (232 frames), social (274 frames), motor (284 frames), gambling (253 frames), language (316 frames), relational (232 frames), social (274 frames), motor (284 frames), gambling (253 frames), language (316 frames), relational (232 frames), social (274 frames).

HTSSR

Before TSR, the whole-brain fMRI signals of each subject were converted to two-dimensional matrix. Then, the matrix $S^1_i$ of the $i$-th subject included seven tasks $(S^1_i = [S^1_{i,E}; S^1_{i,M}; S^1_{i,G}; S^1_{i,L}; S^1_{i,R}; S^1_{i,S}; S^1_{i,W}]) \in \mathbb{R}^{\times (n \times 7)}$, where $S^1_{i,E} \in \mathbb{R}^{\times n}$ with $t$ time points and $n$ voxels. The seven capital subscripts represent 7 different tasks, respectively (E: emotion, M: motor, G: gambling, L: language, R: relational, S: social, and W: work memory). Each column in the matrix was normalized to have zero mean and unit norm. The whole-brain data with multitasks of all training set were spatially concatenated to compose a multisubject fMRI matrix $S^1 = [S^1_1; S^1_2; \ldots; S^1_p] \in \mathbb{R}^{\times (n \times 7 \times p)}$, where $p$ is the number of subjects in training set ($p = 30$; Fig. 1a1). As the online dictionary learning algorithm is an effective way to extract distinct information in original signals (Brett et al., 2002), the algorithm would learn a meaningful dictionary $D$ consisting of $k$ atoms to represent $S$ with the corresponding sparse loading coefficient matrix $x$ ($k \ll n$). Specifically, in TSR, the online dictionary learning algorithm can be used to factorize the multisubject fMRI data $S^1$ into a group-wise temporal dictionary $D^1 \in \mathbb{R}^{\times k_1}$ and reference weight matrix $x^1 (x^1 = [x^1_1; x^1_2; \ldots; x^1_p]) \in \mathbb{R}^{\times \times (n \times 7 \times p)}$, $x^1_i = [x^1_{i,E}; x^1_{i,M}; x^1_{i,G}; x^1_{i,L}; x^1_{i,R}; x^1_{i,S}; x^1_{i,W}] \in \mathbb{R}^{\times \times (7 \times k_1)}$.

The loss function for the dictionary learning algorithm was defined in Equation 1 with a $\lambda$ regularization that impose a sparse constraint to the loading coefficient, where $\lambda$ is a regularization parameter which can balance the residual gradient and sparsity level:

$$\min_1 \frac{1}{2} \| S^1 - D^1 x^1 \|_F^2 + \lambda \| x^1 \|_{1,1}. \quad (1)$$

To prevent $D^1$ from arbitrarily large values which leads to trivial solution of the optimization, its columns $d_1, d_2, \ldots, d_k$ are constrained by Equation 2:

$$C \in \{ D^1 \in \mathbb{R}^{\times k_1}, t \forall j = 1, \ldots, k1, \quad d^T_j d_j \leq 1 \}. \quad (2)$$

To reduce the computational burden, we randomly chose only 10% whole-brain signals in each subject during learning dictionary $D^1$ (Liu et al., 2017). The flowchart is shown in Figure 1b1.

In this work, the dictionary size $k_1$ and value of $\lambda_1$ were determined experimentally and empirically ($k_1 = 200$, $\lambda = 0.05$). After TSR, each atom of resulting dictionary $D^1$ matrix contained the temporal information in the functional brain, while the corresponding loading coefficient matrix $x^1$ contained the spatial distribution of each component (Fig. 1e).

The next major goal was to obtain groupwise spatial features that could reveal the distinctive organization patterns of the fMRI signals under different task stimulation, which was achieved in SSR. In SSR, we combined the reference weight matrices of all subjects obtained in TSR, to obtain one matrix $S^2 = [S^2_1; S^2_2; \ldots; S^2_p] \in \mathbb{R}^{\times \times (n \times 7 \times p)}$, where $S^2 = ([x^1_{i,E}; x^1_{i,M}; x^1_{i,G}; x^1_{i,L}; x^1_{i,R}; x^1_{i,S}; x^1_{i,W}]^T) \in \mathbb{R}^{\times \times (7 \times k_1)}$. Then, $S^2$ would...
be served as the input for SSR to obtain a groupwise spatial dictionary $D^2 \in \mathbb{R}^{n \times k_2}$ and the corresponding loading coefficients $\mathbf{a}^2$. Note that $\mathbf{a}^2 = [\mathbf{a}_{1}^2, \mathbf{a}_{2}^2, \ldots, \mathbf{a}_{p}^2] \in \mathbb{R}^{k_2 \times (K_1 \times 7 \times P)}$, where $\mathbf{a}_{j}^2 = ([\mathbf{x}_{jE}^2]^T, [\mathbf{x}_{jM}^2]^T, [\mathbf{x}_{jL}^2]^T, [\mathbf{x}_{jR}^2]^T, [\mathbf{x}_{jS}^2]^T, (\mathbf{x}_{jW}^2)^T] \in \mathbb{R}^{k_2 \times K_1 \times 7}$. In SSR, we set parameters experimentally and empirically as follows: $k_2 = 50$, $\lambda_2 = 0.1$.

