



# NETWORK NEURO SCIENCE

an open access  journal



Citation: Wang, L., Li, K., & Hu, X. P. (2021). Graph convolution network for fMRI analysis based on connectivity neighborhood. *Network Neuroscience*, 5(1), 83–95. [https://doi.org/10.1162/netn\\_a\\_00171](https://doi.org/10.1162/netn_a_00171)

DOI:  
[https://doi.org/10.1162/netn\\_a\\_00171](https://doi.org/10.1162/netn_a_00171)

Supporting Information:  
[https://github.com/Lebo-Wang/cGCN\\_fMRI](https://github.com/Lebo-Wang/cGCN_fMRI)

Received: 14 May 2020  
Accepted: 24 September 2020

Competing Interests: The authors have declared that no competing interests exist.

Corresponding Author:  
Xiaoping P. Hu  
[xhu@engr.ucr.edu](mailto:xhu@engr.ucr.edu)

Handling Editor:  
Andrew Zalesky

Copyright: © 2020  
Massachusetts Institute of Technology  
Published under a Creative Commons  
Attribution 4.0 International  
(CC BY 4.0) license



## METHODS

# Graph convolutional network for fMRI analysis based on connectivity neighborhood

Lebo Wang<sup>1</sup>, Kaiming Li<sup>2</sup>, and Xiaoping P. Hu<sup>1,2</sup>

<sup>1</sup>Department of Electrical and Computer Engineering, University of California, Riverside, Riverside, CA, USA

<sup>2</sup>Department of Bioengineering, University of California, Riverside, Riverside, CA, USA

**Keywords:** Functional connectivity, Deep learning, Graph convolutional network, Connectivity-based neighborhood

## ABSTRACT

There have been successful applications of deep learning to functional magnetic resonance imaging (fMRI), where fMRI data were mostly considered to be structured grids, and spatial features from Euclidean neighbors were usually extracted by the convolutional neural networks (CNNs) in the computer vision field. Recently, CNN has been extended to graph data and demonstrated superior performance. Here, we define graphs based on functional connectivity and present a connectivity-based graph convolutional network (cGCN) architecture for fMRI analysis. Such an approach allows us to extract spatial features from connectomic neighborhoods rather than from Euclidean ones, consistent with the functional organization of the brain. To evaluate the performance of cGCN, we applied it to two scenarios with resting-state fMRI data. One is individual identification of healthy participants and the other is classification of autistic patients from normal controls. Our results indicate that cGCN can effectively capture functional connectivity features in fMRI analysis for relevant applications.

## INTRODUCTION

In recent years, there has been increasing interest in studying the brain via resting-state functional magnetic resonance imaging (rs-fMRI). Without any task or external stimulus, rs-fMRI can capture the spontaneous brain activity that reflects the functional organization of the brain. Based on rs-fMRI data, functional connectivity (FC) has been calculated using correlations between spatially distant regions within/across functional networks (Rogers, Morgan, Newton, & Gore, 2007; van den Heuvel & Pol, 2010). FC is believed to reflect the functional organization of the brain and has been used as a fingerprint to identify individuals (Finn et al., 2015). Alterations in FC have been associated with psychiatric disorders (Bullmore & Sporns, 2009; Greicius, Supekar, Menon, & Dougherty, 2009; van den Heuvel & Pol, 2010), demonstrating FC's potential as biomarkers in clinical neuroscience. Various methods have been introduced to analyze the spatial FC pattern with fMRI data, including correlation-based approaches (Rogers et al., 2007; van den Heuvel & Pol, 2010), graph-based methods (Keown et al., 2017; Lee, Smyser, & Shimony, 2013; Wang, Zuo, & He, 2010), and matrix factorization techniques (Andersen, Gash, & Avison, 1999; van de Ven, Formisano, Prvulovic, Roeder, & Linden, 2004). The temporal pattern of FC has also been analyzed with several methods, including sliding-window analysis (Sakoğlu et al., 2010), time-frequency analysis (Chang & Glover, 2010), and the Gaussian hidden Markov model (Chen, Langley, Chen, & Hu, 2016). With the increasing

rs-fMRI:  
Resting-state functional magnetic resonance imaging.

FC:  
Functional connectivity.

CNN:  
Convolutional neural network.

availability of large fMRI datasets (Craddock et al., 2013; Miller et al., 2016; Van Essen et al., 2013), deep learning-based methods are becoming more widely used in fMRI analysis.

#### **Deep Learning on fMRI Data**

Deep learning architectures, especially convolutional neural networks (CNN), have achieved remarkable performance in many applications, including image classification (He, Zhang, Ren, & Sun, 2016), image segmentation (Ronneberger, Fischer, & Brox, 2015), and machine translation (Vaswani et al., 2017). These applications mostly utilized automatic feature extraction via an end-to-end training paradigm on structured data (e.g., images and text). Similar strategies were extended to neuroimaging data, and promising results were obtained in several applications, including Alzheimer patient classification (Sarraf & Tofighi, 2016), subcortical brain structure segmentation (Dolz, Desrosiers, & Ayed, 2018), and segmentation of longitudinal structural MRI (Gao et al., 2018). However, there are several limitations to model fMRI data on the perspective of image grids: (a) It is computationally intensive to deal with voxelwise fMRI data via CNN models; (b) Brain activity occurs mostly in cortical and subcortical structures, making convolutions on white matter unnecessary; (c) The fMRI time course from a single voxel is usually very noisy; and (d) spatial features may be confined to a small neighborhood in the Euclidean space, especially in shallow CNN models. Therefore, it is of great necessity to build deep learning models that are more appropriate for the organization of the brain and that can efficiently extract connectomic features beyond a single voxel and its Euclidean neighborhood.

#### **Deep Learning on FC Matrix**

To reduce the spatial complexity and noise in individual voxel data, many studies have turned to approaches based on regions of interest (ROIs). By calculating the correlation coefficients between pairs of ROIs based on the whole scan, the ROI-derived FC matrix reveals the temporal correlation pattern of ROIs. Because of the grid structure of the 2D matrix, the FC matrix shows great compatibility with traditional deep learning models. Therefore, the FC matrix has been directly adopted as inputs of deep learning approaches in several studies. For instance, Suk, Wee, Lee, and Shen (2016) introduced a deep learning architecture to investigate functional dynamics for mild cognitive impairment. Heinsfeld, Franco, Craddock, Buchweitz, and Meneguzzi (2018) identified autism spectrum disorder (ASD) based on autoencoders with the FC matrix as input data. However, the FC matrix describes the linear temporal relationship between ROIs but does not account for the rich temporal dynamic information in the time courses. Our prior work with recurrent neural network (RNN) demonstrated that the spatiotemporal information can be beneficial for individual identifications (Chen & Hu, 2018; L. Wang, Li, Chen, & Hu, 2019). The initial work applied a fully connected model to extract spatial features from ROI data and used RNN for temporal evolutions (Chen & Hu, 2018). In a follow-up study, spatial features among connectomic ROIs were extracted by convolutional layers (L. Wang et al., 2019). It showed that the identification accuracy increased with a larger number of input frames. The superior performance of the latter study suggests that proper convolution between ROIs for spatial features provides extra boosts for identification accuracy. However, it may be suboptimal to apply convolutions based on the predefined neighborhood that is defined by the functional atlas.

