Four-Dimensional Modeling of fMRI Data via Spatio–Temporal Convolutional Neural Networks (ST-CNNs)

Yu Zhao, Xiang Li, Heng Huang, Wei Zhang, Shijie Zhao, Milad Makkie, Mo Zhang, Quanzheng Li, and Tianming Liu, *Senior Member, IEEE*

***Abstract*—Since the human brain functional mechanism has been enabled for investigation by the functional magnetic resonance imaging (fMRI) technology, simultaneous modeling of both the spatial and temporal patterns of brain functional networks from 4-D fMRI data has been a fundamental but still challenging research topic for neuroimaging and medical image analysis fields. Currently, general linear model (GLM), independent component analysis (ICA), sparse dictionary learning, and recently deep learning models, are major methods for fMRI data analysis in either spatial or temporal domains, but there are few joint spatial–temporal methods proposed, as far as we know. As a result, the 4-D nature of fMRI data has not been effectively investigated due to this methodological gap. The recent success of deep learning applications for functional brain decoding and encoding greatly inspired us in this paper to propose a novel framework called spatio–temporal convolutional neural network (ST-CNN) to extract both spatial and temporal characteristics from targeted networks jointly and automatically identify of functional networks. The identification of default mode network (DMN) from fMRI data was used for evaluation of the proposed framework. Results show that only training the framework on one fMRI data set is sufficiently generalizable to identify the DMN from different data sets of different cognitive tasks and resting state. Further investigation of the results shows that the joint-learning scheme can capture the intrinsic relationship between the spatial and temporal characteristics of DMN and thus it ensures the accurate identification of DMN from independent data sets. The ST-CNN model brings new tools and insights for fMRI analysis in cognitive and clinical neuroscience studies.**

***Index Terms*—Deep learning, functional brain networks, functional magnetic resonance imaging (fMRI).**

Manuscript received November 28, 2018; revised April 4, 2019; accepted May 11, 2019. Date of publication May 14, 2019; date of current version September 9, 2020. This work was supported in part by the National Institute of Health under Grant R01 DA-033393 and Grant R01 AG-042599; and in part by the National Science Foundation Career Award under Grant IIS-1149260, Grant CBET-1302089, Grant BCS-1439051, and Grant DBI-1564736. *(Corresponding author: Tianming Liu.)*

Y. Zhao, W. Zhang, M. Makkie, and T. Liu are with the Cortical Architecture Imaging and Discovery Laboratory, Department of Computer Science and Bioimaging Research Center, University of Georgia, Athens, GA 30602 USA (e-mail: tianming.liu@gmail.com).

X. Li is with Massachusetts General Hospital, Harvard Medical School, Boston, MA 02115 USA.

H. Huang and S. Zhao are with the School of Automation, Northwestern Polytechnical University, Xi'an 710072, China.

M. Zhang is with the Center for Data Science, Peking University, Beijing 100871, China.

Q. Li is with Massachusetts General Hospital, Harvard Medical School, Boston, MA 02115 USA, also with the Center for Data Science, Peking University, Beijing 100871, China, and also with the Laboratory for Biomedical Image Analysis, Beijing Institute of Big Data Research, Beijing 100871, China.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCDS.2019.2916916

I. Introduction

INVESTIGATIONS of the human brain's functional mechanism have been enabled by *in-vivo* functional magnetic resonance imaging (fMRI) technology. FMRI decomposition methods (e.g., independent component analysis (ICA) [1], [2], sparse representation [3], and deep learning methods [4], [5]) have significantly facilitated the analytics of the spatial and temporal features in fMRI data [6]. As fMRI data are the acquisition of series of 3-D brain volumes during imaging scan procedure to recording functional temporal dynamics of 3-D spatial brain volumes, the intrinsic spatio–temporal relationships are characterized in the form of 4-D data. Thus, a comprehensive characterization and description of 4-D fMRI data encoding of both spatial and temporal characteristics is significant for the understandings inside the human brain's organizational functional architecture.

In the current literature, methods of spatio–temporal analysis of fMRI data can be categorized into two groups from the perspectives of 3-D spatial or 1-D temporal dimensions of fMRI data. The first group performs the conjugate analysis on single domain, and then performs regression of the variation patterns in the other domain in an alternative manner. For example, temporal ICA [1], [2], [7] extracts the independent non-Gaussian temporal elements in the 4-D fMRI data, and then regresses out the spatial patterns of the corresponding temporal elements. In another recent research, a deep learning-based convolutional autoencoder (CAE) model [8] are explored to characterize temporal features, and corresponding spatial features are generated through regression from temporal features. On the other hand, dictionary learning and sparse representation methods extracts the sparse spatial maps of the fMRI data, while the temporal correspondences of these components are obtained through linear combinational regression. Moreover, the work proposed in [9] utilizing restricted Boltzmann machine (RBM) also focusing analysis on spatial features first and then the characteristics of temporal features.

Other than focusing on single dimension analysis, methods in the second group tend to perform simultaneous

analysis on both spatial and temporal domains. Realizing the intrinsic spatio–temporal interactions within fMRI data, this group aims to perform analysis of the spatio–temporal features of fMRI data comprehensively. For example, the work in [10] exploited hidden process model with spatio–temporal "prototypes" for modeling both domains, and it can disentangle overlapping mental processes evoked by stimuli. However, the spatio–temporal prototypes are limited to a small "region of influence" spatial prior for specific stimulus analysis. Another research also proposed an effective approach using recurrent neural network (RNN) to incorporate temporal dynamics (and between-time-frames correlations) into the intrinsic network (IN) modeling [11]. However, the RNN used in that research [11] is still a prior-like constraint for the ICA analysis. No comprehensive spatio–temporal analysis for whole-brain analysis is available in the above-mentioned research. Thereafter, inspired by better interpretability of the simultaneous intrinsic spatio–temporal modeling concept and the superior performance of deep learning frameworks, we proposed a whole brain level spatio–temporal deep convolutional neural network (ST-CNN) for 4-D fMRI data modeling. We aim to pinpoint or extract the spatial and temporal features of targeted functional networks [e.g., default mode network (DMN) in this paper] directly from the 4-D fMRI data without any template matching/searching processes involved, through the ST-CNN model. The ST-CNN model composes two simultaneous characterizations: 1) the characterization on the spatial pattern of the targeted network from the whole brain signals using a 3-D U-Net [12] and 2) the characterization on the temporal dynamic patterns of the targeted network, using a 1-D CAE [8]. Losses from the two mappings are merged and simultaneously back-propagated to the two networks in an integrated framework, fulfilling simultaneous modeling process of both spatial and temporal domains based on the level of whole brain signals. In the evaluation part, our experimental results show that, the ST-CNN, without hyper-parameter tuning, can extract dynamics of both spatial and temporal pattern of the DMN accurately, even with the presence of remarkable variabilities of cortical structures and functions from different individuals. Further evaluations demonstrated the sufficient generalizability of ST-CNN framework for the network identification task, in that only training the ST-CNN on motor task-evoked fMRI (tfMRI) data set, reproducible results can be achieved on other data sets [other tasks-evoked or resting-state fMRI (rsfMRI)]. ST-CNN can also serve for cognitive or clinical neuroscience studies as a useful tool, with the capability of identifying network in a pin-point way. Furthermore, with the ability of modeling the spatio–temporal variation patterns of the data corresponding to their intrinsic intertwined nature within one integrated framework, ST-CNN shows great potentials to offer refreshing perspectives for understanding human brain functional organization from 4-D fMRI data. It is noted that this paper is an significant extension from a recently accepted MICCAI paper [13]. We extensively evaluated and validated our ST-CNN model with larger data sets from HCP 900 release. Besides, the rsfMRI were also tested via ST-CNN and results show that both the spatial and temporal characteristics can be modeled for the targeted network (DMN).