To derive the loading coefficients for testing set for further classification analysis, firstly, in the TSR stage, the groupwise time-series dictionary matrix $D^1$ obtained during the training stage was used to represent $S_{i\text{test}}^1$ by solving a typical l-1 regularized LASSO problem (Fig. 1b2). In the SSR stage, the dictionary matrix $D^2$ obtained from the training stage was then used to obtain the loading coefficient $\mathbf{a}_{i\text{test}}^2$ of testing set (Fig. 1c2). The acquisition of $\mathbf{a}_{i\text{test}}^2$ was the deterministic LASSO solution as the Equation 3 shows, where $i$ represents 1 or 2. The values of $\lambda_1$ and $\lambda_2$ were set as the same way of the training stage ($\lambda_1 = 0.05$, $\lambda_2 = 0.1$):

$$\min_{\mathbf{a}_{i\text{test}}^2} \frac{1}{2} ||S_{i\text{test}}^t - D\mathbf{a}_{i\text{test}}^2||_F^2 + \lambda ||\mathbf{a}_{i\text{test}}^2||_1, \quad (3)$$

The proposed HTSSR framework reduced the size of original fMRI data dramatically while maintained the intrinsic temporal and spatial information, thus the intrinsic features (the loading coefficient) we exacted via our framework can represent differences of functional brain activity patterns.

Identification of temporal features and representative functional networks

The temporal features and functional networks can be estimated by the proposed framework. In TSR, the $S_{i\text{test}}^1$ can be factorized into $D^1$ and $\mathbf{a}_1^1$, where $i$ represents $i$-th subjects, $t$ represents $t$ kind of task, $\in \Phi = \{E, M, G, L, R, S, W\}$. Then, the transpose of $\mathbf{a}_1^1$ could be factorized into $D^2$ and $\mathbf{a}_i^2$ as the Equation 4 shows. So, we can obtain the Equation 5 as follows:

$$S_{i\text{test}}^1 = (\mathbf{a}_1^1)^T = D^1 \times \mathbf{a}_1^1, \quad \text{Equation } 4$$

$$S_{i\text{test}}^2 = D^1 \times \mathbf{a}_i^2 = D^1 \times (D^2 \times \mathbf{a}_i^2)^T = D^1 \times (\mathbf{a}_i^2)^T \times (D^2)^T. \quad \text{Equation } 5$$

Considering all subjects sharing the same groupwise dictionary, the intrinsic difference of various tasks depends on the loading coefficients $\mathbf{a}_i^2$. Further, since $D^2$ contained intrinsic groupwise spatial patterns, the temporal information of origin signal $S_{i\text{test}}^1$ should exist in the first two items in Equation 5, that is, $D^1$ and $(\mathbf{a}_1^1)^T$. In order to obtain the groupwise temporal pattern of various tasks, we averaged the loading coefficient $(\mathbf{a}_i^2)^T$ in each subject and finally got the Equation 6. That means, task-specific temporal course is the weighted average of the loading coefficient of each subject and the group-wise common temporal dictionary:

$$D_i = D^1 \times \frac{1}{p} \sum_{i=1}^{p} (\mathbf{a}_i^2)^T, \text{ s.t. } t \in \Phi = \{E, M, G, L, R, S, W\}. \quad \text{Equation } 6$$

Based on prior task paradigms, we can obtain the Pearson correlation coefficient between the task paradigms and the task-specific temporal course, which was defined as

$$P_{\text{corr. } i} = \text{corr} (D_i, \text{ TASK}). \quad \text{Equation } 7$$

Essentially, $P_{\text{corr. } i}$ measures the temporal similarity between temporal course of the $j$-th component in $D_i$ and the t-task paradigms stimulus curve, where a larger value means better correspondence between the dictionary and the stimulus. As the common dictionary $D^2$ contains intrinsic groupwise functional patterns derived from SSR, the atoms in $D^2$ could be used to define the functional spatial maps (Fig. 1e).

