TARDRL: Task-Aware Reconstruction for Dynamic Representation Learning of fMRI

Yunxi Zhao[1], Dong Nie[2], Geng Chen[3], Xia Wu[4], Daoqiang Zhang[1], and Xuyun Wen[1(✉)]

[1] College of Artificial Intelligence, Nanjing University of Aeronautics and Astronautics, Key Laboratory of Brain-Machine Intelligence Technology, Ministry of Education, Nanjing, China
[2] Meta Inc., California, United States
[3] School of Computer Science, Northwestern Polytechnical University, Shanxi, China
[4] School of Computer Science and Technology, Beijing Institute of Technology, Beijing, China
wenxuyun@nuaa.edu.cn

Abstract. The mask autoencoder (MAE) is utilized in functional magnetic resonance imaging (fMRI) analysis to construct brain representation learning models and conduct prediction for various fMRI-related tasks (e.g., disease detection). It involves pretraining the model by reconstructing signals of brain regions that are randomly masked at different time segments and subsequently fine-tuning it for prediction tasks. Although the MAE helps to improve prediction performance, directly applying it to fMRI may lead to sub-optimal results for the following reasons: 1) The reconstruction process is not task-aware, meaning the extracted brain representations are unable to sufficiently consider downstream tasks, thereby affecting prediction performance; 2) Random masking of fMRI data ignores that the varying contributions of different brain regions to different prediction tasks. To address these issues, we propose Task-Aware Reconstruction Dynamic Representation Learning (TARDRL). Different from the conventional sequential design, this approach sets up reconstruction and prediction tasks in parallel to learn robust task-aware representations. Based on the parallelized framework, we leverage attention maps from specific tasks to guide the fMRI time series reconstruction, which in turn helps to learn task-aware fMRI representations and improve disease prediction accuracy. Extensive experiments demonstrate that our model outperforms state-of-the-art methods on the ABIDE and ADNI datasets, with high interpretability. The codes are available in the repository.

Keywords: Self-supervised learning · Functional magnetic resonance imaging · Mask autoencoder · Disease diagnosis.

1 Introduction

Functional magnetic resonance imaging (fMRI) is a non-invasive technique to reveal brain activities by measuring blood-oxygen-level-dependent (BOLD) sig-

nals. Given its exceptional spatial resolution, a general practice in neuroimaging communities is mapping the 4D voxel-wise signals into 2D ROI-wise signals with a predefined 3D atlas [19,20,27]. Recently, researchers have embraced supervised deep learning models, extracting information from brain activities and performing clinical prediction, such as disease diagnosis. These supervised methods can be broadly classified into two categories, static and dynamic. Static methods typically assume that the functional interactions between ROIs remain constant, directly modeling the entire signals of ROIs or functional connectivity (FC) derived from signals [11,10,9,14]. In contrast, dynamic approaches aim to explore temporal variations and state transitions in the brain over time [6,12,3].

Nevertheless, these algorithms solely rely on supervised learning unable to learn underlying genuine representations of fMRI, leading to sub-optimal performances in prediction tasks. Even with reduced spatial resolution, neuroimaging datasets still pose a challenge with their high dimensionality and small sample size. In this context, predictive models tend to overfit, resulting in poor performance and severely limiting the potential to obtain interpretable biomarkers [21]. A self-supervised learning framework, mask autoencoder (MAE), has made a splash in natural language processing (NLP) and computer vision (CV) by enabling the training of generalizable large models [2,8]. Inspired by the concept of MAE, one recent study [24] leverages the fMRI data to craft a reconstruction task that masks out signals of randomly chosen ROIs at different time segments and reconstructs them. The pre-trained model is then fine-tuned on a prediction task, by reusing the same data samples along with their labels, leading to improved performance over exclusively doing supervised learning.