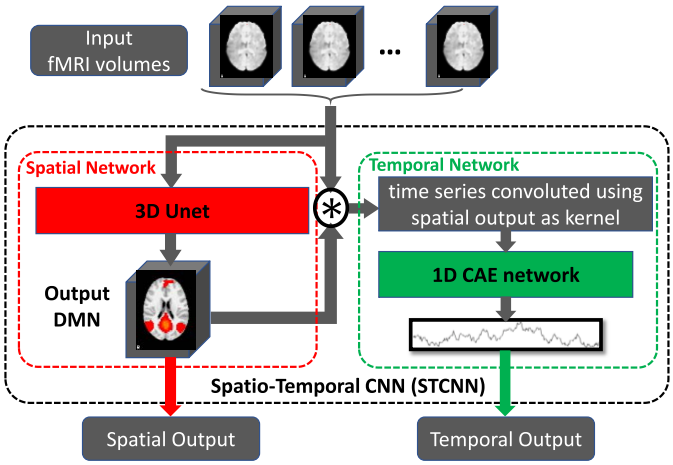

Fig. 1. ST-CNN framework. ST-CNN consists of two subnetworks: spatial network (red part) and temporal network (green part). Combination of the two subnets is defined using "⊛."

II. Method and Materials

Our designed ST-CNN framework takes 4-D fMRI data (either task-evoked or resting-state), and then generates both of spatial map and temporal time series of brain network as outputs simultaneously. Unlike popular CNN structures for natural image classifications (e.g., [14]), the proposed ST-CNN can perform convolution operations in both spatial and temporal domains simultaneously for spatial and temporal features, making our ST-CNN a spatio–temporal convolution framework for 4-D fMRI data modeling. The overview of this ST-CNN framework is illustrated in Fig. 1. To train the ST-CNN, the ground-truth DMN spatial network volumes and DMN temporal dynamics are provided via a dictionary learning and sparse coding method [3], [15], which will be explained in detail in the following sections.

A. Experimental Data and Preprocessing

The experimental data came from the publicly available Human Connectome Project (HCP) data set [16] (900 release) (https://www.humanconnectome.org/study/hcp-young-adult/document/900-subjects-data-release). The 900 subjects release includes behavioral and 3T MR imaging data from over 900 healthy adult participants, which provided a systematic and comprehensive mapping of connectome-scale functional networks over a large population in [17]. The detailed imaging parameters for both task-evoked and resting-state data are referred to [18]. The downloaded data were already preprocessed by a pipeline including: gradient distortion correction, motion correction, field map preprocessing, distortion correction, spline resampling to atlas space, intensity normalization etc. The HCP fMRI preprocessing analysis also uses FEAT in FSL for multiple regression with autocorrelation modeling and prewhitening and spatial smoothing [19]. The preprocessing pipeline is built using FSL [20] and FreeSurfer [21].

For the experiments in this paper, we used motor tfMRI, emotion tfMRI, and rsfMRI data sets from 200 randomly selected ones of the 900 subjects. Only 160 out of 200 subjects' motor tfMRI data sets were used for training purposes and all

the rest 40 subjects' motor tfMRI, 200 subjects' emotion tfMRI, and rsfMRI data sets were used for pure testing to validate the results using our framework. Actually, the 200 subjects for three tasks have intersecting subjects, as some subjects have missed scans for some tasks. A total of 282 subjects were used for motor, emotion and resting state data. We just used 200 subjects from the 282 subjects pool for each task. The ages of the 282 subjects range from 22 to 75, with a mean age range from 27.2 to 31.3. Among 282 subjects, 119 (42%) are male while 163 (58%) are female subjects. After preprocessing of the data sets following the above-mentioned pipeline, all the fMRI data are normalized to 0-1 distribution (normal distribution with 0 mean and 1 standard deviation) as inputs according to

$$\boldsymbol{v_i} = \begin{cases} \frac{v_i - \text{mean}(\boldsymbol{V_{\text{in-mask}}})}{\text{std}(\boldsymbol{V_{\text{in-mask}}})}, & \boldsymbol{v_i} \in \boldsymbol{V_{\text{in-mask}}} \\ 0, & \boldsymbol{v_i} \notin \boldsymbol{V_{\text{in-mask}}} \end{cases} \tag{1}$$

where $\boldsymbol{v_i}$ represents voxel intensity at location i; $\boldsymbol{V_{\text{in-mask}}}$ represents the voxels within the brain mask; mean() is the mean function, while std() is the standard deviation function.

The dictionary learning and sparse representation method [3], [15] was performed to decompose the fMRI data as ground-truth for ST-CNN training for two reasons. First, using individualized DMNs to guide the ST-CNN model to learn from individualized features is the key to achieve the optimized model accommodating for individual variability. As the individual variability of each fMRI scan is huge, the extracted DMNs are very different from each other with individual variances. Thus, using universal ground truth (e.g., some DMN templates) for training is risky as the ST-CNN may just over fit the DMN templates instead of modeling the intrinsic fMRI signals. Second, we chose the work on DMNs extracted by sparse representation over ICA as the works in [22] showed that experimental results demonstrated the sparse representation could better handle network decomposition when the spatial overlap exists between functional network maps. As DMN covers substantial area of regions in the brain, this will likely have spatial overlap with other networks. Following the caveats in [22], sparse representation may have better performance (at least comparable) in constructing and interpreting DMN. The input 4-D fMRI data for dictionary learning and sparse representation was flattened into a 2-D matrix $X \in \Re^{t \times n}$ with t (length of 0-1 normalized time points) rows by n columns, each of which represents one brain voxel out of a total number of n flattened voxels from an individual subject. The output contains one learned dictionary $D \in \Re^{t \times m}$ and a sparse coefficient matrix $\alpha \in \Re^{m \times n}$, with respect to, $X = D \times \alpha + \varepsilon$, as illustrated in [Fig. 2(a)], where ε is the error term and m is the predefined dictionary size, set to 400 in this experiment. The functional networks' temporal dynamics are obtained from the dictionary atoms: $\{d_i \in \Re^{t \times 1} | 1 \le i \le m\}$ and spatial patterns are obtained from the coefficient matrix regressed using a fast implementation of the LARS algorithm [23]: $\{\alpha_i \in \Re^{1 \times n} | 1 \le i \le m\}$ ($\alpha_i \in \Re^{1 \times n}$ are then mapped to 3-D brain volume space as spatial functional network maps). To find the DMN among all the m functional networks, the network with the maximum overlap rate (2) to the well-established DMN templates [24] was taken as its correspondence, whose corresponding dictionary atoms were taken as the DMN temporal dynamics. These

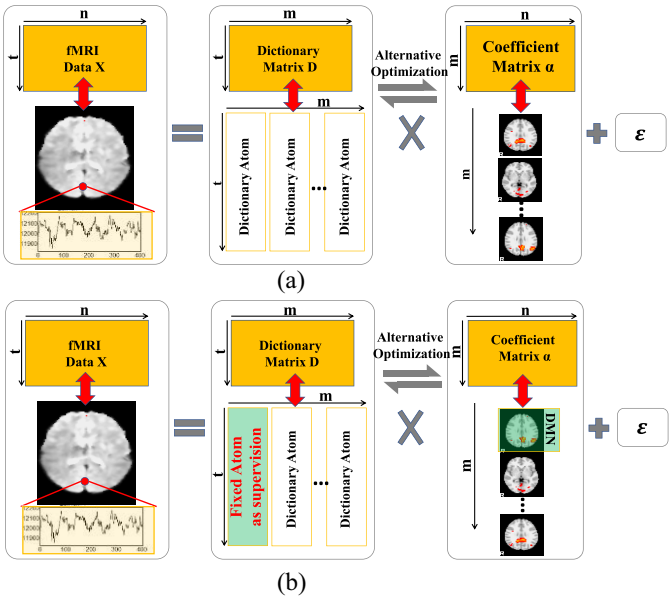

Fig. 2. Illustration of (a) dictionary learning and sparse representation and (b) supervised version. 4-D fMRI data was converted in a 2-D matrix X as input. The decomposed dictionary D contains temporal dynamics in each atom (column) and the coefficient matrix α contains the corresponding spatial maps of the functional networks. For supervised version, one of the dictionary atoms is fixed as the desired temporal dynamic curves. The corresponding coefficient is the corresponding spatial map.