We also identified the spatial matching rate to measure the similarity between spatial patterns derived by our proposed framework and GLM-derived activation patterns. Specifically, the GLM-based activations were performed individually and group-wise using FSL FEAT (http://www.fmrib.ox.ac.uk/fsl/feat5/index.html), and the group-level GLM-based results were used for comparison. The details of GLM analysis can be found in previous literature (Lv et al., 2015b). The overlapping rate with the template was defined quantitatively as Equation 8:

$$R(X, T) = \frac{|X \cap T|}{|T|}, \quad \text{Equation } 8$$

where $X$ is the spatial functional networks of the derived component of our proposed framework and $T$ is the GLM-derived activation template.

SVM-based classification method

To classify multitask brain signals, we first trained the SVM classifier using loading coefficient $\mathbf{a}^2$ derived from training set, as $\mathbf{a}^2$ contained both temporal and spatial information embedded in multitasks fMRI signals. To evaluate the performance of proposed framework, we conducted multitask classification analysis on independent testing set, where the $\mathbf{a}^2_{\text{test}}$ was fed into the trained SVM model to derive the classification rate. Specifically, according to the true label of different seven tasks for each loading coefficient $\mathbf{a}^2_{\text{test}}$, the classification accuracy was calculated by the proportion of samples that are predicted correctly. The SVM classifier was established based on the LIBSVM toolbox (Chang and Lin, 2011) for multitask classification. As the number of features was relatively large, the linear kernel was selected as the SVM kernel. In addition, all the other parameters were set as default values.

Ratio of activation (ROA)-based analysis

The final goal was to find the discriminative features for classification. Inspired by the successes of using ROA in
two types of fMRI signals classification (Zhang et al., 2016), we proposed a novel ROA metric as follows: each row in loading coefficients $\mathbf{x}^2$ represents the active level of the corresponding atoms in $\mathbf{S}^2$, that is, to what extent the $i$-th row in loading coefficients $\mathbf{x}^2$ is activated in the $j$-th atom in $\mathbf{S}^2$ and the ROA of the $i$-th loading coefficients $\mathbf{x}^2$ was defined as follows:

$$\text{ROA}_i = \log \frac{1}{T} \sum_{t=1}^{T} \sum_{k=1}^{T} \left| \frac{\alpha(i,j)}{\alpha(i,j)_0} \right|, \ j \text{th column belongs to task}(t) \cup \ j \text{th column belongs to task}(k)$$

(9)

In Equation 9, $T$ represents task index, which refers to seven in our work. Task (1) to Task (7) represent seven different tasks (emotion, motor, gambling, language, relational, social, and work memory). The ROA was obtained by counting the number of non-zero entries of the rows in $\mathbf{S}^2$ which have been labeled as seven tasks. High ROA value indicates that the corresponding atom in $\mathbf{S}^2$ is highly active in specific task.

In order to verify the components with higher ROA value capture greater capacity in classifying the multitask fMRI signals, we designed an experiment as follows. After sorting the ROA values of all the components (rows in loading coefficients $\mathbf{x}^2$) from high to low, we iteratively employed more rows sorted by their ROA values in $\mathbf{x}^2$ as the feature inputs to train a SVM classifier, that is, the components with higher ROA values would be used to train preferentially. The corresponding components in $\mathbf{x}^2$ of testing set were entered into the trained SVM model to derive the classification accuracy. Here, we adopted the same classification scheme depicted above (SVM-based classification method).

### Code accessibility

The MATLAB code of HTSSR framework and ROA-based analysis described in the paper can be accessed in Extended Data 1 (MATLAB code).

### Results

By applying the proposed HTSSR framework to seven tfMRI data from the HCP dataset, our results revealed that all the tfMRI signals can be effectively differentiated, and the intrinsic spatial/temporal patterns underlying their fundamental differences in signal composition could be characterized by the corresponding loading coefficients $\mathbf{x}^2$. Intriguingly, although we only selected a few components with high ROA values in $\mathbf{x}^2$ as inputs for classification, the seven tasks can be accurately classified, and the average accuracy of 10 independent experiments on different testing sets was 96.67 ± 1.22% (mean ± SD; Fig. 2a). Moreover, our proposed framework cannot only classify seven tasks accurately, but also can effectively identify four types of functional components: task-evoked components, resting-state functional components, integrated functional components, and artifact components. In addition, as the resting state and artifact components were very useful in the clinical populations, we further investigated and discussed their role in multitask classification. Finally, to improve the interpretability of the classification results, we further investigated the underlying network mechanism of the classification capability of each functional component.

