CL-ADDA: Contrastive Learning with Amplitude-Driven Data Augmentation for fMRI-Based Individualized Predictions

Jiangcong Liu[1,2], Le Xu[1], Yun Guan[1], Hao Ma[1], and Lixia Tian[1(✉)]

[1] School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, China
lxtian@bjtu.edu.cn

[2] Beijing Key Laboratory of Traffic Data Analysis and Mining, Beijing Jiaotong University, Beijing 100044, China

Abstract. Effective representations of human brain function are essential for fMRI-based predictions of individual traits and classifications of neuropsychiatric disorders. Contrastive learning techniques can be favorable choices for representations of human brain function, if it were not for their requirement of large batch sizes. In this study, we proposed a novel method, namely, contrastive learning with amplitude-driven data augmentation (CL-ADDA), for effective representations of human brain function and ultimately fMRI-based individualized predictions. *SimSiam*, which sets no requirement on large batches, was used in this study to obtain discriminative representations among subjects to facilitate later predictions of individuals' traits. The fMRI data in this study was augmented based on recent neuroscience findings that fMRI frames with high- and low-amplitude are of quite different functional significance. Accordingly, we generated a positive pair by concatenating the fMRI frames with high-amplitude into one augmented sample and the frames with low-amplitude into another sample. The two augmented samples were used as inputs for CL-ADDA, and individualized predictions were made in an end-to-end way. The performance of the proposed CL-ADDA was evaluated with individualized age and IQ predictions based on a public dataset (Cam-CAN). The experimental results demonstrate that the proposed CL-ADDA can substantially improve the prediction performance as compared to the existing methods.

Keywords: *SimSiam* · Individualized Prediction · fMRI · Functional Connectivity

1 Introduction

In combination with machine learning techniques, functional magnetic resonance imaging (fMRI) has recently been widely used in predictions of individual traits (e.g., age and intelligence quotient (IQ)) [8, 12] and classifications of neuropsychiatric disorders (e.g.,

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/978-3-031-43907-0_37.

H. Greenspan et al. (Eds.): MICCAI 2023, LNCS 14220, pp. 384–393, 2023.
https://doi.org/10.1007/978-3-031-43907-0_37

Alzheimer's disease). Representations of human brain function are essential for such predictions, as effective representations can provide information discriminative among individuals and facilitate the final predictions.

Contrastive learning techniques can be favorable choices for representations of human brain function [1, 3, 4, 10, 11], if the available samples size is large enough for large-batch training (this is not the fact for most fMRI datasets). Contrastive learning can learn effective representations through minimizing/maximizing the distance between similar/dissimilar samples in the representation space [4, 11]. *SimSiam* [5] is a contrastive learning framework that maximizes the similarity between two augmentations of one image. With the involvement of a Siamese structure, *SimSiam* does not rely on large-batch training. Accordingly, *SimSiam* can be a favorable choice for fMRI-based representations of human brain function.

Generation of similar/dissimilar samples is critical for contrastive learning [4]. Among the few studies on fMRI-based individualized predictions in which contrastive learning is involved [7, 9, 13, 21], fMRI data has often been augmented using classic methods in the region of computer vision, such as random erasing, random cropping and adversarial generation. In the pioneering study [21], two highly similar augmented samples were generated for each subject by excerpting two non-overlapping fMRI temporal segments. The highly similar pairs containing redundancy information can lead to poor performance of contrastive learning [2, 24]. Recent neuroscience findings provide an intuitive idea regarding data augmentation for RS-fMRI data. Specifically, fMRI frames with high- and low-amplitude were reported to be of quite different functional significance [23]. Accordingly, a "discrepant-enough" positive pair can be generated by acquiring one sample based on fMRI frames with high-amplitude and the other based on frames with low-amplitude.

In this study, we proposed a framework named contrastive learning with amplitude-driven data augmentation (CL-ADDA) for effective representations of human brain function and ultimately fMRI-based individualized predictions. Two augmented samples of CL-ADDA were generated through excerpting fMRI frames with relatively high amplitude and those with relatively low amplitude. With two augmented samples of the same subject used as inputs, a *SimSiam*-based contrastive learning framework was used to learn effective representations of human brain function. For the consideration that label information can guide *SimSiam* to learn more prediction-relevant representation [15, 25], individualized predictions were performed in an end-to-end way through concatenated fully connected layers.