#### **Graph Convolutional Networks**

Motivated by breakthroughs of deep learning on grid data, efforts have been made to extend CNN to graphs, a natural way to represent many forms of data including fMRI data. Two categories of graph convolutional networks (GCNs)—spectral GCNs and spatial GCNs—have

ROI:  
Region of interest.

ASD:  
Autism spectrum disorder.

RNN:  
Recurrent neural network.

GCN:  
Graph convolutional network.

been proposed. For spectral GCNs, graphs can be decomposed into spectral bases associated with graph-level information according to spectral graph theory (Bruna, Zaremba, Szlam, & LeCun, 2013). In contrast, spatial GCNs imitate the Euclidean convolution on grid data to aggregate spatial features between neighboring nodes. Although spectral GCNs have achieved great success on both structural and functional MRI applications (Gopinath, Desrosiers, & Lombaert, 2019; Hong et al., 2019; Ktena et al., 2018; Parisot et al., 2017), spatial models are preferred over the spectral ones because of their efficiency, generalization, and flexibility (Monti et al., 2017; Wu et al., 2019; Zhang, Cui, & Zhu, 2018), and they have gained increasing interest in the community (Azevedo, Passamonti, Liò, & Toschi, 2020; Gadgil et al., 2020).

cGCN:  
Connectivity-based graph convolutional network.

$k$ -NN:  
 $k$ -nearest neighbors.

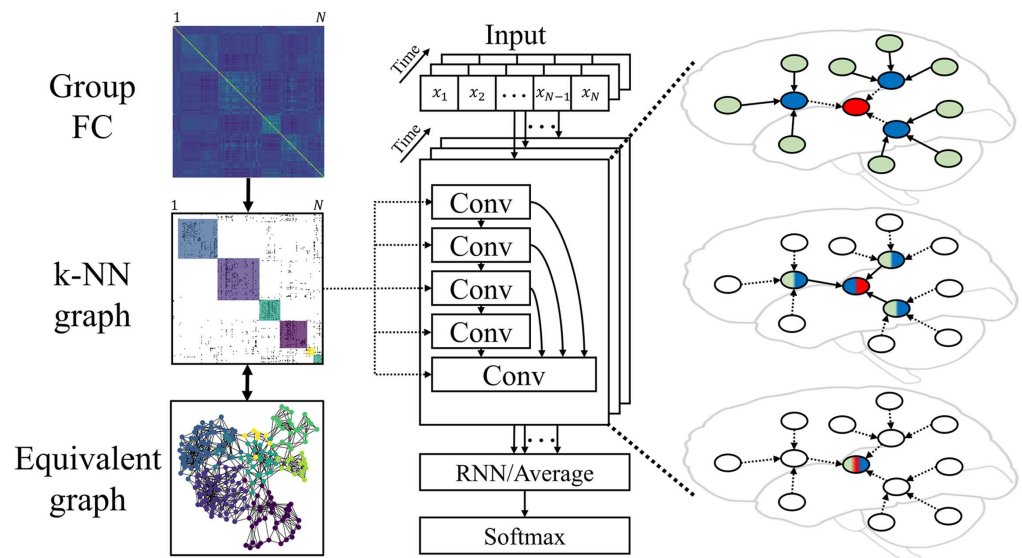
In this paper, we present a connectivity-based graph convolutional network (cGCN), a spatial GCN architecture, for fMRI analysis. The graph representation was defined with the  $k$ -nearest neighbor ( $k$ -NN) graph based on FC matrix. Convolutions were performed on graphs rather than on an image/ROI grid, which allows us to efficiently extract connectomic features of the underlying brain activity. The present model (L. Wang, 2020; [https://github.com/Lebo-Wang/cGCN\\_fmri](https://github.com/Lebo-Wang/cGCN_fmri)) has been applied in two scenarios, and the results indicate that cGCN is effective for fMRI analysis compared with traditional deep learning architectures.

## MATERIALS AND METHODS

### cGCN Overview

BOLD:  
Blood oxygen level-dependent.

The overview of our cGCN architecture is shown in Figure 1. ROIs were considered to be graph nodes with the blood oxygen level-dependent (BOLD) signals of each frame as their attributes. Convolutions were performed within neighbors defined by the  $k$ -NN graph based on the groupwise FC matrix. Specifically, we first obtained the individual FC matrices on training



**Figure 1.** Overview of cGCN. On the left, the graph definition for cGCN was based on the group FC from all training data, which can be further simplified as a  $k$ -nearest neighbors ( $k$ -NN) graph with binarized edges. In the middle, the cGCN architecture consisted of 5 convolutional layers. The convolutional neighborhood was defined by the shared  $k$ -NN graph across convolutional layers, time frames, and subjects. The recurrent neural network (RNN) layer (or the temporal average pooling layer) obtained latent representations from all frames. The final classification was achieved by the Softmax layer. On the right, an intuitive illustration of the spatial graph convolution showed the information aggregation between neighboring nodes.

data and averaged them for the group FC matrix. Based on that, a  $k$ -NN graph was generated by retaining only the top  $k$  edges in terms of their connectivity strength (i.e., average correlation coefficient) for each node. The FC-based  $k$ -NN graph was used to guide the convolutional operations with functional connectivity-based neighborhoods. For simplicity, the same graph was shared by all subjects and at all time frames. Although each node had a few neighbors, the convolution field of each layer was extended by stacking multiple convolutional layers in the architecture. Between convolutional layers, skip connections were added from the prior convolutional layers to the last one, providing multilevel feature fusion for classification and accelerating the model training by alleviating the vanishing gradient problem. Outputs from the convolutional layers were followed by an RNN (or a temporal average pooling) layer to generate temporal evolutions by combining spatial representations from all frames. A Softmax layer was used at the end for the final classification.

### Graph Construction

We represented each frame of fMRI data with a graph. A shared graph structure that reflects the intrinsic functional connectivity was derived from the group FC matrix (Petersen & Sporns, 2015), based on training data only. The ROIs were considered to be nodes of the graph, and the FC connections were considered edges of the graph. To reduce the total number of edges in the graph, a  $k$ -NN graph was obtained by keeping only the top  $k$  connectivity neighbors for each node in terms of the connectivity strength;  $k$  is the hyperparameter related to the topological structure of graphs, which controls the sparsity of the graph. To evaluate the effect of  $k$ , different values of  $k$  (3, 5, 10, and 20) were assessed in our experiments.

### The Edge Function of cGCN

The  $k$ -NN graph  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$  comprises nodes  $\mathcal{V} = \{1, \dots, N\}$  and edges  $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ . Edge  $(i, j)$  represents the directed edge from  $ROI_i$  to  $ROI_j$ . To explicitly model the spatial pattern between ROIs, we chose EdgeConv, an asymmetric edge function (Y. Wang et al., 2018), as the convolutional operation for our cGCN convolutional layers:

$$x'_i = \max_{j:(i,j) \in \mathcal{E}} h_{\Theta}(x_i \parallel x_j - x_i),$$

where  $x_i$  was the BOLD signal of the central node  $i$  and  $x_j$  was that of a connected neighbor node  $j$ . In addition to original features from the central node, the node value difference  $x_j - x_i$  was also appended as the complementary feature.  $h_{\Theta}$  represents trainable weights  $\Theta$  with a nonlinear activation function  $h$ , which was implemented by the multilayer perceptron (MLP) with the rectified linear unit (ReLU). The feature-wise convolution output,  $x'_i$ , was the maximal activation from its top  $k$  neighbors. In this way, spatial features between connectivity-based neighbors were modeled, including the node activity and the coactivation pattern.