### Classification and ROA-based analysis

Our proposed framework can accurately classify seven tasks on testing set and the classification accuracy of 10 experiments ranged from 94.67% to 98.57%, with an average accuracy of 96.67% (Fig. 2a), demonstrating our proposed framework can effectively uncover the inherent differences in composition patterns of multitask fMRI signals. These inherent differences between tasks can be revealed by the loading coefficient, which is distinctive and descriptive enough to classify tfMRI data accurately.

As depicted above (ROA-based analysis), we iteratively fed more components from the loading coefficient $\mathbf{x}^2$ of testing set as the feature inputs for the SVM classifier, where these components were sorted by their ROA values independently derived from training set. The classification results of 10 independent experiments on testing set are shown in Figure 2b. When the number of features used reached at ten, the average accuracy of all the curves increased monotonically and can reach at 80% (55–96%), and then reached 90% with 30 atoms (85–98%). Finally, the accuracy was almost close to 100% as more components were included. As the performance curve exhibited a plateau with more than thirty components, we concluded that the additional components with smaller ROA value contribute little to the differentiation power. The results show that our method can effectively disclose the key components that play great roles in successful classification.

### Task-evoked functional components

The most predominant functional components identified by our framework are the task-evoked functional components, including Emotion, Motor, Gambling, Language, Relational, Social, and Working memory. Specifically, the derived temporal patterns are relatively consistent with the task design paradigms for most of tasks (emotion, motor, language, and WM) although only 10% voxel signals were used during TSR (Fig. 3c) and their associated spatial distributions (Fig. 3a) are also relatively consistent with the results from the groupwise GLM-derived activations (Fig. 3b). In addition, the frequency spectrum of its time courses is highly concentrated on the task design frequency (Fig. 3d). However, despite for relatively consistent functional components derived from proposed method compared with GLM-based results, there still exist some disparities, especially in gambling, and relational tasks (Fig. 3a).

Considering the fact that 10% random voxels may probably introduce these disparities although it reduces the computational burden, we thus conducted the proposed model 10 times to assess the robustness of
proposed model and the consistency of the activation patterns, where different training and testing sets were employed and 10% of voxels of whole brain for each subject were randomly selected for each experiment. The results show that while there were slight differences in spatial activation pattern and strength, the overall functional activations show high consistency across 10 experiments (Extended Data Fig. 3-1). Meanwhile, the revealed temporal patterns of 10 experimental results were generally consistent with task design paradigms across 10 experimental results (Extended Data Fig. 3-2). Furthermore, we computed the Pearson correlation coefficient between the derived time courses and task paradigms (PCCTC), and the overlap rates of the functional activation maps with GLM results for seven tasks and 10 experiments, separately (Table 1). The average Pearson correlation coefficient between the time courses of seven identified task-evoked components and the task paradigms is 0.72 ± 0.05 (mean ± SD). The defined brain activation and temporal patterns of 10 experimental results can be found in Extended Data Figures 3-1, 3-2. In general, these results show that while adopting 10% randomly selected voxels for our sparse decomposition method can slightly affect the derived functional activation patterns, the overall pattern can be very consistent across different tests, further demonstrating the robustness of the proposed framework.

Overall, based on the temporal, spatial, frequency-domain characteristic results, we concluded that our framework could effectively identify task-evoked functional components from large scale combined multitask fMRI data. Additional results of identified task-evoked functional components could be found in Extended Data Figures 3-3, 3-4, 3-5, 3-6, 3-7, 3-8.

### Resting-state functional components

In addition to task-evoked component, our framework can reliably define resting-state networks (RSNs). Several brain networks that have been established previously
were identified, including primary visual network (Fig. 4a), default mode network (DMN; Fig. 4b), cerebellum (Fig. 4c), executive control network (Fig. 4d), left frontoparietal network (lFPN; Fig. 4e), and right FPN (rFPN; Fig. 4f; Damoiseaux et al., 2006). The 3D brain networks were visualized with the BrainNet Viewer (http://www.nitrc.org/projects/bnv/; Xia et al., 2013). Our results further demonstrate that when participants are processing specific cognitive tasks, the RSNs are also consistently “active” (Deco et al., 2013).

