Published in final edited form as:

Med Image Comput Comput Assist Interv. 2023 October ; 14222: 138–148.

doi:10.1007/978-3-031-43898-1_14.

Bidirectional Mapping with Contrastive Learning on Multimodal Neuroimaging Data

Kai Ye[1], **Haoteng Tang**[2], **Siyuan Dai**[1], **Lei Guo**[1], **Johnny Yuehan Liu**[3], **Yalin Wang**[4], **Alex Leow**[5], **Paul M. Thompson**[6], **Heng Huang**[7], **Liang Zhan**[1]

[1]University of Pittsburgh, Pittsburgh, PA 15260, USA

[2]University of Texas Rio Grande Valley, Edinburg, TX 78539, USA

[3]Thomas Jefferson High School for Science and Technology, Alexandria, VA 22312, USA

[4]Arizona State University, Tempe, AZ 85287, USA

[5]University of Illinois at Chicago, Chicago, IL 60612, USA

[6]University of Southern California, Los Angeles, CA 90032, USA

[7]University of Maryland, College Park, MD 20742, USA

Abstract

The modeling of the interaction between brain structure and function using deep learning techniques has yielded remarkable success in identifying potential biomarkers for different clinical phenotypes and brain diseases. However, most existing studies focus on one-way mapping, either projecting brain function to brain structure or inversely. This type of unidirectional mapping approach is limited by the fact that it treats the mapping as a one-way task and neglects the intrinsic unity between these two modalities. Moreover, when dealing with the same biological brain, mapping from structure to function and from function to structure yields dissimilar outcomes, highlighting the likelihood of bias in one-way mapping. To address this issue, we propose a novel bidirectional mapping model, named Bidirectional Mapping with Contrastive Learning (BMCL), to reduce the bias between these two unidirectional mappings via ROI-level contrastive learning. We evaluate our framework on clinical phenotype and neurodegenerative disease predictions using two publicly available datasets (HCP and OASIS). Our results demonstrate the superiority of BMCL compared to several state-of-the-art methods.

Keywords

Bidirectional reconstruction; BOLD signals; Structural networks; Prediction; Biomarkers

1 Introduction

Recent advancements in applying machine learning techniques to MRI-based brain imaging studies have shown substantial progress in predicting neurodegenerative diseases (e.g.,

tanghaoteng@gmail.com, zhan.liang@gmail.com.

Alzheimer's Disease or AD) and clinical phenotypes (e.g., behavior measures), and in uncovering novel biomarkers that are closely related to them [4]. Different MRI techniques can be used to depict different aspects of the brain organization or dynamics [8,19,23]. In general, diffusion MRI can derive brain structural networks that depict the connectivity of white matter tracks among brain regions, which gains system-level insights into the brain structural changes related to brain diseases and those phenotypes [29]. However, the structural networks may not inform us about whether this tract or the regions it connects are "activated" or "not activated" in a specific state. As a complementary counterpart, the functional MRI provides measures of BOLD (blood-oxygen-level-dependent) signals to present activities of brain regions over time [3], but no clue on whether those regions are physically connected or not. Therefore, different brain imaging data provide distinct but complementary information, and separately analyzing the data of each modality will always be suboptimal. In this context, multimodal approaches are being explored to improve prediction accuracy by integrating multiple information sources [9,13,27,31–33]. For example, it has been shown that combining different modalities of data (e.g., image and text) can enhance performance in image classification and clustering tasks [27,33]. In the healthcare field, multimodal machine learning has shown its potential in disease detection and diagnosis [13]. In brain imaging studies, many studies aim to explore multimodal MRI data representations by modeling the communications between functional MRI and its structural counterpart. Most of these studies primarily focus on establishing a unidirectional mapping between these two imaging modalities (i.e., mapping from structural MRI data to the functional counterpart [24,32], or the inverse [16,31]). However, for the same biological brain, these two mappings generate distinct results, which highlights the likelihood of bias in the unidirectional mapping approach.

To address this, we propose a novel bidirectional mapping framework, where the mapping from structural MRI data (i.e., diffusion MRI-derived brain structural network) to the functional counterpart (i.e., BOLD signals) and the inverse mapping are implemented simultaneously. Unlike previous studies [6,15,22,28,32] that employ unidirectional mappings, our approach leverages bidirectional mapping, minimizing the discrepancies in the latent space of each one-way mapping through contrastive learning at the brain region-of-interest level (ROI level). This method subsequently unveils the inherent unity across both imaging modalities. Moreover, our framework is interpretable, where we employ integrated gradients [20] to generate brain saliency maps for interpreting the outcomes of our model. Specifically, the identified top key brain ROIs in the brain saliency maps are closely related to the predicted diseases and clinical phenotypes. Extensive experiments have been conducted to demonstrate the effectiveness and superiority of our proposed method on two publicly available datasets (i.e., the Human Connectome Project (HCP), and Open Access Series of Imaging Studies (OASIS)). In summary, the contributions of this paper can be outlined as follows:

– We propose a novel bidirectional framework to yield multimodal brain MRI representations by modeling the interactions between brain structure and the functional counterpart.