However, we argue that directly applying this learning framework to fMRI may result in sub-optimal results. Firstly, the reconstruction process is not task-aware, meaning that representations learned from reconstruction are not aware of the prediction task. Consequently, the learned representations may not be fully leveraged to achieve optimal performance in the prediction task. Secondly, the random masking for fMRI data may not be effective, ignoring a critical phenomenon in neuroimaging that depending on different prediction tasks, brain regions exhibit varying contributions. For instance, in autism spectrum disorder (ASD) diagnosis, the attention scores in the sensory-motor network (SMN) are highest, indicating that the regions in SMN play a critical role in ASD diagnosis [1]. Moreover, numerous studies have found that individuals with cognitive impairment exhibit abnormalities in default mode network (DMN) [5,13]. Based on these observations, we propose a hypothesis that *learning representations by reconstructing signals from important ROIs at different time segments will yield improved performance in prediction tasks over reconstruction on randomly chosen ROIs.*

To this end, we propose TARDRL, **T**ask-**A**ware **R**econstruction **D**ynamic **R**presentation **L**earning, a novel algorithm to improve prediction performance with task-aware reconstruction. Different from the conventional *sequential* design on the reconstruction and prediction, we set up the reconstruction and prediction task in a *parallel* paradigm to learn robust task-aware representations (training

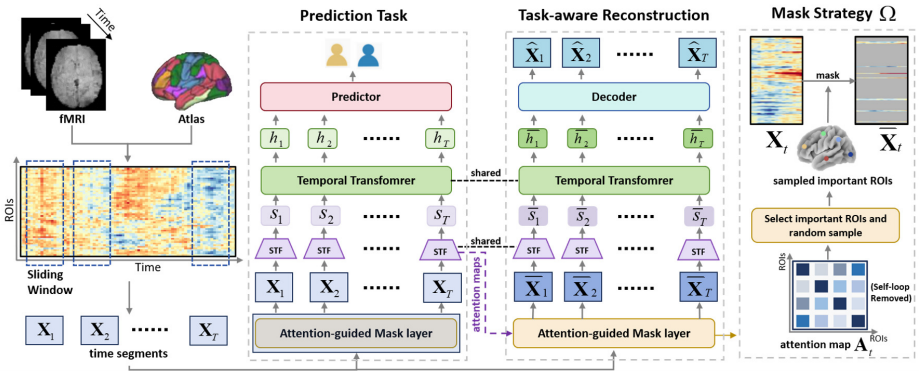

Fig. 1. Overview of Task-Aware Reconstruction Dynamic Representation Learning.

TARDRL in a multi-task learning way). Based on the parallelized framework, we adopt an attention-guided masking strategy to mask ROIs important to the prediction task. Specifically, as shown in Fig.1, it uses the attention maps generated by STF during prediction and determines the set of ROIs to mask. During reconstruction, signals of these ROIs are masked out and reconstructed, facilitating the reconstruction in a task-aware manner. Our extensive experiments on the public ABIDE dataset and ADNI dataset show significant improvements brought by TARDRL. Furthermore, our in-depth analysis validates the interpretability of our model.

2 Method

Fig.1 illustrates the architecture of TARDRL, where tasks of prediction and reconstruction are trained simultaneously. TARDRL comprises four major components: an attention-guided mask layer only activated during reconstruction, a shared encoder composed of spatial transformer (STF) and temporal transformer (TTF), a predictor for prediction tasks, and a decoder for task-aware reconstruction.

2.1 Problem Definition

The ROI-wise signals $\mathbf{X} \in \mathbb{R}^{N \times M}$ with a target class label $y \in [0, \dots C-1]$ is extracted by a pre-defined atlas consisting of N ROIs, where M is the number of time points. The set of ROIs is denoted as $\mathcal{V}$. Values of each ROI are standardized across time. The matrix $\mathbf{X}$ is partitioned into T segments along the temporal dimension via non-overlapping sliding windows of length τ, obtaining a set of time segments $\{\mathbf{X}_t\}_{t=1}^{T}$, where $\mathbf{X}_t \in \mathbb{R}^{N \times \tau}$ and $M = \tau T$. In prediction, TARDRL takes fMRI segments as input and produces prediction results. Additionally, it also generates the self-loop removed attention maps $\{\mathbf{A}_t\}_{t=1}^{T}, \mathbf{A}_t \in \mathbb{R}^{N \times N}$, which are used to select important ROIs. In task-aware

reconstruction, the activated mask layer masks out the signals of important ROIs at different time segments, according to an attention-guided masking strategy Ω, i.e. $\left(\{\mathbf{X}_t\}_{t=1}^T, \{\mathbf{A}_t\}_{t=1}^T\right) \xrightarrow{\Omega} \{\bar{\mathbf{X}}_t\}_{t=1}^T$. Then, TARDRL outputs the reconstructed data $\left\{\hat{\mathbf{X}}_t\right\}_{t=1}^T$ based on masked fMRI signals.