### Integrated functional component

Besides task evolved and resting-state functional components, some complex interconnected networks so-called integrated functional component can be found in our work as well. Figure 5a shows a bilateral FPN, which might indicate the interaction between IFPN and rFPN. The frontoparietal network (FPN) is critical for our ability to coordinate behavior in a rapid, accurate, and flexible goal-driven manner (Marek and Dosenbach, 2018). Figure 5b illustrates a network blended with a DMN, dorsolateral prefrontal cortex (dLPFC) and frontopolar area. Some studies demonstrated that the dLPFC has robust fMRI functional connectivity and reciprocal anatomic connections with the posterior DMN core regions: posterior parietal cortex (PPC) and posterior cingulate cortex (PCC) in marmoset (Liu et al., 2019). This complex network shown in Figure 5b may be associated with mental processes that require rigorous computation, control, and

### Table 1: The average Pearson correlation coefficients of the time courses (PCCTC) and the overlap rates of the functional networks of seven tasks for 10 experiments (mean ± SD)

| Task  | Event | Emotion | Motor | Gambling | Language | Relational | Social | WM |
|-------|-------|---------|-------|----------|----------|------------|--------|----|
|       | 1     | 0.81 ± 0.05 | 0.91 ± 0.02 | 0.51 ± 0.09 | 0.87 ± 0.02 | 0.63 ± 0.04 | 0.69 ± 0.10 | 0.66 ± 0.06 |
|       | 2     | 0.89 ± 0.05 | 0.70 ± 0.09 | 0.49 ± 0.13 | 0.62 ± 0.14 | 0.54 ± 0.13 | 0.79 ± 0.11 | 0.63 ± 0.08 |
decision-making. Figure 5c shows another complex network named salience network (SN), which plays a crucial role in identifying the most biologically and cognitively relevant events for adaptive guiding attention and behavior, and constitutes a key interface for cognitive, homeostatic, motivational, and affective systems (Seeley et al., 2007). These integrated functional components were activated during the task and could reflect the interactions between different brain regions/network, indicating that our framework cannot only define traditional task-evoked and resting-state functional components, but also reveal the interconnections between brain regions/networks.

**Artifact-related component**

Our framework cannot only define meaningful networks, but also detect artifact-related components related to head movement (Fig. 6a), white-matter (Fig. 6b), cardiac-related (Fig. 6c), and MRI acquisition/reconstruction related (Fig. 6d). Head movement and cardiac artifact-related components are mainly caused by physiology and subject motion during MRI acquisition (Fig. 6a,c; Griffanti et al., 2014). In addition, white-matter and MRI acquisition/reconstruction artifact-related components could be caused by the MRI hardware or software (Fig. 6b,d; Griffanti et al., 2014; Salimi-Khorshidi...
et al., 2014). Detecting and separating these artifact-related component make sense for finding meaningful networks.

The functional role of resting state and artifact components in multitask classification

We here further investigated the role of resting state and artifact components in multitask classification. Specifically, we excluded these two kinds of components from all the defined components of training set respectively, and put the rest components into the SVM model to train the classifier, resulting in four types of component groups used for classification ("all the components," "excluding artifact components," "excluding resting state components," and "excluding both artifact and resting state components"). Afterwards, we selected and imported these four types of components of the testing set into the trained model to obtain classification rate, respectively. Overall, we conducted this experiment five times on different training and testing sets. The results show that the classification rates are relatively high and there is little difference in the classification rate among different cases (Extended Data Fig. 2–1). In two testing sets, excluding resting state components results in lower accuracy, but the effect is not significant. The main reason for this slight effect is that these two types of components account for a small proportion of all defined components (only 10 components in total). Thus, task-related components play the most important role in classification analysis.

The underlying network mechanism of key components with high classification accuracy

To further explore the neural implications of key components with greater classification capacity, we investigated whether there is significant correlation between the classification accuracy and the overlap rate of each component, where the overlap rate of component is defined as the spatial matching rate with GLM-derived activation patterns or RSNs templates. As shown in Figure 7, y-axis represents the accuracy of using each independent component in $x^2_{test}$ for multitask classification reflecting their classification capacity, and x-axis refers to the spatial overlap rate of corresponding atom in dictionary $D^2$. Each red point represents the component in loading coefficient $x^2$ derived from SSR stage. Note that the classification accuracy is significantly correlated with the overlap rate of each component ($R^2 = 0.37, p = 3.14e-06$; Fig. 7). These results thus suggest that a strengthened overlap rate predicts greater classification capacity of a functional component, indicating the underlying network mechanism of classification ability for derived functional components and gaining the interpretability of the proposed framework.

