IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 22, 2020, accepted March 21, 2020, date of publication April 1, 2020, date of current version April 24, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2984948

Supervised Brain Network Learning Based on Deep Recurrent Neural Networks

SHIJIE ZHAO[1], YAN CUI[2], LINWEI HUANG[1], LI XIE[2], YAOWU CHEN[2], JUNWEI HAN[1], (Senior Member, IEEE), LEI GUO[1], SHU ZHANG[3], TIANMING LIU[4], (Senior Member, IEEE), AND JINGLEI LV[5]

[1]School of Automation, Northwestern Polytechnical University, Xi'an 710072, China
[2]College of Biomedical Engineering and Instrument Science, Zhejiang University, Hangzhou 310027, China
[3]School of Computer Science, Northwestern Polytechnical University, Xi'an 710072, China
[4]Department of Computer Science, University of Georgia, Athens, GA 30602, USA
[5]Sydney Imaging and School of Biomedical Engineering, The University of Sydney, Camperdown, NSW 2006, Australia

Corresponding author: Shu Zhang (shuzhang8967@163.com)

This work was supported in part by the Guangdong Provincial Key Laboratory of Medical Image Processing under Grant 2017B030314133, in part by the Key Research and Development Program of Guangdong Province under Grant 2019B010110001, in part by the National Natural Science Foundation of China under Grant 61806167, Grant 61936007, and Grant U1801265, in part by the China Postdoctoral Science Foundation under Grant 2019T120945 and Grant 2017M613206, in part by the Fundamental Research Funds for the Central Universities under Grant 3102019PJ005 and Grant 3102019QD1005, in part by the Natural Science Basic Research Plan in Shaanxi Province of China under Grant 2019JQ-630, and in part by the research funds for interdisciplinary subject, NWPU.

ABSTRACT Task-based functional magnetic resonance imaging (tfMRI) is a widely used neuroimaging technique in exploring brain networks and functions associated with cognitive behaviors. Traditionally, the general linear model (GLM) is the most popular method in tfMRI data analysis due to its simpleness and robustness. This model-driven method adopts a canonical hemodynamic response function (HRF) and its various derivatives to construct regressors in the design matrix and estimate changes in the tfMRI data. However, a possible limitation of current model-driven methods is that the HRF is fixed and non-adaptive which may overlook other diverse and concurrent brain networks. In order to overcome these limitations, we proposed a novel hybrid framework, supervised brain network learning based on deep recurrent neural networks (SUDRNN), to reconstruct the diverse and concurrent functional brain networks. Specifically, this hybrid framework first takes advantage of the great capacity of deep recurrent neural networks (DRNN) in modeling sequential data to learn the diverse regressors from real tfMRI data. After that, it utilizes the effective supervised dictionary learning (SDL) method to reconstruct both the task-related functional brain networks and other latent brain networks simultaneously. Extensive experiment results on different tfMRI datasets from Human connectome project (HCP) demonstrated the superiority of the proposed framework.

INDEX TERMS Task fMRI, brain networks, recurrent neural network, supervised dictionary learning.

I. INTRODUCTION

Functional magnetic resonance imaging (fMRI) is one of the most popular noninvasive neuroimaging techniques in the study of neuroscience, experimental psychology, and brain disorders [1]–[3]. Specifically, a large amount of these studies adopted task-based fMRI (tfMRI) paradigm, in which participants are required to perform predefined tasks during the scan sessions [1], [2]. With the help of tfMRI paradigm, the regularity and variability of brain functions have been greatly improved [4]–[6]. After decades of active research, there has been accumulating evidence [5], [7] suggesting that the brain function is realized by multiple diverse and concurrent functional brain networks and these networks are spatially distributed across brain regions [8]. Meanwhile, there have been consistent effort in developing brain network reconstruction and modeling techniques including the general linear model (GLM)[9], [10], principal component analysis (PCA) [11], independent component analysis(ICA) [12], [13], sparse representation based methods(SR) [14]–[18]. Among all of these methods, GLM [9], [10] is the most popular methodology in tfMRI data analysis. The basic idea of this model-driven method is utilizing the hemodynamic response function (HRF) [19]–[21] and the experimental paradigm to construct

The associate editor coordinating the review of this manuscript and approving it for publication was Jiachen Yang.

regressors and make inferences about brain activity patterns. The HRF function describes the haemodynamic response of the instantaneous burst of activation in the fMRI blood oxygen level-dependent (BOLD) signal [19], [20]. However, the HRF is fixed and non-adaptive [22] in the standard model-driven method. Actually, the haemodynamic response across different brain areas exhibits considerable variability in different aspects [23]–[25], such as time-to-peak, response delay, and amplitude. Thus the fixed HRF, a one-size-fits-all strategy, cannot fully adapt to the different brain area response variations, which may result in inaccurate detection performance [23], [25]. Besides, model-driven methods focus on the major task-evoked brain activity patterns and largely ignores other spontaneous brain activities [26].

To overcome these draw backs, data-driven methods have been proposed, including principal component analysis (PCA) [11], independent component analysis(ICA) [12], [13], sparse representation based methods(SR)[18], [27], [28]. These methods do not make any assumptions about the shape of HRF and have been very popular in resting state fMRI analysis. The basic idea of this group methods is that they consider the fMRI time series measured at each voxel as a mixture of brain activity patterns [12], [13], [18]. Therefore, these methods can be characterized as factor models [16] with different specific constrains. For example, the sparsity constraint has derived sparse representation and dictionary learning based methods [18], [27], [28], and the independence constraint has derived ICA based methods [12], [13]. Deep learning based methods [6], [29] have also been proposed for fMRI analysis. Power *et al.* [6] proposed a restricted Boltzmann machine framework to investigate the complex structures in fMRI data. In [29], a deep convolutional autoencoder (DCAE) model was proposed to identify the accurate brain activities in fMRI data. Although these methods greatly helped us improved the understanding of brain networks, possible limitations still existed. There is no correspondence between different components/ dictionaries since they are learned in a purely data-driven way which made it difficult to make comparisons between different subjects and groups.

Recently, recurrent neural networks (RNN) have achieved more and more attention in fMRI data analysis [30]–[32]. An important characteristic of the RNN model is that they make their predictions based on not only the information available at a given time, but also the information that was available in the past from multi-layers which is quite similar with human brain. Güçlü *et al.* proposed a recurrent neural network (RNN) based encoding response model [31], and demonstrated the superiority of the proposed RNN model in estimating temporal dependencies under natural stimulus fMRI data. In [32], a RNN based brain state recognition model is proposed to effectively separate different brain state. In [33], we proposed a deep recurrent neural network (DRNN) model to identify group-wise task-related functional brain networks and brain response patterns in tfMRI data. Although this group-wise activation detection framework is limited in characterizing individual specific brain networks, the identified brain response patterns are diverse and meaningful [33]. These response patterns naturally account for the temporal dependencies of input stimulus and reflect the possible variations of brain response to input stimulus. However, whether these meaningful response patterns could be served as regressors in model-driven method has never been studied.

To improve the brain network detection performance, a few supervised dictionary learning based methods [26], [28], [34] are proposed. In [26], [34], supervised dictionary learning methods were proposed to detect task evoked brain networks and spontaneous brain networks simultaneously. However, these regressors are still simply borrowed from GLM method or generated solely from theory assumptions lacking flexibility and adaptability [26], [34]. Motivated by these impressive works in recurrent neural networks and supervised dictionary learning studies, we proposed a novel hybrid framework, supervised brain network learning based on deep recurrent neural networks (SUDRNN), to reconstruct the diverse and concurrent functional brain networks from tfMRI data. The basic idea of SUDRNN framework is combing the superior ability of RNN to characterize brain response patterns and the ability of supervised learning in brain network reconstruction which largely overcomes limitations in either model-driven or data-driven brain network models. Specifically, we utilized our recently developed deep recurrent neural network model [33] to learn adaptable regressors from real tfMRI data and employed supervised dictionary learning method to identify both task-related and spontaneous brain networks simultaneously. It should be noticed that, although preliminary work [33] demonstrates its ability in data-drivenly estimating group-wise task-related functional brain networks and brain response patterns, DRNN-derived response patterns has never been adopted as regressors for model-driven method. An important characteristic of these DRNN-derived regressors is that they are learned from real tfMRI data and well adapt to the specific brain response variations. After adaptable regressors are obtained, a supervised dictionary learning method (SDL) [26] is employed to reconstruct the spatial distributions corresponding to these regressors and other spontaneous brain network components under tfMRI conditions. Experiments on independent tfMRI datasets have shown that meaningful data-driven regressors can be effectively and robustly derived by with the SUDRNN model and the corresponding spatial maps of these regressors are consistent and meaningful across different subjects. Extensive experiment result demonstrated the superiority of proposed SUDRNN framework in identifying diverse and concurrent functional brain networks from tfMRI data.