2.2 TARDRL

Shared Encoder. The encoder is utilized as the backbone of TARDRL for dynamic fMRI representation learning, which consists of STF and TTF. The STF plays a crucial role in capturing functional interconnections between ROIs, while the temporal dynamics are captured by the TTF. The transformer encoder [22] is a key component of STF and TTF and has a multi-head-attention module to model the spatial/temporal dependencies among ROIs/segments. Both STF and TTF consist of multiple transformer blocks and position embeddings are added for all tokens. Since both tasks share the parameters of this encoder, for simplicity, we choose forward propagation during prediction as an example to specify this encoder in this section. Our STF firstly embeds each fMRI segment $\mathbf{X}_t$ by a linear projection, and then processes the resulting set via a series of Transformer blocks. The output from transformer blocks forms a matrix $\mathbf{X}_t^s \in \mathbb{R}^{N \times d_{\text{spat}}}$, where each row represent the features learned for each ROI at time segment t. Subsequently, the average pooling layer aggregates all of the ROI-wise vectors into a single vector $s_t \in \mathbb{R}^{d_{\text{spat}}}$. Finally, TTF maps the input sequence of embeddings $\{s_t\}_{t=1}^T$ generated by the STF to a sequence of latent representations $\{h_t\}_{t=1}^T, h_t \in \mathbb{R}^{d_{\text{temp}}}$.

Attention-guided Mask Layer. Considering that the importance of ROIs to prediction tasks may vary across different time segments, we eschew random reconstruction of fMRI data in favor of a strategy, Ω, which identifies and masks important ROIs instead. Specifically, to define the notion of important ROIs, we use attention maps generated during the forward propagation of STF in prediction. The attention maps indicate the weights between brain regions and to some extent reflect the importance of brain regions. For $\mathbf{X}_t$, we compute the final attention map $\mathbf{A}_t \in \mathbb{R}^{N \times N}$ by averaging the attention maps generated by each head of each layer of STF, while also removing the self-loop (setting the values at diagonal positions to 0). We compute $\alpha_t \in \mathbb{R}^N$, where $\alpha_t(j) = \frac{\sum_{i=1}^N \mathbf{A}_t(i,j)}{\sum_{i=1}^N \sum_{j=1}^N \mathbf{A}_t(i,j)}$ for $j = 1, 2, \ldots, N$. We quantify the importance of each ROI at time t by its magnitude in α_t. Then, the ROIs corresponding to the top k values in α_t are selected ($k = \lfloor \delta N \rfloor$ with mask ratio $\delta \in (0, 1)$) and masked out for the purpose of reconstruction. However, as the same training data is fed at each epoch, the model may consistently mask out the same set of ROIs for each fMRI segment of samples throughout the entire training process, leading to sub-optimal performance in reconstruction. To address it, we ensure that for all fMRI segments of samples, at each epoch the model explores a random set of ROIs

among the important ones. Specifically, we randomly sample $\lfloor \mu k \rfloor$ ROIs among k important ROIs, where μ is a hyperparameter and $0 < \mu < 1$. The masked signals $\{\bar{\mathbf{X}}_t\}_{t=1}^{T}$ can be derived from $\{\mathbf{X}_t\}_{t=1}^{T}$ and $\{\mathbf{A}_t\}_{t=1}^{T}$ via Ω. The masked ROIs are replaced with a mask token which is a learnable d_{spat} dimensional vector that indicates the presence of a missing ROI signal.

Prediction and Reconstruction. The shared encoder operates on $\{\mathbf{X}_t\}_{t=1}^{T}$ and $\{\bar{\mathbf{X}}_t\}_{t=1}^{T}$ and outputs corresponding sequences of latent representations $\{h_t\}_{t=1}^{T}$ and $\{\bar{h}_t\}_{t=1}^{T}$, respectively. In the prediction task, the predictor begins by computing the overall fMRI representation h_{global} by average pooling across $\{h_t\}_{t=1}^{T}$, followed by passing h_{global} through two fully connected layers to produce the final prediction. The supervised Cross Entropy loss $\mathcal{L}_{\text{CE}}$ is utilized for this prediction task. In task-aware reconstruction, the decoder firstly adds positional encoding to $\{\bar{h}_t\}_{t=1}^{T}$, then processes them through multiple transformer blocks, and finally reconstructs the fMRI segments $\left\{\hat{\mathbf{X}}_t\right\}_{t=1}^{T}$. The label for this task is the raw input data $\{\mathbf{X}_t\}_{t=1}^{T}$. To ensure accurate reconstruction, we calculate the Mean Square Error (MSE) for masked ROIs as follows:

$$\mathcal{L}_{\text{masked}} = \frac{1}{\tau T \left|\Phi_t\right|} \sum_{t=1}^{T} \sum_{i \in \Phi_t} \left\| \hat{\mathbf{X}}_t(i) - \mathbf{X}_t(i) \right\|^2, \tag{1}$$

where Φ_t is the set of masked ROIs in the segment t, $\mathbf{X}_t(i)$ is the t-th fMRI segment of the i-th ROI and $\hat{\mathbf{X}}_t(i)$ is the reconstructed one. Given that the fMRI is a type of time-series data with a strong correlation across time, we also consider the loss of reconstructing unmasked parts, as follows:

$$\mathcal{L}_{\text{unmasked}} = \frac{1}{\tau T \left|\mathcal{V} \backslash \Phi_t\right|} \sum_{t=1}^{T} \sum_{i \in \mathcal{V} \backslash \Phi_t} \left\| \hat{\mathbf{X}}_t(i) - \mathbf{X}_t(i) \right\|^2. \tag{2}$$

The combined reconstruction loss $\mathcal{L}_{\text{rec}}$ is a weighted sum of $\mathcal{L}_{\text{masked}}$ and $\mathcal{L}_{\text{unmasked}}$, given by $\mathcal{L}_{\text{rec}} = \lambda \mathcal{L}_{\text{masked}} + (1-\lambda)\mathcal{L}_{\text{unmasked}}$. With $\mathcal{L}_{\text{CE}}$, the total loss becomes $\mathcal{L}_{\text{Total}} = \mathcal{L}_{\text{CE}} + \eta \mathcal{L}_{\text{rec}}$. In this work, the λ and η are set to 0.75 and 1, respectively.

3 Experiments and Results

3.1 Experiments Settings

Datasets. We conduct experiments on two publicly available real-world fMRI datasets. (a) **ABIDE**: This dataset consists of rs-fMRI data collected from 17 international sites. It contains fMRI data of 1009 subjects, with 516 (51.14%) being Autism spectrum disorder (ASD) patients. We select the first 100 time points for each subject. Considering the issue of multiple sites, a stratified sampling strategy is used for train-validation-test data split [10]. (b) **ADNI**: This

dataset comprises a total of 510 samples of data from 143 subjects, which can be divided into 4 categories based on the degree of cognitive impairment (NC, EMCI, LMCI, and AD). Each sample comprises 140 time points. In this paper, we mainly focus on the binary classification task of NC vs. (EMCI & LMCI & AD), with 370 (72.5%) samples from patients with cognitive impairment. For ABIDE and ADNI, the region definitions are based on the Craddock 200 atlas [4] and AAL 90 atlas [18], respectively.

Implementation. The non-overlapping sliding windows are set at a fixed size of 20. For the ABIDE dataset, samples are split into 5 segments, and for ADNI, they are split into 7 segments. The δ is set to 0.5, i.e., k is 100 and 45 for ABIDE and ADNI, respectively. We randomly split 70% of the dataset for training, 10% for validation, and use the remaining 20% as the test set. The model is trained with 200 epochs by using an early stopping strategy. The epoch with the highest AUROC performance on the validation set is used for performance comparison on the test set. All reported performances are the average of 5 random runs on the test set with the standard deviation. Additional settings are available in the code repository.

Table 1. Comparisons of different methods on ABIDE and ADNI (mean plus/minus std). The first and second-best results are **bold** and underlined, respectively. The superscript * denotes that models are trained in a non-fully supervised way.

Type	Methods	ABIDE		ADNI	
		Accuracy(%)	AUROC	Accuracy(%)	AUROC
Static	MLP	58.55±4.08	0.6025±0.0326	62.45±2.79	0.6013±0.0292
	BrainnetCNN	66.54±3.63	0.7399±0.0267	68.70±3.10	0.6424±0.0141
	FBNETGNN	61.80±2.17	0.5943±0.0277	70.20±2.56	0.6949±0.0302
	BCGCN	55.40±2.07	0.5647±0.0672	66.67±1.96	0.5541±0.0082
	VanillaTF	67.72±1.22	0.7191±0.0170	68.24±2.90	0.5652±0.0138
	BrainNetTF	69.52±1.58	0.7565±0.0449	71.00±3.80	0.7478±0.0543
Dynamic	RNN	60.80±1.79	0.6817±0.0221	68.12±3.58	0.7136±0.0227
	STGCN	66.20±1.26	0.6974±0.0303	70.98±1.64	0.6436±0.0227
	STAGIN	62.40±2.07	0.6221±0.0334	70.59±2.77	0.6738±0.0104
	DBGSL	56.49±2.51	0.5731±0.0398	67.05±2.15	0.5698±0.0373
	FDG-BrainMAE*	70.22±1.03	0.7269±0.0179	73.28±2.25	0.7608±0.0261
	TARDRL*	**71.83±3.34**	**0.7884±0.0549**	**76.08±2.91**	**0.8007±0.0255**

