SCMT: SEMANTIC-GUIDED CROSS-MODAL TRANSFER FOR MULTIMODAL BRAIN NETWORK LEARNING

Anonymous authors
Paper under double-blind review

ABSTRACT

Learning from functional and structural brain networks is constrained by the scarcity of paired neuroimaging data. Larger functional-only cohorts offer additional information, but differences in modalities and representations make direct transfer challenging. We propose **SCMT**, a semantic-guided cross-modal transfer framework that uses atlas-defined brain regions as a shared interface. SCMT calibrates language-derived regional descriptions with source functional connectivity and diagnosis supervision, then transfers regional affinities and a healthy-connectivity reference to multimodal learning. A dual-branch reconstruction model combines subject-specific structure–function relationships with an auxiliary functional view guided by the transferred priors. An annealed consistency objective progressively shifts emphasis from population guidance toward the observed paired data. Experiments on ABIDE I and ADNI show improved classification over the base reconstruction model under label-blind transductive adaptation. Fusion and source-scale analyses further characterize the role of transferred information. These findings support atlas-level transfer as an approach to learning multimodal brain representations from limited paired data. *Code will be made publicly available upon acceptance.*

1 INTRODUCTION

Structural and functional brain networks describe complementary aspects of neural organization. Structural connectivity (SC) derived from DTI reflects anatomical pathways, whereas functional connectivity (FC) derived from fMRI measures statistical synchronization between brain regions (Park & Friston, 2013; Sporns, 2013). Jointly modeling them can reveal alterations that are difficult to identify from either modality alone. In practice, acquiring paired fMRI and DTI requires each subject to complete both imaging protocols, making multimodal cohorts costly to construct and typically limited in scale. This scarcity is particularly challenging for flexible graph models that must learn FC–SC correspondence from a small number of paired observations.

A natural direction is to use information from relatively abundant fMRI-only cohorts. Existing pretraining methods exploit brain networks through masked modeling, graph self-supervision, and Transformer objectives (Wen et al., 2023; Hu et al., 2024; Yang et al., 2024b). However, directly transferring complete encoder parameters or subject-level representations is difficult when the source functional model and target multimodal model differ in architecture, modality composition, and representation space. We therefore study a unimodal-to-multimodal transfer setting: a larger fMRI-only source cohort supplies population information, while a small, non-overlapping paired cohort supplies the target structure–function relationships.

Atlas-defined regions provide a shared interface between these settings. Their anatomical identities are consistent across subjects, and their indices can be matched between functional and structural connectomes. Rather than transferring a source patient's embedding, we transfer relations between regions. Language-derived anatomical-functional descriptions provide an initial representation of each ROI; source functional learning then calibrates these descriptions to the observed connectivity. This produces a compact atlas-level prior that can be used by a separately trained target encoder.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

The remaining challenge is to combine population information with individual variation. A fixed prior imposed throughout target learning could obscure subject-specific connectivity. SCMT addresses this through dual cross-modal reconstruction. A base branch learns from the observed FC and SC, while a second branch uses the transferred affinity and a healthy-connectivity reference to construct an auxiliary functional view. Consistency with that view regularizes target learning, and an annealed weight gradually shifts emphasis toward the paired data. At prediction time, frozen encoders process the original inputs and their representations are fused.

We evaluate SCMT on paired ABIDE I and ADNI cohorts. Target reconstruction is label-blind and transductive: it can access target connectivity before the downstream classification folds are evaluated. The study includes site-disjoint source-to-target transfer within ABIDE and transfer of the same source priors to ADNI. The reported gains are larger on ABIDE; ADNI provides a more limited positive result under this protocol. Our contributions are:

- An atlas-level transfer interface that carries calibrated ROI relations and a healthy-connectivity reference from functional source learning to multimodal target reconstruction.
- A dual-branch reconstruction strategy that combines population guidance with subject-specific FC–SC correspondence through an annealed consistency objective.
- Evaluation of classification, representation fusion, and source-pretraining scale on two cohorts, with additional initialization, efficiency, and ROI analyses characterizing the complete framework.

2 Related Work

Multimodal brain-network methods learn from paired connectomes through graph fusion, cross-modal interaction, or mutual learning (Liu et al., 2022; Yang et al., 2024a; Li et al., 2025; Feng et al., 2026; Yang et al., 2026). Brain pretraining and language-guided representations offer complementary sources of information (Wen et al., 2023; Hu et al., 2024; Yang et al., 2024b; Xu et al., 2025). SCMT connects these directions by transferring calibrated ROI relations to a separately trained multimodal model. Its source representation combines language, connectivity, and diagnostic supervision. Appendix A provides the full discussion of these research directions and their relation to SCMT.