Discussion

In this study, we proposed a framework using HTSSR to examine fundamental differences in multitask fMRI signal composition patterns that can effectively classify seven task fMRI signals with an average accuracy of 96.67%. In addition, our framework also identified interpretable and distinctive spatial-temporal components critical for differentiating diverse fMRI signals, and disclosed the neural implications of these key components. Furthermore, our framework can effectively detect various networks...
including task-related components, RSNs, integrated complex functional components and artifact-related components, further suggesting the effectiveness of proposed method.

Considering the challenges and pitfalls of existing studies, our framework provided a suitable way to classify multitask fMRI signals and characterize the key components for classification and their underlying network mechanism. First, regarding to enormous number of voxels in group-wise fMRI signals, our framework only randomly selected 10% of whole-brain signals for each subject in TSR stage, which greatly reduced the matrix dimensions of millions of data points for group-wise signals, and dramatically lessenened the computational burden while preserving the inherent features of tfMRI signals. Consequently, with only 10% signals adopted as training sample, our framework can still effectively and efficiently differentiate the group-wise multitask fMRI signals. Second, in terms of tackling the issues of interindividual variability and the noise of fMRI signals, our framework successfully derived meaningful functional activation patterns by extracting group-wise common dictionary from fMRI signals of all the subjects with great interindivudual variability. Meanwhile, the two-stage sparse representations framework can effectively remove most noises. Third, according to the limited sample size of task-based fMRI datasets and clinical populations, our framework is effective on decoding multitasks fMRI signals for small cohort datasets. Finally, our study further defined the critical components in multitask classification and their neural implications. Specifically, our results uncovered the significant correlation between classification accuracy and the overlap rate with well-defined network templates of each component, indicating the underlying neural mechanism of key components with great classification capacity.

Despite the promising classification performance of the HTSSR framework, there also exist some limitations. First, while most of the functional components corresponding to the events of task designs have been detected by our framework, there are still a few activation maps that have not been found compared with GLM-based results, such as 0BK_FACE design of WM. Second, some functional components and associated temporal patterns derived by our framework are not perfectly consistent with those defined by GLM-based method. For instance, compared with the original design, time course of relational task has an additional task block (Extended Data Fig. 3-9c), and its associated activation map misses some regions such as the dorsal anterior cingulate cortex and inferior prefrontal gyrus (Extended Data Fig. 3-9a). One reason causing these disparities is that to find the key features for task classification, the number of dictionary atoms in the second stage of our method was set to 50, resulting in only 50 functional components defined. In contrast, the atom number was usually set to 400 in previous task-based activation identification studies using sparse decomposition method (Lv et al., 2015a; Zhao et al., 2018). In addition, for the classification purpose, our method concatenated and aggregated the tfMRI signals of seven tasks together for the model training, which leads to truncations of some tasks, instead of applying the sparse decomposition method to single and complete task fMRI data alone (Lv et al., 2015a). The truncation would lead to incomplete task designs (Extended Data Fig. 1-1), thus impacting the characterization of task-related functional activations. For example, four events were excluded in WM task because of the data truncation, the activations of which would not be detected inevitably. On the other hand, another reason might be that we randomly selected 10% voxels of each subject for model training and testing. Thus, while previous functional activations detection studies using sparse representation framework manifest great consistency with GLM-based results, the limited number of dictionary atoms, truncations of tfMRI signals, and 10% randomly selected voxels for group-wise training of our method might result in disparities between functional components defined and GLM-based results. Nevertheless, despite the existing disparities, the main purpose of our study was to develop an efficient and effective framework for multitask classification and uncover the interpretable and discriminative fMRI composition patterns.

Third, regarding to the parameters setting issue of HTSSR framework, another limitation of our study is manual setting of sparse parameters \( \lambda \) for two stage sparse representation. As there is no golden criterion for the selection of \( \lambda \) in sparse representation algorithm, we here systematically varied the \( \lambda \) settings and assessed their impact on the classification performance using separate training and testing sets. According to the parameter setting in previous FBNs identification studies using sparse representation methods (Lv et al., 2015a; Zhang et al., 2017; Ge et al., 2018), we assessed the impact of parameter settings by systematically varying \( \lambda_1 \) (0.05, 0.1, 0.5) and \( \lambda_2 \) (0.05, 0.1, 0.5). To avoid information leakage, we randomly selected 30 subjects from the whole dataset as training set and trained the model using training set alone with different combinations of \( \lambda_1 \) and \( \lambda_2 \), and then conducted the classification analysis on the testing set composed of remaining subjects, using all parameters derived from the trained model. The classification results with different parameter settings are shown in Table 2. The classification rates for most combinations of \( \lambda_1 \) and \( \lambda_2 \) are consistently high, except for the \( \lambda_1 = 0.5 \). By manually inspecting all the functional activations derived with \( \lambda_1 = 0.5 \) (Extended Data Fig. 2-2), we find that the functional activation maps become very sparse with no meaningful activation patterns under these parameter settings. Thus, when the sparsity penalty was set too large, the key functional components in differentiating the multitask signals cannot be characterized, consequently leading to poor classification performance. However, the overall classification