Our major contributions are as follows:

- *SimSiam* was utilized to learn representations of human brain function.
- A neuroscience-oriented amplitude-driven data augmentation method was introduced to generate positive pairs.
- Predictions were made in an end-to-end way to improve the generalizability of the predictive models.
- CL-ADDA outperformed a variety of state-of-the-art methods for fMRI-based individualized predictions.

2 Method

2.1 Overall Workflow of CL-ADDA

Figure 1 shows the workflow of the proposed CL-ADDA. fMRI data of one subject is first excerpted into two segments, one composed of frames with high-amplitude and the other composed of frames with low-amplitude. Two functional connectivity (FC) maps (FC_{high} and FC_{low}) can then be obtained based on the augmented samples, and the two FC maps are used as inputs for the contrastive learning module. The *SimSiam* structure is used to perform contrastive learning in this study, and the classic convolution in *SimSiam* is replaced by row and column convolutions to adapt to FC maps. *SimSiam* learns representations of brain function (r_{high} and r_{low}) based on the two FC maps, using encoders (F) with shared parameters. Individualized predictions can be made through a predictor (Φ) (three fully connected layers in this study) based on the learned representations. The whole model is trained though simultaneously minimizing the distance between the two representations (r_{high} and r_{low}) and the difference between the predicted and actual label (individual traits in this study).

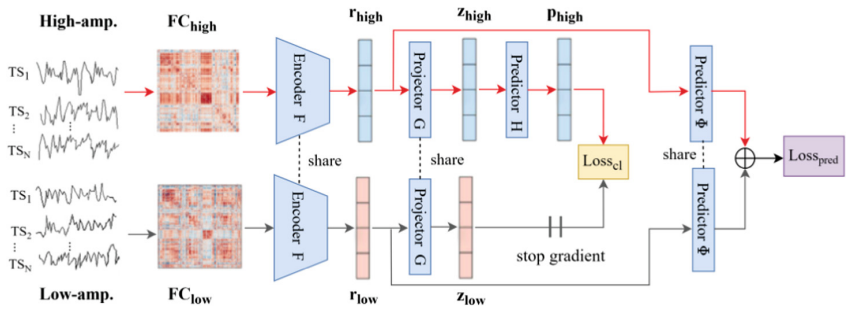

Fig. 1. Workflow of CL-ADDA. The high- and low-amplitude time series are generated from the same subject. TS - time series; FC - functional connectivity; amp. - amplitude.

2.2 Amplitude-Driven Data Augmentation

Figure 2 provides an illustration of the amplitude-driven data augmentation method. In this study, we generated two augmented samples for each subject based on the amplitude of fMRI frames following [23]. Specifically, we first defined N ROIs and extracted the mean time series of each ROI. We then z-scored each time series and obtained frame-wise co-fluctuation of each ROI pair as follows:

$$e_{jk} = \left[e^1_{jk}, e^2_{jk}, ..., e^T_{jk}\right],\ e^t_{jk} = x^t_j \cdot x^t_k \tag{1}$$

where e^t_{jk} is the co-fluctuation of ROIs-j and k at time t; x^t_j $\left(x^t_k\right)$ is the z-scored fMRI signal amplitude of ROI-j (-k) at time t; e_{jk} is obtained the co-fluctuation time series, and T is the length of fMRI time series. Accordingly, a $C^2_N \times T$ co-fluctuation matrix

$E = [e_{12}; e_{13}; ...e_{1N}; e_{23}; e_{24}; ...e_{2N}; ...e_{(N-1)N}]$ can be obtained. We computed the root sum square (RSS) of the co-fluctuation matrix E at each time point and finally generated one high-amplitude sample by excerpting the top 50% of frames with high co-fluctuation RSS, and one low-amplitude sample by concatenating the remaining frames (with low co-fluctuation RSS).

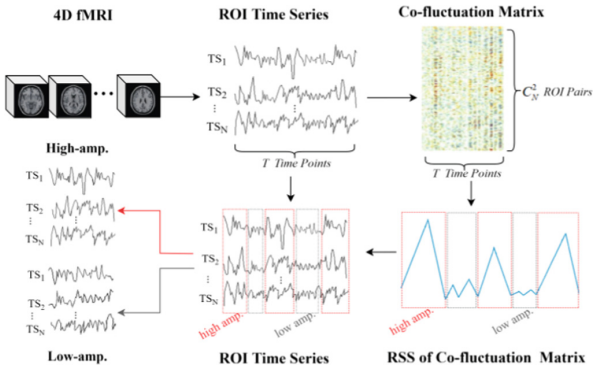

Fig. 2. Illustration of the amplitude-driven data augmentation method. TS - time series; amp. - amplitude; RSS - root sum square.