### Experiments and Settings

Two supervised classification experiments were carried out to evaluate the performance of our proposed architecture. In the first experiment, we used the “100 unrelated subjects” dataset (54 females, age: 22–36) released by the Human Connectome Project (HCP) and aimed to identify them based on their rs-fMRI data. Each subject had four rs-fMRI sessions (each with 1,200 volumes) scanned on two days (Van Essen et al., 2013). The fMRI data were preprocessed by the HCP minimal preprocessing pipeline (Glasser et al., 2013) and denoised by ICA-FIX (Salimi-Khorshidi et al., 2014) to remove spatial artifacts and motion-related fluctuations.

Surface-based registration was performed with the MSM-ALL method (Robinson et al., 2014). We utilized 236 ROIs based on the Power atlas (Power et al., 2011). The two sessions from Day 1 were employed as the training dataset, and the two sessions from Day 2 were used as the validation dataset and test dataset, respectively. The best model was chosen according to the best validation accuracy and final identification accuracy was assessed on the test dataset. For the training and validation datasets, fMRI time courses were cut into 100-frame clips. The final identification performance was measured with different numbers of frames (from 1,200 frames to a single frame) on the test dataset. To evaluate the contribution of the connectivity-based neighborhood for convolutions, we also ascertained the performance of the same cGCN architecture with random graphs (random GCNs).

In the second experiment, we used cGCN to classify ASD patients from healthy controls on the ABIDE (Autism Brain Imaging Data Exchange) dataset (Di Martino et al., 2014). There are 1,057 subjects (525 ASD subjects and 532 neurotypical controls) from 17 imaging sites. All fMRI data were preprocessed by the Connectome Computation System pipeline (Zuo et al., 2013), with bandpass filtering (0.01–0.1 Hz) and without global signal regression. The Craddock 200 atlas (Craddock, James, Holtzheimer, Hu, & Mayberg, 2012) was utilized to extract ROI signals. We adopted both leave-one-site-out and tenfold cross-validations. For the leave-one-site-out cross-validation, data from each site were independently tested with the model trained on data from other sites, which evaluated the model on heterogeneous datasets considering the site-specific variation. The tenfold cross-validation mixed all data together and split them into different folds by keeping the proportions of sites and diagnostic groups across folds. The site-specific heterogeneity was overlooked by the random partition of subsamples.

We carried out the two experiments in Keras (Chollet, 2015) using TensorFlow as the back end (Abadi et al., 2016). Adam optimizer was applied to update model weights with adaptive learning rates (Kingma & Ba, 2014). Stepwise learning rate decay was also used if the validation accuracy stopped increasing, with the smallest learning rate of  $1e-6$ . The model was evaluated on the validation dataset after each training epoch, and the model parameters were saved only if better validation accuracy was achieved. During training, the L2 regularization was used to avoid overfitting. The final performance was reported with the highest accuracy among different L2 values (0.1, 0.01, 0.001, and 0.0001).

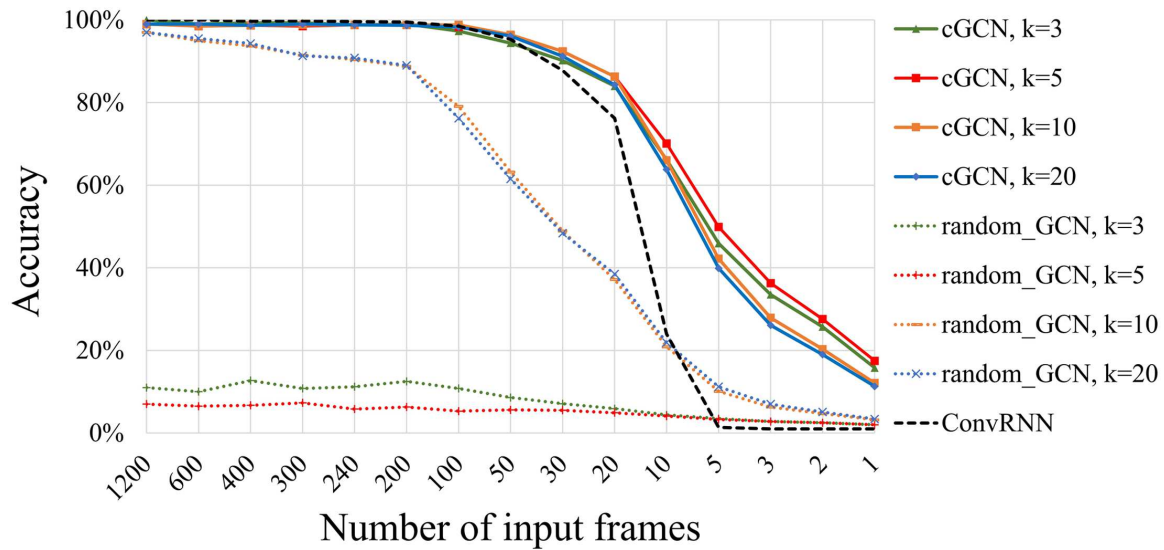
### **Visualization**

The occlusion method (Zeiler & Fergus, 2014) was utilized to visualize the important ROIs for classification in both experiments. With trained model parameters, each ROI of input data was zeroed out one at a time, and the performance degradation under the same model configuration was considered as the contribution of the ROI to the classification task. The occlusion pattern was mapped onto the cortical surface for visualization. We projected ROI data to the cortical surface and obtained the surface renditions. For the individual identification task, the occlusion pattern was generated and averaged on the test dataset. For ASD classification, we utilized the pretrained models from the leave-one-site-out cross-validation and averaged the pattern of performance degradation on each leave-out dataset to get the visualization pattern.

## **RESULTS**

### ***Performance Related to the Number of Neighbors***

The hyperparameter  $k$  determines the number of edges in graphs. The classification performance was evaluated for a range of  $k$  values (3, 5, 10, and 20). In Figure 2, we show the performance of individual identification. The cGCN with  $k$  of 5 achieved the highest



**Figure 2.** Performance on individual identification regarding the number of neighbors and the number of input frames. The highest mean classification accuracy was achieved with  $k = 5$ . With 20 frames or fewer as input data, cGCN achieved significantly higher classification accuracy compared with random GCNs and the convolutional RNN (ConvRNN) model in our prior study (L. Wang et al., 2019). Particularly with 5 frames, cGCNs obtained the identifying accuracy of over 49% with  $k = 5$ , which was much higher than random GCNs and ConvRNN. It demonstrated that spatiotemporal features from each frame of fMRI data was successfully extracted by cGCNs for the individual identification task.