3.2 Results

Comparisons with SOTA methods. We compare our method with state-of-the-art deep learning-based methods, which can roughly be categorized into two types: static-FC and dynamic-FC methods. More specifically, static-FC methods

are MLP, BrainNetCNN [11], FBNETGNN [9], BCGCN [14], BrainNetTF [10] and its variant VanillaTF [10]. Dynamic-FC methods include RNN, STGCN [6], STAGIN [12], and DBGSL [3], FDG-BrainMAE [24]. FDG-BrainMAE is firstly pretrained in the reconstruction task with a variable mask ratio, and then fine-tuned in the prediction task. The comparison result is shown in Table.1. We can see that the self-supervised models achieve higher performance than models trained in a supervised way, indicating that MAE empowers deep learning models to acquire a genuine representation of fMRI. Among all the methods, TARDRL performs best across all datasets. Specifically, for ABIDE dataset, TARDRL outperforms the second-best baseline FDG-BrainMAE in terms of accuracy by 1.61% whereas on ADNI the FDG-BrainMAE lags with a considerable deficit of 2.8%. The superior performance of TARDRL is attributed to its attention-guided masking strategy. Additionally, TARDRL, FDG-BrainMAE, and BrainNetTF are transformer-based models, indicating the powerful representation learning capabilities of the transformer.

Ablation. We report the ablation study results in Table 2, which further justifies the attention-guided mask strategy Ω and emphasizes the enhancement through the incorporation of self-supervised learning (reconstruction). DRL uses the same architecture as the prediction task in TARDRL and is trained through supervised learning. TARDRL-Random disables the strategy Ω, thereby randomly selecting ROIs to mask and reconstruct. Besides, TARDRL-Opp uses the exact opposite strategy to Ω, which is to mask off low-attended ROIs. We observe that DRL's performances are comparatively inferior to all other self-supervised learning methods except for TARDRL-Opp. Furthermore, removing the strategy Ω or using the opposite of it, both come a degradation in performance. This underscores that the significance of reconstructing signals from the important ROIs, enabling the model to capture more information relevant to prediction tasks.

Table 2. Ablation study of our proposed method (mean plus/minus std).

Methods	ABIDE		ADNI	
	Accuracy(%)	**AUROC**	**Accuracy(%)**	**AUROC**
DRL	65.39±1.51	0.6580±0.0286	68.06±1.31	0.6711±0.0121
TARDRL-Random	69.72±1.32	0.7472±0.0078	70.95±2.65	0.7240±0.0565
TARDRL-Opp	67.50±2.39	0.6732±0.0304	67.83±3.05	0.6951±0.0424
TARDRL	**71.83±3.34**	**0.7884±0.0549**	**76.08±2.91**	**0.8007±0.0255**

Interpretability. To identify the discriminative brain regions associated with the end task, we create a brain region score vector using temporally ROI weights $z = \frac{1}{T}\sum_{t=1}^{T} \alpha_t \in \mathbb{R}^N$. The discriminative ROIs are determined by retaining all

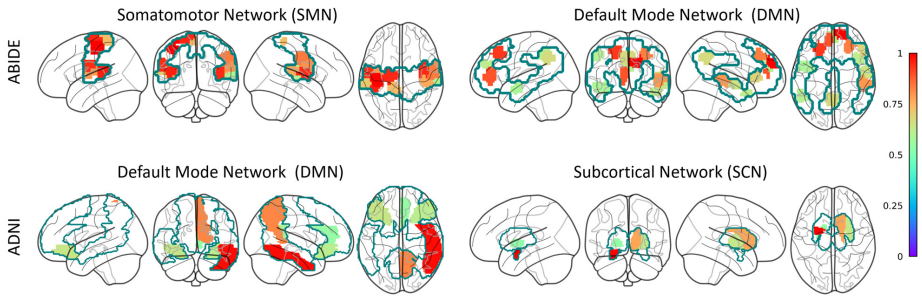

Fig. 2. The visualization of discriminative ROIs belonging to the top 2 functional networks. The teal contour outlines the area of the corresponding functional network.

