Explainable Contrastive Multiview Graph Representation of Brain, Mind, and Behavior

Chongyue Zhao[1], Liang Zhan[1], Paul M. Thompson[2], and Heng Huang[1(✉)]

[1] Department of Electrical and Computer Engineering, University of Pittsburgh, Pittsburgh, PA, USA
liang.zhan@pitt.edu, heng.huang@pitt.edu

[2] Imaging Genetics Center, University of Southern California, Los Angeles, CA, USA
pthomp@usc.edu

Abstract. Understanding the intrinsic patterns of human brain is important to make inferences about the mind and brain-behavior association. Electrophysiological methods (i.e. MEG/EEG) provide direct measures of neural activity without the effect of vascular confounds. The blood oxygenated level-dependent (BOLD) signal of functional MRI (fMRI) reveals the spatial and temporal brain activity across different brain regions. However, it is unclear how to associate the high temporal resolution Electrophysiological measures with high spatial resolution fMRI signals. Here, we present a novel interpretable model for coupling the structure and function activity of brain based on heterogeneous contrastive graph representation. The proposed method is able to link manifest variables of the brain (i.e. MEG, MRI, fMRI and behavior performance) and quantify the intrinsic coupling strength of different modal signals. The proposed method learns the heterogeneous node and graph representations by contrasting the structural and temporal views through the mind to multimodal brain data. The first experiment with 1200 subjects from Human connectome Project (HCP) shows that the proposed method outperforms the existing approaches in predicting individual gender and enabling the location of the importance of brain regions with sex difference. The second experiment associates the structure and temporal views between the low-level sensory regions and high-level cognitive ones. The experimental results demonstrate that the dependence of structural and temporal views varied spatially through different modal variants. The proposed method enables the heterogeneous biomarkers explanation for different brain measurements.

Keywords: Brain dynamics · Spatio-temporal graphs · Explanations on graphs

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/978-3-031-16431-6_34.

L. Wang et al. (Eds.): MICCAI 2022, LNCS 13431, pp. 356–365, 2022.
https://doi.org/10.1007/978-3-031-16431-6_34

1 Introduction

The brain activity remains latent construct that could not be directly measured with present technologies [18]. Non-invasive electrophysiology such as Magneto- and electo- phencephalography (M/EEG) shows insights into many healthy and diseased brain activity at millisecond but lacks spatial resolution. Additional modalities with higher spatial resolution at millimeter such as Magnetic resonance imaging (MRI), functional MRI(fMRI) and positron emission tomography (PET) have paved the way to human connectomics in clinical practice. However, this kind of technique is sluggish to reveal neuronal activity. The challenge of fusing non-invasive brain measurements is that different technique provides either high spatial or temporal resolution but not both [3,11,21]. The development of fMRI, MEG and MRI made it possible to obtain the system level of structural connectomics and functional connectomics. Many methods have been proposed to associate the connectomics from different modalities. Simple and direct correlational approaches have been commonly used to link SC and FC [9,10,22]. With the prior of SC, dynamic casual model could explain the functional signals in terms of excitatory and inhibitory interactions [5,17]. Graph models allow the extraction of system level connectivity properties associated with brain changes in the life cycle, such as attention and control networks related to late adolescence and aging process, the strength and organization of function connectivity related to neurological diseases and intrinsic brain activity of behavior performance during resting and task state [1,2,4]. Recently, graph harmonic analysis with Laplacian embedding and spectral clustering have been utilized for revealing brain organization [16]. Basically, the graph harmonic model use harmonic components to summarize the spatial patterns with the nodes of the graph. With the structurally informed components, the relationship among structural connectivity, functional connectivity and behavior performance could be decomposed.

Literature on previous graph based methods that utilizes graph theoretical metrics to summarize the function connectivity ignores the high-order interactions between ROIs [12,14,19]. The existing methods are not very suitable for the integration of structural connectivity, functional connectivity and behavior performance for the following reasons:

Lack of Individual and Group-Level Explanation: existing methods especially for fMRI analysis assume that the nodes in the same brain graphs are translation invariant. Ignoring the correspondence of nodes of different brain ROIs limits the explanation in individual and group level.