II. MATERIALS AND METHODS

A. OVERVIEW

Fig 1 is the flowchart of the proposed SUDRNN framework. Firstly, a deep recurrent neural network (DRNN) is adopted to derive brain response regressors. The DRNN consists of an input layer, two RNN layers, a fully connected layer and

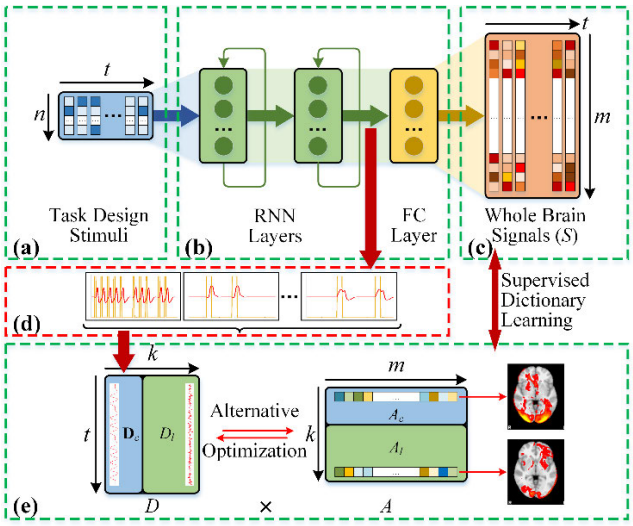


FIGURE 1. Illustration of the proposed SUDRNN framework. (a) Stimulus matrix. (b) Deep recurrent neural network layers. (c) TfMRI signal matrix. (d) Output of the top RNN layer represents the deep recurrent neural network derived regressors. (e) Supervised dictionary learning and sparse representation for identifying spatial maps of functional brain networks. Blue part of the dictionary represents the constrained regressors and green part represents data-driven dictionary items.

an output layer (Fig 1a-c). During the training stage, task paradigm information is gathered into a stimulus matrix (n stimuli with t time series, Fig 1a) and propagated through the model. The whole brain tfMRI signals are aggregated into a big signal matrix (m voxels' signals with t time series, Fig 1c) to optimize the reconstructed whole brain signals. After model convergence, we replaced the stimulus matrix with a testing one that is keeping one selected vector unchanged and setting the others to zeros, and propagated it through the trained model. Then the output of each unit in the top RNN layer represents a typical brain response pattern (e.g., as shown in Fig 1d) which could serve as a regressor in the following analysis. In the standard GLM model, the regressors are constructed by a convolution operation of the task design stimulus with the hypothesized hemodynamic response function (HRF). While in the proposed model, the regressors (e.g., as shown in Fig. 1d) are derived from the output of the top RNN layer through the model that was trained on whole brain tfMRI signals that represents the real data-driven brain activities. The DRNN-derived regressors naturally accounts for the possible brain response variations and may help reveal more diverse and concurrent functional brain networks.

After obtained the diverse and flexible regressors, we adopted a supervised dictionary learning method [26] to simultaneously reconstruct the task-related and spontaneous functional brain networks. We choose supervised dictionary learning method rather than GLM method for several reasons. First, dictionary learning and sparse representation methods usually learn an over-completed dictionary from real datasets and then sparsely selected the most related dictionaries to reconstruct the signals [18], [26]. In such a way, the corresponding coefficient matrix corresponding to the dictionaries has better sensitivity. Besides, the sparse representation is more suitable to the biological findings [35], [36]. What's more, a supervised dictionary learning fashion make it possible to simultaneously identify the task-related and spontaneous brain networks. In the following sections, we will briefly introduce the data acquisition and pre-processing and the basic theories of deep recurrent neural networks and supervised brain network learning based on deep recurrent neural network.

B. DATA ACQUISITION AND PRE-PROCESSING

To advance the understanding of brain structures and functions based on fMRI data, great effort have been devoted in acquiring and processing fMRI data with high spatial-temporal resolution and across different subjects. One of such successful effort is the Human Connectome Project (HCP) [37]–[39]. The HCP project offers seven independent tfMRI datasets (e.g. motor, working memory, emotion, gambling, language, relational, and social) [37]. Emotion, relational, and social tfMRI datasets are adopted as test beds in this study. Briefly, in motor task, subjects are instructed to tap their tap their left or right fingers, squeeze their left or right toes, or move their tongue with visual cue. Participants are required to match emotional faces or different shapes in emotion task. In relational task, subjects are asked to tell the difference between objects. In social task, participants were first presented with a short video and then make a decision about whether the objects had a social interaction or not. More detailed task design information is available in [37].

The detailed acquisition parameters of these HCP tfMRI data are as follows: 220 mm FOV, in-plane FOV: 208 × 180 mm, flip angle = 52, BW = 2290 Hz/Px, 2 × 2 × 2 mm spatial resolution, 90 × 104 matrix, 72 slices, TR = 0.72s, TE = 33.1ms. The preprocessing of the tfMRI datasets includes skull removal, motion correction, slice time correction, spatial smoothing, and global drift removal (high-pass filtering) [40]. All these preprocessing steps were implemented in FSL FEAT [41]. More detailed data acquisition protocol and preprocessing procedures are detailed in literatures [40]. All of these individual fMRI datasets are first registered [42] to the MNI common space for further study. For comparison reason, the GLM activation detection results are also performed using FSL FEAT [41].

C. DEEP RECURRENT NEURAL NETWORK AND DATA-DRIVEN REGRESSORS

Recurrent neural network (RNN) is a category of neural network that has feedback loops or memory cell which help better recognize temporal patterns in data. RNNs have gained great success in speech recognition [43], machine translation [44] and language modeling [45] works. RNNs can make their predictions based on not only the information available at current time points, but also the information at previous time points [46], [47] which quite coincide the characteristic of brain response activities. Actually, brain response proce-

dure is a complex and hierarchical process across multiple time scales [48]. Therefore, we proposed to utilize DRNN model automatically learn the task-based regressors from real tfMRI data. Specifically, RNNs are stacked layer by layer to construct DRNN model and the stacked RNNs automatically creates different time scales and temporal hierarchy as demonstrated in previous studies [46].

We define a DRNN with L layers and n_i neurons per layer. Suppose the input sequence is denoted as $\left(\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \ldots, \boldsymbol{x}^{(t)}\right)$ where each data point is a real-valued vector and the target sequence is denoted as $\left(\boldsymbol{y}^{(1)}, \boldsymbol{y}^{(2)}, \ldots, \boldsymbol{y}^{(t)}\right)$ and the hidden state of i^{th} layer is denoted as $\boldsymbol{h}_i^{(t)}$. In order to simplify notations of the indices of nodes and sequence steps, we use superscripts for time steps and subscripts for layer index. The predicted output of DRNN model is as follows:

$$\hat{\boldsymbol{y}}^t = \sigma(\boldsymbol{V}\boldsymbol{h}_i^t + \boldsymbol{b}_i) \tag{1}$$

where $\hat{\boldsymbol{y}}^t$ is the output of the top hidden layer and $\boldsymbol{V}$ is the weight matrix between the hidden layer and output. $\boldsymbol{b}_i$ is the bias parameters containing the offset of each node.

Among different RNN architectures, the long short-term memory (LSTM) [49] is one of the most popular memory units of RNNs, which has excellent performance in capturing long temporal dependencies. The hidden states in the LSTM unit can be defined as:

$$\boldsymbol{h}^t = \boldsymbol{o}^t \odot \tanh(\boldsymbol{c}^t) \tag{2}$$
$$\boldsymbol{o}^t = \sigma(\boldsymbol{U}_o\boldsymbol{h}^{t-1} + \boldsymbol{W}_o\boldsymbol{x}^t + \boldsymbol{b}_o) \tag{3}$$

where $\boldsymbol{o}^t$ is the output gate activities, $\boldsymbol{c}^t$ is the cell state, and $\odot$ denotes elementwise multiplication. Previous time point's information is stored in the cell state $\boldsymbol{c}^t$ and the output gate controls what information will be retrieved from the cell state. The cell state of an LSTM unit can be further defined as:

$$\boldsymbol{f}^t = \sigma\left(\boldsymbol{U}_f\boldsymbol{h}^{t-1} + \boldsymbol{W}_f\boldsymbol{x}^t + \boldsymbol{b}_f\right) \tag{4}$$
$$\boldsymbol{i}^t = \sigma(\boldsymbol{U}_i\boldsymbol{h}^{t-1} + \boldsymbol{W}_i\boldsymbol{x}^t + \boldsymbol{b}_i) \tag{5}$$
$$\tilde{\boldsymbol{c}}^t = \tanh(\boldsymbol{U}_c\boldsymbol{h}^{t-1} + \boldsymbol{W}_c\boldsymbol{x}^t + \boldsymbol{b}_c) \tag{6}$$
$$\boldsymbol{c}^t = \boldsymbol{f}^t\boldsymbol{c}^{t-1} + \boldsymbol{i}^t\tilde{\boldsymbol{c}}^t \tag{7}$$

where $\boldsymbol{f}^t$ represents the forget gate activities, $\boldsymbol{i}^t$ represents the input gate activities. Forget gate controls what old information will be thrown away from $\boldsymbol{c}^t$. Input gate controls the new information will be kept in the cell states. $\tilde{\boldsymbol{c}}^t$ is an auxiliary variable decides what new candidate values could be added to the state. Besides, $\boldsymbol{U}_f, \boldsymbol{U}_i, \boldsymbol{U}_c$ and $\boldsymbol{W}_f, \boldsymbol{W}_i, \boldsymbol{W}_c$ are the corresponding weights between different layers and $\boldsymbol{b}_f, \boldsymbol{b}_i, \boldsymbol{b}_c$ are the corresponding biases. These parameters determined the behavior of the gates and shared over time.

