SPATIAL-TEMPORAL CONVOLUTIONAL ATTENTION FOR MAPPING FUNCTIONAL BRAIN NETWORKS

Yiheng Liu¹, Enjie Ge¹, Ning Qiang¹, Tianming Liu² and Bao Ge¹∗

¹School of Physics & Information Technology, Shaanxi Normal University, Xi’an, China
²Department of Computer Science, University of Georgia, GA, USA; * Corresponding Author.

ABSTRACT

Using functional magnetic resonance imaging (fMRI) and deep learning to explore functional brain networks (FBNs) has attracted many researchers. However, most of these studies still rely on temporal correlation between sources and voxel signals, lacking exploration of the dynamics of brain function. Due to the prevalent local correlations in volumes, FBNs can be directly generated in the spatial domain using spatial-wise attention (SA) in a self-supervised manner, resulting in higher spatial similarity with templates compared to classical methods. Therefore, we propose a novel Spatial-Temporal Convolutional Attention (STCA) model to dynamically discover FBNs using sliding windows. We validate the performance of our proposed method on the HCP-rest dataset, showing that STCA can be used to dynamically discover FBNs, offering a novel approach to better understand the human brain.

Index Terms— fMRI, Attention Mechanism, Functional Brain Network, Brain Function Dynamic

1. INTRODUCTION

Functional magnetic resonance imaging (fMRI) is a non-invasive imaging method to estimate brain function which indirectly detect the activities of brain function by measuring the blood-oxygen level dependence (BOLD) [1]. To explore the functional brain networks (FBNs), various analytical approaches have been proposed, each providing a different model for mapping brain functional pattern. Most of the previous methods for mapping FBNs using fMRI are based on the temporal correlation between the source and voxel signals, such as independent component analysis (ICA) [2] and sparse dictionary learning (SDL) [3]. Many of these approaches can be viewed as a blind source separation problem, the fMRI signals are modeled as a linear combination of the sources and the temporal patterns [4]. Although significant results have been achieved by these methods, the process of extracting the sources is still in a linearity way, which makes these classical methods limited by their shallow nature.

Recently, to overcome the shallow nature of the linear models, various of deep learning based methods have been proposed to discover the FBNs. Most of these methods are based on the autoencoders, they use different autoencoders to extract the sources in an self-supervised manner, and then use the generative linear model, such as LASSO to generate the FBNs [5]. In general, these deep learning based methods can indeed extract better encoder representations as the sources than the classical methods, such as ICA and SDL, but still generate FBNs in a linear and independent manner, with the sources extraction and the FBNs generation as 2 separate steps. Generating the FBNs in such way is time-consuming and does not fully utilize the advantages of deep learning, and cannot directly generate the FBNs with deep learning. In fact, the FBNs should be able to be represented at the feature extraction stage if we directly extract the spatial features, and generating the FBNs in a linear and independent way would require more longer source signals to increase confidence. At the same time, there is still a lack of methods for using deep learning to generate the dynamic FBNs. The mainstream methods usually adopt sliding window technology to generate the dynamic FBNs, they usually set the window size and the sliding step, and then input the data into commonly used classical models, such as ICA [6] and SDL [7]. In addition to sliding window approaches, there are also some methods tried to discover the spatial pattern in dynamic brain function, such as co-activation patterns (CAPs) [8]. The most widely used CAP model assumes that one CAP is active at each time point, and each CAP represents a spatial pattern of the source, usually clustering the CAP based on spatial similarity. Such a view oversimplified about the brain function and can only provides a limited amount of CAP.

To solve these problems, we proposed a novel deep learning based method called Spatial-Temporal Convolutional Attention (STCA) to directly generate the high quality dynamic FBNs by the sliding window technology. The STCA model is based on the convolutional neural network (CNN) architecture and trained in a self-supervised manner. The CNN architecture can make full use of the FBN’s prior knowledge which is the widespread local correlation in the fMRI signals, it is suitable for modeling the fMRI signals naturally. To validate the performance of the STCA model, we perform the experiments on the HCP-rest dataset which is a resting-state fMRI (rsfMRI) dataset. The re-
sult demonstrate that STCA can generate dynamic FBNs more better than the mainstream methods, such as ICA and SDL, the generated FBNs have higher spatial similarity to the templates. Source code of this work can be found in https://github.com/SNNUBIAI/STCAE.