DMN temporal dynamics and spatial maps were then used as the training and validation/comparison sets for our framework

$$\text{overlap rate} = \sum_{k=1}^{|V|} \frac{\min(\boldsymbol{V_k}, \boldsymbol{W_k})}{(\boldsymbol{V_k} + \boldsymbol{W_k})/2} \tag{2}$$

where $\boldsymbol{V_k}$ and $\boldsymbol{W_k}$ are the activation scores of voxel k in the spatial maps V and W, respectively. $|\boldsymbol{V}|$ is the total number of the voxels in the spatial map.

B. ST-CNN

As shown in Fig. 1, the ST-CNN framework consists of two parts: a spatial part and a temporal part. Unlike traditional ICA or dictionary learning and sparse representation methods, the inputs are 4-D fMRI data without flattening the 3-D volumes into a vector thus sacrificing the spatial geometric relationship between each voxel. Furthermore, rather than conjugating the updates between the spatial and temporal outputs, the ST-CNN use a spatio–temporal combination joint to process the spatio–temporal relationship inside the original input 4-D fMRI data and output the spatial and temporal results simultaneously in a pin-point way for the specific function network (DMN in this case).

1) Spatial Network: The spatial network is basically inspired by the 2-D Unet [12] designed for 2-D biomedical image segmentation. The key innovation of the 2-D Unet is the feature preserving from the contracting path to the extending path of the Unet structure. The preservation of the features generated from the contracting path will be fed back (copied) directly to the expansive path to assist the accurate segmentation by providing the original image feature information. The autoencoder-like contracting and expansive

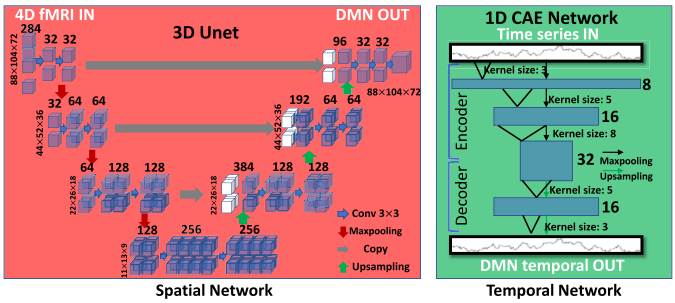

Fig. 3. Spatial network and temporal network structure inside ST-CNN.

structure makes Unet an end-to-end (image-to-image) framework for image segmentation, which is suitable to be modified as an end-to-end image pixel-level regression framework while still maintaining the feature of accurate reconstruction of the original input images. Intending to preserve the 3-D spatial context information of the fMRI, we finally adapted the 2-D image segmentation Unet to a 3-D image regression Unet as our spatial network.

By extending and adapting the 2-D classification Unet to a 3-D regression network as shown in Fig. 3, we can take 4-D fMRI data as input, each 3-D brain volume of which along the time points is assigned with one independent channel and regress/output DMN's spatial map. Basically, this 3-D Unet is a symmetrical structure with a contracting path and an expansive path. The contracting path is a structure with successive convolutional layers, alternating with the pooling layers (red arrows in Fig. 3 in spatial network). The expansive path is arguably symmetric with the contracting path with convolutional layers alternating with up-sampling layers. This 3-D u-shaped CNN structure is a fully convolutional network (FCN) without fully connected layers. For the contracting path among the structure, it follows the canonical CNN architecture, which repeated two 3×3 convolution operation, each followed by a rectified linear unit (ReLU), batch normalization layer and a down sampling max-pooling layer of pooling kernel size of 2. We will then double the size of the feature map channels following the down sampling process. The expansive path symmetrizes contracting path, except that the max-pooling layers are replaced with up-sampling layers, and that the number of the feature map channels are halved (except that the output layer has 1 channel as 1 3-D map output) after each up-sampling step. There are connections between contracting path and expansive path by copying feature maps from contracting path to expansive path to preserve features and contexts from the original input images. The loss function for training the spatial network is mean squared error (MSE) to resemble the targeted training spatial maps of DMN.

2) Temporal Network: The temporal network (Fig. 3 temporal network) is inspired from the excellent performance of a 1-D CAE network to deal with the time series for fMRI modeling [8]. In this paper, we adopt a 6-layers depth 1-D CAE to deal with the temporal features of the input fMRI. As shown in Fig. 3 temporal network part, the 1-D CAE is also a symmetric structure with a contractive and an expansive path. The contractive path, namely, the encoder starts by taking

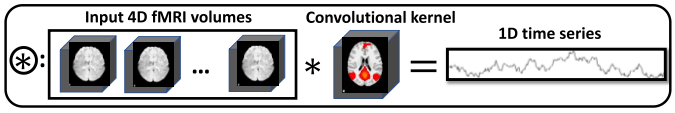

Fig. 4. Spatio-temporal combination joint illustration. The spatial output will be used a convolutional kernel applied to the original input 4-D fMRI data. The output of this combination joint is a time series reflecting the temporal dynamics from the fMRI data associated with the spatial map.

input 1-D time series and convolving them using size 3 convolutional kernel, which yields eight feature map channels, followed by a pooling layer for down-sampling. Then a size 5 convolutional layer is cascaded, which yields 16 feature map channels, also followed by a pooling layer for down-sampling. The last part of the encoder consists of a size 8 convolutional layer, which yields feature maps of 32 channels. The expansive path, namely, the decoder, takes the features output by the encoder as input and mirrors symmetrically the encoder to reconstruct the input time series. The loss function for training temporal 1-D CAE network is the following negative Pearson correlation to resemble the temporal dynamics of the training DMN:

$$\text{Temporal loss} = -\frac{N\sum_{i=1}^{N} x_i y_i - \sum_{i=1}^{N} x_i \sum_{i=1}^{N} y_i}{\sqrt{\left(N\sum_{i=1}^{N} x_i^2 - \left(\sum_{1}^{N} x_i\right)^2\right)\left(N\sum_{i=1}^{N} y_i^2 - \left(\sum_{i=1}^{N} y_i\right)^2\right)}} \tag{3}$$

where x and y are the output time series and ground-truth time series, and N is the length of the time series. ST-CNN incorporates this 1-D CAE to generate temporal dynamics of the DMN from the fMRI data.

C. Convolutional Spatio-Temporal Combination Joint

Since we already built networks for both spatial and temporal analysis, the next question is how to connect those two parts as an entire framework. That is, the relationship between spatial and temporal features needed to be characterized. Intuitively, as the activated regions (see the red regions of the convolutional kernel in Fig. 4) shall contain concurrent signals, we designed a convolutional spatio–temporal combination joint (Fig. 4). Through this joint, the concurrent signal features will be fused corresponding to the spatial map, meanwhile preserving an FCN structure. In detail, this combination joint in Fig. 4 connects spatial network and temporal network through a convolution operator. The combination takes the 4-D fMRI data and the 3-D DMN generated from spatial network as input. The spatial network generated 3-D output is taken as a 3-D convolutional kernel to convolve with each 3-D volume along the time frames of the input 4-D fMRI data in a valid way without any paddings

$$ts \in \mathbb{R}^{T\times 1} = \{t_1, t_2, \ldots, t_T \,|\, t_i = V_i * \text{DMN} \in \mathbb{R}\} \tag{4}$$

where t_i represents the single convolutional value at each time frame, V_i represents each 3-D volume scanned at time frame i. DMN represents the 3-D spatial map, which is also used in the combination joint as convolution kernel. T is the total

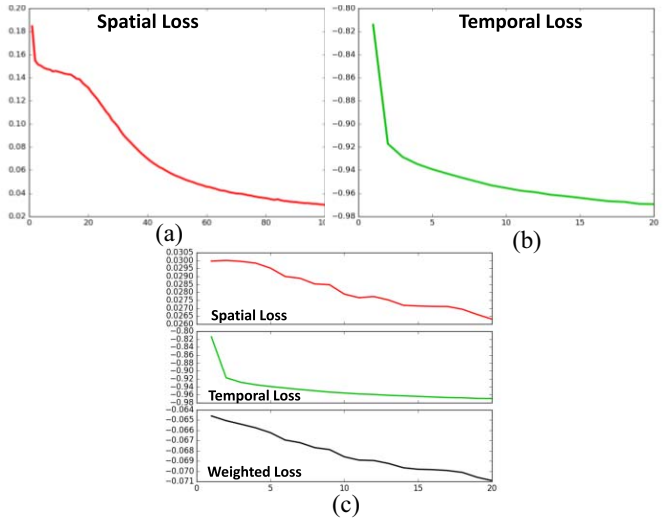

Fig. 5. Training loss curves. (a) Stage one for spatial network training. (b) Stage two for temporal network training. (c) Stage three for fine-tuning.

length of the time series, which is 284 in this paper. This valid no-padding convolutional operation will yield a single value for each time frame, resulting in a series of values as time series for the estimated DMN, namely, *ts*, which will be the input for temporal 1-D CAE, as above-mentioned.