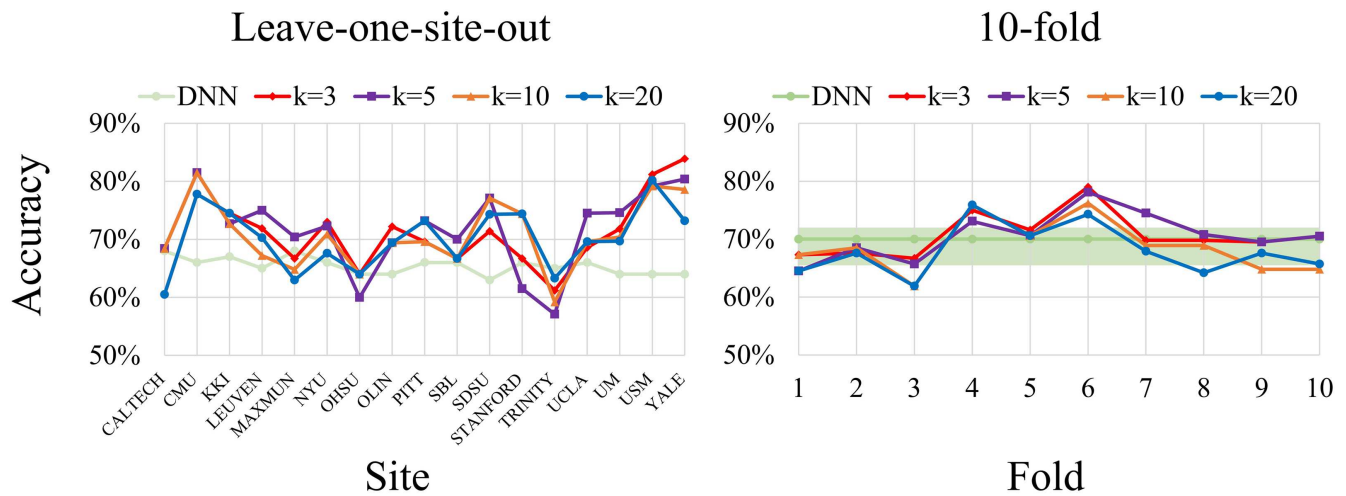
identification accuracy on average. Increasing the  $k$  value beyond 5 ( $k = 10$  or  $20$ ) led to diminished performance. The performance of random GCNs was significantly lower than cGCN models, especially for small  $k$  values ( $k = 3$  or  $5$ ).

For ASD classification, the performance with leave-one-site-out and tenfold cross-validation are depicted in Figure 3. With the leave-one-site-out cross-validation, the highest mean accuracy was 71.6% when  $k = 5$ , and the lowest mean accuracy was 70.1% when  $k = 20$ . For the tenfold cross-validation, the highest mean accuracy was 70.7% when  $k = 3$ , and the lowest mean accuracy was 68.0% for cGCN with  $k = 20$ . On both cross-validations, cGCN with large  $k$  values ( $k = 10$  or  $20$ ) had relatively lower classification accuracy.

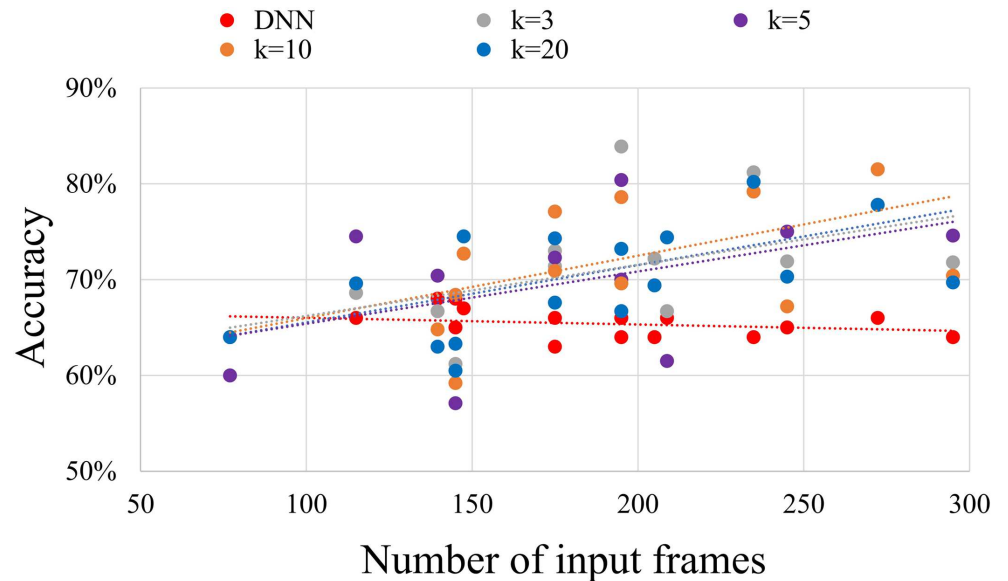
#### Performance Related to the Number of Input Frames

For the individual identification application, we evaluated cGCNs and random GCNs with different numbers of frames as inputs. With cGCN models, the improved performance of individual identification was shown with increasing numbers of fMRI frames. As shown in Figure 2, cGCN achieved better performance with a larger number of input frames, and the performance gradually saturated at approximately 100 frames. Random GCNs exhibited a similar pattern for large  $k$  values ( $k = 10$  or  $20$ ), but with accuracy much lower than that of cGCNs. However, the performance of convolutional RNN (ConvRNN) dropped dramatically with reducing frames, close to the random guess (1%) with fewer than 5 frames. In contrast, cGCN still obtained an identifying accuracy of  $>49\%$  with 5 frames when  $k = 5$ .

Similarly, for the ASD classification task shown in Figure 4, better classification accuracy can be achieved with a larger number of input frames for different imaging sites in the leave-one-site-out cross-validation. As a contrast, the number of input frames did not significantly affect the classification accuracy of the deep neural network (DNN) model, in which FC matrices were used as inputs without considering the temporal dimension (Heinsfeld et al., 2018).



**Figure 3.** Performance on ABIDE datasets with the leave-one-site-out cross-validation and the tenfold cross-validation. The best average accuracy with the leave-one-site-out cross-validation on ABIDE dataset was 71.6% (min: 57.1%, max: 81.5%) when  $k = 5$ . As a comparison, the DNN model achieved mean accuracy of 65.4% (min: 63%, max: 68%; Heinsfeld et al., 2018). Except for some imaging sites (CALTECH, MAX\_MUN, OHSU, and TRINITY), cGCNs obtained distinct performance improvement compared with the DNN model. The best average accuracy with the tenfold cross-validation was 70.7% (min: 66.7%, max: 79.0%) when  $k = 3$ , which outperformed the DNN model of 70% (min: 66%, max: 71%).



**Figure 4.** The relationship between the classification accuracy and the number of frames as inputs for the ASD classification task with the leave-one-site-out cross-validation. The average classification accuracy increased linearly with the number of input frames, while the number of input frames did not affect the classification accuracy for the deep neural network model with FC matrices as inputs (Heinsfeld et al., 2018).