2. METHODS

2.1. Dataset and preprocessing

In this paper, we adopt the Human Connectome Project (HCP) 900 Subjects Data Release as the experiments data. The HCP-rest [9] is a publicly accessible rsfMRI dataset (available on https://db.humanconnectome.org). Each individual contains 1200 time points, all of these data have been officially preprocessed and registered to the MNI152 4 × 4 × 4 standard template space. We masked the fMRI data and resampled it to a size of 49 × 58 × 47.

2.2. Spatial-Temporal Convolutional Attention (STCA)

The Fig 1 summarizes the proposed Spatial-Temporal Convolutional Attention (STCA) based pipeline for modeling the FBNs. Before going into the architectural details, we will briefly explain the design idea. Since the existing mainstream practice is to extract the sources through the encoder and then generate the FBNs with a generative linear model. Extracting sources and constructing FBN are two separate steps, the local correlation in the spatial and temporal pattern. The local correlation in the spatial and temporal pattern. The local correlation in the spatial and temporal pattern. The local correlation in the spatial and temporal pattern.

The goal of our task is to train a neural network:

\[ f : X_{\text{dyn}} \rightarrow S_{\text{dyn}} \]  

(1)

where \( X_{\text{dyn}} = (X_1, ..., X_t) \) is the sequence of the fMRI volumes with \( t \) time points and \( S_{\text{dyn}} \) is the weight maps of the volumes \( X_{\text{dyn}} \) with number of the weight maps \( N \), the depth of the weight map \( D \), the height of the weight map \( H \) and the width of the weight map \( W \). The STCA is an independent attention module, which cannot directly generate the weight maps in the absence of the labeled data. It needs to be trained in a self-supervised manner to guide the attention module to learn the FBNs. Autoencoders are commonly used self-supervised training methods, which compress the high-dimensional information to low-dimensional information to extract the sources. Guided by the autoencoder, the STCA learns to focus on the spatial patterns of the source to help the encoder generate better encoder representations as the sources, so that the spatial pattern of the sources can be represented at the feature extraction stage.

The STCA consists of a 3D convolutional layer, a channel-wise attention (CA) module and a spatial-wise attention (SA) module shown in Fig 1. The convolutional autoencoder is just to guide the SA module and does not involve in the evaluate stage. The convolutional layer and CA module are both designed to help extract more useful features, only the SA module is used to locate the regions where the FBNs are. We represent the GAP as global average pooling which will transform the 3D volumes to 1D vector, \( W \) as the fully connected layer, \( \text{sigmoid} \) as the activation function that can map the features to \([0, 1]\). The CA can be represented as:

\[ CA = \text{sigmoid}(\sigma(\sigma(GAP(f^l)W_1)W_2)) \]  

(2)

\[ \tilde{f}^l = f^l \circ CA \]  

(3)

We represent the 3D convolutional layer’s output features as \( f^l \in \mathbb{R}^{C \times D \times W \times H} \), \( \text{conv3D} \) as the 3D convolutional layer, \( \sigma \) as the activation function \( \text{GELU} \), \( \tilde{f}^l \) as final output of the SA module. The SA can be represented as:

\[ SA = \text{conv3D}(\sigma(\text{conv3D}(f^l))) \]  

(4)

\[ \tilde{f}^l = f^l \circ \text{sigmoid}(SA) \]  

(5)

2.3. Generation FBNs based on the STCA

We adopt the sliding window technique to discover the dynamic FBNs. Here, the window size is set as 40, stride is set as 1. Like the classical method, such as ICA and SDL, both training and testing are performed in the same individual. When training the STCA, the loss function of the autoencoder is mean square error (MSE), the learning rate is set as 0.0001, the optimizer is Adam and the epoch is set as 1. We convert SA to z-score, and the threshold is set as 3.6.

3. EXPERIMENTAL RESULTS

3.1. Overview of the weight maps derived by STCA

The representative weight maps derived from the STCA are shown in the upper panel of the Fig 2, these weight maps are all convert into z-score, the thresholded maps shown in the lower panel. It can be seen that regions activated with high weights can be easily explained by the known functional networks, such as visual network, default mode network, etc.
This result demonstrate that the spatial pattern of the sources can indeed be captured by the spatial-wise attention mechanism in the feature extraction stage which means that we provide a new paradigm for discovering the functional networks.

In Fig 3 and Table 1, the 10 RSN templates are in line 4. The results indicate that the proposed STCA can generate better FBNs than ICA.