Incomplete and Missing Data in Clinical Data Collection: due to the scanner availability and patient demands, it is impossible to do multimodal assessment for all patients. Incomplete or missing data hinders the potential of multimodal usage. There are very few databases that provides public access to MEG, MRI and fMRI of the same subjects.

Violation of the Brain Dynamics Information: the existing joint model uses the linearity assumption among latent variables from different modal measurements. The effective usage includes subject specific integration (structural connectivity), modal specific association (i.e. fMRI and MEG). However, the brain is highly dynamic and the linearity assumption is not applicable in many cases.

In the paper, we build a novel heterogeneous contrast subgraphs representation learning based method to exploit the coupling of structural and functional connectivity from different brain modality. The proposed method has the following advantages: 1) The proposed heterogeneous graph representation learning method utilizes the contrastive learning to explore the coupling information of different modality. The proposed method with the semantic attention model enables the complex and dynamics link between structural and functional connectivity within each modal measures. The proposed method is capable of modeling heterogeneous spatio-temporal dynamics and learn the contrast graph structures simultaneously.

2) The proposed method uses a causal explanation model to improve the individual and group-level interpretability. The explanation approach helps to locate the significant brain region with sex difference and neurodevelopment. The experimental results for gender classification with 1200 subjects from HCP have shown the performance of the proposed method with incomplete multimodal data (fMRI and MEG).

3) The proposed method utilizes graph convolution theory to link the brain structure and function. The experimental results with meta-analysis reveal the strength of structural-functional coupling patterns among functional connectivity, structural connectivity and behavior performance.

2 Problem Formulation

To associate heterogeneous multimodal brain measurements, a heterogeneous graph representation learning with semantic attention is introduced based on fMRI and filtered MEG data. Next we introduce dynamical neural graph encoder framework to associate the spatial and temporal patterns from structural and functional connectivity of multimodal brain measurements. Then, we give the details of the multi-view contrastive graph representation learning. The contrastive graph learning method makes sure the maximization of mutual information of the node representation from one view and the graph representation from another view. Finally, we discuss the interpretable causal explanations for the proposed method on graph. The overall framework is illustrated in Fig. 1.

Heterogeneous Graph Representation Learning. We use a graph $\mathcal{G}_i = (\mathcal{V}, \mathcal{E}_i)$ to represent the heterogeneous graph representation, with the node type $\mathcal{V} = [v_{t,i} | t = 1, ..., T; i = 1, ..., N]$ with N brain ROIs and T time points. The edge mapping $\mathcal{E}$ represents the connection of different brain ROIs in spatial and temporal domain. The aim of contrastive graph representation is to explore the spatial and temporal pattern of the fMRI and MEG data. Given two multimodal graph $\mathcal{G}_\mathcal{A}$ and $\mathcal{G}_\mathcal{B}$ with different time points $T_\mathcal{A}$ and $T_\mathcal{B}$ and their correspondence multivariate value $X_\mathcal{A} = [x_{t_1}^\mathcal{A}, x_{t_2}^\mathcal{A}, ..., x_{t_\mathcal{A}}^\mathcal{A}]$ and $X_\mathcal{B} = [x_{t_1}^\mathcal{B}, x_{t_2}^\mathcal{B}, ..., x_{t_\mathcal{B}}^\mathcal{B}]$. To integrate manifest variables of the brain with different spatial and temporal resolution, a dynamical neural graph encoder is proposed to explore the spatio-temporal dynamics. The data augmentation mechanism introduces a $\Phi_{i}{}_{i=1}^{P}$ of P