– We use contrastive learning to extract the intrinsic unity of both modalities.

– The experimental results on two publicly available datasets demonstrate the superiority of our proposed method in predicting neurodegenerative diseases and clinical phenotypes. Furthermore, the interpretability analysis highlights that our method provides biologically meaningful insights.

2 Method

The proposed bidirectional mapping framework (Fig. 1) comprises two encoder-decoder structures. One constructs BOLD signals from structural networks, while the other performs the inverse mapping. A ROI-level's contrastive learning is utilized between the encoder and decoder to minimize the distinction of the latent spaces within two reconstruction mappings. Finally, a multilayer perceptron (MLP) is utilized for task predictions. It's worth mentioning that instead of using the functional connectivity matrix, we directly utilize BOLD signals for bidirectional mapping. We believe this approach is reasonable as it allows us to capture the dynamic nature of the brain through the BOLD time sequence. Using the functional connectivity matrix may potentially disrupt this dynamic information due to the calculations of correlations. Furthermore, our experiments indicate that our encoder can directly model the temporal relations between different brain regions from the BOLD signals, eliminating the need to construct functional networks.

Preliminaries.

A structural brain network is an attributed and weighted graph $\mathcal{G} = (A, H)$ with N nodes, where $H \in \mathbb{R}^{N \times d}$ is the node feature matrix, and $A \in \mathbb{R}^{N \times N}$ is the adjacency matrix where $a_{i,j} \in \mathbb{R}$ represents the edge weight between node i and node j. Meanwhile, we utilize $X_{\mathcal{B}} \in \mathbb{R}^{N \times T}$ to represent the BOLD signal matrix derived from functional MRI data of each subject, where each brain ROI has a time series BOLD signal with T points.

Reconstruction.

For the reconstruction task, we deploy an encoder-decoder architecture and utilize the L_1 loss function. Particularly, we use a multi-layer feed-forward neural network as the encoder and decoder. Our method differs from previous studies [21,32], where the encoder and decoder do not necessitate a GNN-based framework, allowing us to directly utilize the adjacency matrix A of structural networks as the inputs. Previous studies randomly initialize the node features (i.e., H) for the GNN input, since it is difficult to find informative brain node features that provide valuable information from the HCP and OASIS datasets. Hence, we propose a reconstruction framework that detours using the node feature matrix. Our framework is bidirectional, where we simultaneously conduct structural network and BOLD signal reconstruction. Here, we have latent representations $Z_{\mathcal{B}} = Encoder_B(X_{\mathcal{B}})$ and $Z_{\mathcal{S}} = Encoder_S(A)$ for BOLD signals and structural networks, respectively.

ROI-Level's Contrastive Representation Learning.

With latent representation $Z_{\mathcal{B}} \in \mathbb{R}^{N \times d_B}$ generated from BOLD signal and $Z_{\mathcal{S}} \in \mathbb{R}^{N \times d_S}$ from structural networks, we then conduct ROI-level's contrastive learning to associate the static structural and dynamic functional patterns of multimodal brain measurements. The

contrastive learning loss aims to minimize the distinctions between latent representations from two modalities. To this end, we first utilize linear layers to project $Z_{\mathcal{B}}$ and $Z_{\mathcal{S}}$ to the common space, where we obtain $Z'_{\mathcal{B}} = WZ_{\mathcal{B}} + b$, $Z'_{\mathcal{B}} \in \mathbb{R}^{N \times d}$ and similarly, $Z'_{\mathcal{S}} \in \mathbb{R}^{N \times d}$. We use $\left(z_i^B, z_i^S\right)_{i=1\cdots N}$ to denote representations from the same ROI, where z_i^B and z_i^S are elements of $Z'_{\mathcal{B}}$ and $Z'_{\mathcal{S}}$, respectively. For the same brain ROI, the static structural representation and the dynamic functional counterpart are expected to share a maximum similarity. Conversely, for the pairs that do not match, represented as $\left(z_i^B, z_j^S\right)_{i \neq j}$, these are drawn from different ROIs and should share a minimum similarity.