regions falling within the top 25% across all subjects in the test set, and they are mapped with the definition of resting-state functional networks of Yeo et al [26]. For each functional network, we compute the proportion of the cumulative scores from its constituent discriminative ROIs, as shown in Table 3 and plot the discriminative ROIs belonging to the top 2 functional networks in Fig.2. For ABIDE, the SMN exhibits the highest contributions, which aligns with studies revealing altered sensory and motor processing in autistics [7,15,16]. Additionally, 24.37% of the highest scores are within the default mode network (DMN), a key network that is consistently observed in studies on autism [17,23]. For ADNI, 28.87% of the highest scores are attributed to the DMN, indicating its essential role in detecting the fundamental decline of cognitive function, in line with existing research [5,13]. Notably, the brain region located in the amygdala, a component of the subcortical network (SCN), stands out with a remarkably high score. This observation is aligned with neuroscience findings that the amygdala is intricately related to cognitive function impairment [25].

Table 3. The proportion of the cumulative scores from each functional network's constituent discriminative ROIs.

	DMN	**SMN**	**VN**	**FPN**	**DAN**	**VAN**	**LN**	**SCN**
ABIDE	24.37%	27.51%	16.16%	11.70%	10.58%	6.11%	1.91%	1.66%
ADNI	28.87%	7.80%	4.11%	16.35%	0%	8.46%	14.16%	20.25%

4 Conclusion

In this paper, we propose TARDRL, a novel model designed to enhance prediction performance through task-aware reconstruction from fMRI. In particular,

we set up the reconstruction and prediction task in parallel to learn robust task-aware representations. Besides, we design an attention-guided mask strategy to identify ROIs important to end tasks. Experiments on two real-world fMRI datasets demonstrate that our model achieves state-of-the-art results for brain disease prediction. Besides, its interpretability makes it a valuable tool for uncovering the mechanisms within the brain. We only consider disease detection as our prediction task. In fact, TARDRL is a universal framework that can be easily extended to more classification tasks such as emotion recognition and sex classification, as well as regression tasks such as age prediction and behavioral prediction. For future works, we will validate the feasibility of our model on other prediction tasks with different datasets.

Acknowledgments. This study was in part supported by the National Natural Science Foundation of China (Grant No. 62136004), the Fundamental Research Funds for the Central Universities (Grant No. NZ2024040), and the China Postdoctoral Science Foundation funded project (Grant No. 2021TQ0150 and No. 2021M701699).

Disclosure of Interests. No potential disclosure of interests was reported by the authors.

References

1. Bannadabhavi, A., Lee, S., Deng, W., Ying, R., Li, X.: Community-aware transformer for autism prediction in fmri connectome. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 287–297. Springer (2023)
2. Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J.D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., et al.: Language models are few-shot learners. Advances in neural information processing systems **33**, 1877–1901 (2020)
3. Campbell, A., Zippo, A.G., Passamonti, L., Toschi, N., Lio, P.: Dyndepnet: Learning time-varying dependency structures from fmri data via dynamic graph structure learning. In: ICML 3rd Workshop on Interpretable Machine Learning in Healthcare (IMLH) (2023)
4. Craddock, R.C., James, G.A., Holtzheimer III, P.E., Hu, X.P., Mayberg, H.S.: A whole brain fmri atlas generated via spatially constrained spectral clustering. Human brain mapping **33**(8), 1914–1928 (2012)
5. Dennis, E.L., Thompson, P.M.: Functional brain connectivity using fmri in aging and alzheimer's disease. Neuropsychology review **24**, 49–62 (2014)
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: Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part VII 23. pp. 528–538. Springer (2020)
7. Gowen, E., Hamilton, A.: Motor abilities in autism: a review using a computational context. Journal of autism and developmental disorders **43**, 323–344 (2013)
8. He, K., Chen, X., Xie, S., Li, Y., Dollár, P., Girshick, R.: Masked autoencoders are scalable vision learners. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 16000–16009 (2022)