We also employed the Gated Recurrent Units (GRU) [50] for further study. Different with LSTM, GRU unit combines input gates with forget gates and merges cell states with hidden states. The GRU unit hidden states are defined as follows:

$$\boldsymbol{h}^t = \left(1 - \boldsymbol{z}^t\right) \odot \boldsymbol{h}^{t-1} + \boldsymbol{z}^t \odot \tilde{\boldsymbol{h}}^t \tag{8}$$

$$\boldsymbol{z}^t = \sigma(\boldsymbol{U}_z\boldsymbol{h}^{t-1} + \boldsymbol{W}_z\boldsymbol{x}^t + \boldsymbol{b}_z) \tag{9}$$
$$\boldsymbol{r}^t = \sigma(\boldsymbol{U}_r\boldsymbol{h}^{t-1} + \boldsymbol{W}_r\boldsymbol{x}^t + \boldsymbol{b}_r) \tag{10}$$
$$\tilde{\boldsymbol{h}}^t = \tanh\left(\boldsymbol{U}_h\left(\boldsymbol{r}^t * \boldsymbol{h}^{t-1}\right) + \boldsymbol{W}_h\boldsymbol{x}^t + \boldsymbol{b}_h\right) \tag{11}$$

where $\boldsymbol{z}^t$ represents update gate activities, $\boldsymbol{r}^t$ represents reset gate activities and $\tilde{\boldsymbol{h}}^t$ is an auxiliary variable. Similar with the parameters in LSTM units, $\boldsymbol{U}_z, \boldsymbol{U}_r, \boldsymbol{U}_h$ and $\boldsymbol{W}_z, \boldsymbol{W}_r, \boldsymbol{W}_h$ represent the corresponding weights between layers and $\boldsymbol{b}_z, \boldsymbol{b}_r, \boldsymbol{b}_h$ represent the biases and all these parameters together determined the behavior of the gates.

During the training stage, the task paradigm information is separated in different time points and added into the model iteratively. At the top hidden layer, the output is connected to the whole brain signals through a fully connected layer. Thus, the DRNN model forms a dynamic simulation model of brain activities. After model convergence, we replaced the stimulus matrix with a testing one that is keeping one selected vector unchanged and setting the others to zeros, and propagated it through the trained model. The output of in the top RNN layer represents the possible regressors corresponding different stimulus information. In this way, we can obtain a group of regressors $\boldsymbol{R}$ learned from real tfMRI data.

D. SUPERVISED BRAIN NETWORK LEARNING BASED ON DRNN

After learned the meaningful and adaptive regressors from real tfMRI data, we employed the supervised dictionary learning method [26] to identify the diverse and concurrent functional brain networks in tfMRI data, thus forming a supervised brain network learning based on deep recurrent neural networks (SUDRNN) framework. The basic idea of SUDRNN is to learn a hybrid dictionary through changing the dictionary learning procedure and then simultaneously reconstruct the task related brain networks and intrinsic brain networks utilizing the hybrid dictionary. Specifically, the learned regressors from DRNN model is fixed as a constant part of the dictionary and the other part is automatically learned from the tfMRI data. Importantly, the SDL method can utilize the most suitable dictionary item/ regressors to reconstruct the tfMRI data which improved the sensitivity.

Given a tfMRI signal matrix $\boldsymbol{S}\epsilon\mathbb{R}^{t\times n}$ and the learned regressor group $\boldsymbol{R}$, SUDRNN aims to represent each signal in $\boldsymbol{S}$ with a sparse linear combination of atoms in an over-completed, and hybrid dictionary basis $\boldsymbol{D}$, i.e., $= \boldsymbol{D} \times \alpha_i$ and $\boldsymbol{S} = \boldsymbol{D} \times \boldsymbol{A}$. Specifically, t rerpresents the number of fMRI time points and n represents the tfMRI signals from n voxels, and $\boldsymbol{A} = (\alpha_1, \alpha_2, \ldots.\alpha_n)$ is the coefficient matrix. Specifically,

$$\boldsymbol{D} = [\mathbf{D}_c, \boldsymbol{D}_l]\,\epsilon\mathbb{R}^{t\times k}, \quad \mathbf{D}_c\epsilon\mathbb{R}^{t\times k_c},\ \boldsymbol{D}_l\epsilon\mathbb{R}^{t\times k_l} \tag{12}$$

where $\mathbf{D}_\mathbf{c}$ is the learned regressor group $\boldsymbol{R}$ from DRNN model and $\boldsymbol{D}_l$ is the random generated dictionary atoms. During dictionary learning procedure, $\mathbf{D}_\mathbf{c}$ will be fixed and $\boldsymbol{D}_l$ will automatically update to fit to the data. k_c is the fixed atom number in $\mathbf{D}_c$ and k_l is the learned dictionary atom number

in $\boldsymbol{D}_l$, respectively. In SUDRNN, the empirical cost function could be modeled as the averaging loss of regression of n signals.

$$\ell\left(s_i, [\mathbf{D}_c, \boldsymbol{D}_l]\right) \triangleq \frac{1}{2}||s_i - [\mathbf{D}_c, \boldsymbol{D}_l]\,\alpha_i||_2^2 + \lambda||\alpha_i||_1 \quad (13)$$

$\ell\left(s_i, [\mathbf{D}_c, \boldsymbol{D}_l]\right)$ represents the reconstruction error in sparse representation of the signals. s_i represents the signal vector and α_i represents the reconstruction weight vector. The regularization parameter λ defines the regression residual and sparsity level.

To prevent $\boldsymbol{D}$ from arbitrary value, the following constrains C is introduced in (15) and the whole problem can be summarized in (16).

$$C \triangleq \left\{\boldsymbol{D}\epsilon\mathbb{R}^{t\times k} s.t\ \forall j{=}\ 1, \ldots k, d_j^T d_j \leq 1\right\} \quad (14)$$

$$\underset{\boldsymbol{D}\in C, \boldsymbol{A}\in\mathbb{R}^{k\times n}}{min} \frac{1}{2}||\boldsymbol{S} - [\mathbf{D}_c, \boldsymbol{D}_l]\,\boldsymbol{A}||_2^2 + \lambda||\boldsymbol{A}||_{1,1} \quad (15)$$

To solve this problem, we modified the codes in the online dictionary learning package [51] and form the SUDRNN framework in this paper. The proposed SUDRNN framework is summarized as Alg.1.

Algorithm 1 SUDRNN Framework

Input: $\boldsymbol{S} = (s_1, s_2, \ldots, s_n)\epsilon\mathbb{R}^{t\times n}$, number of iteration T.
Stage 1: Learned the adaptive regressor group $\boldsymbol{R}$ using DRNN model.
Stage 2: Generate the initial dictionary $\boldsymbol{D}_0 = [\mathbf{D}_c, \boldsymbol{D}_{l0}]\,\epsilon\mathbb{R}^{t\times k}$ ($\mathbf{D}_c$ is the DRNN derived regressors $\boldsymbol{R}$, $\boldsymbol{D}_{l0}$ is randomly generated matrix.
Stage 3:
for z = 1 to T
i = z%n (T > n)
Draw s_i from S
Sparse coding using Least-angle Regression:

$$\alpha_z \triangleq \underset{\alpha_{z}\in\mathbb{R}^{m}}{\text{argmin}} \tfrac{1}{2}\left|\left|s_z - \mathbf{D}_{(z-1)}\alpha_z\right|\right|_2^2 + \lambda\,||\alpha_z||_1$$

Update $\boldsymbol{D}_{(z)}$ but keep the extended regressor group $\mathbf{D}_c$ constant:

$$D_z \triangleq \underset{D_z\in C}{\text{argmin}} \tfrac{1}{2}\,||s_z - \boldsymbol{D}\alpha_z||_2^2 + \lambda\,||\alpha_z||_1$$
$$(\boldsymbol{D}_{(z-1)}{=}(\mathbf{D}_c, \boldsymbol{D}_{l(z-1)}))$$

end for
until convergence
Return $\boldsymbol{D}$ and $\boldsymbol{A}$.

E. IDENTIFICATION OF FUNCTIONAL BRAIN NETWORKS

After dictionary learning and sparse representation, each column in dictionary $\boldsymbol{D}$ represents a typical brain activity patterns and the corresponding row vector in coefficient matrix $\boldsymbol{A}$ represents the spatial distribution of the dictionary item. With the help of preserved index information, each row vector in coefficient matrix $\boldsymbol{A}$ could be mapped back into the volume space as shown in Fig.1e. Since the dictionary $\boldsymbol{D}$ could be separated as constant part $\mathbf{D}_c$ and learned part $\boldsymbol{D}_l$, $\boldsymbol{A}$ naturally divided into task related brain networks $\boldsymbol{A}_c$ corresponding to fixed DRNN derived regressors $\mathbf{D}_c$ and data-driven networks $\boldsymbol{A}_l$ corresponding to $\boldsymbol{D}_l$. As for fixed DRNN derived regressors $\mathbf{D}_c$, it is straightforward to map all the task related brain networks from $\boldsymbol{A}_c$ for each subject in each tfMRI data.

These data-driven networks $\boldsymbol{A}_l$ are learned in an unsupervised way and it is hard to direct interpret these networks from $\boldsymbol{A}_l$. To solve this problem, a spatial matching method is employed to compare the similarity between the well-established intrinsic connectivity network (ICN) templates [52] in the literature and the row vectors in $\boldsymbol{A}_l$ to detect meaningful intrinsic connectivity networks. The spatial similarity is calculated as:

$$R\,(X, T) = \frac{|X \cap T|}{|T|} \quad (16)$$

T is the ICN template and X is the row vector in $\boldsymbol{A}_l$.