2.3 Contrastive Learning on Functional Connectivity Maps

We constructed the contrastive learning model based mainly on *SimSiam*. Setting no requirement on large batches, *SimSiam* can be a favorable choice for fMRI-based representation learning [5]. As shown in Fig. 1, the *SimSiam* structure in this study consists of two branches of the same encoders (F) with shared weights for encoding the two input FC maps in parallel, followed by the same nonlinear projectors (G) with share weights for further processing the brain function representations from the encoders, and one predictor (H) on one branch for transforming the output of the branch to match it to the output of the other branch. A stop-gradient operation is applied on the branch without predictor to avoid model collapsing [5].

For the consideration that spatial locality does not exist among adjacent elements on FC maps [14], row and column convolutions were used to construct the backbone of the encoder (F) and extract effective information from the FC maps. Specifically, each encoder (F) in this study consists of one row convolution layer (with $C_r@1 \times N$ row convolution filter, C_r is the channel number, N is the ROI number) and one column convolution layer (with $C_c@N \times 1$ column convolution filters, C_c is the channel number, outputs $C_c@1 \times 1$ representation). Information throughout the brain is expected to be integrated with the use of row and column convolutions. The row and column convolution layers in CL-ADDA are each followed by a batch normalization layer and a *Leaky_ReLU* layer.

2.4 Individualized Prediction and Loss Function

As shown in Fig. 1, individualized prediction in this study was implemented using two parallel predictors (Φ) with shared weights. Each predictor ((Φ)) consists of three fully connected layers, which transform the C_c outputs from the corresponding encoder (F) (learned representations of brain function, r_{high} and $r_{low} \in R^{C_c}$ in Fig. 1) into the predicted individual trait for the branch. For model training, the final prediction is made based on a weighted sum of the predictions from the two branches as follows:

$$\hat{y} = \alpha \cdot \Phi\left(F\left(\xi_{\text{high}}\right)\right) + (1-\alpha) \cdot \Phi(F(\xi_{\text{low}})) \tag{2}$$

where $\xi \in \mathrm{R}^{\mathrm{N}\times\mathrm{N}}$ denotes a FC map, F denotes the encoder, and α is a hyper-parameter. In the model testing stage, prediction can directly be made as $\hat{y} = \Phi(F(\xi))$, where ξ can be a FC map calculated based on the whole fMRI scan (rather than augmented data).

The whole loss function for CL-ADDA includes two parts: contrastive loss and prediction loss. Contrastive learning minimizes the negative cosine similarity between the two branches:

$$\begin{aligned} \mathrm{D}(\mathrm{p}_{high}, \mathrm{z}_{\text{low}}) &= -\frac{p_{high}}{||p_{high}||_2} \cdot \frac{z_{low}}{||z_{low}||_2} \\ \mathrm{D}(\mathrm{p}_{low}, \mathrm{z}_{\text{high}}) &= -\frac{p_{low}}{||p_{low}||_2} \cdot \frac{z_{high}}{||z_{high}||_2} \end{aligned} \tag{3}$$

where $p_{low} = H(G(F(\xi_{low})))$, $p_{high} = H(G(F(\xi_{high})))$, $z_{low} = G(F(\xi_{low}))$, $z_{high} = G(F(\xi_{\text{high}}))$, G denotes the projector in Fig. 1, and H denotes the predictor in the contrastive learning (Fig. 1); $\|.\|_2$ is L2-norm.

Following [25], we defined contrastive loss as:

$$Loss_{cl} = \frac{1}{2}D(p_{low}, stopgrad(z_{high})) + \frac{1}{2}D(p_{high}, stopgrad(z_{low})) \tag{4}$$

L1 loss was used as the prediction loss:

$$Loss_{pred} = ||y - \hat{y}||_1 \tag{5}$$

where y and $\hat{y}$ are the actual and predicted labels, respectively. The total loss was defined as follows:

$$Loss = \lambda \cdot Loss_{pred} + (1-\lambda) \cdot Loss_{cl} \tag{6}$$

where λ is hyper-parameter.