**Comparison**

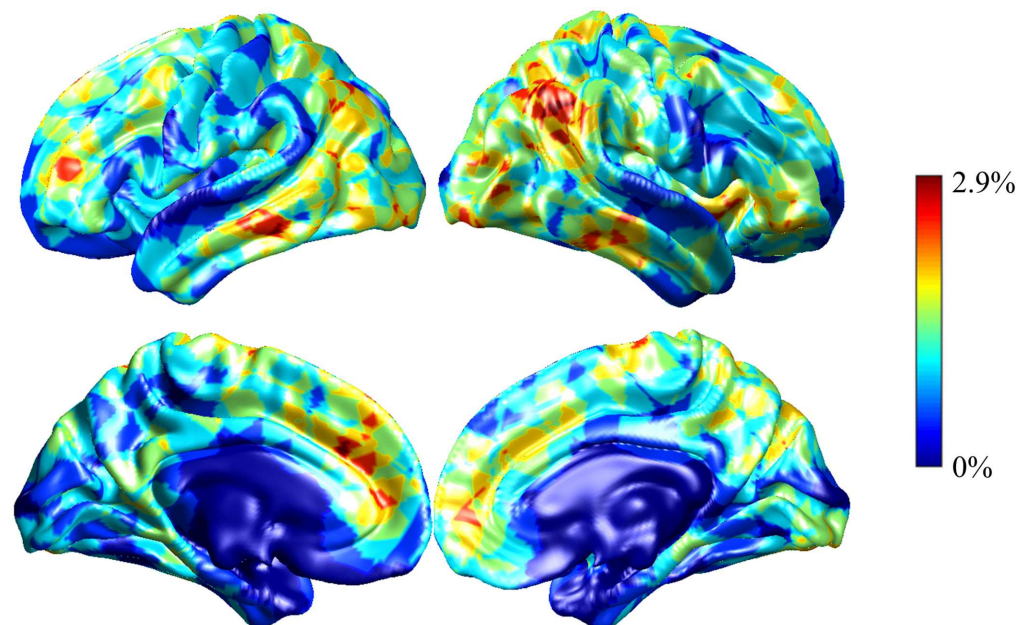
For the individual identification task, cGCN achieved an accuracy of 98.8% with  $k = 10$  when the number of input frames was fixed to 100. It outperformed several prior architectures, including the RNN model of 94.4% (Chen & Hu, 2018), the ConvRNN model of 98.5% (L. Wang et al., 2019), and the traditional correlation method of around 70% (Finn et al., 2015). In

particular, cGCNs showed overwhelming superiority over the traditional CNN model with fewer than 20 frames. As shown in Figure 2, the previous ConvRNN model with only 5 frames of input data led to an accuracy close to the random guess (1%).

For the ASD classification task with the leave-one-site-out cross-validation, the best mean classification accuracy was 71.6% (min: 57.1%, max: 81.5%) when  $k = 5$  as shown in Figure 3. Except for some imaging sites (CALTECH, MAX\_MUN, OHSU, and TRINITY), our cGCNs obtained significant performance improvement compared with the DNN model, whose mean accuracy was 65.4% (min: 63%, max: 68%; Heinsfeld et al., 2018). With the tenfold cross-validation, our best average performance was 70.7% (min: 66.7%, max: 79.0%) when  $k = 3$ , which is better than the RNN model of 68.5% (Dvornek, Ventola, Pelphrey, & Duncan, 2017) and the DNN model of 70% (min: 66%, max: 71%; Heinsfeld et al., 2018).

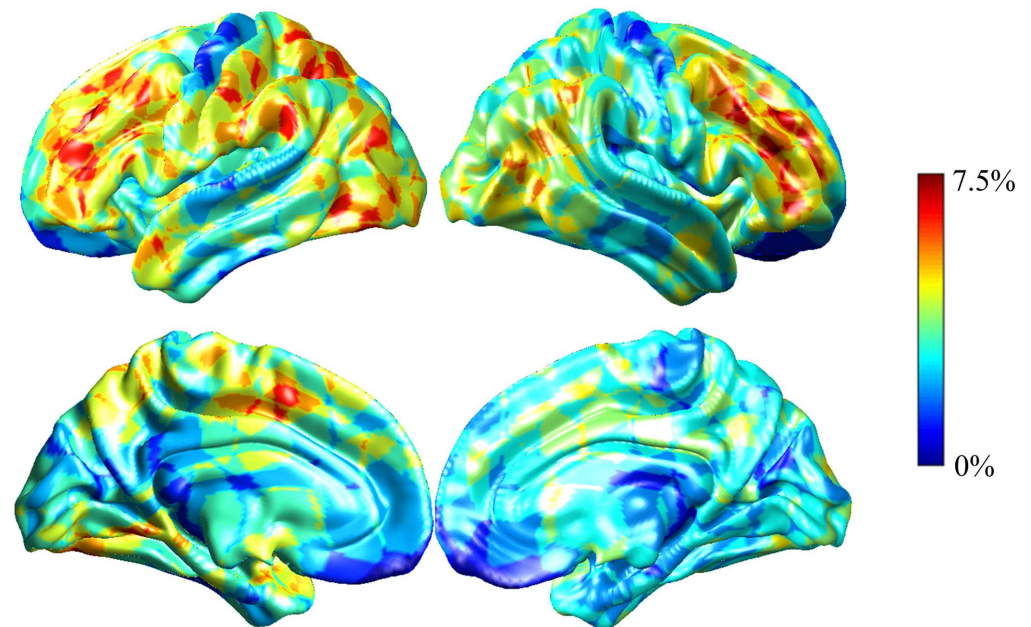
### Visualization

We applied the occlusion method to identify informative regions for two classification tasks and visualized them in Figures 5 and 6. For the individual identification task, the smallest accuracy degradation was 0.6% when  $k = 20$  and the largest accuracy drop was 2.9% when  $k = 3$ . Significant performance degradation was observed when some ROIs were individually occluded, while some regions did not suffer from any performance degradation. The salient regions for cGCN models with different  $k$  values were similar. The significant resting-state networks were default mode network (DMN), frontoparietal network (FPN), as well as visual network (VN).



**Figure 5.** Visualization of the performance degradation through the single-ROI occlusion for the individual identification. The red region reflected large performance degradation if corresponding ROIs were occluded. The performance degradation reflects the relative contribution of each ROI. When  $k = 20$ , the smallest performance drop was only 0.6%, while the largest performance drop of 2.9% was obtained with  $k = 3$ . Default mode network, frontoparietal network, and visual network contributed more to the individual identification.





**Figure 6.** Visualization of the performance degradation through the single-ROI occlusion for the ASD classification. We averaged performance drop on all models used in the leave-one-site-out cross-validation. The maximum performance drop was 7.5% when  $k = 3$ , and the smallest performance drop was 6.0% when  $k = 20$ . The color of the region reflected the performance degradation if corresponding ROI was occluded. The salient regions related to the ASD classification included frontoparietal network, default mode network, and ventral attention network.

The visualization of cGCN for the ASD classification task is shown in Figure 6. The largest performance drop was 7.5% when  $k = 3$ , and the smallest performance drop was 6.0% when  $k = 20$ . The salient regions for cGCN models with different  $k$  values were similar. The salient networks identified by cGCN included FPN, DMN, and ventral attention network.

## DISCUSSION

We presented cGCN architecture for fMRI analysis and applied cGCN on two classification tasks with rs-fMRI data, that is, the individual identification and classification of ASD, respectively. The superior performance compared with prior studies demonstrated that cGCN can effectively capture spatial features of fMRI data within connectomic neighbors.