III. RESULTS

A. DRNN MODEL IMPLEMENTATION

In order to test the proposed SUDRNN framework, we adopted the well-established tfMRI datasets of HCP 900 subjects release as a test bed. To be specific, there are 822 subjects after removing a few poor quality or incomplete subjects. To further verify the stability of the model, the motor tfMRI dataset of HCP 900 Subjects Release was randomly divided into two subsets, each containing 411 valid individuals for model implementation test. Then the same DRNN model was trained independently in each dataset. In our implementation, the DRNN model consists of an input layer, two RNN layer, an output layer and a fully connected layer. Each RNN layer includes 30 LSTM units. The LSTM cells were initialized by the default Xavier uniform initializer [53] in TensorFlow. All the cell initial state, the weight and bias of the fully connected layer were initialized to zeros. The learning rate was set to 0.004, and the decay factor is 0.25 every five epochs. The mean square error (MSE) between the whole brain signals and its reconstruction was adopted as the loss function and the Adam optimizer [54] with its default parameters ($\beta 2 = 0.999, \epsilon{=}\ 10{-}8, 1 = 0.9$) was applied for optimizing the parameters. The DRNN model was trained on dual NVIDIA GTX 1080Ti GPU cards for 20 epochs.

Fig 2 shows the training loss on the motor tfMRI datasets of HCP 900 subjects release and its split subsets. Clearly, the DRNN model on both the whole HCP 900 datasets and the split datasets will reach convergence at a small value after a few epochs. Besides, the training loss on independent split datasets were almost the same and the results are quite similar which implied the SUDRNN model is robust and reproducible. However, more training data (the whole dataset) will slightly improve the convergence speed and avoid overfitting better. Therefore, the following results are obtained from the whole HCP 900 release datasets with 822 subjects.

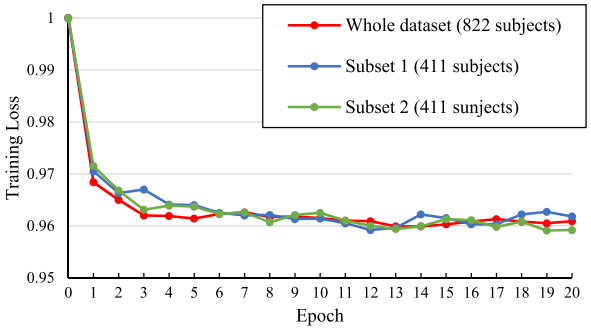


FIGURE 2. Training losses on the whole dataset of HCP 900 subjects release (822 subjects), and two split subsets of HCP 900 subjects release with half data (411 subjects).

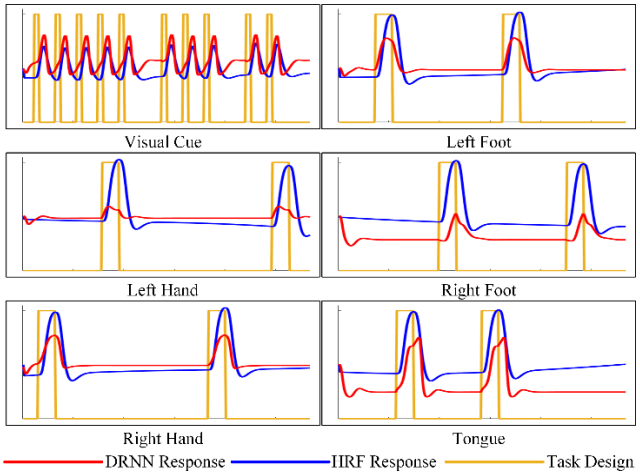


FIGURE 3. Parts of the DRNN-derived data-driven regressors in motor tfMRI dataset. Yellow lines represent task design stimuli, red lines represent DRNN-derived data-driven regressors, and blue lines represent the response patterns by GLM method.

B. IDENTIFIED TYPICAL REGRESSORS AND FUNCTIONAL BRAIN NETWORKS

After DRNN model training and testing procedure, diverse and concurrent task regressors could be obtained from the top RNN layer. In order to explore meaningful functional brain activities, a few temporal responses which are similar to the HRFs were selected for further study as DRNN-derived data-driven regressors. Fig 3 illustrates part of the regressors identified in the motor tfMRI dataset of HCP 900 subjects release. From this figure, we can see that these regressors are quite similar with the typical theoretical HRF response regressors adopted in GLM model and typical HRF response regressors could be easily identified in DRNN model.

After the DRNN-derived data-driven regressors were obtained, we applied the supervised dictionary learning and sparse coding method by fixing the pre-defined part of the dictionary to learn the spatial maps of these regressors. Fig 4 shows two examples of the learned spatial maps of these regressors with dictionary size $k = 400$ and regularization parameter $\lambda = 0.05$. In this figure, the first row shows the visual cue and left foot regressors from motor task which

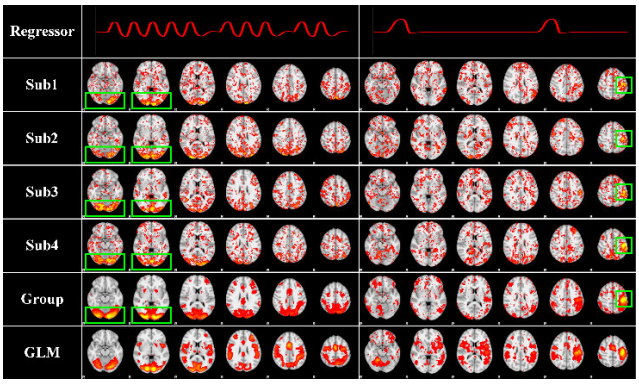


FIGURE 4. Examples of learned individual and group spatial maps by the DRNN-derived data-driven regressors. Regressors are the learned visual cue and left foot regressor from motor tfMRI dataset.

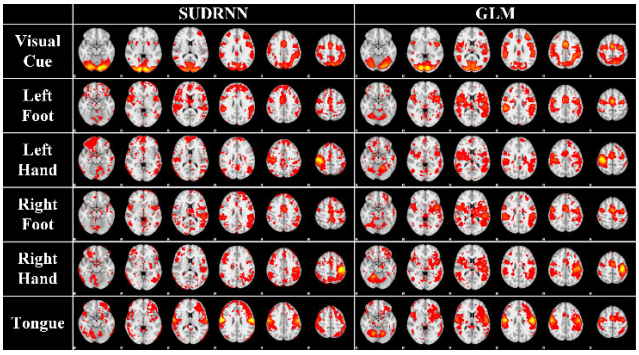


FIGURE 5. Identified typical functional brain networks by SUDRNN and GLM in HCP motor tfMRI dataset. The corresponding regressors are illustrated in Fig 3.

are randomly selected as constrained regressors. The second to the fifth rows illustrate the identified functional brain networks in four randomly selected subjects in motor tfMRI dataset. After that, the sixth row is the group activation map of these regressors in motor tfMRI dataset. From these result, we can see that with the help of these supervised regressors, meaningful and consistent functional brain networks could be well identified by SDL method. They are consistent across different subjects especially in the peak activated areas (areas highlighted in green box). For comparison, we also calculated the group-wise GLM activation maps in the last row. The group activation maps calculated by SUDRNN and GLM are quite similar and the peak activated areas are well matched. In Fig 5, we illustrated all the functional brain networks identified by the proposed SUDRNN framework and GLM method with these typical regressors. These identified functional brain networks with these regressors in both methods are similar and well matched which suggests the SUDRNN method could identify the typical functional brain networks as well as traditional GLM model.

C. DIVERSE BRAIN NETWORKS WITH DRNN-DERIVED REGRESSORS

In our proposed SUDRNN framework, regressors are the output of units in the top RNN layer of DRNN model and

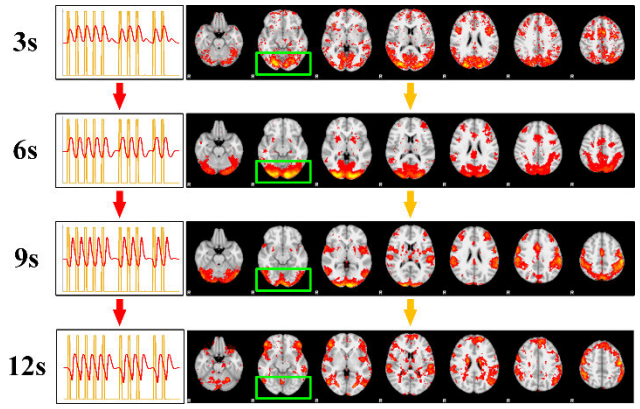


FIGURE 6. Visual cue related regressors with multiple times delay and corresponding functional brain networks in HCP motor task.