To formally build up the ROI-level's contrastive loss, it is intuitive to construct positive samples and negative ones based on the match of ROIs. Specifically, we construct $\left(z_i^B, z_i^S\right)_{i=1\cdots N}$ as positive sample pair, and $\left(z_i^B, z_j^S\right)_{i \neq j}$ as negative sample pair. And our contrastive loss can be formulated as follow:

$$\begin{aligned}
\mathcal{L}_{C1} &= -\mathbb{E}\left[\underset{i=1\cdots N}{\log} \frac{Similarity\left(z_i^B, z_i^S\right)}{\sum_{j=1}^{N} Similarity\left(z_i^B, z_j^S\right)}\right] \\
\mathcal{L}_{C2} &= -\mathbb{E}\left[\underset{i=1\cdots N}{\log} \frac{Similarity\left(z_i^S, z_i^B\right)}{\sum_{j=1}^{N} Similarity\left(z_i^S, z_j^B\right)}\right] \\
\mathcal{L}_{contrast} &= \mathcal{L}_{C1} + \mathcal{L}_{C2}
\end{aligned} \tag{1}$$

where $Similarity(\cdot)$ is substantiated as cosine similarity.

Loss Functions.

The loss functions within our proposed framework are summarized here. Besides the reconstruction loss ($\mathcal{L}_{rec}$) and the ROI-level's contrastive loss ($\mathcal{L}_{contrast}$), we utilize cross-entropy loss ($\mathcal{L}_{supervised} = \mathcal{L}_{cross-entropy}$) for classification tasks, and L_1 loss ($\mathcal{L}_{supervised} = \mathcal{L}_{mean-absolute-error}$) for regression tasks, respectively. In summary, the loss function can be described as:

$$\mathcal{L} = \eta_1 \mathcal{L}_{contrast} + \eta_2 \mathcal{L}_{rec} + \eta_3 \mathcal{L}_{supervised}, \tag{2}$$

where η_1, η_2 and η_3 are loss weights.

3 Experiments

3.1 Data Description and Preprocessing

Two publicly available datasets were used to evaluate our framework. The first includes data from 1206 young healthy subjects (mean age 28.19 ± 7.15, 657 women) from the Human Connectome Project [25] (HCP). The second includes 1326 subjects (mean age = 70.42 ± 8.95, 738 women) from the Open Access Series of Imaging Studies (OASIS)

dataset [12]. Details of each dataset may be found on their official websites. CONN [26] and FSL [10] were used to reconstruct the functional and structural networks, respectively. For the HCP data, both networks have a dimension of 82 × 82 based on 82 ROIs defined using FreeSurfer (V6.0) [7]. For the OASIS data, both networks have a dimension of 132 × 132 based on the Harvard-Oxford Atlas and AAL Atlas. We deliberately chose different network resolutions for HCP and OASIS, to evaluate whether the performance of our new framework is affected by the network dimension or atlas. The source code is available at: https://github.com/FlynnYe/BMCL.

3.2 Experimental Setup and Evaluation Metrics

We randomly split each dataset into 5 disjoint sets for 5-fold cross-validations, and all the results are reported in *mean* (*s.t.d.*) across 5 folds. To evaluate the performance of each model, we utilize accuracy, precision score, and F_1 score for classification tasks, and mean absolute error (MAE) for regression tasks. The learning rate is set as 1×10^{-4} and 1×10^{-3} for classification and regression tasks, respectively. The loss weights (i.e., η_1, η_2, and η_3) are set equally as 1/3. To demonstrate the superiority of our method in cross-modal learning, bidirectional mapping, and ROI-level's contrastive learning, we select four baselines including 2 single-modal graph learning methods (i.e., DIFFPOOL [30] and SAGPOOL [14]), as well as 2 multimodal methods (i.e., VGAE [11] and DSBGM [22]) for all tasks. We use both functional brain networks, in which edge weights are defined as the *Pearson Correlation* between BOLD signals, and brain structural networks as input for baseline methods. The functional brain networks are signed graphs including positive and negative edge weights, however, the DIFFPOOL, SAGPOOL, and VGAE can only take unsigned graphs (i.e., graphs only include positive edges) as input. Therefore, we convert the functional brain networks to unsigned graphs by using the absolute values of the edge weights.

3.3 BOLD Signal and Structural Network Reconstruction

We train the model in a task-free manner where no task-specific supervised loss is involved. The MAE values between the edge weights in the ground-truth and reconstructed structural networks are 0.0413 ± 0.0009 and 0.0309 ± 0.0015 under 5-fold cross-validation on the HCP and OASIS, respectively. The MAE values between ground-truth and reconstructed BOLD signals are 0.0049 ± 0.0001 and 0.0734 ± 0.0016 on the HCP and OASIS, respectively. The reconstruction results on HCP are visualized in Fig. 2.