3 Experiments and Results

3.1 Dataset

The resting-state fMRI data included in the dataset collected and released by the Cambridge Centre for Ageing and Neuroscience (Cam-CAN) [19] was used in this study. The public dataset contains multi-modal data from a large cohort of adult lifespan population-based samples. After removing the subjects with excessive head motions

(translation/rotation more than 2.0 mm/2.0° in/around any of the x, y, or z directions) throughout the scan, 600 subjects remained (18–87, 53.900 ± 18.549 years), and IQ scores of 568 of them were available (11–44, 32.032 ± 6.762). We preprocessed the data to remove the spatial and temporal artifacts and register the images to standard space (MNI) using FSL. We defined 200 ROIs based on independent component analysis and extracted the ROI time series through regressing the spatial maps of the independent components released by the Human Connectome Project (HCP) [20] against the preprocessed fMRI data using the *dual_regress* command included in FSL. Later analyses were all based on the extracted ROI time series.

3.2 The Performance of CL-ADDA

Age and IQ predictions were taken as test cases to evaluate the performance of the proposed method, based on the Cam-CAN dataset. Amplitude-driven data augmentation was performed on the 200 ROI time series of each subject, and two FC maps were obtained based on two augmented samples for each subject. The two augmented FC maps were used as inputs for *SimSiam* for representation learning and later individualized predictions. We performed 1000 epochs of model training, with the batch size set to 128. For the consideration that high-amplitude FC maps (FC_{high}) may carry more detailed information about individuals' brain function, we empirically weight the predictions based on FC_{high} more, by setting the α in Eq. (2) to 0.8 The hyper-parameter λ in Eq. (6) was set to 0.5. Adam optimizer with a learning rate of 0.001 was used.

Ten-fold cross-validation was used to evaluate the performance of CL-ADDA, and Pearson's correlation coefficient (r) and mean absolute error (MAE) between the predicted and actual labels were used to quantitatively measure this performance. The results show that CL-ADDA performed well on both age and IQ predictions, as indicated by an r-value (MAE) of 0.886 (6.992 years) for age prediction, and 0.620 (4.531) for IQ prediction.

3.3 Comparison Experiments

We compared the performance of our proposed method with six deep learning methods for fMRI-based individualized predictions, namely, spatial-temporal graph convolutional network (ST-GCN) [22], RNN based on gated recurrent units (GRU) [6], pooling regularized graph neural network (PR-GNN) [16], brain graph neural network (BrainGNN) [17], simple fully convolutional network (SFCN) [18], and BrainNetCNN [14]. Each of the six methods has been reported to perform well on fMRI-based individualized predictions, or even provide state-of-the-art results. Each method was implemented based on its online code, with the hyper-parameters set according to its original paper.

Table 1 is a list of prediction accuracies based on the seven methods (including CL-ADDA). According to Table 1, CL-ADDA outperformed the methods for comparison by large margins. For instance, compared with the second best (ST-GCN), CL-ADDA demonstrated an r-value increase of 0.085 for IQ prediction.

Table 1. Age and IQ prediction accuracies based on different deep learning methods.

Age Prediction							
Method	ST-GCN	GRU	PR-GNN	BrainGNN	SFCN	BrainNetCNN	CL-ADDA (ours)
r	0.848	0.789	0.808	0.811	0.738	0.711	**0.886**
MAE (years)	7.702	9.733	9.347	9.008	9.815	11.566	**6.992**
IQ Prediction							
Method	ST-GCN	GRU	PR-GNN	BrainGNN	SFCN	BrainNetCNN	CL-ADDA (ours)
r	0.535	0.458	0.488	0.515	0.402	0.412	**0.620**
MAE	4.602	5.572	5.263	4.614	5.53	6.935	**4.531**

3.4 Ablation Experiments

To evaluate the effectiveness of the proposed amplitude-driven data augmentation strategy, we performed age and IQ predictions with the data augmented using classic methods, and the strategy of excerpting non-overlapping segments as proposed in [21]. For classic augmentation, two 175 × 175 augmented FC maps were generated by applying random cropping and Gaussian blurring on the 200 × 200 FC matrix calculated based on the whole fMRI scan [5]. For the non-overlapping segment excerpting strategy, we generated two 200 × 200 augmented FC matrices based on the two non-overlapping fMRI data segments excerpted from the same scan [21]. According to Table 2, the proposed amplitude-driven data augmentation is obviously superior to the other two data augmentation strategies.