D. Model Training Scheme

As introduced in the combination joint part, the temporal network relies on the DMN output of the spatial network, and we have designed a 3-step training strategy for efficiency. At the first stage, we train spatial network only; the second stage, we freeze the weights of the spatial network and train temporal network only; as a simultaneous spatio–temporal framework, during the third stage, the entire ST-CNN is fine-tuned simultaneously. From empirical practice, we observed that the temporal network loss is around ten times less than the spatial network loss, thus we designed a weighted total loss (10:1 for spatial:temporal) as the ST-CNN loss.

For the first stage, the spatial network was trained for 100 epochs until the spatial loss reaches 0.03 [Fig. 5(a)], which indicated the output MSE with the training DMN is very small. For the second stage, the temporal network was trained for 20 epochs when the temporal loss reached -0.97 [Fig. 5(b)], which indicated the Pearson correlation coefficient between the temporal output and the ground-truth is 0.97 (highly correlated). For the third stage [Fig. 5(c)], we can see co-operative refinement of both spatial and temporal network. The gradient descent optimizer is Adadelta [25].

In order to demonstrate the generalizability to different task-evoked data and resting-state data of the ST-CNN, our training data set was only based on the motor task tfMRI from 160 subjects. The rest 40 subjects' motor task tfMRI data and all the 200 subjects' emotion task tfMRI and rsfMRI were used for pure testing purpose.

E. Evaluation and Validation

To evaluate the performance of the framework, both the spatial and temporal similarities were quantified with the well-established DMN template spatial map and the corresponding time series of the DMNs decomposed from each individual. For the spatial similarity measurement, the overlap rate (2) between the output spatial map and the ground truth map was used, while the temporal similarity was measured by the Pearson correlation coefficient [the negative value of the temporal loss in (3)]. Qualitative evaluation will also be done as the "ground-truth" DMN decomposed from dictionary learning and sparse coding and identified with spatial overlap scheme may not be perfectly reliable as "true" DMN. In the result section, we will show some qualitative cases where the dictionary learning and sparse coding failed to generate DMN while our ST-CNN can successfully pinpoint the DMN.

Furthermore, to validate that the output of the ST-CNN models the correct spatio–temporal relationship from the 4-D fMRI data rather than overfitting the DMN without modeling the spatio–temporal relationships, we performed a supervised dictionary learning (SDL) and sparse representation method [26] [Fig. 2(b)] to check whether our ST-CNN framework generate spatio–temporal outputs by successfully modeling the intrinsic spatio–temporal characteristics within the 4-D fMRI data. The SDL and sparse representation method [26] takes the temporal output of the ST-CNN as the temporal supervision of the dictionary [Fig. 2(b) green part in dictionary], and reconstruct the corresponding spatial maps [Fig. 2(b) green part in coefficient matrix) based on the supervised dictionary to generate the corresponding spatial maps to the supervision. The spatial overlap rate was utilized to check the similarity between the spatial maps generated by ST-CNN and the SDL and sparse representation. By this way, we can confirm with confidence that our ST-CNN produces intrinsic spatio–temporal dynamics of the DMN.

III. Results

In this section, result analysis and performance evaluation of the spatio–temporal output of the ST-CNN from testing data sets are presented. ST-CNN was trained on the motor tfMRI data from 160 subjects, while the testing data sets includes motor tfMRI data from 40 different subjects, emotion tfMRI and rsfMRI from the corresponding 200 subjects in HCP Q900 release. In summary, testing results showed that ST-CNN can perform DMN identification with simultaneous intrinsic spatial and temporal characterization.

A. DMN Spatio-Temporal Identification

ST-CNN is first tested on the same task (motor) tfMRI data from 40 different subjects. Visualizations of the identified DMN from three sample subjects are shown in Fig. 6. Spatial output of ST-CNN resembled ground-truth spatial maps decomposed by dictionary learning and sparse representations (SR for brevity) (overlap rate all larger than 0.2). While the spatial pattern of DMN template was never provided to ST-CNN (only subject-wise decomposition results were used for training), ST-CNN outputs are more similar or at least comparable to dictionary learning method which used DMN template as input (Table I). As reported in [27], networks with spatial overlap rate larger than 0.1 will be considered similar to each other.

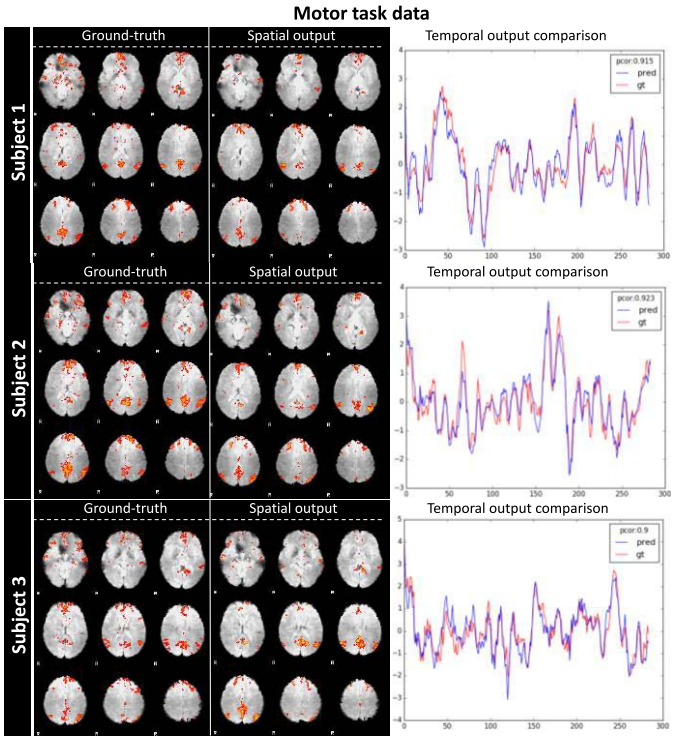

Fig. 6. DMN identification with spatio–temporal co-learning by ST-CNN. Randomly selected three subjects are visualized with their ground-truth DMN spatial map (decomposed by dictionary learning and sparse representation and identified by spatial overlap with DMN template), spatial output map from ST-CNN, and temporal dynamics of both SR and ST-CNN. For temporal dynamics, the blue curve is the output of ST-CNN and the red curve is the ground-truth dictionary atom corresponding to the ground-truth DMN spatial map. The Pearson correlation of the two curves are also displayed in the top right corner: 0.915, 0.923, and 0.9, respectively.

TABLE I
DMN IDENTIFICATION QUANTITATIVE ANALYSIS FOR THREE SAMPLE SUBJECTS. HIGHER VALUES FOR SPATIAL OVERLAP RATES WITH DMN TEMPLATES ARE HIGHLIGHTED IN BOLD TEXTS

Subject	Spatial overlap between SR and ST-CNN	Spatial overlap with DMN template		Temporal similarity (Pearson correlation)
		Sparse Representation	ST-CNN	
Subject 1	0.248	**0.121**	0.120	0.915
Subject 2	0.255	0.120	**0.122**	0.923
Subject 3	0.238	0.128	**0.133**	0.900

Results of motor tfMRI from 40 subjects can be found at: http://hafni.cs.uga.edu/DMN_dynamic/HCP_900/MOTOR/.