3.4 Disease and Sex Classification

We conduct Alzheimer's disease (AD) classification on the OASIS dataset, and sex classification on the HCP dataset. As shown in Table 1, our proposed BMCL can achieve the best results in accuracy, precision, and F_1 score for both tasks among all methods. For example, in the AD classification, our model outperforms the baselines with at least 4.2%, 5.8% and 4.0% increases in accuracy, precision and F_1 scores, respectively. In general, multimodal methods can outperform single-model methods. The superiority of our bidirectional BMCL model, compared to the unidirectional methods, attributes to the

fact that our BMCL reduces the distinction between the latent spaces generated by two unidirectional mappings through ROI-level's contrastive learning.

3.5 ASR and MMSE Regression

Mini-Mental State Exam (MMSE) is a quantitative measure of cognitive status in adults, and Adult Self-Report scale (ASR) [1] is to measure the adult's behavior. As shown in Table 2, our proposed BMCL model outperforms all baselines in terms of MAE values. The regression results also demonstrate the superiority of bidirectional mapping and the importance of ROI-level's contrastive learning, which is consistent with the results in the classification tasks.

3.6 Ablation Study

To demonstrate the significance of bidirectional mapping, we remove a part of our proposed BMCL model to yield two unidirectional mappings (i.e., either mapping from structural network to BOLD signal, or mapping inversely). As shown in the bottom three rows in Table 1 and Table 2, the prediction results are declined when we remove each directional mapping, which clearly demonstrates the importance of bidirectional mapping.

3.7 Interpretability

The 10 key brain regions (Fig. 3) associated with AD (from OASIS) and with each sex (from HCP) are identified using the brain saliency map. The salient regions for AD are concentrated in cerebelum (i.e., cerebelum 3 right and left, cerebelum 8 left, cerebelum crus2 right) and middle Temporal gyrus (i.e., the posterior division left and right, as well as the temporooccipital right of middle temporal gyrus), which have been verified as core AD biomarkers in literature [2,18]. Similarly, 10 key regions (Fig. 3) are identified for regression tasks (i.e., 3 ASR from HCP and MMSE from OASIS). Interestingly, several brain regions (including left and right accumbens areas, cortex left hemisphere cuneus and insula, as well as cortex right hemisphere posteriorcingulate and parahippocampal) are consistently identified across 3 ASR scales (i.e., aggression, rule-break, and intrusive). This finding is supported by [21], which suggests that similar ASR exhibits common or similar biomarkers. Also, these regions have been reported as important biomarkers for aggressive-related behaviors in literature [5,17].

4 Conclusions

We propose a new multimodal data mining framework, named BMCL, to learn the representation from two modality data through bidirectional mapping between them. The elaborated ROI-level contrastive learning in BMCL can reduce the distinction and eliminate biases between two one-way mappings. Our results on two publicly available datasets show that BMCL outperforms all baselines, which demonstrates the superiority of bidirectional mapping with ROI-level contrastive learning. Beyond these, our model can identify key brain regions highly related to different clinical phenotypes and brain diseases, which demonstrates that our framework is interpretable and the results are biologically meaningful. The contrastive learning method, while emphasizing the alignment of features from different modalities, may inadvertently neglect the unique characteristics inherent to each modality.

Moving forward, we intend to refine our method by aiming for a balance between the alignment of modalities and the preservation of modality-specific information. Additionally, the pre-selection of important features or the consideration of subnetworks holds promising for further research.

Acknowledgments.

This study was partially supported by NIH (R01AG071243, R01MH125928, R21AG065942, R01EY032125, and U01AG068057) and NSF (IIS 2045848 and IIS 1837956).