A major contribution of the present work is that rather than performing convolution on structured image grids, ROI-based fMRI data are considered graphs based on FC, and cGCN is carried out between connectomic neighbors on the graph. Previous studies utilized fully connected layers for the spatial feature extraction and ignored the brain's functional organization. In addition, it is difficult to train fully connected models with good generalization because of overfitting. Our prior paper demonstrated that it was feasible to capture spatial patterns on a small batch of ROIs from the same FC network by convolutions (L. Wang et al., 2019), although ROIs were artificially arranged in a predefined order according to the atlas. In the present work, the convolutional neighborhood was defined by the group-level FC matrix. Each ROI was involved in the convolution with its top connectivity-based neighbors, and high-level features were extracted by stacking convolutional layers hierarchically. The superior performance of cGCN over random GCN and previously used methods demonstrated that

connectivity-based neighborhood for convolutions was a significant advantage for the cGCN architecture.

We adopted the  $k$ -NN graph rather than hard thresholding FC to reduce the size of the neighborhood and effects of noise in connectivity (Liu, Nalci, & Falahpour, 2017; Murphy & Fox, 2017). The advantages of  $k$ -NN graphs lie in the following aspects. First,  $k$ -NN graphs naturally have good local homogeneity with the same number of edges originating from each node, appropriate for convolutions. Second, hierarchical feature extraction can be easily achieved with stacking layers under the guidance of the  $k$ -NN graph.

Like the convolutional kernel size in the traditional CNN models,  $k$  explicitly determines the convolutional region on graphs, as well as the computational complexity for the cGCN model. We tested the performance of our cGCN architecture with different  $k$  values. The best performance of the individual identification was achieved when  $k = 5$ , while the highest classification accuracy on ASD with two types of cross-validations was achieved when  $k = 3$  or 5. For both tasks, increasing  $k$  did not significantly improve the classification accuracy. One possible reason is that convolution on too many neighbors might fail to generate local features with good generalization. The same situation also happened in the modeling of 3D objects (Y. Wang et al., 2018). Therefore, we suggest a  $k$  value of 5 or less for good performance, although the exact optimal value may be application dependent.

Unlike  $k$ , increasing the number of input frames significantly improved the performance for both tasks. This effect was especially obvious with a few input frames in the beginning and gradually leveled off with over 100 frames for individual identification. This result indicates that cGCN can efficiently extract spatial patterns between functionally “adjacent” nodes from each frame of fMRI data in spite of substantial intersubject and intersession variabilities that arise because of hardware variations, and physiological variations (McGonigle et al., 2000). Thus, it is beneficial to keep the temporal axis of the fMRI data and apply framewise feature extraction based on fMRI time course rather than the FC matrix.

There are some appealing properties by using the asymmetric edge function for convolutions according to Y. Wang et al. (2018). First, the spatial features of both global information  $x_i$  and local interactions  $x_j - x_i$  are captured by the asymmetric edge function. Second, the long-distance characterization of spatial features between multi-hop neighborhoods can be captured by the stacked convolutional structures, based on the topological structure of the graph.

Furthermore, the single-ROI occlusion experiments revealed the salient regions related to individual identification and classification of ASD tasks, which are in agreement with those seen in previous studies (Finn et al., 2015; Heinsfeld et al., 2018; Nielsen et al., 2013; L. Wang et al., 2019). For the individual identification task, the performance drop was much smaller than that of the ConvRNN model (L. Wang et al., 2019), suggesting that the graph-based representation of fMRI data builds robust relationships between ROIs, and missing values of any ROI might be compensated for by its functional neighbors based on the topological structure of graphs.

Some limitations should be noted for future work. First, the graph topology is invariant from layer to layer, frame to frame, and subject to subject. Considering the significant variability in FC, future work shall adopt dynamic update across convolutional layers, time frames, and subjects. Second, the graph in the present work reflects the brain’s organization only from the functional perspective. Other image modalities, such as diffusion MRI, could be utilized

to build graphs that reflect the brain's structural connectivity. Third, intuitive visualization of temporal features captured by cGCN is difficult. This warrants further investigation in future studies.

## CONCLUSION

In this paper, we describe a connectome-defined neighborhood for graph convolution to extract connectomic features from rs-fMRI data for classification. Our model allows for spatial feature extraction within connectomic neighborhoods rather than Euclidian ones. Significant improvement on individual identification and ASD classification tasks suggests that the cGCN model is effective in capturing connectomic features from fMRI data and is promising for fMRI analysis.

## AUTHOR CONTRIBUTIONS

Lebo Wang: Conceptualization; Methodology; Software; Validation; Visualization; Writing - Original Draft. Kaiming Li: Formal analysis; Methodology; Writing - Original Draft; Writing - Review & Editing. Xiaoping Hu: Conceptualization; Formal analysis; Methodology; Project administration; Resources; Writing - Original Draft; Writing - Review & Editing.

## REFERENCES

- Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., . . . Isard, M. (2016). Tensorflow: A system for large-scale machine learning. *Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI, 2016)*, 265–283.
- Andersen, A. H., Gash, D. M., & Avison, M. J. (1999). Principal component analysis of the dynamic response measured by fMRI: A generalized linear systems framework. *Magnetic Resonance Imaging*, 17(6), 795–815. DOI: [https://doi.org/10.1016/S0730-725X\(99\)00028-4](https://doi.org/10.1016/S0730-725X(99)00028-4)
- Azevedo, T., Passamonti, L., Liò, P., & Toschi, N. (2020). Towards a predictive spatio-temporal representation of brain data. *arXiv: 2003.03290*.
- Bruna, J., Zaremba, W., Szlam, A., & LeCun, Y. (2013). Spectral networks and locally connected networks on graphs. *arXiv:1312.6203*.
- Bullmore, E., & Sporns, O. (2009). Complex brain networks: Graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience*, 10(3), 186. DOI: <https://doi.org/10.1038/nrn2575>, PMID: 19190637
- Chang, C., & Glover, G. H. (2010). Time–frequency dynamics of resting-state brain connectivity measured with fMRI. *NeuroImage*, 50(1), 81–98. DOI: <https://doi.org/10.1016/j.neuroimage.2009.12.011>, PMID: 20006716, PMCID: PMC2827259
- Chen, S., & Hu, X. P. (2018). Individual identification using functional brain fingerprint detected by recurrent neural network. *Brain Connectivity*, 8(4), 197–204. DOI: <https://doi.org/10.1089/brain.2017.0561>, PMID: 29634323
- Chen, S., Langley, J., Chen, X., & Hu, X. (2016). Spatiotemporal modeling of brain dynamics using resting-state functional magnetic resonance imaging with Gaussian hidden Markov model. *Brain Connectivity*, 6(4), 326–334. DOI: <https://doi.org/10.1089/brain.2015.0398>, PMID: 27008543
- Chollet, F. (2015). Keras, GitHub, <https://github.com/fchollet/keras>
- Craddock, C., Benhajali, Y., Chu, C., Chouinard, F., Evans, A., Jakab, A., . . . Milham, M. (2013). The Neuro Bureau Pre-processing Initiative: Open sharing of preprocessed neuroimaging data and derivatives. *Frontiers in Neuroinformatics, Conference Abstract: Neuroinformatics 2013*. DOI: <https://doi.org/10.3389/conf.fninf.2013.09.00041>
- Craddock, R. C., James, G. A., Holtzheimer III, P. E., Hu, X. P., & Mayberg, H. S. (2012). A whole brain fMRI atlas generated via spatially constrained spectral clustering. *Human Brain Mapping*, 33(8), 1914–1928. DOI: <https://doi.org/10.1002/hbm.21333>, PMID: 21769991, PMCID: PMC3838923
- Di Martino, A., Yan, C.-G., Li, Q., Denio, E., Castellanos, F. X., Alaerts, K., . . . Milham, M. P. (2014). The autism brain imaging data exchange: Towards a large-scale evaluation of the intrinsic brain architecture in autism. *Molecular Psychiatry*, 19(6), 659. DOI: <https://doi.org/10.1038/mp.2013.78>, PMID: 23774715, PMCID: PMC4162310
- Dolz, J., Desrosiers, C., & Ayed, I. B. (2018). 3D fully convolutional networks for subcortical segmentation in MRI: A large-scale study. *NeuroImage*, 170, 456–470. DOI: <https://doi.org/10.1016/j.neuroimage.2017.04.039>, PMID: 28450139
- Dvornek, N. C., Ventola, P., Pelphrey, K. A., & Duncan, J. S. (2017). Identifying autism from resting-state fMRI using long short-term memory networks. *Machine Learning in Medical Imaging, 10541*, 362–370. DOI: [https://doi.org/10.1007/978-3-319-67389-9\\_42](https://doi.org/10.1007/978-3-319-67389-9_42), PMID: 29104967, PMCID: PMC5669262
- Finn, E. S., Shen, X., Scheinost, D., Rosenberg, M. D., Huang, J., Chun, M. M., . . . Constable, R. T. (2015). Functional connectome fingerprinting: Identifying individuals using patterns of brain connectivity. *Nature Neuroscience*, 18(11), 1664. DOI: <https://doi.org/10.1038/nn.4135>, PMID: 26457551, PMCID: PMC5008686