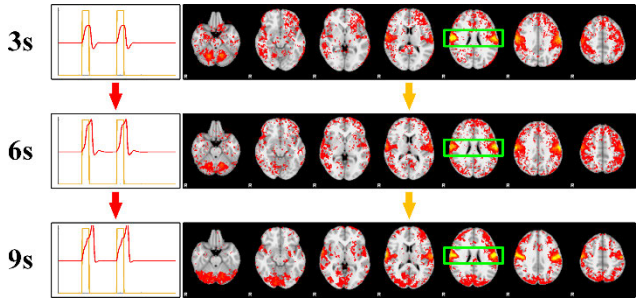


FIGURE 7. Tongue related regressors with multiple times delay and corresponding brain networks in HCP motor task.

the DRNN model is trained with the real tfMRI data and stimulus task design paradigm. Thus, these regressors naturally reflects the real diverse and concurrent brain activity patterns under task condition. Specifically, benefited from the excellent temporal pattern recognition ability of RNN layers, SUDRNN could not only identify typical theoretical HRF response regressors and brain networks as detailed in above section but also multiple times delay regressors and brain networks simultaneously. For instance, in motor tfMRI dataset of HCP 900 subjects release, a few DRNN-Derived regressors consists of multiple delay times. Fig 6 shows the identified DRNN-derived visual cue related regressors with different time delays in left column and the corresponding spatial maps in right column. Similarly, the left column of Fig 7 illustrates the identified DRNN-derived tongue related regressors and the right column shows the corresponding spatial maps. An interesting result could be observed. As shown in of Fig 6 and Fig 7, part of the task design stimuli (such as visual cue and tongue stimuli) may evoke different time delay brain activity patterns. The corresponding brain networks of these activity patterns keep relatively stable within a period of time. However, the peak activated areas may decrease and transform to adjacent areas as the time delay goes forward. This observation is benefit from the data-driven regressors reflect the complex real brain activity patterns and it's hard to find with previous model-driven methods.

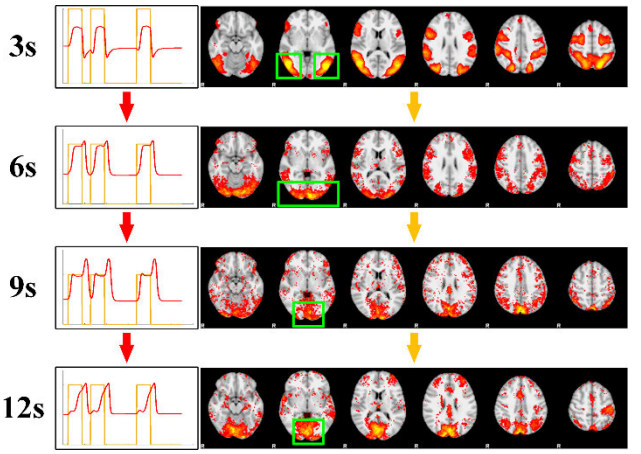


FIGURE 8. Social event related regressors with multiple times delay and corresponding brain networks in HCP social task.

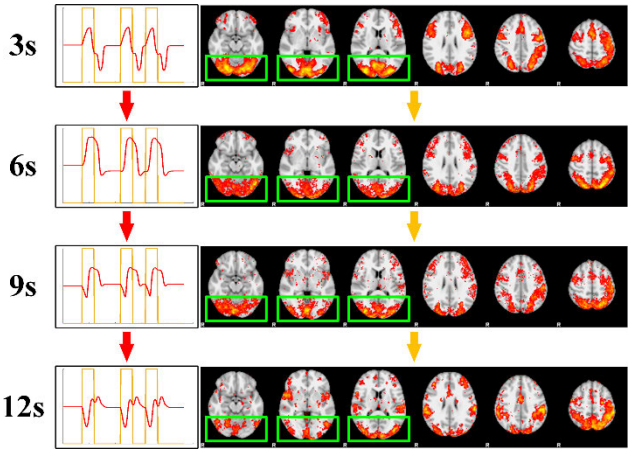


FIGURE 9. Relational event related regressors with multiple times delay and corresponding brain networks in HCP relational task.

To test the universality of this phenomenon, we further test the proposed SUDRNN framework on other independent HCP tfMRI datasets including the social, relational and emotion tfMRI datasets. The learned data-driven regressors and spatial maps from HCP social and relational tfMRI datasets are illustrated in Fig 8, and Fig 9, respectively. As expected, different time delay brain activity patterns are observed and these regressors are also correlated with the theoretical HRF response patterns. Similarly, the activated areas in these spatial maps also vary and transform as time goes forward. These results further demonstrated the proposed SUDRNN method is robust and reproducible across different tfMRI datasets and can reveal meaningful task-related functional brain networks which could not be observed with previous methods. More importantly, data-driven regressors from DRNN model provide us new insight to comprehensively understand the task-related functional brain networks and the possible interactions between them.

D. IDENTIFIED CONCURRENT SPONTANEOUS BRAIN NETWORKS

With the help of the data-driven part dictionary D_l, the proposed SUDRNN framework is able to simultaneously

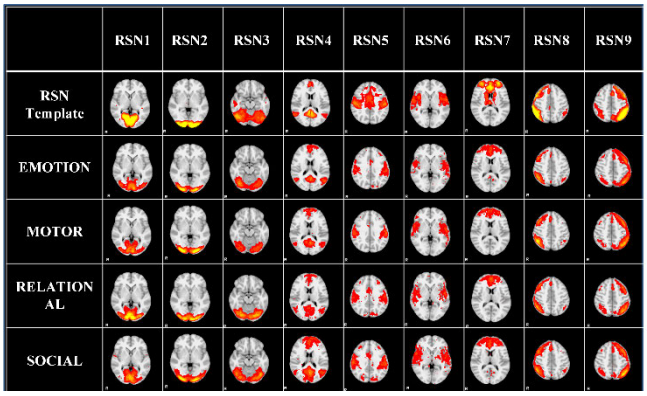


FIGURE 10. The identified RSN networks in adopted HCP 900 subjects release tfMRI datasets.

identify the concurrent spontaneous brain networks or resting state networks (RSN) in tfMRI data. However, these dictionary items are learned in an unsupervised way which made it difficult to group-wisely interpret the brain activity patterns from D_l. To solve this problem, we adopted the well-established RSN templates [52] to aid the analysis. Specifically, we calculated the spatial overlap rate between the corresponding spatial maps of these data-driven dictionary items in D_l and the well-established RSN templates in the literature [52]. The spatial overlap rate is defined as the percentage of the overlapping area between the data-driven spatial maps and the RSN templates which characterized in Eq. (16). We examined the all the data-driven dictionary items in D_l across all the involved HCP tfMRI datasets and consistently identified these RSN networks in different HCP tfMRI datasets. Fig 10 shows the identified nine consistent RSN networks in adopted HCP 900 subjects release tfMRI datasets and the RSN templates are put in the second row for comparison. For better visualization result, only the most informative slice overlaid on the MNI152 template image, is shown as the spatial map of the specific RSN. It is easy to see that the identified group-wise RSNs in different tfMRI datasets are quite similar with RSN templates. We further quantitatively measured the spatial overlap rate between the identified RSNs in different tfMRI datasets and the RSN template and the quantitative measurement result is shown in TABLE1. As expected, the spatial overlap value is kept at a high level across different tasks and RSNs. From these result, we can see that the identified spontaneous brain networks (RSNs) is consistent and similar to the RSN templates across different tasks from both spatial patterns and quantitative measurements. Our results demonstrated that SUDRNN framework is capable of detecting meaningful spontaneous functional brain networks as well as task-related functional brain networks at the same time.

E. PARAMETER TUNING OF PROPOSED SUDRNN FRAMEWORK

The proposed SUDRNN framework could be divide into a DRNN model and a supervised dictionary learning part. Therefore, parameters in these models may have influence on

TABLE 1. Spatial overlap rate of the RSN templates and the group averaged RSNs identified in HCP 900 subjects release tfMRI datasets.

	EMOTION	MOTOR	RELATIONAL	SOCIAL
RSN1	0.90	0.90	0.89	0.87
RSN2	0.88	0.86	0.87	0.89
RSN3	0.90	0.90	0.90	0.90
RSN4	0.82	0.86	0.78	0.82
RSN5	0.72	0.83	0.81	0.84
RSN6	0.89	0.85	0.68	0.84
RSN7	0.91	0.91	0.75	0.79
RSN8	0.83	0.84	0.82	0.81
RSN9	0.80	0.85	0.87	0.79

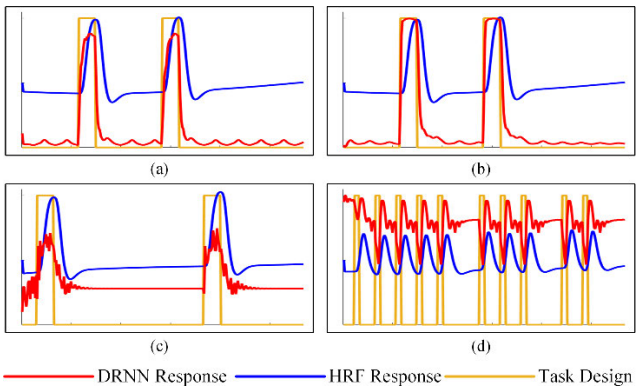


FIGURE 11. Brain response patterns with basic RNN units.

the SUDRNN model. In this section, we will briefly discuss a few important parameters including the RNN unit type, the sparsity parameter λ, and the dictionary size k.

1) PARAMETERS IN DRNN MODEL

An important parameter in DRNN model is the RNN cell unit. In this section, the effect of different RNN cell units, which includes the basic RNN unit, LSTM and GRU unit, to the performance of the model are compared through experiments. The basic RNN unit is relatively limited to reconstruct whole-brain tfMRI signals due to its weakness in capture temporal patterns [46]. This is because that the basic RNN unit makes predictions only depending on the input signal at the current moment and the previous moment, without the ability to perceive the long-term dependent information. Fig 11 illustrates the major results of basic RNN unit on motor tfMRI dataset. Briefly, most of the brain response activity patterns extracted by the basic RNN unit will closely follow the task stimulus curve (Fig 11(a, b)), and the time-delay brain response regressors are rarely observed. Besides, a few brain response curves extracted by the basic RNN unit has obvious jitter and burr (Fig 11 (c, d)) in change points, making it difficult to extract smooth response activity curve. A possible explanation is that the basic RNN unit can't effectively deal with long-term information in the tfMRI data.