After examining all the testing results from motor tfMRI data, we found ST-CNN perform superiorly than SR in the following aspects: ST-CNN can identify DMN in a pinpoint way, rather than relying on spatial overlap measurement (such as SR/ICA). Besides, as the spatial–temporal dynamics of DMN within fMRI data are simultaneously captured by ST-CNN, and it can more accurately identify DMN comparing with unsupervised approaches such as SR (which relies on the sparsity prior of fMRI data). We measured the spatial overlap rate between the results by ST-CNN/SR and DMN template in motor fMRI data from all 40 testing subjects. The result shows that ST-CNN achieved a noticeably higher mean spatial overlap rate and lower standard deviation with DMN template (0.124 ± 0.016) comparing with SR (0.107 ± 0.042). In addition, temporal dynamics of the identified DMNs by ST-CNN shows high Pearson correlation (averagely 0.758 across 40 subjects) with temporal dynamics of ground-truth DMNs.

TABLE II
ST-CNN SUPERIOR PERFORMANCE IN DMN IDENTIFICATION

Case	Spatial overlap between SR and ST-CNN output	Spatial overlap with DMN template		Temporal similarity (Pearson correlation)
		Sparse Representation	ST-CNN	
Case 1	0.166	0.044	0.106	0.805
Case 2	0.059	0.080	0.135	0.268
Case 3	0.067	0.068	0.109	0.113

In addition to the fact that ST-CNN outperformed SR on average, we also observed cases where ST-CNN generated obviously better DMN maps than SR (Fig. 7, quantitative results are shown in Table II). These cases are particularly interesting, as the ST-CNN is trained based on the results of SR. Thus, if ST-CNN can obtain correct DMN identification where SR fails (which is not uncommon due to various factors, as illustrated below), that would be an indication for the superior generalizability of ST-CNN over its training data.

In case 1, only a partial DMN was identified by SR, while posterior cingulate cortex (PCC) and inferior parietal lobe (IPL) were partially inactivated. On the contrary, ST-CNN identified these two regions correctly. Temporal dynamics of the results from two models are similar (Pearson correlation 0.805), as major regions in DMN were still preserved by SR.

In case 2, DMN identified by SR show a mixed spatial pattern of DMN and later visual network [24], possibly caused by the interdigitated functional area [22] that affected decomposition results (i.e., two networks are not decomposable based on current parameter setting of SR). Again, DMN identified by ST-CNN maintained most of the related functional networks. Correspondingly, temporal dynamics of SR show a significant out-of-phase spike comparing with network identified by ST-CNN, decreasing the Pearson correlation between these two to 0.268. This is likely caused by the extra involvement of later visual network in SR result.

Case 3, SR DMN identification failure. As cortical microcircuits overlap and interdigitate with each other [28], rather than being independent and segregated in space, the medial visual regions and DMN are spatially overlapped. Either incurred by the failure of dictionary learning and sparse representation method or the failure of spatial overlap-based DMN identification process, the SR DMN turned out to be a medial visual network. As a result, the ST-CNN predicted temporal dynamics for DMN is quite different from the one corresponding to the medial visual network (Pearson correlation 0.113).

B. Generalizability for Other Task fMRI Data

We trained our ST-CNN on motor tfMRI data, since the pure testing on motor tfMRI data is not adequate for demonstrating the generalizability of the proposed framework as one can argue that the trained ST-CNN is overfitting to motor task data. Therefore, without any further training after purely training on

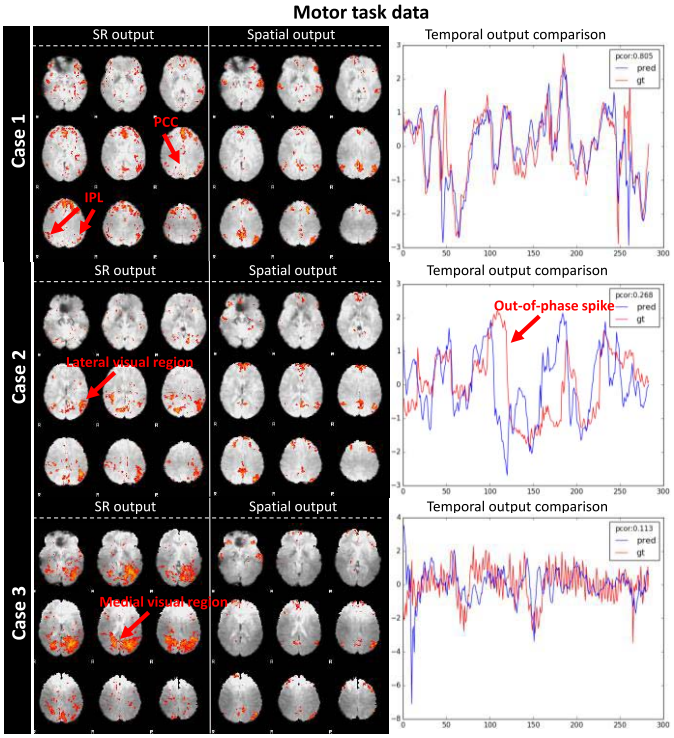

Fig. 7. Superior DMN identification ability of ST-CNN than dictionary learning and sparse representation (SR). Case 1, ground-truth DMN partial spatial pattern. Case 2, ground-truth DMN mixed spatial pattern. Case 3, ground-truth DMN identification failure.

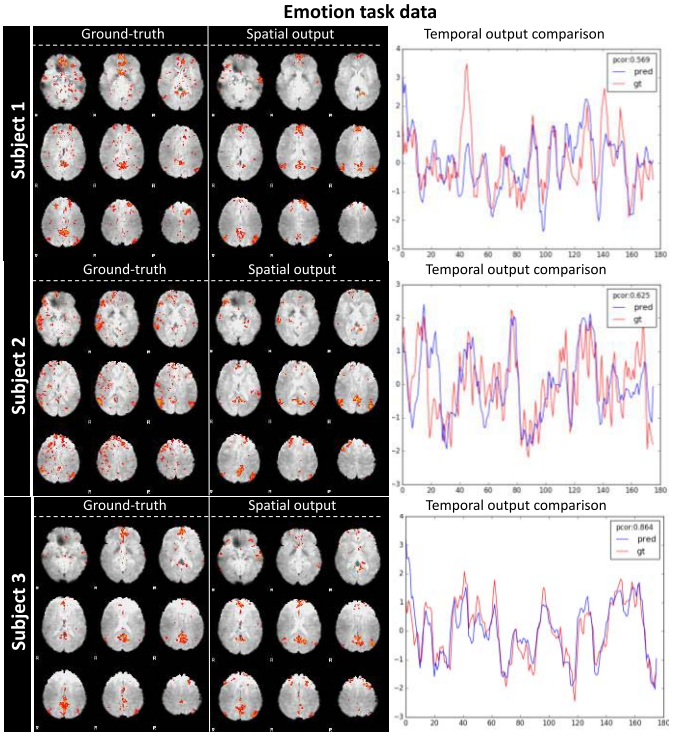

Fig. 8. DMN spatial–temporal identification generalizability to emotion task. The spatial and temporal output of three subjects are shown above. Ground-truth DMN spatial map (decomposed by dictionary learning and sparse representation and identified by spatial overlap with DMN template), spatial output map from ST-CNN, and temporal dynamics of both dictionary learning and sparse representation and ST-CNN are listed for each column, respectively. For temporal dynamics, the blue curve is the output of ST-CNN and the red curve is the ground-truth dictionary atom corresponding to the ground-truth DMN spatial map. The Pearson correlation of the two curves are also displayed in the top right corner.

TABLE III
DMN IDENTIFICATION QUANTITATIVE ANALYSIS BOTH SPATIALLY (SPATIAL OVERLAP RATE, MEAN±STD) AND TEMPORALLY (PEARSON CORRELATION) FOR HCP 900 RELEASE DATA

Datasets	Spatial overlap with DMN template		Temporal similarity (Pearson correlation)
	Sparse Representation	ST-CNN	
MOTOR (40 subjects)	0.107±0.042	**0.124±0.016**	0.758
EMOTION (200 subjects)	0.102±0.041	**0.115±0.026**	0.751
RSN (200 subjects)	0.109±0.044	**0.118±0.030**	0.725

motor data, we deployed a test on emotion tfMRI data to test the generalizability of ST-CNN for other tasks.