3.2. Dynamic FBNs in resting state fMRI

The proposed STCA can be combined with sliding window technique to discover dynamic FBNs. Here, the window size is set as 40, stride is set as 1. We plot some representative functional network’s dynamic change process shown in Fig 4, which shows that some FBNs’ transferring processes. We can easily find that the RSN fade in or fade out over time. Quantitatively, we also plot the spatial similarity change between the FBNs generated by each window and 10 RSN templates.

Fig. 1. The architecture of the Spatial-Temporal Convolutional Attention. The STCA used the spatial and channel-wise attention in the spatial and temporal of the fMRI signals to discover the FBNs. The convolutional autoencoder module is used to guide the attention modules to focus on the activated regions in the training stage, it is not needed in the evaluate stage. The spatial-wise attention module’s weight map are transformed to z-score values, which represent the degree of functional activation on space, that is, the FBNs are derived. The results are close to the known FBNs derived by the classical methods, such as ICA and SDL.

Fig. 2. The weight and thresholded maps derived by STCA.

Fig. 3. Comparision with the classical methods, such as ICA and SDL. The templates are based on [10]
Table 1. IoU with templates in an individual in HCP-rest dataset.

| Template | ICA   | SDL   | STCA  |
|----------|-------|-------|-------|
| RSN 1    | 0.6273| 0.6053| 0.6505|
| RSN 2    | 0.4260| 0.4554| 0.5695|
| RSN 3    | 0.2414| 0.3340| 0.3446|
| RSN 4    | 0.2889| 0.3446| 0.4250|
| RSN 5    | 0.1932| 0.2635| 0.4474|
| RSN 6    | 0.0888| 0.2279| 0.3900|
| RSN 7    | 0.1179| 0.1992| 0.4284|
| RSN 8    | 0.0540| 0.1951| 0.3159|
| RSN 9    | 0.1843| 0.1890| 0.2628|
| RSN 10   | 0.2436| 0.1920| 0.2597|

over time, which are shown in Fig 5, we can also see the fade in and fade out process in form of IoU values.

Fig. 4. Some representative FBN’s dynamic change process.

Fig. 5. The spatial similarity change between the FBNs generated by each window and 10 RSN templates over time. The horizontal axis is the change of the time window, and the vertical axis is the IoU with the RSN template.

4. CONCLUSION

The proposed STCA model provides a novel framework for constructing FBNs using spatial-wise attention, which demonstrates superior performance in generating FBNs and provides a new tool for studying dynamic FBNs. In the future, we plan to investigate brain spatial state transitions using this model and conduct new experiments on task-based fMRI (tfMRI). Additionally, we aim to apply the proposed STCA to functional connectivity (FC) analysis and disease classification.

5. ACKNOWLEDGMENTS

This research was supported by the National Natural Science Foundation of China (No. 61976131).

6. REFERENCES

[1] Nikos K Logothetis, “What we can do and what we cannot do with fmri,” Nature, vol. 453, no. 7197, pp. 869–878, 2008.

[2] M. J. McKeown, “Detection of consistently task-related activations in fmri data with hybrid independent component analysis,” Neuroimage, vol. 11, no. 1, pp. 24–35, 2000.

[3] Jinglei Lv et al., “Sparse representation of whole-brain fmri signals for identification of functional networks,” Medical Image Analysis, vol. 20, no. 1, pp. 112–134, 2015.

[4] Armin Iraji et al., “Space: a missing piece of the dynamic puzzle,” Trends in Cognitive Sciences, vol. 24, no. 2, pp. 135–149, 2020.

[5] Mengshen He et al., “Multi-head attention-based masked sequence model for mapping functional brain networks,” in International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, 2022, pp. 295–304.

[6] Vesa Kiviniemi et al., “A sliding time-window ica reveals spatial variability of the default mode network in time,” Brain connectivity, vol. 1, no. 4, pp. 339–347, 2011.

[7] Bao Ge et al., “Discovering and characterizing dynamic functional brain networks in task fmri,” Brain Imaging and Behavior, vol. 14, no. 5, pp. 1660–1673, 2020.

[8] Xiao Liu, Catie Chang, and Jeff H Duyn, “Decomposition of spontaneous brain activity into distinct fmri co-activation patterns,” Frontiers in systems neuroscience, vol. 7, pp. 101, 2013.

[9] David C Van Essen et al., “The wu-minn human connectome project: an overview,” Neuroimage, vol. 80, pp. 62–79, 2013.

[10] S. M. Smith et al., “Correspondence of the brain’s functional architecture during activation and rest,” Proceedings of the National Academy of Sciences, vol. 106, no. 31, pp. 13040–13045, 2009.