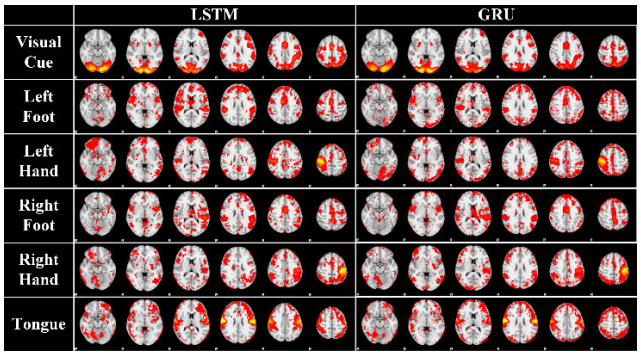


FIGURE 12. Identified typical functional Brain networks using LSTM and GRU units.

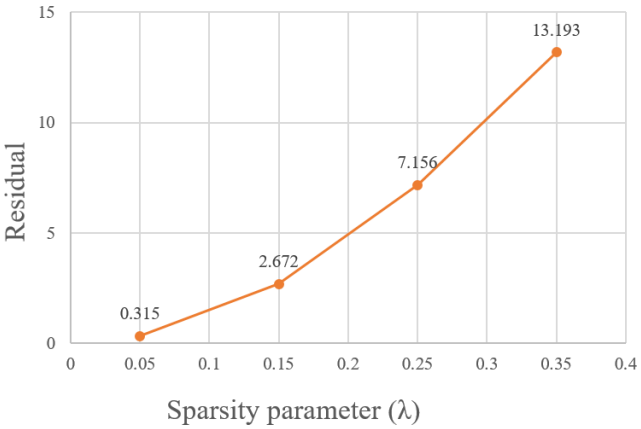


FIGURE 13. Regression residuals under different regularization parameters λ.

Conversely, the LSTM unit and the GRU unit, which are extensions to the basic RNN unit and are good at capturing the long-term dependent information in sequential signals, have similar and good performance. As shown in Fig. 12, the typical functional brain networks identified by LSTM and GRU unit in motor tfMRI dataset are quite similar and reproducible. Therefore, considering the good performance of LSTM unit, we selected the LSTM unit as the RNN units in our implementation.

2) PARAMETERS IN SDL MODEL

Two important parameters in SDL model are the sparsity parameter λ and dictionary size k. However, how to optimize these parameters is still an open question in dictionary learning and sparse coding field. In our implementation, we set these parameters experimentally. Figure 13 shows the mean training residuals of different regularization parameters λ on the HCP motion tfMRI dataset with the fixed dictionary size 400. The residuals start to raise quickly from the 0.05. Regularization parameters λ are mainly used to control the sparsity of the coefficient matrix. The larger the λ value, the greater the influence of regularization terms in Eq. (15) on the loss function and the larger the regression residual of training. The larger the λ value, the sparser the coefficient

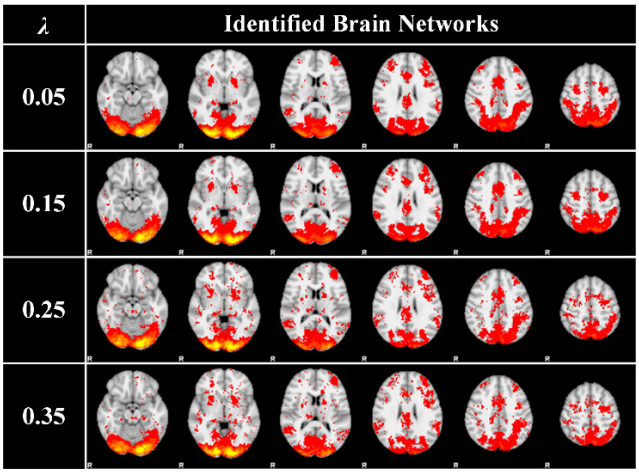


FIGURE 14. Identified functional brain networks under different regularization parameters λ. The visual cue regressor corresponding network in motor tfMRI is adopted as an example.

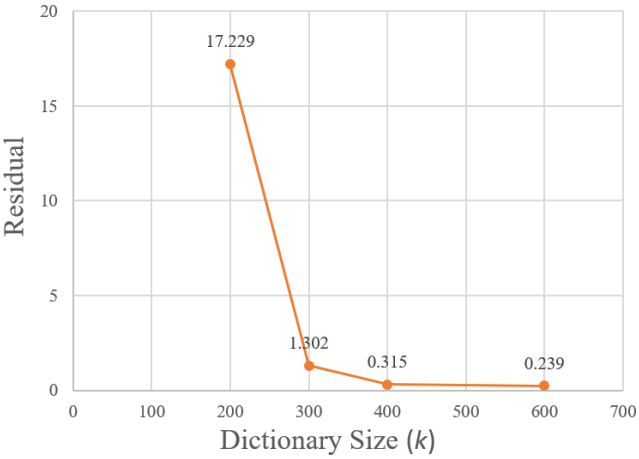


FIGURE 15. Regression residuals under different dictionary size k in motor tfMRI dataset.

matrix and the sparser of the voxels in the identified functional brain networks. However, the within a certain range of λ settings, the identified functional brain networks are still kept consistent as shown in Fig. 14. This also demonstrated the SUDRNN is robust and reproducible with a range of parameter settings. After considering the reconstruction residual and the voxel density in brain networks, we set the sparsity parameter λ as 0.05 in our implement.

In another experiment, we fixed the sparsity parameter λ as 0.05, and alternated the dictionary size from 200 to 600. Fig. 15 shows the reconstruction residuals in different dictionary size under motor tfMRI datasets. From this figure, we can see that the reconstruction residuals drop quickly from 200 to 400 and then become relatively stable. We also test the influence to the identified functional brain networks and take the visual cue regressor in motor tfMRI as an example. The identified brain networks with different dictionary size settings are shown in Fig. 16. From Fig. 16, we can see that with the increase of dictionary size, the identified functional brain networks become sparser. However, the identified functional

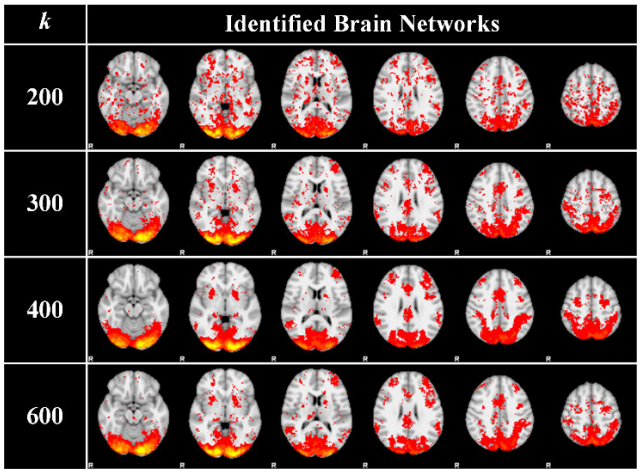


FIGURE 16. Identified functional brain networks under different dictionary size k. The visual cue regressor corresponding network in motor tfMRI is adopted as an example.

brain networks are relatively stable with these dictionary size settings. In order to avoid introducing too much redundant information or extracting too sparse functional brain networks, we set the dictionary size as 400.

IV. DISCUSSION AND CONCLUSION

In this work, we proposed a novel hybrid framework, supervised brain network learning based on deep recurrent neural networks (SUDRNN), to explore the diverse and concurrent functional brain networks for tfMRI data. Specifically, we proposed to adopt deep recurrent neural network (DRNN) to automatically learn the data-driven regressors. These learned regressors not only includes typical theoretical HRF response regressors (as shown in Fig 3) adopted in traditional GLM model, but also multiple time delay regressors which holistically revealed the possible task-related regressors (such as Fig 6 to Fig 9)in tfMRI data. After that, an effective supervised dictionary learning and sparse representation method [26]was adopted to reconstruct the corresponding spatial maps of these regressors. Extensive experiments on the motor tfMRI of HCP 900 subjects release (more than 800 subjects) have demonstrated the superiority of the proposed framework in identifying diverse and concurrent brain networks. Briefly, the proposed framework could not only identify typical task evoked brain networks like traditional GLM models, but also many other concurrent brain networks such as different time delay networks and RSN brain networks simultaneously. Compared with traditional methods, the proposed SUDRNN framework could identify more complete task-related functional brain networks and these network related regressors are automatically learned from the real tfMRI data. What's more, with the help of these task related regressors, it is easy to analyze the task related brain networks across different subjects. We further test the proposed SUDRNN framework on three different HCP tfMRI datasets (social, relational and motor task). Experiment results further validated the robustness and superiority of the proposed model. In general, our proposed framework can identify the diverse and concurrent functional brain networks and contributed novel insights into the holistic functional brain activities under task performance.