| \( \lambda_1 \) | \( \lambda_2 \) | Classification Rate |
|---|---|---|
| 0.05 | 0.05 | 99.05% |
| 0.05 | 0.1 | 98.57% |
| 0.05 | 0.5 | 99.52% |
| 0.1 | 0.05 | 99.52% |
| 0.1 | 0.1 | 99.52% |
| 0.1 | 0.5 | 99.05% |
| 0.5 | 0.05 | 14.29% |
| 0.5 | 0.1 | 14.29% |
| 0.5 | 0.5 | 14.29% |
performance was quite stable and satisfactory with reasonable \( \lambda \) settings. Therefore, in our study, the \( \lambda 1 \) and \( \lambda 2 \) were set to 0.05 and 0.1, respectively. In the future, we would like to further develop an automatic optimization strategy for parameter setting of the proposed method.

Overall, our proposed framework provided an effective and interpretable tool for classifying multitask fMRI data. In the future, this framework can be easily applied to a wide range of neuroimaging research with a small dataset, such as mental state classification or brain disorders diagnosis.

References

Barch DM, Burgess GC, Harms MP, Petersen SE, Schlaggar BL, Wager TD, Xue J (2013) Identifying functional networks via sparse coding of whole brain fMRI signals. Annual International IEEE EMBS Conference on Neural Engineering, pp 778–781. San Diego, California, 6–8 November, IEEE, Los Alamitos, CA.

Lv J, Jiang X, Li X, Zhu D, Chen H, Zhang T, Zhang S, Hu X, Han J, Huang H, Zhang J, Guo L, Liu T (2015a) Sparse representation of whole-brain fMRI signals for identification of functional networks. Med Image Anal 20:112–134.

Lv J, Jiang X, Li X, Zhu D, Zhang S, Zhao S, Chen H, Zhang T, Hu X, Han J, Ye J, Guo L, Liu T (2015b) Holistic atlases of functional networks and interactions reveal reciprocal organizational architecture of cortical function. IEEE Trans Biomed Eng 62:1120–1131.

Mueller S, Wang D, Fox MD, Yeo BT, Sepulcre J, Sabuncu MR, Shafae R, Lu J, Liu H (2013) Individual variability in functional connectivity architecture of the human brain. Neuron 77:586–595.

Olofshaven BA, Field DJ (2004) Sparse coding of sensory inputs. Curr Opin Neurobiol 14:481–487.

Parhi BSKK (2019) Predicting tasks from task-fMRI using blind source separation. Asilomar Conference on Signals, Systems, and Computers, pp 2201–2205. Pacific Grove, CA, USA, 3-6 Nov, IEEE, Los Alamitos, CA.

Poldrack RA, Halchenko YO, Hanson SJ (2009) Decoding the large-scale structure of brain function by classifying mental states across individuals. Psychol Sci 20:1364–1372.

Rubin TN, Koyejo O, Gorgolewski KJ, Jones MN, Poldrack RA, Yarkoni T (2017) Decoding brain activity using a large-scale probabilistic functional-anatomical atlas of human cognition. PLoS Comput Biol 13:e1005649.

Salimi-Khorshidi G, Douaud G, Beckmann CF, Glasser MF, Griffanti L, Smith SM (2014) Automatic denoising of functional MRI data: combining independent component analysis and hierarchical fusion of classifiers. Neuroimage 90:449–468.

Marek S, Dosenbach NUF (2018) The frontoparietal network: function, electrophysiology, and importance of individual precision mapping. Dialogues Clin Neurosci 20:133–140.

Seeley WW, Menon V, Schatzberg AF, Keller J, Glover GH, Kenn A, Reiss AL, Greicius MD (2007) Dissociable intrinsic connectivity networks for salience processing and executive control. J Neurosci 27:2349–2356.

Xu J, Jiang X, Jiang Z, Nguchu BA, Zhou Y, Wang Y, Wang H, Li Y, Zhu Y, Wu F, Gao JH, Qiu B (2020) Decoding and mapping task states of the human brain via deep learning. Hum Brain Mapp 41:1505–1519.