Similarly, we show the ST-CNN prediction for emotion task from three randomly selected subjects in Fig. 8. As we can see, the spatial output of ST-CNN successfully identified DMN spatial maps. Using the dictionary learning and sparse representation DMN output as ground-truth, we can also see the temporal output of ST-CNN is highly correlated with ground-truth. All 200 emotion tfMRI testing results are referred to http://hafni.cs.uga.edu/DMN_dynamic/HCP_900/EMOTION/.

As analyzed in Table III, the mean spatial overlap rate of the ST-CNN outputs with well-established DMN template is still clearly larger than the sparse representation outputs, and the larger standard deviation of the sparse representation results than ST-CNN results also demonstrate that ST-CNN is much more accurate and robust in DMN identification. Still, the temporal similarity preserved the same level (mean Pearson correlation 0.751) as in motor data set. Both spatial and temporal quantitative and qualitative results support that our ST-CNN trained on one specific task has robust generalizability to other tasks.

C. Generalizability for Resting-State fMRI Data

The DMN is vastly known for its presence during resting state, namely, default mode [29], [30]. DMN will also establish or internally orient tasks, which means during task-evoked states, DMN is also present [30], [31]. Correspondences of DMN during activation and rest were also found as full repertoire of functional networks utilized by the brain in task-evoked states is continuously and dynamically active [24]. According to literature DMN related research studies [29]–[31], DMN tends to be an IN that constantly exists inside human brain no matter it is healthy brain or diseased brain [32]–[34]. Following this logic and the generalizability of the trained ST-CNN to other task data, we further tested our ST-CNN trained on task-evoked data for resting state DMN modeling, which is another important reason we designed ST-CNN to pinpoint DMN.

Similar to motor and emotion results sections, we randomly pick three subjects' results as a qualitative illustration in Fig. 9 and put all results for DMN spatio–temporal dynamics outputs for rsfMRI at http://hafni.cs.uga.edu/DMN_dynamic/HCP_900/RSN/. As

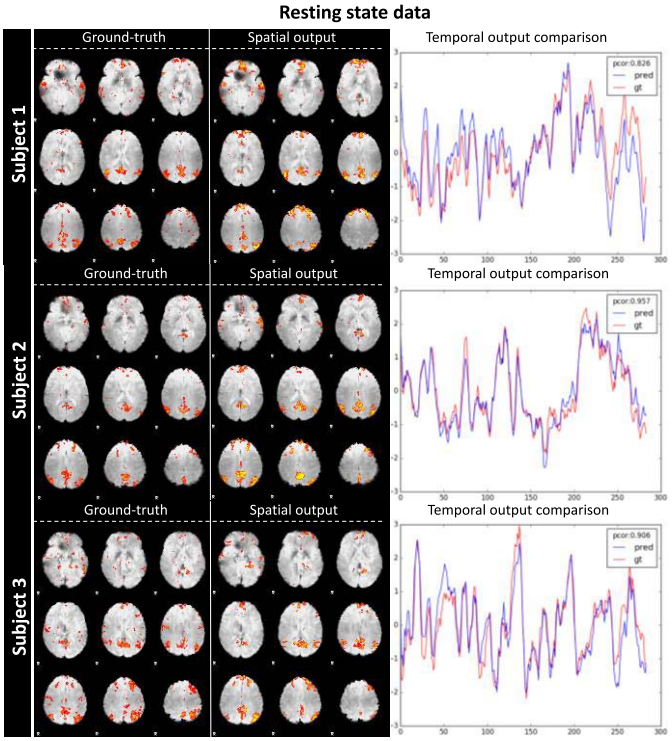

Fig. 9. DMN's spatial–temporal identification generalizability to resting state fMRI data. The spatial and temporal output of three subjects are shown above. Ground-truth DMN spatial map (decomposed by dictionary learning and sparse representation and identified by spatial overlap with DMN template), spatial output map from ST-CNN, and temporal dynamics of both dictionary learning and sparse representation and ST-CNN are listed for each column, respectively. For temporal dynamics, the blue curve is the output of ST-CNN and the red curve is the ground-truth dictionary atom corresponding to the ground-truth DMN spatial map. The Pearson correlation of the two curves are also displayed in the top right corner.

shown in Fig. 9, the DMN spatial pattern is accurately captured by our ST-CNN, with posterior cingulate cortex (PCC), medial prefrontal cortex (mPFC), and IPL activated. The ground-truth DMN spatial maps decomposed by dictionary learning and sparse representation clearly have high similarity with ST-CNN outputs. It is intriguing that we still achieved high spatial overlap rate for DMN in rsfMRI. As shown in Table III, the average spatial overlap rate of ST-CNN output with well-established DMN templates is higher than the outputs from dictionary learning and sparse representation. The main reason is similar to the analysis for motor data, that is, dictionary learning and sparse representation method has limited power for DMN interpretation and identifying DMN using spatial overlap rate from hundreds of networks is not very robust.

From the testing results from resting state data, we can conclude that our ST-CNN can successfully model the intrinsic dynamics of the DMN from fMRI data and identify DMN in a pinpoint way for different tasks as well as resting state data. Only being trained using one task data set (motor task), the generalizability of the ST-CNN can be demonstrated for other task data and resting-state data. This result cross-validated the correspondence of DMN in both task-evoked state and resting state and suggested the ST-CNN can capture the spatial and temporal dynamics intrinsically from any given fMRI data.

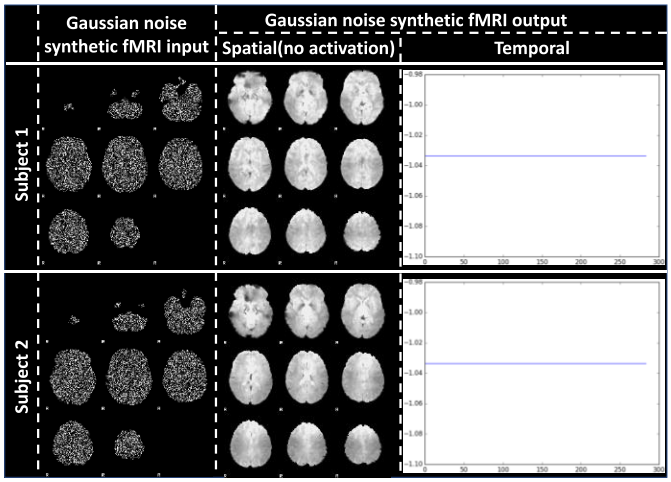

Fig. 10. ST-CNN output for synthetic fMRI data input with Gaussian noise. Visualization of synthetic fMRI data with Gaussian noise on first column. The synthetic data and real fMRI data have the same brain shape as boundary mask. Spatial output: only brain background is shown, not activation for the output. Temporal output: no active curve corresponding to spatial maps with no activation.

D. Robustness to Noise

To further test the robustness to noise of the ST-CNN and to further demonstrate that the ST-CNN is not just overfitting DMN from "brain shaped" signals, we synthesized fMRI data using Gaussian noises (first column in Fig. 10) within the brain mask to test our ST-CNN framework.

The testing results showed that the trained ST-CNN is not sensitive to noise and there is no output temporal signals and no activation for the spatial maps, as shown in Fig. 10. The results confirmed the robustness to noise of our trained ST-CNN. Further, it also demonstrated that our trained ST-CNN is not overfitting DMN according to the brain shapes, rather modeling the intrinsic signals from the fMRI data. With the confidence in the intrinsic 4-D modeling from the fMRI data of ST-CNN, we further checked whether ST-CNN can model the intrinsic spatial and temporal relationships from the fMRI data in the next section.

E. Spatial and Temporal Relationship

As network spatial pattern and temporal dynamics are intertwined with each other, it is interesting to examine the relationship between spatial and temporal domains of functional networks. In order to validate that the ST-CNN can well capture the intrinsic spatial and temporal relationship from fMRI data, we performed an SDL method [26] (introduced in the evaluation and validation section) onto the fMRI data by taking ST-CNN temporal output as input supervision to reconstruct the corresponding spatial response of that temporal supervision to check whether the spatial response is DMN or not.