References

1. Achenbach TM, McConaughy S, Ivanova M, Rescorla L: Manual for the ASEBA brief problem monitor (BPM), vol. 33. ASEBA, Burlington, VT (2011)
2. Airavaara M, Pletnikova O, Doyle ME, Zhang YE, Troncoso JC, Liu QR: Identification of novel GDNF isoforms and cis-antisense GDNFOS gene and their regulation in human middle temporal gyrus of Alzheimer disease. J. Biol. Chem 286(52), 45093–45102 (2011) [PubMed: 22081608]
3. Bathelt J, O'Reilly H, Clayden JD, Cross JH, de Haan M: Functional brain network organisation of children between 2 and 5 years derived from reconstructed activity of cortical sources of high-density EEG recordings. Neuroimage 82, 595–604 (2013) [PubMed: 23769920]
4. Calhoun VD, Sui J: Multimodal fusion of brain imaging data: a key to finding the missing link (s) in complex mental illness. Biol. Psychiatry Cogn. Neurosci. Neuroimaging 1(3), 230–244 (2016) [PubMed: 27347565]
5. Couppis MH, Kennedy CH: The rewarding effect of aggression is reduced by nucleus accumbens dopamine receptor antagonism in mice. Psychopharmacology 197, 449–456 (2008) [PubMed: 18193405]
6. Cui H, et al. : BrainGB: a benchmark for brain network analysis with graph neural networks. IEEE Trans. Med. Imaging (2022)
7. Fischl B: Freesurfer. Neuroimage 62(2), 774–781 (2012) [PubMed: 22248573]
8. Fornito A, Zalesky A, Bullmore E: Fundamentals of Brain Network Analysis. Academic Press, Cambridge (2016)
9. Freeman D, Ha D, Metz L: Learning to predict without looking ahead: world models without forward prediction. In: Advances in Neural Information Processing Systems, vol. 32 (2019)
10. Jenkinson M, Beckmann CF, Behrens TE, Woolrich MW, Smith SM: FSL. Neuroimage 62(2), 782–790 (2012) [PubMed: 21979382]
11. Kipf TN, Welling M: Variational graph auto-encoders. arXiv preprint arXiv:1611.07308 (2016)
12. LaMontagne PJ, et al. : Oasis-3: longitudinal neuroimaging, clinical, and cognitive dataset for normal aging and Alzheimer disease. MedRxiv pp. 2019–2012 (2019)
13. Lee G, Nho K, Kang B, Sohn KA, Kim D: Predicting Alzheimer's disease progression using multi-modal deep learning approach. Sci. Rep 9(1), 1952 (2019) [PubMed: 30760848]
14. Lee J, Lee I, Kang J: Self-attention graph pooling. In: International Conference on Machine Learning, pp. 3734–3743. PMLR (2019)
15. Li X, et al. : BrainGNN: interpretable brain graph neural network for fMRI analysis. Med. Image Anal 74, 102233 (2021) [PubMed: 34655865]
16. Liu Y, Ge E, Qiang N, Liu T, Ge B: Spatial-temporal convolutional attention for mapping functional brain networks. arXiv preprint arXiv:2211.02315 (2022)
17. Peterson CK, Shackman AJ, Harmon-Jones E: The role of asymmetrical frontal cortical activity in aggression. Psychophysiology 45(1), 86–92 (2008) [PubMed: 17850239]
18. Piras IS, et al. : Transcriptome changes in the Alzheimer's disease middle temporal gyrus: importance of RNA metabolism and mitochondria-associated membrane genes. J. Alzheimers Dis 70(3), 691–713 (2019) [PubMed: 31256118]

19. Qi S, Meesters S, Nicolay K, ter Haar Romeny BM, Ossenblok P: The influence of construction methodology on structural brain network measures: a review. J. Neurosci. Methods 253, 170–182 (2015) [PubMed: 26129743]
20. Sundararajan M, Taly A, Yan Q: Axiomatic attribution for deep networks. In: International Conference on Machine Learning, pp. 3319–3328. PMLR (2017)
21. Tang H, et al.: Hierarchical brain embedding using explainable graph learning. In: 2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI), pp. 1–5. IEEE (2022)
22. Tang H, et al. : Signed graph representation learning for functional-to-structural brain network mapping. Med. Image Anal 83, 102674 (2023) [PubMed: 36442294]
23. Tang H, Ma G, Guo L, Fu X, Huang H, Zhan L: Contrastive brain network learning via hierarchical signed graph pooling model. IEEE Trans. Neural Netw. Learn. Syst (2022)
24. Tewarie P, et al. : Mapping functional brain networks from the structural connectome: relating the series expansion and eigenmode approaches. Neuroimage 216, 116805 (2020) [PubMed: 32335264]
25. Van Essen DC, et al. : The Wu-Minn human connectome project: an overview. Neuroimage 80, 62–79 (2013) [PubMed: 23684880]
26. Whitfield-Gabrieli S, Nieto-Castanon A: Conn: a functional connectivity toolbox for correlated and anticorrelated brain networks. Brain Connect 2(3), 125–141 (2012) [PubMed: 22642651]
27. Xu K, et al.: Show, attend and tell: neural image caption generation with visual attention. In: International Conference on Machine Learning, pp. 2048–2057. PMLR (2015)
28. Yan J, et al. : Modeling spatio-temporal patterns of holistic functional brain networks via multi-head guided attention graph neural networks (multi-head GaGNNs). Med. Image Anal 80, 102518 (2022) [PubMed: 35749981]
29. Yeh CH, Jones DK, Liang X, Descoteaux M, Connelly A: Mapping structural connectivity using diffusion MRI: challenges and opportunities. J. Magn. Reson. Imaging 53(6), 1666–1682 (2021) [PubMed: 32557893]
30. Ying Z, You J, Morris C, Ren X, Hamilton W, Leskovec J: Hierarchical graph representation learning with differentiable pooling. In: Advances in Neural Information Processing Systems, vol. 31 (2018)
31. Zhang L, Wang L, Zhu D, Initiative ADN, et al. : Predicting brain structural network using functional connectivity. Med. Image Anal 79, 102463 (2022) [PubMed: 35490597]
32. Zhang W, Zhan L, Thompson P, Wang Y: Deep representation learning for multimodal brain networks. In: Martel AL, et al. (eds.) MICCAI 2020. LNCS, vol. 12267, pp. 613–624. Springer, Cham (2020). 10.1007/978-3-030-59728-3_60
33. Zhen L, Hu P, Wang X, Peng D: Deep supervised cross-modal retrieval. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 10394–10403 (2019)