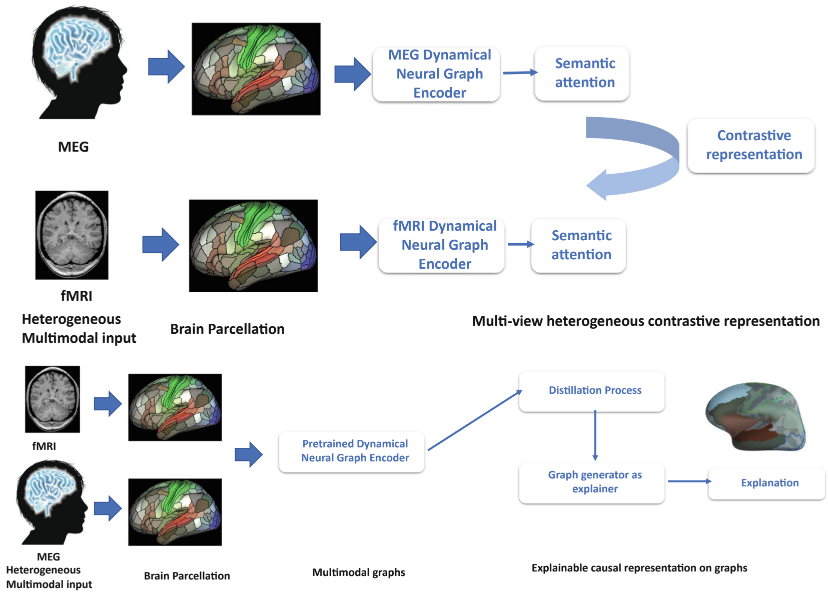

Fig. 1. Schematic illustration of the explainable contrastive graph representation with heterogeneous brain measurements. Top: training process, bottom: explanation process.

set of multiview heterogeneous graph for each modality, where P represents the total view of the brain measurement after data augmentation (i.e. the number of filter bands for MEG). We could define its corresponding adjacency matrices as $[A_{\Phi_i}]_{i=1}^P$. The node representation within each view is defined as $Z_{\Phi_i}^{\mathcal{A}}$ and $Z_{\Phi_i}^{\mathcal{B}}$. Then we use the heterogeneous latent attention module to aggregate the graph representation for each modal measurement $H^{\mathcal{A}} = f_{\psi}(L_{att}(Z_{\Phi_1}^{\mathcal{A}}, Z_{\Phi_2}^{\mathcal{A}}, ..., Z_{\Phi_P}^{\mathcal{A}}))$.

2.1 Dynamical Neural Graph Encoder

We define the brain dynamical state with N neurons as $\dot{z}(t) = f(z, l, t)$, where $z(t) = [z_1(t), z_2(t), ..., z_N(t)]^T$ represents the internal states of N neuron nodes at time t. $f(\cdot)$ denotes the nonlinear dynamical function of each node. And $l(t) = [l_1(t), l_2(t), ..., l_S(t)]^T$ represents the external stimuli for S neurons.

Within each single modality, we define a continuous neural-graph differential equation as follows,

$$\dot{Z}(t) = f_{G_{t_k}}(t, Z_t, \theta_t) \quad and \quad Z_t^+ \quad = \mathcal{L}_{G_{t_k}}^j(Z_t, X_t) \tag{1}$$

where f_G, $\mathcal{L}_G^j$ are graph encoder networks. Z_t^+ is introduced to represent the value after discrete operation. Z_t^+ could represent the state 'jump' for brain measurements such as task fMRI.

Heterogeneous Graph Output with Semantic Attention. Within each brain modal measurement, we could obtain the a set of heterogeneous representation $[Z_{\Phi_i}]_{i=1}^{P}$. Then, we use a semantic level attention layer to associate the cross view heterogeneous graph representation L_{att} with the learned weights $\beta_{\Phi_1}, \beta_{\Phi_2}, ..., \beta_{\Phi_p} = Latt(Z_{\Phi_1}, Z_{\Phi_2}, ..., Z_{\Phi_p})$.