In this paper, we have focused on estimating meaningful task-related regressors and identifying more completed functional brain networks with the proposed SUDRNN framework. However, there are still a few questions need to be solved in the further exploration. For instance, although we have demonstrated LSTM and GRU cell unit is robust and reliable to learn the meaningful task-related regressors, more automatic parameter setting methods such as Neural Architecture Search [55] should be further explored to better tuning the model parameters. The parameters in dictionary learning part, such as the sparsity parameter λ and dictionary size k, also need to be determined each time. More automatically parameter tuning methods should be developed in the future. Besides, we have focused on characterization of the theoretical HRF related regressors and well-established RSN brain networks with the limited neuroscience knowledge. However, it should be noticed that there are still other potentially meaningful regressors and spontaneous brain network components need to be examined in the future. Furthermore, although tfMRI datasets of HCP 900 subject release are excellent test beds for brain image analysis methods, we should adopt more datasets and methods to further validate our methods.

In general, we have proposed a novel and hybrid tfMRI brain network analysis framework. This framework could automatically learn the task-related brain response regressors rather than theoretically estimate the task regressors. With the proposed SUDRNN framework, a few meaningful multiple time-delay regressors are revealed and corresponding functional brain networks are identified which help us better understand the functional brain activity patterns under task performance. Besides, both spontaneous brain networks and task-related brain networks could be simultaneously identified which makes it a comprehensive tfMRI brain network analysis framework. Motivated by these exciting results, we will apply it to more tfMRI datasets to learn the possible alterations of functional brain networks and interactions in brain disorders, such as for Alzheimer's disease and Autism Spectrum Disorder.

REFERENCES

[1] K. J. Friston, "Modalities, modes, and models in functional neuroimaging," *Science*, vol. 326, no. 5951, pp. 399–403, Oct. 2009.
[2] N. K. Logothetis, "What we can do and what we cannot do with fMRI," *Nature*, vol. 453, no. 7197, pp. 869–878, Jun. 2008.
[3] C. J. Stam and J. C. Reijneveld, "Graph theoretical analysis of complex networks in the brain," *Nonlinear Biomed. Phys.*, vol. 1, no. 1, pp. 1–3, Dec. 2007.
[4] Y. Zhao, Q. Dong, H. Chen, A. Iraji, Y. Li, M. Makkie, Z. Kou, and T. Liu, "Constructing fine-granularity functional brain network atlases via deep convolutional autoencoder," *Med. Image Anal.*, vol. 42, pp. 200–211, Dec. 2017.
[5] J. D. Power, A. L. Cohen, S. M. Nelson, G. S. Wig, K. A. Barnes, J. A. Church, A. C. Vogel, T. O. Laumann, F. M. Miezin, B. L. Schlaggar, and S. E. Petersen, "Functional network organization of the human brain," *Neuron*, vol. 72, no. 4, pp. 665–678, Nov. 2011.
[6] X. Hu, H. Huang, B. Peng, J. Han, N. Liu, J. Lv, L. Guo, C. Guo, and T. Liu, "Latent source mining in FMRI via restricted Boltzmann machine," *Hum. Brain Mapping*, vol. 39, no. 6, pp. 2368–2380, Jun. 2018.

[7] M. D. Fox, A. Z. Snyder, J. L. Vincent, M. Corbetta, D. C. V. Essen, and M. E. Raichle, "The human brain is intrinsically organized into dynamic, anticorrelated functional networks," *Proc. Nat. Acad. Sci. USA*, vol. 102, pp. 9673–9678, May 2005.

[8] E. Bullmore and O. Sporns, "Complex brain networks: Graph theoretical analysis of structural and functional systems," *Nature Rev. Neurosci.*, vol. 10, no. 3, pp. 186–198, Mar. 2009.

[9] K. J. Friston, A. P. Holmes, K. J. Worsley, J.-P. Poline, C. D. Frith, and R. S. J. Frackowiak, "Statistical parametric maps in functional imaging: A general linear approach," *Hum. Brain Mapping*, vol. 2, no. 4, pp. 189–210, 1994.

[10] K. J. Friston, P. Fletcher, O. Josephs, A. Holmes, M. D. Rugg, and R. Turner, "Event-related fMRI: Characterizing differential responses," *NeuroImage*, vol. 7, no. 1, pp. 30–40, Jan. 1998.

[11] R. Viviani, G. Grön, and M. Spitzer, "Functional principal component analysis of fMRI data," *Hum. Brain Mapping*, vol. 24, no. 2, pp. 109–129, Feb. 2005.

[12] M. J. McKeown and T. J. Sejnowski, "Independent component analysis of fMRI data: Examining the assumptions," *Hum. Brain Mapping*, vol. 6, nos. 5–6, pp. 368–372, 1998.

[13] V. D. Calhoun, T. Adali, G. D. Pearlson, and J. J. Pekar, "Spatial and temporal independent component analysis of functional MRI data containing a pair of task-related waveforms," *Hum. Brain Mapping*, vol. 13, no. 1, pp. 43–53, May 2001.

[14] K. Lee, S. Tak, and J. Chul Ye, "A data-driven sparse GLM for fMRI analysis using sparse dictionary learning with MDL criterion," *IEEE Trans. Med. Imag.*, vol. 30, no. 5, pp. 1076–1089, May 2011.

[15] Y.-B. Lee, J. Lee, S. Tak, K. Lee, D. L. Na, S. W. Seo, Y. Jeong, and J. C. Ye, "Sparse SPM: Group sparse-dictionary learning in SPM framework for resting-state functional connectivity MRI analysis," *NeuroImage*, vol. 125, pp. 1032–1045, Jan. 2016.

[16] A.-K. Seghouane and A. Iqbal, "Sequential dictionary learning from correlated data: Application to fMRI data analysis," *IEEE Trans. Image Process.*, vol. 26, no. 6, pp. 3002–3015, Jun. 2017.

[17] A.-K. Seghouane and A. Iqbal, "Basis expansion approaches for regularized sequential dictionary learning algorithms with enforced sparsity for fMRI data analysis," *IEEE Trans. Med. Imag.*, vol. 36, no. 9, pp. 1796–1807, Sep. 2017.

[18] J. Lv, X. Jiang, X. Li, D. Zhu, S. Zhang, S. Zhao, H. Chen, T. Zhang, X. Hu, J. Han, J. Ye, L. Guo, and T. Liu, "Holistic atlases of functional networks and interactions reveal reciprocal organizational architecture of cortical function," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 4, pp. 1120–1131, Apr. 2015.

[19] G. H. Glover, "Deconvolution of impulse response in event-related BOLD fMRI1," *NeuroImage*, vol. 9, no. 4, pp. 416–429, Apr. 1999.

[20] R. L. Buckner, "Event-related fMRI and the hemodynamic response," *Hum. Brain Mapping*, vol. 6, nos. 5–6, pp. 373–377, 1998.

[21] M. S. Cohen, "Parametric analysis of fMRI data using linear systems methods," *NeuroImage*, vol. 6, no. 2, pp. 93–103, Aug. 1997.

[22] A. Veloz, C. Moraga, A. Weinstein, L. Hernandez-Garcia, S. Chabert, R. Salas, R. Riveros, C. Bennett, and H. Allende, "Fuzzy general linear modeling for functional magnetic resonance imaging analysis," *IEEE Trans. Fuzzy Syst.*, vol. 28, no. 1, pp. 100–111, Jan. 2020.

[23] D. A. Handwerker, J. Gonzalez-Castillo, M. D'Esposito, and P. A. Bandettini, "The continuing challenge of understanding and modeling hemodynamic variation in fMRI," *NeuroImage*, vol. 62, no. 2, pp. 1017–1023, Aug. 2012.

[24] J. J. Purcell and B. Rapp, "Local response heterogeneity indexes experience-based neural differentiation in reading," *NeuroImage*, vol. 183, pp. 200–211, Dec. 2018.

[25] M. A. Lindquist, J. M. Loh, L. Y. Atlas, and T. D. Wager, "Modeling the hemodynamic response function in fMRI: Efficiency, bias and mismodeling," *NeuroImage*, vol. 45, no. 1, pp. S187–S198, Mar. 2009.

[26] S. Zhao, J. Han, J. Lv, X. Jiang, X. Hu, Y. Zhao, B. Ge, L. Guo, and T. Liu, "Supervised dictionary learning for inferring concurrent brain networks," *IEEE Trans. Med. Imag.*, vol. 34, no. 10, pp. 2036–2045, Oct. 2015.

[27] J. Lv, X. Jiang, X. Li, D. Zhu, S. Zhao, T. Zhang, X. Hu, J. Han, L. Guo, Z. Li, C. Coles, X. Hu, and T. Liu, "Assessing effects of prenatal alcohol exposure using group-wise sparse representation of fMRI data," *Psychiatry Res., Neuroimaging*, vol. 233, no. 2, pp. 254–268, Aug. 2015.

[28] J. Lv, B. Lin, Q. Li, W. Zhang, Y. Zhao, X. Jiang, L. Guo, J. Han, X. Hu, C. Guo, J. Ye, and T. Liu, "Task fMRI data analysis based on supervised stochastic coordinate coding," *Med. Image Anal.*, vol. 38, pp. 1–16, May 2017.

[29] H. Huang, X. Hu, Y. Zhao, M. Makkie, Q. Dong, and S. Zhao, "Modeling task fMRI data via deep convolutional autoencoder," *IEEE Trans. Med. Imag.*, vol. 37, pp. 1551–1561, Jan. 2018.

[30] Y. Cui, S. Zhao, Y. Chen, J. Han, L. Guo, L. Xie, and T. Liu, "Modeling brain diverse and complex hemodynamic response patterns via deep recurrent autoencoder," *IEEE Trans. Cognit. Develop. Syst.*, early access, Oct. 25, 2019.