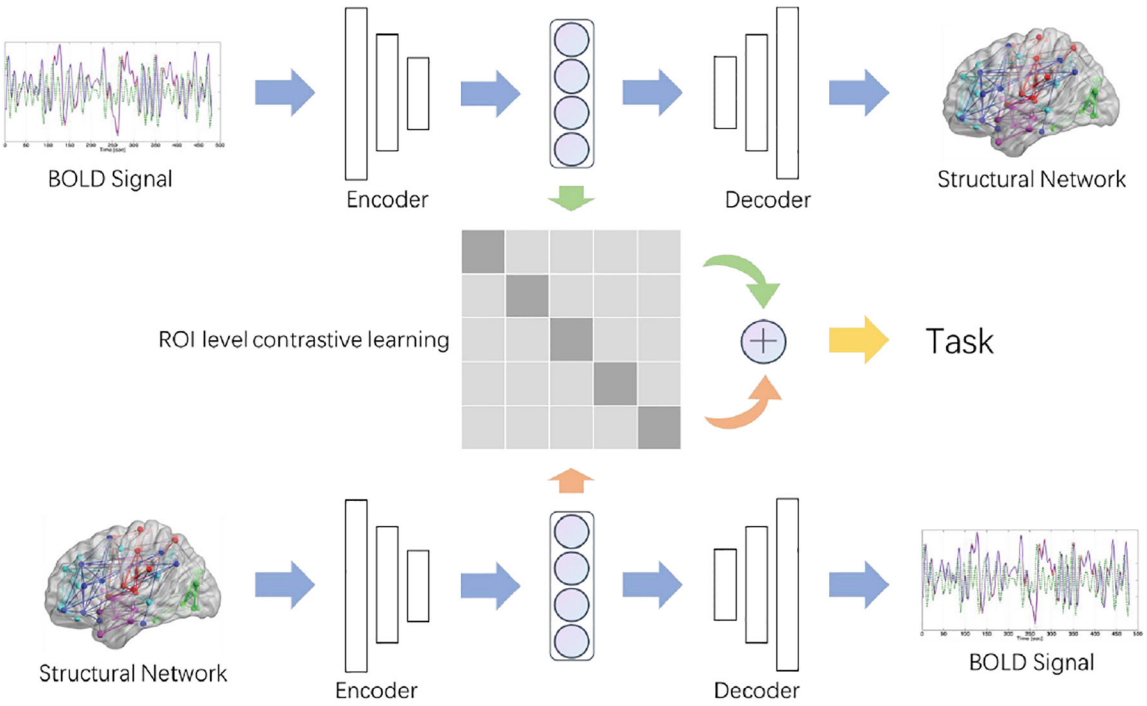
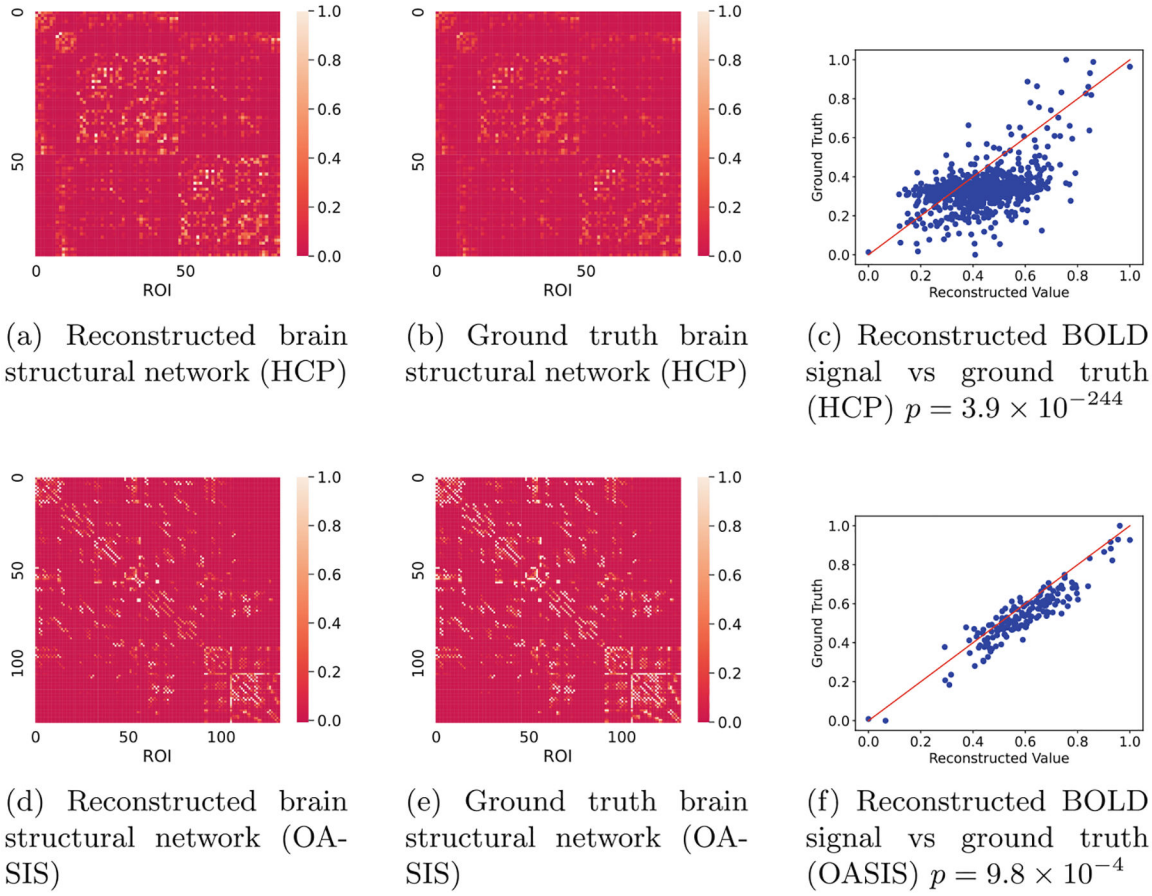