We performed SDL on all three test data sets: 1) 40 motor task data; 2) 200 emotion task data; and 3) 200 resting state data. We randomly picked one exemplar result per data set to briefly illustrate the validation result for ST-CNN in Fig. 11. The first column of Fig. 11 shows temporal dynamics of ST-CNN, which was used to be the fixed dictionary atom as the supervision, and the corresponding spatial outputs

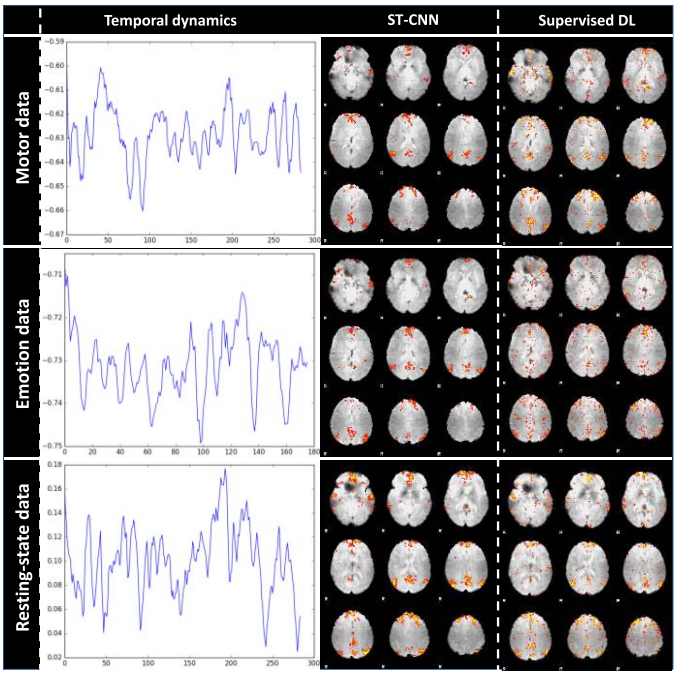

Fig. 11. Validation of ST-CNN performing supervised dictionary learning (supervised DL) method. The first column shows temporal dynamics of ST-CNN. Using that as input, we performed supervised DL to generate the corresponding spatial response (third column). The second column shows ST-CNN spatial maps, which resemble the supervised DL results. Both spatial map results display the DMN spatial pattern.

TABLE IV
VALIDATION OF ST-CNN PERFORMING SDL. THE SPATIAL OVERLAP RATE BETWEEN ST-CNN SPATIAL OUTPUTS AND SDL SPATIAL OUTPUTS USING ST-CNN TEMPORAL OUTPUTS AS INPUT SUPERVISION. THE STATISTICS SHOWN BELOW ARE MEAN VALUE ± STANDARD DEVIATION

Dataset	MOTOR	EMOTION	RS
Spatial overlap rate between ST-CNN and SDL ouputs (Mean±std)	0.294±0.072	0.304±0.056	0.336±0.075

of the SDL are shown in the third column in Fig. 11. The ST-CNN spatial outputs are in the second column of Fig. 11. We can clearly see that the ST-CNN spatial maps resemble the SDL spatial maps matching DMN spatial pattern, which means that ST-CNN captured intrinsic relationships between temporal and spatial dynamics of DMN. We also put all validation results for further reference: motor data validation: http://hafni.cs.uga.edu/DMN_dynamic/HCP_900/validation/MOTOR/; emotion data validation: http://hafni.cs.uga.edu/DMN_dynamic/HCP_900/validation/EMOTION/; resting-state data validation: http://hafni.cs.uga.edu/DMN_dynamic/HCP_900/validation/RSN/. Quantitatively, we calculated the spatial overlap rate as the similarity metric to measure how similar the ST-CNN outputs resemble the SDL outputs. As shown in Table IV, the mean spatial overlap rates are all larger than 0.2, which can be considered strongly similar according to [27]. The results statistically demonstrated the high similarity between the ST-CNN and the SDL results, which indicates that our proposed ST-CNN is effectively modeling the intrinsic spatio–temporal relationship of DMN from fMRI data. However, to achieve the same results, the SDL methods or other equivalent methods need to take prior knowledge such as the temporal input as supervision to obtain the spatial output, which might hamper the application of such methods given that we do not have any prior knowledge of a specific data. On the contrary, once ST-CNN is properly trained, it can produce both temporal and spatial dynamics of DMN without any form of prior information, which paves a much broader way for applications.

All the above qualitative and quantitative results demonstrated that our proposed ST-CNN can model the intertwined intrinsic spatio–temporal dynamics from 4-D fMRI data, no matter task-evoked or resting-state data.

IV. DISCUSSION

In this paper, we proposed a novel ST-CNN to model and analyze 4-D fMRI data and simultaneously generate DMN spatial and temporal dynamics in a pinpoint way. This spatio–temporal deep learning framework provided a new tool and insight for 4-D fMRI analysis in future cognitive and clinical research studies. By utilizing the proposed ST-CNN, we aimed to solve the two challenging problems in fMRI analysis research: 1) spatio–temporal intrinsic 4-D analysis for specific functional network (DMN in this paper) and 2) functional network identification directly from fMRI data after only basic preprocessing (e.g., gradient distortion correction, motion correction, field map preprocessing, distortion correction, spline resampling to atlas space, intensity normalization etc.). As we already discussed, the spatio–temporal and 4-D simultaneous analysis for fMRI data is still an open question and many current functional network identification methods are still based on data decomposition technique (e.g., ICA, dictionary learning, and sparse coding [35]–[37]). Those techniques have randomized index for the extracted DMN among all the extracted functional networks, which will impose a burden for the DMN identification process, while ST-CNN is trained specifically for DMN, which will yield the targeted DMN directly as output without any ambiguities. Now in the proposed ST-CNN, these two challenging open questions can be effectively handled at the same time and the DMN identification process is much more robust and reliable than traditional fMRI data decomposition and network identification techniques. More importantly, the reproducibility of the ST-CNN is clearly demonstrated by training ST-CNN on one task-evoked data set and applied to other task-evoked data sets as well as resting-state data set. As for DMN regression, it is logically more reasonable to use rsfMRI data for both training and testing. However, this is a relatively simpler task as training and testing are both performed on the same type of the data set. Considering the ST-CNN framework, which is proposed to regress the DMN spatial map and temporal response within that region, while concurrent temporal response can also pose a penalty for falsely regressed spatial regions, it really does not quite matter whether the DMN temporal response is positively or negatively correlated with the task design since the task design is not even utilized in training ST-CNN, as long as the temporal response is concurrent regarding to the DMN spatial response. Besides,

correspondence between task-evoked state and resting-state has been found to be established in [24]. Therefore, the generalizability of the ST-CNN for different tasks/resting states fMRI data is also demonstrated. The robustness to noise and nonoverfitting analysis further exhibited the robustness of the ST-CNN framework. With further validation on the relationship between spatial and temporal outputs, we further confirmed the effectiveness of the proposed ST-CNN.

In the future work, we will focus on extending the current framework on pinpointing more functional networks from raw 4-D fMRI data, which can be further applied on brain disease data sets for better understanding of abnormal brain activity. As indicated by the current research results [27], [38], comprehensive resting-state networks including high-order and low-order networks [39] are necessary for brain disease analysis. Since the ST-CNN model is quite robust and reproducible for various types of data only across a small range of hyperparameter settings, we plan to use a neural architecture search scheme to investigate the optimal architecture of the ST-CNN for different types of data and applications. Other simultaneous spatio–temporal fMRI analysis models can also be inspired from ST-CNN to accelerate the investigation of the brain's functional architecture.

ACKNOWLEDGMENT

The authors would like to thank the Human Connectome Project for sharing their invaluable fMRI data sets.

REFERENCES

[1] D. M. Cole, S. M. Smith, and C. F. Beckmann, "Advances and pitfalls in the analysis and interpretation of resting-state fMRI data," *Front. Syst. Neurosci.*, vol. 4, p. 8, Feb. 2010.