We further evaluated the effectiveness of contrastive learning, as well as the end-to-end individualized prediction strategy. Specifically, (1) we performed individualized age and IQ predictions based on a network composed of one encoder (F) and one predictor (Φ) to imitate a network with contrastive learning removed. (2) We pre-trained CL-ADDA and then predicted age and IQ using the representations based on the pre-trained CL-ADDA to imitate abandoning the end-to-end individualized prediction strategy. According to Table 3, both contrastive learning and the end-to-end individualized prediction strategy were critical for CL-ADDA. The results indicate that supervised contrastive learning can be a favorable choice for neuroimage-based individualized predictions and neuropsychiatric disease classifications.

Table 2. Age and IQ prediction accuracies based on different data augmentation strategies.

Age Prediction			
Method	Classic Augmentation	Non-Overlapping Segments	Amplitude-Driven (ours)
r	0.349	0.831	**0.886**
MAE (years)	14.849	9.167	**6.992**
IQ Prediction			
Method	Classic Augmentation	Non-overlapping Segments	Amplitude-Driven (ours)
r	0.288	0.589	**0.620**
MAE	5.333	5.034	**4.531**

Table 3. Age and IQ prediction accuracies based on CL-ADDA with different modules being removed.

Age Prediction			
Method	with Individualized Prediction Removed	with Contrastive Learning Removed	CL-ADDA (ours)
r	0.782	0.873	**0.886**
MAE (years)	9.467	7.377	**6.992**
IQ Prediction			
Method	with Individualized Prediction Removed	with Contrastive Learning Removed	CL-ADDA
r	0.573	0.589	**0.620**
MAE	4.779	5.042	**4.531**

4 Conclusion

In this study, we proposed CL-ADDA for effective representation learning and ultimately precise fMRI-based individualized predictions. Originating from a recent neuroscientific finding, the proposed amplitude-driven data augmentation method provides the contrastive learning module discrepant-enough positive pairs for effective representation learning. *SimSiam*-based contrastive learning enables effective representation learning on fMRI dataset including limited samples. We evaluated the performance of CL-ADDA with age and IQ predictions based on a public dataset, and the experiments demonstrate that CL-ADDA achieved state-of-the-art predictions for both age and IQ.

Acknowledgement. We thank investigators from Cambridge Centre for Ageing and Neuroscience for sharing the public dataset.

References

1. Bachman, P., Hjelm, R.D., Buchwalter, W.: Learning representations by maximizing mutual information across views. Adv. Neural Inf. Process. Syst. **32** (2019)
2. Barlow, H.B.: Possible principles underlying the transformation of sensory messages. Sens. Commun. **1**(01), 217–233 (1961)
3. Caron, M., Misra, I., Mairal, J., Goyal, P., Bojanowski, P., et al.: Unsupervised learning of visual features by contrasting cluster assignments. Adv. Neural. Inf. Process. Syst. **33**, 9912–9924 (2020)
4. Chen, T., Kornblith, S., Norouzi, M., Hinton, G.: A simple framework for contrastive learning of visual representations. In: International Conference on Machine Learning, pp. 1597–1607. PMLR (2020)
5. Chen, X., He, K.: Exploring simple siamese representation learning. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 15750–15758 (2021)
6. Chung, J., Gulcehre, C., Cho, K., Bengio, Y.: Empirical evaluation of gated recurrent neural networks on sequence modeling. arXiv preprint arXiv:1412.3555 (2014)
7. Dufumier, B., Gori, P., Victor, J., Grigis, A., Wessa, M., Brambilla, P., Favre, P., Polosan, M., McDonald, C., Piguet, C.M., Phillips, M., Eyler, L., Duchesnay, E.: Contrastive learning with continuous proxy meta-data for 3D MRI classification. In: de Bruijne, M., Cattin, P.C., Cotin, S., Padoy, N., Speidel, S., Zheng, Y., Essert, C. (eds.) MICCAI 2021. LNCS, vol. 12902, pp. 58–68. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87196-3_6
8. 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., Abolmaesumi, P., Stoyanov, D., Mateus, D., Zuluaga, M.A., Zhou, S.K., Racoceanu, D., Joskowicz, L. (eds.) MICCAI 2020. LNCS, vol. 12267, pp. 528–538. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59728-3_52
9. Grigis, A., Gomez, C., Tasserie, J., Ambroise, C., Frouin, V., et al.: Predicting cortical signatures of consciousness using dynamic functional connectivity graph-convolutional neural networks. BioRxiv, pp. 2020–2005 (2020)
10. Grill, J.B., Strub, F., Altché, F., Tallec, C., Richemond, P., et al.: Bootstrap your own latent-a new approach to self-supervised learning. Adv. Neural. Inf. Process. Syst. **33**, 21271–21284 (2020)
11. He, K., Fan, H., Wu, Y., Xie, S., Girshick, R.: Momentum contrast for unsupervised visual representation learning. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 9729–9738 (2020)
12. He, T., Kong, R., Holmes, A.J., Nguyen, M., Sabuncu, M.R., et al.: Deep neural networks and kernel regression achieve comparable accuracies for functional connectivity prediction of behavior and demographics. Neuroimage **206**, 116276 (2020)
13. Hsieh, W.T., Lefort-Besnard, J., Yang, H.C., Kuo, L.W., Lee, C.C.: Behavior score-embedded brain encoder network for improved classification of Alzheimer disease using resting state fMRI. In: International Conference of the IEEE Engineering in Medicine & Biology Society, pp. 5486–5489. IEEE (2020)
14. Kawahara, J., Brown, C.J., Miller, S.P., Booth, B.G., Chau, V., et al.: BrainNetCNN: convolutional neural networks for brain networks; towards predicting neurodevelopment. Neuroimage **146**, 1038–1049 (2017)
15. Li, J., Zhao, G., Tao, Y., Zhai, P., Chen, H., et al.: Multi-task contrastive learning for automatic CT and X-ray diagnosis of COVID-19. Pattern Recogn. **114**, 107848 (2021)
16. Li, X., Zhou, Y., Dvornek, N.C., Zhang, M., Zhuang, J., Ventola, P., Duncan, J.S.: Pooling regularized graph neural network for fmri biomarker analysis. In: Martel, A.L., Abolmaesumi, P., Stoyanov, D., Mateus, D., Zuluaga, M.A., Zhou, S.K., Racoceanu, D., Joskowicz, L. (eds.)