- Gadgil, S., Zhao, Q., Adeli, E., Pfefferbaum, A., Sullivan, E. V., & Pohl, K. M. (2020). Spatio-temporal graph convolution for functional MRI analysis. *arXiv:2003.10613*. DOI: [https://doi.org/10.1007/978-3-030-59728-3\\_52](https://doi.org/10.1007/978-3-030-59728-3_52), PMID: 33257918, PMCID: PMC7700758
- Gao, Y., Phillips, J. M., Zheng, Y., Min, R., Fletcher, P. T., & Gerig, G. (2018). Fully convolutional structured LSTM networks for joint 4D medical image segmentation. *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, 1104–1108. DOI: <https://doi.org/10.1109/ISBI.2018.8363764>
- Glasser, M. F., Sotiropoulos, S. N., Wilson, J. A., Coalson, T. S., Fischl, B., Andersson, J. L., . . . Polimeni, J. R. (2013). The minimal preprocessing pipelines for the Human Connectome Project. *NeuroImage*, *80*, 105–124. DOI: <https://doi.org/10.1016/j.neuroimage.2013.04.127>, PMID: 23668970, PMCID: PMC3720813
- Gopinath, K., Desrosiers, C., & Lombaert, H. (2019). Graph convolutions on spectral embeddings for cortical surface parcellation. *Medical Image Analysis*, *54*, 297–305. DOI: <https://doi.org/10.1016/j.media.2019.03.012>, PMID: 30974398
- Greicius, M. D., Supekar, K., Menon, V., & Dougherty, R. F. (2009). Resting-state functional connectivity reflects structural connectivity in the default mode network. *Cerebral Cortex*, *19*(1), 72–78. DOI: <https://doi.org/10.1093/cercor/bhn059>, PMID: 18403396, PMCID: PMC2605172
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 770–778. DOI: <https://doi.org/10.1109/CVPR.2016.90>
- Heinsfeld, A. S., Franco, A. R., Craddock, R. C., Buchweitz, A., & Meneguzzi, F. (2018). Identification of autism spectrum disorder using deep learning and the ABIDE dataset. *NeuroImage: Clinical*, *17*, 16–23. DOI: <https://doi.org/10.1016/j.nicl.2017.08.017>, PMID: 29034163, PMCID: PMC5635344
- Hong, Y., Kim, J., Chen, G., Lin, W., Yap, P.-T., & Shen, D. (2019). Longitudinal prediction of infant diffusion MRI data via graph convolutional adversarial networks. *IEEE Transactions on Medical Imaging*, *38*(12), 2717–2725. DOI: <https://doi.org/10.1109/TMI.2019.2911203>, PMID: 30990424, PMCID: PMC6935161
- Keown, C. L., Datko, M. C., Chen, C. P., Maximo, J. O., Jahedi, A., & Müller, R.-A. (2017). Network organization is globally atypical in autism: A graph theory study of intrinsic functional connectivity. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *2*(1), 66–75. DOI: <https://doi.org/10.1016/j.bpsc.2016.07.008>, PMID: 28944305, PMCID: PMC5607014
- Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv:1412.6980*.
- Ktena, S. I., Parisot, S., Ferrante, E., Rajchl, M., Lee, M., Glocker, B., & Rueckert, D. (2018). Metric learning with spectral graph convolutions on brain connectivity networks. *NeuroImage*, *169*, 431–442. DOI: <https://doi.org/10.1016/j.neuroimage.2017.12.052>, PMID: 29278772
- Lee, M. H., Smyser, C. D., & Shimony, J. S. (2013). Resting-state fMRI: A review of methods and clinical applications. *American Journal of Neuroradiology*, *34*(10), 1866–1872. DOI: <https://doi.org/10.3174/ajnr.A3263>, PMID: 22936095, PMCID: PMC4035703
- Liu, T. T., Nalci, A., & Falahpour, M. (2017). The global signal in fMRI: Nuisance or Information? *NeuroImage*, *150*, 213–229. DOI: <https://doi.org/10.1016/j.neuroimage.2017.02.036>, PMID: 28213118, PMCID: PMC5406229
- McGonigle, D. J., Howseman, A. M., Athwal, B. S., Friston, K. J., Frackowiak, R., & Holmes, A. P. (2000). Variability in fMRI: An examination of intersession differences. *NeuroImage*, *11*(6), 708–734. DOI: <https://doi.org/10.1006/nimg.2000.0562>, PMID: 10860798
- Miller, K. L., Alfaro-Almagro, F., Bangerter, N. K., Thomas, D. L., Yacoub, E., Xu, J., . . . Andersson, J. L. R. (2016). Multimodal population brain imaging in the UK Biobank prospective epidemiological study. *Nature Neuroscience*, *19*(11), 1523–1536. DOI: <https://doi.org/10.1038/nn.4393>, PMID: 27643430, PMCID: PMC5086094
- Monti, F., Boscaini, D., Masci, J., Rodola, E., Svoboda, J., & Bronstein, M. M. (2017). Geometric deep learning on graphs and manifolds using mixture model CNNs. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 5425–5434. DOI: <https://doi.org/10.1109/CVPR.2017.576>
- Murphy, K., & Fox, M. D. (2017). Towards a consensus regarding global signal regression for resting state functional connectivity MRI. *NeuroImage*, *154*, 169–173. DOI: <https://doi.org/10.1016/j.neuroimage.2016.11.052>, PMID: 27888059, PMCID: PMC5489207
- Nielsen, J. A., Zielinski, B. A., Fletcher, P. T., Alexander, A. L., Lange, N., Bigler, E. D., . . . Anderson, J. S. (2013). Multisite functional connectivity MRI classification of autism: ABIDE results. *Frontiers in Human Neuroscience*, *7*, 599. DOI: <https://doi.org/10.3389/fnhum.2013.00599>, PMID: 24093016, PMCID: PMC3782703
- Parisot, S., Ktena, S. I., Ferrante, E., Lee, M., Moreno, R. G., Glocker, B., & Rueckert, D. (2017). Spectral graph convolutions for population-based disease prediction. *Medical Image Computing and Computer-Assisted Interventions*. DOI: [https://doi.org/10.1007/978-3-319-66179-7\\_21](https://doi.org/10.1007/978-3-319-66179-7_21)
- Petersen, S. E., & Sporns, O. (2015). Brain networks and cognitive architectures. *Neuron*, *88*(1), 207–219. DOI: <https://doi.org/10.1016/j.neuron.2015.09.027>, PMID: 26447582, PMCID: PMC4598639
- Power, J. D., Cohen, A. L., Nelson, S. M., Wig, G. S., Barnes, K. A., Church, J. A., . . . Schlaggar, B. L. (2011). Functional network organization of the human brain. *Neuron*, *72*(4), 665–678. DOI: <https://doi.org/10.1016/j.neuron.2011.09.006>, PMID: 22099467, PMCID: PMC3222858
- Robinson, E. C., Jbabdi, S., Glasser, M. F., Andersson, J., Burgess, G. C., Harms, M. P., . . . Jenkinson, M. (2014). MSM: A new flexible framework for Multimodal Surface Matching. *NeuroImage*, *100*, 414–426. DOI: <https://doi.org/10.1016/j.neuroimage.2014.05.069>, PMID: 24939340, PMCID: PMC4190319
- Rogers, B. P., Morgan, V. L., Newton, A. T., & Gore, J. C. (2007). Assessing functional connectivity in the human brain by fMRI. *Magnetic Resonance Imaging*, *25*(10), 1347–1357. DOI: <https://doi.org/10.1016/j.mri.2007.03.007>, PMID: 17499467, PMCID: PMC2169499
- Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. DOI: [https://doi.org/10.1007/978-3-319-24574-4\\_28](https://doi.org/10.1007/978-3-319-24574-4_28)