Fig. 1.

The pipeline of Bidirectional Mapping with Contrastive Learning (BMCL). The brain structural network and BOLD signals are initially processed by two separate encoders for representation learning. Afterward, ROI-level contrastive learning is applied to these extracted representations, facilitating their alignment in a common space. These derived representations are then utilized for downstream prediction tasks.



(a) Reconstructed brain structural network (HCP)

(b) Ground truth brain structural network (HCP)

(c) Reconstructed BOLD signal vs ground truth (HCP) $p = 3.9 \times 10^{-244}$

(d) Reconstructed brain structural network (OASIS)

(e) Ground truth brain structural network (OASIS)

(f) Reconstructed BOLD signal vs ground truth (OASIS) $p = 9.8 \times 10^{-4}$

Fig. 2.
Bidirectional reconstruction results on the HCP and OASIS dataset.

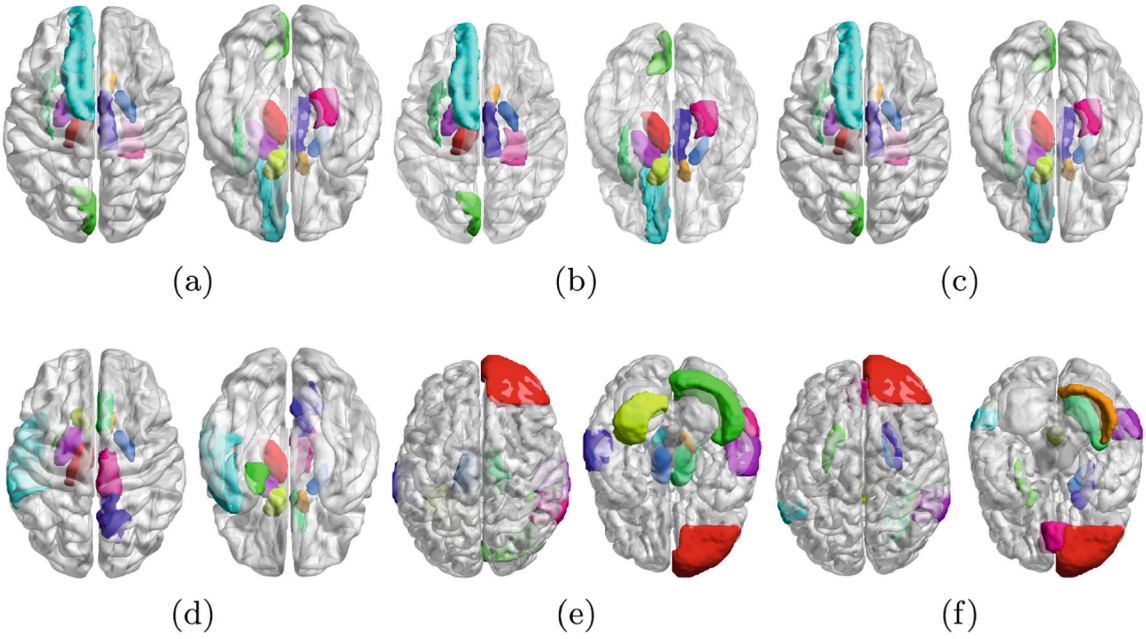


Fig. 3.

Saliency maps to identify top 10 regions associated with (a) intrusiveness, (b) aggression, (c) rule-break, (d) sex, (e) AD and (f) MMSE, respectively.

Table 1.

The results for sex classification on HCP and AD classification on OASIS.

Method	HCP (gender)			OASIS (disease)		
	Acc	**Pre**	**F1**	**Acc**	**Pre**	**F1**
DIFFPOOL† w/ F	67.77 (3.56)	65.25 (2.65)	68.82 (1.72)	68.97 (1.34)	66.03 (3.36)	69.24 (1.83)
SAGPOOL† w/ F	70.95 (2.88)	69.83 (1.85)	71.44 (1.29)	65.65 (2.01)	63.33 (1.95)	67.27 (2.09)
DIFFPOOL† w/ S	58.71 (4.62)	30.96 (4.73)	40.6 (5.17)	86.04 (2.65)	64.92 (4.16)	74.01 (3.64)