[2] M. J. McKeown, L. K. Hansen, and T. J. Sejnowsk, "Independent component analysis of functional MRI: What is signal and what is noise?" *Current Opin. Neurobiol.*, vol. 13, no. 5, pp. 620–629, 2003.

[3] J. Lv *et al.*, "Sparse representation of whole-brain fMRI signals for identification of functional networks," *Med. Image Anal.*, vol. 20, no. 1, pp. 112–134, Feb. 2015.

[4] G. Litjens *et al.*, "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, Dec. 2017.

[5] Y. Zhao *et al.*, "Constructing fine-granularity functional brain network atlases via deep convolutional autoencoder," *Med. Image Anal.*, vol. 42, pp. 200–211, Dec. 2017.

[6] D. J. Heeger and D. Ress, "What does fMRI tell us about neuronal activity?" *Nat. Rev. Neurosci.*, vol. 3, no. 2, pp. 142–151, Feb. 2002.

[7] S. M. Smith *et al.*, "Temporally-independent functional modes of spontaneous brain activity," *Proc. Nat. Acad. Sci. USA*, vol. 109, no. 8, pp. 3131–3136, Feb. 2012.

[8] H. Huang *et al.*, "Modeling task fMRI data via deep convolutional autoencoder," *IEEE Trans. Med. Imag.*, vol. 37, no. 7, pp. 1551–1561, Jul. 2018.

[9] R. D. Hjelm, V. D. Calhoun, R. Salakhutdinov, E. A. Allen, T. Adali, and S. M. Plis, "Restricted Boltzmann machines for neuroimaging: An application in identifying intrinsic networks," *Neuroimage*, vol. 96, pp. 245–260, Aug. 2014.

[10] Y. Shen, S. D. Mayhew, Z. Kourtzi, and P. Tiòo, "Spatial–temporal modelling of fMRI data through spatially regularized mixture of hidden process models," *Neuroimage*, vol. 84, pp. 657–671, Aug. 2014.

[11] R. D. Hjelm, S. M. Plis, and V. Calhoun, "Recurrent neural networks for spatiotemporal dynamics of intrinsic networks from fMRI data," in *Proc. NIPS Brains Bits*, 2016.

[12] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput. Assisted Intervent.*, 2015, pp. 234–241.

[13] Y. Zhao *et al.*, "Modeling 4D fMRI data via spatial–temporal convolutional neural networks (ST-CNN)," in *Proc. Med. Image Comput. Comput. Assisted Intervent. Soc.*, 2018, pp. 181–189.

[14] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.* 2012, pp. 1097–1105.

[15] S. Zhao *et al.*, "Decoding auditory saliency from brain activity patterns during free listening to naturalistic audio excerpts," *Neuroinformatics*, vol. 16, nos. 3–4, pp. 309–324, Feb. 2018.

[16] D. C. Van Essen *et al.*, "The WU-minn human connectome project: An overview," *Neuroimage*, vol. 80, pp. 62–79, Oct. 2013.

[17] D. M. Barch *et al.*, "Function in the human connectome: Task-fMRI and individual differences in behavior," *Neuroimage*, vol. 80, pp. 169–189, Oct. 2013.

[18] (2015). *WU-Minn HCP 900 Subjects Data Release: Reference Manual*. [Online]. Available: https://www.humanconnectome.org/storage/app/media/documentation/s900/HCP_S900_Release_Reference_Manual.pdf

[19] M. W. Woolrich, B. D. Ripley, M. Brady, and S. M. Smith, "Temporal autocorrelation in univariate linear modeling of fMRI data," *Neuroimage*, vol. 14, no. 6, pp. 1370–1386, Dec. 2001.

[20] M. Jenkinson, C. F. Beckmann, T. E. Behrens, M. W. Woolrich, and S. M. Smith, "FSL," *Neuroimage*, vol. 62, no. 2, pp. 782–790, 2012.

[21] A. M. Dale, B. Fischl, and M. I. Sereno, "Cortical surface-based analysis," *Neuroimage*, vol. 9, no. 2, pp. 179–194, Feb. 1999.

[22] W. Zhang *et al.*, "Experimental comparisons of sparse dictionary learning and independent component analysis for brain network inference from fMRI data," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 1, pp. 289–299, Jan. 2019.

[23] R. Tibshirani, I. Johnstone, T. Hastie, and B. Efron, "Least angle regression," *Ann. Stat.*, vol. 32, no. 2, pp. 407–499, Apr. 2004.

[24] S. M. Smith *et al.*, "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.

[25] M. D. Zeiler, "ADADELTA: An adaptive learning rate method," *arXiv:1212.5701*, Dec. 2012.

[26] S. Zhao *et al.*, "Supervised dictionary learning for inferring concurrent brain networks," *IEEE Trans. Med. Imag.*, vol. 34, no. 10, pp. 2036–2045, Oct. 2015.

[27] Y. Zhao *et al.*, "Connectome-scale group-wise consistent resting-state network analysis in autism spectrum disorder," *NeuroImage. Clin.*, vol. 12, pp. 23–33, Jun. 2016.

[28] K. D. Harris and T. D. Mrsic-Flogel, "Cortical connectivity and sensory coding," *Nature*, vol. 503, no. 7474, pp. 51–58, Nov. 2013.

[29] M. E. Raichle, A. M. MacLeod, A. Z. Snyder, W. J. Powers, D. A. Gusnard, and G. L. Shulman, "A default mode of brain function," *Proc. Nat. Acad. Sci. USA*, vol. 98, no. 2, pp. 676–682, Jan. 2001.

[30] M. D. Greicius, B. Krasnow, A. L. Reiss, and V. Menon, "Functional connectivity in the resting brain: A network analysis of the default mode hypothesis," *Proc. Nat. Acad. Sci. USA*, vol. 100, no. 1, pp. 253–258, Jan. 2003.

[31] D. Vatansever, D. K. Menon, and E. A. Stamatakis, "Default mode contributions to automated information processing," *Proc. Nat. Acad. Sci. USA*, vol. 114, no. 48, pp. 12821–12826, Nov. 2017.

[32] M. Jung *et al.*, "Default mode network in young male adults with autism spectrum disorder: Relationship with autism spectrum traits," *Mol. Autism*, vol. 5, no. 1, p. 35, 2014.

[33] A. Iraji *et al.*, "Resting state functional connectivity in mild traumatic brain injury at the acute stage: Independent component and seed-based analyses," *J. Neurotrauma*, vol. 32, no. 14, pp. 1031–1045, Jul. 2015.

[34] M. D. Greicius, G. Srivastava, A. L. Reiss, and V. Menon, "Default-mode network activity distinguishes Alzheimer's disease from healthy aging: Evidence from functional MRI," *Proc. Nat. Acad. Sci. USA*, vol. 101, no. 13, pp. 4637–4642, Mar. 2004.

[35] L. Griffanti *et al.*, "Hand classification of fMRI ICA noise components," *Neuroimage*, vol. 154, pp. 188–205, Jul. 2017.

[36] J. Tohka, K. Foerde, A. R. Aron, S. M. Tom, A. W. Toga, and R. A. Poldrack, "Automatic independent component labeling for artifact removal in fMRI," *Neuroimage*, vol. 39, no. 3, pp. 1227–1245, Feb. 2008.

[37] Y. Zhao *et al.*, "Automatic recognition of fMRI-derived functional networks using 3D convolutional neural networks," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 9, pp. 1975–1984, Sep. 2017.

[38] Y. Zhao, F. Ge, S. Zhang, and T. Liu, "3D deep convolutional neural network revealed the value of brain network overlap in differentiating autism spectrum disorder from healthy controls," in *Proc. MICCAI*, 2018, pp. 172–180.

[39] W. Liao *et al.*, "Preservation effect: Cigarette smoking acts on the dynamic of influences among unifying neuropsychiatric triple networks in schizophrenia," *Schizophrenia Bull.*, Dec. 2018. [Online]. Available: https://doi.org/10.1093/schbul/sby184