3 Methods

Figure 1 summarizes the three stages of SCMT. Source pretraining learns an ROI affinity and healthy FC reference; target reconstruction uses them to regularize a dual-branch multimodal model; downstream classification combines the frozen representations of the original FC and SC. The source encoder and target encoders are distinct: the transferred objects are atlas-level matrices rather than source model parameters.

3.1 Preliminaries

The brain is parcellated into V regions of interest (ROIs) based on the AAL atlas. For subject i, the functional connectivity (FC) matrix $\mathbf{X}_i^f \in \mathbb{R}^{V\times V}$ is defined by the Pearson correlation coefficients between ROI BOLD time series, and the structural connectivity (SC) matrix $\mathbf{X}_i^s \in \mathbb{R}^{V\times V}$ is constructed via DTI tractography. We represent these modality-specific matrices as $\mathbf{X}_i^m$, where $m \in \{f, s\}$.

To capture high-order topological relationships, we instantiate a Top-k spatial hypergraph $\mathbf{H}_i^m = \mathcal{G}(\mathbf{X}_i^m)$ for each modality, where each node forms hyperedges with its k-nearest functional or structural neighbors. Spatial feature updates are performed via standard hypergraph neural networks (Feng et al., 2019).

3.2 Semantic Functional Pretraining

Semantic functional pretraining uses the larger fMRI-only cohort to calibrate anatomical-functional ROI descriptions against observed connectivity. The descriptions provide a common regional vo-

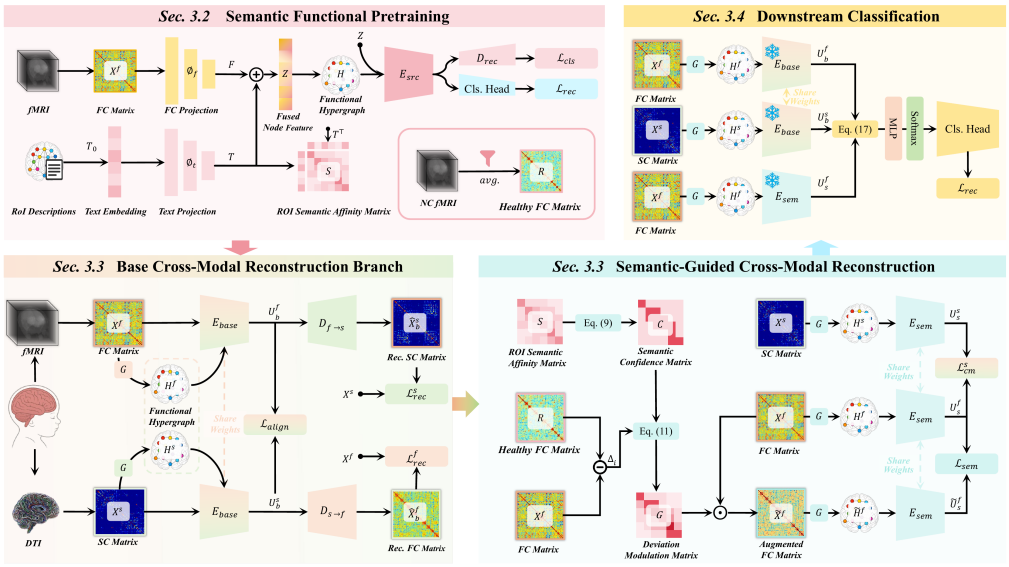


Figure 1: **Overview of the proposed SCMT**. SCMT first calibrates LLM-derived atlas semantics using a larger fMRI-only source cohort and transfers the resulting prior to guide dual FC-SC reconstruction in the small paired target cohort. The semantic constraint is gradually annealed, allowing subject-specific multimodal evidence to dominate the final representation.

cabulary, whereas source learning adapts their representation to the functional data. The resulting relation matrix is indexed by ROI identity and can therefore be used independently of the source encoder architecture. This stage also estimates a healthy-connectivity reference for measuring each target subject's functional deviations.

For each ROI $v \in \{1, \dots, V\}$, LLM-derived ROI anatomical-functional descriptions are encoded offline by a frozen text encoder, producing initial semantic embeddings $\mathbf{T}_0 \in \mathbb{R}^{V \times d_l}$. Given the FC matrix $\mathbf{X}_i^f$ of source subject i, textual and functional features are projected and fused as:

$$\begin{aligned} \mathbf{T} &= \mathrm{Norm}(\phi_t(\mathbf{T}_0)), \qquad \mathbf{F}_i = \phi_f(\mathbf{X}_i^f), \\ \mathbf{Z}_i &= \mathrm{LN}(\mathbf{F}_i + \mathbf{T}), \end{aligned} \tag{1}$$