- Sakoğlu, Ü., Pearlson, G. D., Kiehl, K. A., Wang, Y. M., Michael, A. M., & Calhoun, V. D. (2010). A method for evaluating dynamic functional network connectivity and task-modulation: Application to schizophrenia. *Magnetic Resonance Materials in Physics, Biology and Medicine*, 23(5–6), 351–366. **DOI:** <https://doi.org/10.1007/s10334-010-0197-8>, **PMID:** 20162320, **PMCID:** PMC2891285
- Salimi-Khorshidi, G., Douaud, G., Beckmann, C. F., Glasser, M. F., Griffanti, L., & Smith, S. M. (2014). Automatic denoising of functional MRI data: Combining independent component analysis and hierarchical fusion of classifiers. *NeuroImage*, 90, 449–468. **DOI:** <https://doi.org/10.1016/j.neuroimage.2013.11.046>, **PMID:** 24389422, **PMCID:** PMC4019210
- Sarraf, S., & Tofighi, G. (2016). DeepAD: Alzheimer’s disease classification via deep convolutional neural networks using MRI and fMRI. *bioRxiv:070441*. **DOI:** <https://doi.org/10.1101/070441>
- Suk, H.-I., Wee, C.-Y., Lee, S.-W., & Shen, D. (2016). State-space model with deep learning for functional dynamics estimation in resting-state fMRI. *NeuroImage*, 129, 292–307. **DOI:** <https://doi.org/10.1016/j.neuroimage.2016.01.005>, **PMID:** 26774612, **PMCID:** PMC5437848
- van den Heuvel, M. P., & Pol, H. E. H. (2010). Exploring the brain network: A review on resting-state fMRI functional connectivity. *European Neuropsychopharmacology*, 20(8), 519–534. **DOI:** <https://doi.org/10.1016/j.euroneuro.2010.03.008>, **PMID:** 20471808
- van de Ven, V. G., Formisano, E., Prvulovic, D., Roeder, C. H., & Linden, D. E. J. (2004). Functional connectivity as revealed by spatial independent component analysis of fMRI measurements during rest. *Human Brain Mapping*, 22(3), 165–178. **DOI:** <https://doi.org/10.1002/hbm.20022>, **PMID:** 15195284, **PMCID:** PMC6872001
- Van Essen, D. C., Smith, S. M., Barch, D. M., Behrens, T. E. J., Yacoub, E., Ugurbil, K., & WU-Minn HCP Consortium. (2013). The WU-Minn Human Connectome Project: An overview. *NeuroImage*, 80, 62–79. **DOI:** <https://doi.org/10.1016/j.neuroimage.2013.05.041>, **PMID:** 23684880, **PMCID:** PMC3724347
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., . . . Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*.
- Wang, J., Zuo, X., & He, Y. (2010). Graph-based network analysis of resting-state functional MRI. *Frontiers in Systems Neuroscience*, 4, 16. **DOI:** <https://doi.org/10.3389/fnsys.2010.00016>, **PMID:** 20589099, **PMCID:** PMC2893007
- Wang, L. (2020). cGCN fMRI, GitHub, [https://github.com/Lebo-Wang/cGCN\\_fMRI](https://github.com/Lebo-Wang/cGCN_fMRI)
- Wang, L., Li, K., Chen, X., & Hu, X. P. (2019). Application of convolutional recurrent neural network for individual recognition based on resting state fMRI data. *Frontiers in Neuroscience*. **DOI:** <https://doi.org/10.3389/fnins.2019.00434>, **PMID:** 31118882, **PMCID:** PMC6504790
- Wang, Y., Sun, Y., Liu, Z., Sarma, S. E., Bronstein, M. M., & Solomon, J. M. (2018). Dynamic graph CNN for learning on point clouds. *arXiv:1801.07829*. **DOI:** <https://doi.org/10.1145/3326362>
- Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., & Yu, P. S. (2019). A comprehensive survey on graph neural networks. *arXiv:1901.00596*. **DOI:** <https://doi.org/10.1109/TNNLS.2020.2978386>, **PMID:** 32217482
- Zeiler, M. D., & Fergus, R. (2014). Visualizing and understanding convolutional networks. *European Conference on Computer Vision*, 818–833. **DOI:** [https://doi.org/10.1007/978-3-319-10590-1\\_53](https://doi.org/10.1007/978-3-319-10590-1_53)
- Zhang, Z., Cui, P., & Zhu, W. (2018). Deep learning on graphs: A survey. *arXiv:1812.04202*.
- Zuo, X.-N., Xu, T., Jiang, L., Yang, Z., Cao, X.-Y., He, Y., . . . Milham, M. P. (2013). Toward reliable characterization of functional homogeneity in the human brain: Preprocessing, scan duration, imaging resolution and computational space. *NeuroImage*, 65, 374–386. **DOI:** <https://doi.org/10.1016/j.neuroimage.2012.10.017>, **PMID:** 23085497, **PMCID:** PMC3609711