MICCAI 2020. LNCS, vol. 12267, pp. 625–635. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59728-3_61
17. Li, X., Zhou, Y., Dvornek, N., Zhang, M., Gao, S., et al.: Braingnn: interpretable brain graph neural network for FMRI analysis. Med. Image Anal. **74**, 102233 (2021)
18. Peng, H., Gong, W., Beckmann, C.F., Vedaldi, A., Smith, S.M.: Accurate brain age prediction with lightweight deep neural networks. Med. Image Anal. **68**, 101871 (2021)
19. Taylor, J.R., Williams, N., Cusack, R., Auer, T., Shafto, M.A., et al.: The Cambridge Centre for Ageing and Neuroscience (Cam-CAN) data repository: Structural and functional MRI, MEG, and cognitive data from a cross-sectional adult lifespan sample. Neuroimage **144**, 262–269 (2017)
20. Van Essen, D.C., Ugurbil, K., Auerbach, E., Barch, D., Behrens, T.E., et al.: The human connectome project: a data acquisition perspective. Neuroimage **62**(4), 2222–2231 (2012)
21. Wang, X., Yao, L., Rekik, I., Zhang, Y.: Contrastive functional connectivity graph learning for population-based fMRI classification. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds.) Medical Image Computing and Computer Assisted Intervention – MICCAI 2022. MICCAI 2022. Lecture Notes in Computer Science, vol. 13431, pp. 221–230. Springer, Cham.https://doi.org/10.1007/978-3-031-16431-6_21
22. Yan, S., Xiong, Y., Lin, D.: Spatial temporal graph convolutional networks for skeleton-based action recognition. In: Proceedings of the AAAI Conference on Artificial Intelligence vol. 32, no. 1 (2018)
23. Zamani Esfahlani, F., Jo, Y., Faskowitz, J., Byrge, L., Kennedy, D.P., et al.: High-amplitude cofluctuations in cortical activity drive functional connectivity. Proc. Natl. Acad. Sci. **117**(45), 28393–28401 (2020)
24. Zbontar, J., Jing, L., Misra, I., LeCun, Y., Deny, S.: Barlow twins: Self-supervised learning via redundancy reduction. In: International Conference on Machine Learning, pp. 12310–12320. PMLR (2021)
25. Zhao, Z., Liu, H.: Semi-supervised feature selection via spectral analysis. In: Proceedings of the SIAM international conference on data mining. pp. 641−646. Society for Industrial and Applied Mathematics (2007)