SAGPOOL† w/ S	61.06 (4.58)	32.79 (3.54)	42.64 (3.78)	88.48 (2.51)	68.71 (3.92)	77.33 (3.44)
VGAE	73.59(2.42)	74.43 (1.84)	76.25 (1.49)	64.68 (2.49)	62.57 (2.19)	65.85 (1.91)
DSBGM	82.19 (2.01)	85.35 (1.99)	84.71(2.37)	78.92 (1.38)	79.81 (1.41)	80.22(2.25)
BMCL w/o F	93.68 (2.88)	91.71 (2.19)	92.31 (2.31)	90.09 (2.65)	71.26 (4.16)	83.61 (3.64)
BMCL w/o S	69.54 (1.77)	68.61 (1.71)	56.82 (2.60)	89.66 (2.93)	73.35 (3.15)	79.52 (3.29)
BMCL	**94.83 (1.35)**	**93.47 (3.65)**	**93.21 (1.98)**	**92.23 (0.62)**	**84.47(2.14)**	**83.38(0.76)**

The best results are highlighted in **bold** font. Methods marked with † are unimodal methods.

Table 2.

The experimental results for ASR regression on HCP and MMSE regression on OASIS.

Method	HCP (aggression)	HCP (rule-break)	HCP (intrusive)	OASIS (MMSE)
DIFFPOOL† w/ F	2.39 (0.021)	2.26 (0.0092)	2.47 (0.15)	1.77 (0.56)
SAGPOOL† w/ F	3.07 (0.062)	2.88 (0.0022)	3.47 (0.029)	1.73 (0.79)
DIFFPOOL† w/ S	1.78 (0.268)	1.12 (0.473)	0.61 (0.3335)	2.13 (15.5941)
SAGPOOL† w/ S	1.82 (0.2674)	1.13 (0.3672)	0.63 (0.2608)	0.53 (0.2125)
VGAE	1.74(0.019)	1.37(0.051)	0.67 (0.022)	1.27 (0.25)
DSBGM	1.71 (0.11)	1.21 (0.24)	0.65 (0.026)	0.87 (0.18)
BMCL w/o F	1.98 (0.2688)	1.12 (0.3508)	0.62 (0.3145)	0.49 (0.1908)
BMCL w/o S	2.03 (0.2045)	1.11 (0.3704)	0.63 (0.3839)	0.50 (0.2008)
BMCL	**1.68 (0.2374)**	**1.05 (0.5046)**	**0.58 (0.3377)**	**0.45 (0.1726)**

The best results are highlighted in **bold** font. Methods marked with † are unimodal methods.