We define the importance of each view graph representation as,

$$e_{\Phi_i} = \frac{1}{N}\sum_{n=1}^{N} tanh(\vec{q}^T \cdot [W_{sem} \cdot \vec{Z}_{n,\Phi_i} + \vec{b}]) \quad and \quad \beta_{\Phi_i} = softmax(e_{\Phi_i}), \quad (2)$$

where W_{sem} denotes the linear transformation and $\vec{q}$ is learnable. The heterogeneous node representation could be denoted by $H = f_{\psi}(\sum_{i=1}^{p} \beta_{\Phi_i} Z_{\Phi_i})$.

Next, we apply the readout function to the aggregated heterogeneous representation of each view with the shared projection head $f_{\phi}(.) \in R^{d_h}$. In the experiment, we use an MLP with two hidden layers as the projection head. We get the projected representations $\vec{h}_g^{\mathcal{A}}$ and $\vec{h}_g^{\mathcal{B}}$. For each view, the node representation are concatenated as $\vec{h}_g = \sigma(\mathop{\|}\limits_{l=1}^{L} [\sum_{i=1}^{n} \vec{h}_i^l]W)$.

2.2 Mutual Information Based Training Process

In the training process, we maximize the mutual information between the node representation of one modal and the graph representation of another modal, i.e. the node representation of fMRI and the graph representaion of MEG and vice versa. The objective is defined with contrastive learning as follows,

$$\max_{\theta,\phi,\psi} \frac{1}{|G|}\sum_{g\in G}\frac{1}{|g|}\sum_{i=1}^{|g|}[MI(\vec{h}_i^{\mathcal{A}}, \vec{h}_g^{\mathcal{B}}) + MI(\vec{h}_g^{\mathcal{A}}, \vec{h}_i^{\mathcal{B}})], \quad (3)$$

where θ, ϕ, ψ represent the parameters of heterogeneous dynamical neural graph encoder and projection head. $|G|$ denotes the total numbers of graph. $|g|$ is the number of nodes. MI is denoted as the dot production $MI(\vec{h}_i^{\mathcal{A}}, \vec{h}_g^{\mathcal{B}}) = \vec{h}_i^{\mathcal{A}} \cdot (\vec{h}_g^{\mathcal{B}})^T$.

2.3 Explainable Causal Representation on Graphs

To highlight the importance of the brain ROIs, we introduce the explainable causal representation to encourage the reasonable node selection process. We train an explanation model to explain the multmodal graph representation approach based on granger causality. The explanation process are divided into two steps, the distillation process and explainer training process.

In distillation process, we use a subgraph G_s to represent the main cause of the target prediction y. The explainable causal representation does not require the re-training of the dynamical graph encoder which could lower the computation complexity. We use $\delta_{G\backslash e_j}$ to represent the prediction error exclude the edge e_j. The model error is defined as $\Delta_{\delta,e_j} = \delta_{G\backslash e_j} - \delta_G$.

With the ground truth label y, we define the model error as the loss difference $\mathcal{L}(y, \hat{y}_G)$.

$$\delta_G = \mathcal{L}(y, \hat{y}_G) \quad and \quad \delta_{G \setminus e_j} = \mathcal{L}(y, \hat{y}_{G \setminus e_j}), \tag{4}$$

Given the causal contributions of each edge, we could sort the top-K most relevant edges for model explanation. After the model distillation process, we will train a new explainer model based on graph convolutional layer.

$$Z = GCN(A, Z) \quad and \quad \hat{A} = \sigma(ZZ^T), \tag{5}$$

where each value in A represent the contribution of specific edge to prediction. The explanation model generate $\hat{A}$ as an explanation mask. We show more details of the explainable causal model in Supplementary material.