Wang J, Zhang X, Li X, Zhu D, Chen H, Zhang T, Zhang S, Hu X, Han J, Huang H, Zhang J, Guo L, Liu T (2017) FMRI network classification based on sparse coding and graph theory. Med Image Anal 20:112–134.

Lv J (2013) Identifying functional networks via sparse coding of whole brain fMRI signals. Annual International IEEE EMBS Conference on Neural Engineering, pp 778–781. San Diego, California, 6–8 November, IEEE, Los Alamitos, CA.

Lv J, Jiang X, Li X, Zhu D, Chen H, Zhang T, Zhang S, Hu X, Han J, Huang H, Zhang J, Guo L, Liu T (2015a) Sparse representation of whole-brain fMRI signals for identification of functional networks. Med Image Anal 20:112–134.

Lv J, Jiang X, Li X, Zhu D, Zhang S, Zhao S, Chen H, Zhang T, Hu X, Han J, Ye J, Guo L, Liu T (2015b) Holistic atlases of functional networks and interactions reveal reciprocal organizational architecture of cortical function. IEEE Trans Biomed Eng 62:1120–1131.

Mueller S, Wang D, Fox MD, Yeo BT, Sepulcre J, Sabuncu MR, Shafae R, Lu J, Liu H (2013) Individual variability in functional connectivity architecture of the human brain. Neuron 77:586–595.

Olofshaven BA, Field DJ (2004) Sparse coding of sensory inputs. Curr Opin Neurobiol 14:481–487.

Parhi BSKK (2019) Predicting tasks from task-fMRI using blind source separation. Asilomar Conference on Signals, Systems, and Computers, pp 2201–2205. Pacific Grove, CA, USA, 3-6 Nov, IEEE, Los Alamitos, CA.

Poldrack RA, Halchenko YO, Hanson SJ (2009) Decoding the large-scale structure of brain function by classifying mental states across individuals. Psychol Sci 20:1364–1372.

Rubin TN, Koyejo O, Gorgolewski KJ, Jones MN, Poldrack RA, Yarkoni T (2017) Decoding brain activity using a large-scale probabilistic functional-anatomical atlas of human cognition. PLoS Comput Biol 13:e1005649.

Salimi-Khorshidi G, Douaud G, Beckmann CF, Glasser MF, Griffanti L, Smith SM (2014) Automatic denoising of functional MRI data: combining independent component analysis and hierarchical fusion of classifiers. Neuroimage 90:449–468.

Marek S, Dosenbach NUF (2018) The frontoparietal network: function, electrophysiology, and importance of individual precision mapping. Dialogues Clin Neurosci 20:133–140.

Seeley WW, Menon V, Schatzberg AF, Keller J, Glover GH, Kenn A, Reiss AL, Greicius MD (2007) Dissociable intrinsic connectivity networks for salience processing and executive control. J Neurosci 27:2349–2356.

Xu J, Jiang X, Jiang Z, Nguchu BA, Zhou Y, Wang Y, Wang H, Li Y, Zhu Y, Wu F, Gao JH, Qiu B (2020) Decoding and mapping task states of the human brain via deep learning. Hum Brain Mapp 41:1505–1519.

Wen D, Wei Z, Zhou Y, Li G, Zhang X, Han W (2018) Deep learning methods to process fMRI data and their application in the diagnosis of cognitive impairment: a brief overview and our opinion. Front Neuroinform 12:23.

Wright J, Ma Y, Mairal J, Sapiro G, Huang TS, Yan S (2010) Sparse representation for computer vision and pattern recognition. Proc IEEE 98:1031–1044.

Xia M, Wang J, He Y (2013) BrainNet Viewer: a network visualization tool for human brain connectomics. PLoS One 8:e68910.

Xu J, Potenza MN, Calhoun VD (2013) Spatial ICA reveals functional activity hidden from traditional fMRI GLM-based analyses. Front Neurosci 7:154.

Zhang S, Li X, Lv J, Jiang X, Guo L, Liu T (2016) Characterizing and differ-entiating task-based and resting state fMRI signals via two-stage sparse representations. Brain Imaging Behav 10:21–32.

Zhang S, Li X, Guo L, Liu T (2017) Exploring human brain activation via nested sparse coding and functional operators. International Symposium on Biomedical Imaging (ISBI 2017) 2017 IEEE 14th, Melbourne, VIC, Australia, 18–21 April, IEEE, Los Alamitos, CA.

Zhang S, Li X, Hu X, Jiang X, Lv J, Zhang T, Zhang S, Guo L, Liu T (2018) Extendable supervised dictionary learning for exploring di-verse and concurrent brain activities in task-based fMRI. Brain Imaging Behav 12:743–757.