[31] U. Güçlü and M. A. J. van Gerven, "Modeling the dynamics of human brain activity with recurrent neural networks," *Frontiers Comput. Neurosci.*, vol. 11, pp. 1–7, Feb. 2017.

[32] H. Wang, S. Zhao, Q. Dong, Y. Cui, Y. Chen, J. Han, L. Xie, and T. Liu, "Recognizing brain states using deep sparse recurrent neural network," *IEEE Trans. Med. Imag.*, vol. 38, no. 4, pp. 1058–1068, Apr. 2019.

[33] Y. Cui, S. Zhao, H. Wang, L. Xie, Y. Chen, J. Han, L. Guo, F. Zhou, and T. Liu, "Identifying brain networks at multiple time scales via deep recurrent neural network," *IEEE J. Biomed. Health Informat.*, vol. 23, no. 6, pp. 2515–2525, Nov. 2019.

[34] S. Zhao, J. Han, X. Hu, X. Jiang, J. Lv, T. Zhang, S. Zhang, L. Guo, and T. Liu, "Extendable supervised dictionary learning for exploring diverse and concurrent brain activities in task-based fMRI," *Brain Imag. Behav.*, vol. 12, no. 3, pp. 743–757, Jun. 2018.

[35] B. A. Olshausen and D. J. Field, "Sparse coding with an overcomplete basis set: A strategy employed by v1?" *Vis. Res.*, vol. 37, no. 23, pp. 3311–3325, Dec. 1997.

[36] B. A. Olshausen and D. J. Field, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature*, vol. 381, no. 6583, pp. 607–609, Jun. 1996.

[37] D. M. Barch, G. C. Burgess, M. P. Harms, S. E. Petersen, B. L. Schlaggar, M. Corbetta, M. F. Glasser, S. Curtiss, S. Dixit, C. Feldt, D. Nolan, E. Bryant, T. Hartley, O. Footer, J. M. Bjork, R. Poldrack, S. Smith, H. Johansen-Berg, A. Z. Snyder, and D. C. Van Essen, "Function in the human connectome: Task-fMRI and individual differences in behavior," *NeuroImage*, vol. 80, pp. 169–189, Oct. 2013.

[38] S. M. Smith, C. F. Beckmann, J. L. R. Andersson, E. J. Auerbach, J. D. Bijsterbosch, and G. Douaud, "Resting-state fMRI in the human connectome project," *NeuroImage*, vol. 80, pp. 144–168, Oct. 2013.

[39] D. C. Van Essen, S. M. Smith, D. M. Barch, T. E. J. Behrens, E. Yacoub, and K. Ugurbil, "The WU-minn human connectome project: An overview," *NeuroImage*, vol. 80, pp. 62–79, Oct. 2013.

[40] M. F. Glasser, S. N. Sotiropoulos, J. A. Wilson, T. S. Coalson, B. Fischl, J. L. Andersson, J. Xu, S. Jbabdi, M. Webster, J. R. Polimeni, D. C. Van Essen, and M. Jenkinson, "The minimal preprocessing pipelines for the human connectome project," *NeuroImage*, vol. 80, pp. 105–124, Oct. 2013.

[41] S. Smith, P. R. Bannister, C. Beckmann, M. Brady, S. Clare, D. Flitney, P. Hansen, M. Jenkinson, D. Leibovici, B. Ripley, M. Woolrich, and Y. Zhang, "FSL: New tools for functional and structural brain image analysis," *NeuroImage*, vol. 13, no. 6, p. 249, Jun. 2001.

[42] M. Jenkinson and S. Smith, "A global optimisation method for robust affine registration of brain images," *Med. Image Anal.*, vol. 5, no. 2, pp. 143–156, Jun. 2001.

[43] A. Graves, A.-R. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, May 2013, pp. 6645–6649.

[44] K. Cho, B. van Merrienboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," 2014, *arXiv:1406.1078*. [Online]. Available: http://arxiv.org/abs/1406.1078

[45] I. Sutskever, J. Martens, and G. E. Hinton, "Generating text with recurrent neural networks," in *Proc. Int. Conf. Mach. Learn.*, 2011, pp. 1017–1024.

[46] M. Hermans and B. Schrauwen, "Training and analyzing deep recurrent neural networks," in *Proc. Neural Inf. Process. Syst.*, 2013, pp. 190–198.

[47] Z. C. Lipton, J. Berkowitz, and C. Elkan, "A critical review of recurrent neural networks for sequence learning," 2015, *arXiv:1506.00019*. [Online]. Available: http://arxiv.org/abs/1506.00019

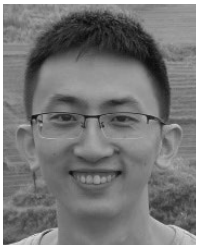
[48] C. J. Honey, R. Kotter, M. Breakspear, and O. Sporns, "Network structure of cerebral cortex shapes functional connectivity on multiple time scales," *Proc. Nat. Acad. Sci. USA*, vol. 104, no. 24, pp. 10240–10245, Jun. 2007.

[49] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[50] K. Cho, B. van Merrienboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using RNN Encoder–Decoder for statistical machine translation," in *Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP)*, 2014, pp. 1724–1734.
[51] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online learning for matrix factorization and sparse coding," *J. Mach. Learn. Res.*, vol. 11, pp. 19–60, Mar. 2010.
[52] S. M. Smith, P. T. Fox, K. L. Miller, D. C. Glahn, P. M. Fox, C. E. Mackay, N. Filippini, K. E. Watkins, R. Toro, A. R. Laird, and C. F. Beckmann, "Correspondence of the brain's functional architecture during activation and rest," *Proc. Nat. Acad. Sci. USA*, vol. 106, no. 31, pp. 13040–13045, Aug. 2009.
[53] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proc. Int. Conf. Artif. Intell. Statist.*, 2010, pp. 249–256.
[54] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*. [Online]. Available: https://arxiv.org/abs/1412.6980
[55] B. Zoph and Q. V. Le, "Neural architecture search with reinforcement learning," in *Proc. Int. Conf. Learn. Representations*, 2017, pp. 1–16

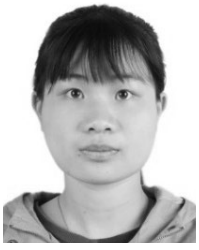
SHIJIE ZHAO received the Ph.D. degree in pattern recognition and intelligent systems from Northwestern Polytechnical University, Xi'an, Shaanxi, China.

He is currently an Assistant Professor with the School of Automation, Northwestern Polytechnical University. His research interests include brain imaging, cognitive computing, multimedia, machine learning, and deep learning.



YAN CUI received the bachelor's degree in electrical engineering and automation from Zhejiang University, Hangzhou, China, in 2011, where he is currently pursuing the Ph.D. degree in electronic information technology and instruments.

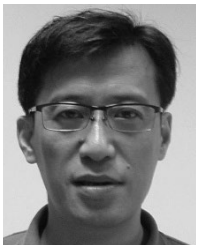
His research interests include brain imaging, machine learning, and embedded systems.



LINWEI HUANG received the bachelor's degree in automation from the School of Automation, Northwestern Polytechnical University, Xi'an, Shannxi, China, in 2019, where she is currently pursuing the master's degree.

Her research interests focuses on brain imaging, machine learning, and deep learning.

LI XIE is currently an Associate Professor with the College of Biomedical Engineering and Instrument Science, Zhejiang University, Hangzhou, China. His research interests include brain imaging, and control and signal processing.



YAOWU CHEN is currently a Qiushi Distinguished Professor with the College of Biomedical Engineering and Instrument Science, Zhejiang University, Hangzhou, China. His research interests include network multimedia, embedded systems, and electronic instrumentation system.

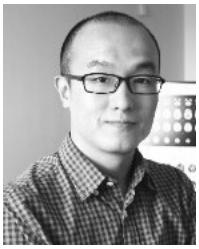
JUNWEI HAN (Senior Member, IEEE) is currently a Full Professor with the School of Automation, Northwestern Polytechnical University, Xi'an, China. His research interests include computer vision, multimedia processing, and brain imaging analysis.

LEI GUO is currently a Full Professor with the School of Automation, Northwestern Polytechnical University, Xi'an, China. His research interests include computer vision, multimedia processing, and brain imaging analysis.

SHU ZHANG received the Ph.D. degree in computer science from the University of Georgia, USA, in 2018.

He is currently a Professor with the School of Computer Science, Northwestern Polytechnical University, Xi'an, China. His research interests include biomedical image analysis, brain image analysis, deep learning and machine learning algorithms, and artificial intelligence.

TIANMING LIU (Senior Member, IEEE) has been a Distinguished Research Professor, since 2017, and a Full Professor of computer science, since 2015, with UGA. He is also an affiliated faculty with the UGA Bioimaging Research Center (BIRC), UGA Institute of Bioinformatics (IOB), UGA Neuroscience PhD Program, and UGA Institute of Artificial Intelligence (IAI). His research interest includes biomedical image analysis, neuroimaging, imaging informatics, healthcare informatics, and bioinformatics.



JINGLEI LV received the Ph.D. degree in engineering from the School of Automation, Northwestern Polytechnical University, Xi'an, China, in 2016.

He is currently a Senior Lecturer with the School of Biomedical Engineering, The University of Sydney, Sydney. He is also a Senior Researcher with the Sydney Imaging, Brain and Mind Center, The University of Sydney. His research interest focuses on developing neuroimaging method to better understand the fundamental neuroscience and mental disorders.