3 Experiments

In the experiments, we use the public available s1200 dataset from Human Connectome Project (HCP), which contains 1096 young adults. Additionally, about 95 subjects have resting-state and/or task MEG (tMEG) data. The resting-state fMRI is pre-processed following the minimal preprocessing pipeline [8]. Then the pre-processed data is registered into a standard cortical surface using MSMAll [8]. The cortical surface was parcellated into N = 22 major ROIs [7]. In addition, the averaged time course of each ROI is normalized using z-score. The resting-state MEG has been pre-processed using ICA to remove out artefacts related to head and eye movement. Sensor-space data were down-sampled 300 Hz using anti-aliasing filter. Next the MEG data were source-reconstructed with a scalar beamformer and registered into the standard space of the Montreal Neuroimaging Institute (MNI). Data were then filtered into 1–30 Hz and beamformed onto 6 mm grid. The parcellation atlas and z-score normalization method of MEG are similar to resting-state fMRI.

3.1 Sex Classification

We first test the performance of the proposed method with sex classification task using HCP data. We adopt 5-fold cross validation on the 1091 subjects. We compare the proposed method with several state-of-the-art methods such as Long-Short-Term Memory (LSTM) [13], graph convolution LSTM (GC-LSTM) [20] and spatio-temporal graph convolution network (ST-GCN) [6]. The hidden state of LSTM was set to 256. A simple Multi-Layer Perception (MLP) with 2 hidden layers and ReLU activation is also included as the baseline method. For the proposed method, we report two kinds of sex classification accuracy. The first single model uses only the fMRI to explore the dynamics within the brain ROIs. The multimodal based method integrates both fMRI and MEG to exploit the spatio-temporal dynamics and achieves better sex classification performance.

The accuracy of sex classification is shown in Table 1. Comparing with the baseline method, the proposed method could learn the dynamic contrast graph representation between fMRI and MEG. The proposed method could take advantage of the high spatial resolution of fMRI and high temporal resolution of MEG to achieve the highest sex classification performance of 85.2%. The importance of the brain regions that contributes to the sex classification is show in Fig. 2. The causal explanation module provides us a new way to find individual-level and group-level biomarkers for sex difference.

Table 1. Sex classification accuracy with different baseline models

Method	Accuracy
LSTM	0.808(0.033)
GC-LSTM	0.811(0.075)
MLP	0.770(0.051)
ST-GCN	0.839(0.044)
Proposed method with only fMRI	0.827(0.061)
Proposed method with multimodal	0.852(0.046)

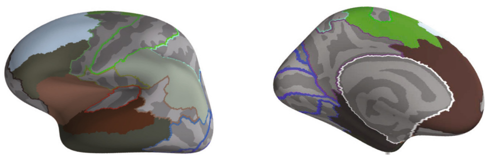

Fig. 2. Top K important brain regions for sex classification

3.2 Brain Activity Decomposed with Functional and Structural Connectivity

In the second experiment, we use the structural-decoupling index which reveals the function and structure relationship to measure the energies of high pass decoupled activity versus low pass coupled activity per brain ROIs. The average structural-decoupling index for surrogate (with or without SC) and function signals is shown in Fig. 3. Without the SC prior knowledge, the surrogate shows significant decoupling patterns. While the knowledge of SC increases the coupling pattern in functional signals. Compared with the functional time courses, the high-level cognition network detaches from the SC. We also use the NeuroSynth meta-analysis on the same topic in [15] to assess the structural-decoupling index. As shown in Fig. 4, the structural-decoupling index associates the behaviorally relevant gradient based on FC data. We could find a macroscale gradient of regions related to low to high level cognition with the learned graph representation. Due to the fact that the functional connectivity comes from MEG data in the second experiment, we compare the gradient learned from the graph representation with the original MEG data. For example, the terms related to acting and perceiving such as "visual perception", "multisensory processing", "reading" and "motor/eye movement" are grouped into the top end. The terms related to complex cognition such as "autobiographical memory", "emotion" and "reward-based decision making" are characterized into the other end. Similar organization phenomenon could be found in the previous research [15,16]. However, the gradient learned by original MEG data lacks the pattern of system organization.