9. Kan, X., Cui, H., Lukemire, J., Guo, Y., Yang, C.: Fbnetgen: Task-aware gnn-based fmri analysis via functional brain network generation. In: International Conference on Medical Imaging with Deep Learning. pp. 618–637. PMLR (2022)
10. Kan, X., Dai, W., Cui, H., Zhang, Z., Guo, Y., Yang, C.: Brain network transformer. Advances in Neural Information Processing Systems **35**, 25586–25599 (2022)
11. Kawahara, J., Brown, C.J., Miller, S.P., Booth, B.G., Chau, V., Grunau, R.E., Zwicker, J.G., Hamarneh, G.: Brainnetcnn: Convolutional neural networks for brain networks; towards predicting neurodevelopment. NeuroImage **146**, 1038–1049 (2017)
12. Kim, B.H., Ye, J.C., Kim, J.J.: Learning dynamic graph representation of brain connectome with spatio-temporal attention. Advances in Neural Information Processing Systems **34**, 4314–4327 (2021)
13. Li, H.J., Hou, X.H., Liu, H.H., Yue, C.L., He, Y., Zuo, X.N.: Toward systems neuroscience in mild cognitive impairment and alzheimer's disease: A meta-analysis of 75 fmri studies. Human brain mapping **36**(3), 1217–1232 (2015)
14. Li, Y., Zhang, X., Nie, J., Zhang, G., Fang, R., Xu, X., Wu, Z., Hu, D., Wang, L., Zhang, H., et al.: Brain connectivity based graph convolutional networks and its application to infant age prediction. IEEE Transactions on Medical Imaging **41**(10), 2764–2776 (2022)
15. Marco, E.J., Hinkley, L.B., Hill, S.S., Nagarajan, S.S.: Sensory processing in autism: a review of neurophysiologic findings. Pediatric research **69**(8), 48–54 (2011)
16. Mostofsky, S.H., Ewen, J.B.: Altered connectivity and action model formation in autism is autism. The Neuroscientist **17**(4), 437–448 (2011)
17. Padmanabhan, A., Lynch, C.J., Schaer, M., Menon, V.: The default mode network in autism. Biological Psychiatry: Cognitive Neuroscience and Neuroimaging **2**(6), 476–486 (2017)
18. Rolls, E.T., Huang, C.C., Lin, C.P., Feng, J., Joliot, M.: Automated anatomical labelling atlas 3. Neuroimage **206**, 116189 (2020)
19. Simpson, S.L., Bowman, F.D., Laurienti, P.J.: Analyzing complex functional brain networks: fusing statistics and network science to understand the brain. Statistics surveys **7**, 1 (2013)
20. Smith, S.M., Miller, K.L., Salimi-Khorshidi, G., Webster, M., Beckmann, C.F., Nichols, T.E., Ramsey, J.D., Woolrich, M.W.: Network modelling methods for fmri. Neuroimage **54**(2), 875–891 (2011)
21. Thomas, A., Ré, C., Poldrack, R.: Self-supervised learning of brain dynamics from broad neuroimaging data. Advances in Neural Information Processing Systems **35**, 21255–21269 (2022)
22. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. Advances in neural information processing systems **30** (2017)
23. Washington, S.D., Gordon, E.M., Brar, J., Warburton, S., Sawyer, A.T., Wolfe, A., Mease-Ference, E.R., Girton, L., Hailu, A., Mbwana, J., et al.: Dysmaturation of the default mode network in autism. Human brain mapping **35**(4), 1284–1296 (2014)
24. Yang, Y., Mao, Y., Liu, X.: Learning transferrable and interpretable representation for brain network (2024), https://openreview.net/forum?id=ajG8vLTHh5
25. Yao, H., Liu, Y., Zhou, B., Zhang, Z., An, N., Wang, P., Wang, L., Zhang, X., Jiang, T.: Decreased functional connectivity of the amygdala in alzheimer's disease

revealed by resting-state fmri. European journal of radiology **82**(9), 1531–1538 (2013)
26. Yeo, B.T., Krienen, F.M., Sepulcre, J., Sabuncu, M.R., Lashkari, D., Hollinshead, M., Roffman, J.L., Smoller, J.W., Zöllei, L., Polimeni, J.R., et al.: The organization of the human cerebral cortex estimated by intrinsic functional connectivity. Journal of neurophysiology (2011)
27. Zhang, J., Zhou, L., Wang, L., Liu, M., Shen, D.: Diffusion kernel attention network for brain disorder classification. IEEE Transactions on Medical Imaging **41**(10), 2814–2827 (2022)