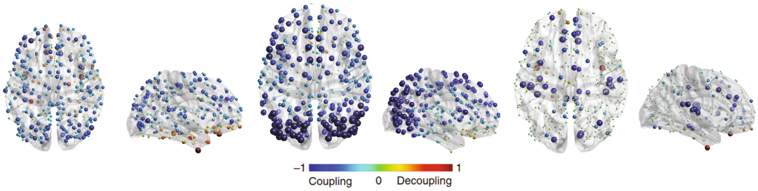

Fig. 3. Structural decoupling index shows brain activity between function and structure. Left: Surrogate brain activity without structural connectome. Middle: surrogate brain activity with structural connectome. Right: brain activity with decoupling difference to the surrogate

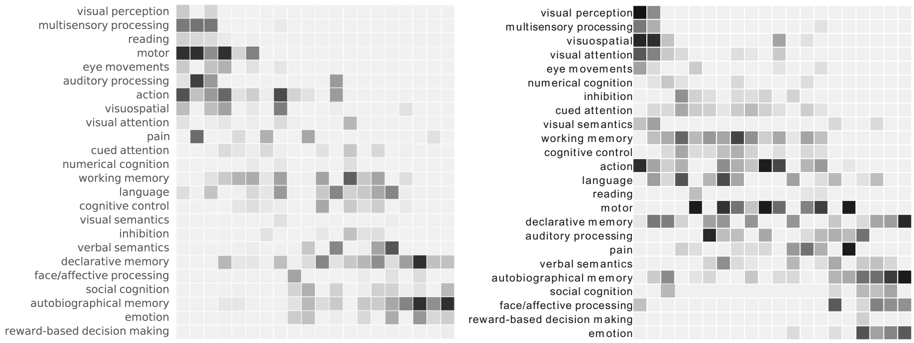

Fig. 4. Behaviorally relevant gradient shows brain organization with Structural decoupling index. Left: the learned graph representation. Right: original MEG data

4 Conclusions

Brain activity is shaped by the anatomical structure. In the paper, we propose an explainable contrastive graph representation learning based model to associate heterogeneous brain measurements such as MRI, functional MRI, MEG and behavior performance. The framework allows key advantages to concentrate brain, mind and behavior in cognitive neuroscience. The proposed method outperforms the state-of-the-art methods in gender classification using fMRI and MEG data. Moreover, the framework could localize the important brain region with sex difference through a causal explanation model. The second experiment with meta analysis demonstrates that the structure-function coupling pattern with the learned contrast graph representation. Future work that links the function connectivity with other modal data (i.e. gene expression and microstructure properties) could be easily adapted to our framework.

Acknowledgement. This work was partially supported by NIH R01AG071243, R01MH125928, R01AG049371, U01AG068057, and NSF IIS 2045848, 1845666, 1852606, 1838627, 1837956, 1956002, IIA 2040588.

References

1. Bassett, D.S., Sporns, O.: Network neuroscience. Nat. Neurosci. **20**(3), 353–364 (2017)
2. Bullmore, E., Sporns, O.: Complex brain networks: graph theoretical analysis of structural and functional systems. Nat. Rev. Neurosci. **10**(3), 186–198 (2009)
3. Dale, A.M., Halgren, E.: Spatiotemporal mapping of brain activity by integration of multiple imaging modalities. Curr. Opin. Neurobiol. **11**(2), 202–208 (2001)
4. Fornito, A., Zalesky, A., Breakspear, M.: Graph analysis of the human connectome: promise, progress, and pitfalls. Neuroimage **80**, 426–444 (2013)
5. Friston, K.J., Harrison, L., Penny, W.: Dynamic causal modelling. Neuroimage **19**(4), 1273–1302 (2003)
6. Gadgil, S., Zhao, Q., Pfefferbaum, A., Sullivan, E.V., Adeli, E., Pohl, K.M.: Spatio-Temporal Graph Convolution for Resting-State fMRI Analysis. In: Martel, A.L., et al. (eds.) MICCAI 2020. LNCS, vol. 12267, pp. 528–538. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59728-3_52
7. Glasser, M.F., Coalson, T.S., Robinson, E.C., Hacker, C.D., Harwell, J., Yacoub, E., Ugurbil, K., Andersson, J., Beckmann, C.F., Jenkinson, M., et al.: A multi-modal parcellation of human cerebral cortex. Nature **536**(7615), 171–178 (2016)
8. Glasser, M.F., et al.: The minimal preprocessing pipelines for the human connectome project. Neuroimage **80**, 105–124 (2013)
9. Hagmann, P., Cammoun, L., Gigandet, X., Meuli, R., Honey, C.J., Wedeen, V.J., Sporns, O.: Mapping the structural core of human cerebral cortex. PLoS Biol. **6**(7), e159 (2008)
10. Honey, C.J., et al.: Predicting human resting-state functional connectivity from structural connectivity. Proc. Natl. Acad. Sci. **106**(6), 2035–2040 (2009)
11. Jorge, J., Van der Zwaag, W., Figueiredo, P.: EEG-fMRI integration for the study of human brain function. Neuroimage **102**, 24–34 (2014)

12. Kazi, A., et al.: InceptionGCN: receptive field aware graph convolutional network for disease prediction. In: Chung, A.C.S., Gee, J.C., Yushkevich, P.A., Bao, S. (eds.) IPMI 2019. LNCS, vol. 11492, pp. 73–85. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-20351-1_6
13. Li, H., Fan, Y.: Brain decoding from functional MRI using long short-term memory recurrent neural networks. In: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds.) MICCAI 2018. LNCS, vol. 11072, pp. 320–328. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00931-1_37
14. Li, X., Dvornek, N.C., Zhou, Y., Zhuang, J., Ventola, P., Duncan, J.S.: Graph neural network for interpreting task-fMRI biomarkers. In: Shen, D., et al. (eds.) MICCAI 2019. LNCS, vol. 11768, pp. 485–493. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32254-0_54
15. Margulies, D.S., et al.: Situating the default-mode network along a principal gradient of macroscale cortical organization. Proc. Natl. Acad. Sci. **113**(44), 12574–12579 (2016)
16. Preti, M.G., Van De Ville, D.: Decoupling of brain function from structure reveals regional behavioral specialization in humans. Nat. Commun. **10**(1), 1–7 (2019)
17. Stephan, K.E., Tittgemeyer, M., Knösche, T.R., Moran, R.J., Friston, K.J.: Tractography-based priors for dynamic causal models. Neuroimage **47**(4), 1628–1638 (2009)
18. Turner, B.M., Palestro, J.J., Miletić, S., Forstmann, B.U.: Advances in techniques for imposing reciprocity in brain-behavior relations. Neurosci. Biobehav. Rev. **102**, 327–336 (2019)
19. Yan, Y., Zhu, J., Duda, M., Solarz, E., Sripada, C., Koutra, D.: GroupINN: grouping-based interpretable neural network for classification of limited, noisy brain data. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp. 772–782 (2019)
20. Yu, B., Yin, H., Zhu, Z.: Spatio-temporal graph convolutional networks: a deep learning framework for traffic forecasting. arXiv preprint arXiv:1709.04875 (2017)
21. Zhao, C., Gao, X., Emery, W.J., Wang, Y., Li, J.: An integrated spatio-spectral-temporal sparse representation method for fusing remote-sensing images with different resolutions. IEEE Trans. Geosci. Remote Sens. **56**(6), 3358–3370 (2018)
22. Zhao, C., Li, H., Jiao, Z., Du, T., Fan, Y.: A 3D convolutional encapsulated long short-term memory (3DConv-LSTM) model for denoising fMRI data. In: Martel, A.L., et al. (eds.) MICCAI 2020. LNCS, vol. 12267, pp. 479–488. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59728-3_47