where ϕ_t and ϕ_f are projection heads into a shared d-dimensional space. In particular, ϕ_f maps each ROI's connectivity profile to a node feature, while ϕ_t maps its text embedding to the same feature dimension. Their sum combines subject-specific functional information with an ROI representation shared across source subjects, and LN denotes layer normalization. The fused node features $\mathbf{Z}_i$ are used to construct the functional hypergraph $\mathbf{H}_i^f = \mathcal{G}(\mathbf{Z}_i)$, which is then encoded by the source hypergraph encoder E_{src}:

$$\mathbf{U}_i^f = E_{\mathrm{src}}(\mathbf{Z}_i, \mathbf{H}_i^f). \tag{2}$$

Disease classification and masked FC reconstruction objectives are jointly optimized to calibrate the language-functional fusion pathway:

$$\mathcal{L}_{\mathrm{src}} = \lambda_{\mathrm{cls}}\mathcal{L}_{\mathrm{cls}} + \lambda_{\mathrm{rec}}\mathcal{L}_{\mathrm{rec}}, \tag{3}$$

where $\mathcal{L}_{\mathrm{cls}}$ denotes standard cross-entropy loss for disease classification, and $\mathcal{L}_{\mathrm{rec}}$ is the masked FC reconstruction loss computed via squared Frobenius norm over masked functional connectivity entries.

Through this optimization, the language-derived embedding $\mathbf{T}$ is transformed into a functionally grounded semantic representation aligned with population-level connectivity organization. After pretraining, we construct the ROI semantic affinity matrix as:

$$S_{uv} = \frac{\mathbf{T_u}^\top \mathbf{T_v}}{\|\mathbf{T_u}\|_2 \|\mathbf{T_v}\|_2}, \tag{4}$$

where $\mathbf{S} \in \mathbb{R}^{V \times V}$ represents the semantic relationships between ROI pairs. Meanwhile, the normative functional reference is calculated as:

$$\mathbf{R} = \frac{1}{|\mathcal{I}_{\mathrm{HC}}|} \sum_{i \in \mathcal{I}_{\mathrm{HC}}} \mathbf{X}_i^f, \tag{5}$$

where $\mathcal{I}_{\mathrm{HC}}$ contains source healthy controls. Because source training also uses disease labels, the transferred affinity is task-calibrated rather than disease-independent. The healthy mean is a source-cohort reference, not a validated target-population normative model.

After pretraining, source-specific components are discarded, and only $\mathbf{S}$ and $\mathbf{R}$ are retained. The two matrices serve different purposes: $\mathbf{S}$ ranks relations between regions, while $\mathbf{R}$ gives a reference connectivity value for each ROI pair. Together they define an auxiliary view of a target subject rather than supplying a source subject's representation. Target encoders are trained separately; neither source encoder parameters nor source patient embeddings are transferred.

3.3 Semantic-Guided Dual Cross-Modal Reconstruction

For each target subject, paired FC and SC matrices define hypergraphs $\mathbf{H}_i^m = \mathcal{G}(\mathbf{X}_i^m)$ using the Top-k construction above. The two reconstruction branches share the aim of learning FC–SC correspondence but receive different training constraints. The base branch learns directly from the paired observations. The semantic branch has a separate encoder and adds consistency with a prior-guided functional view. Keeping both branches allows the final representation to retain a pathway trained without the transferred prior.

Base Cross-Modal Reconstruction Branch. The base branch learns modality-shared structural-functional correspondence without external semantic constraints. A shared base encoder E_{base} maps both modalities into a common cross-modal latent space:

$$\mathbf{U}_{i,b}^f = E_{\mathrm{base}}(\mathbf{X}_i^f, \mathbf{H}_i^f), \qquad \mathbf{U}_{i,b}^s = E_{\mathrm{base}}(\mathbf{X}_i^s, \mathbf{H}_i^s). \tag{6}$$

Two independent direction-specific decoders are employed for bidirectional cross-modal translation:

$$\widehat{\mathbf{X}}_i^s = D_b^{f \to s}(\mathbf{U}_{i,b}^f), \qquad \widehat{\mathbf{X}}_i^f = D_b^{s \to f}(\mathbf{U}_{i,b}^s). \tag{7}$$

The base objective combines bidirectional reconstruction with cross-modal latent alignment:

$$\mathcal{L}_{\mathrm{cm}}^b = \sum_{m \in \{f,s\}} \mathcal{L}_{\mathrm{rec}}^m(\widehat{\mathbf{X}}_i^m, \mathbf{X}_i^m) + \beta \mathcal{L}_{\mathrm{align}}(U_{i,b}^s, U_{i,b}^f), \tag{8}$$

For each modality, reconstruction uses squared Frobenius error between the predicted and observed connectivity matrices. The alignment term uses squared ℓ_2 distance between functional and structural latent embeddings, with β controlling its contribution. Bidirectional reconstruction asks each modality to retain information relevant to the other, while latent alignment encourages a common representation of the paired subject. No source affinity is used in this branch.

Semantic-Guided Deviation-Aware Reconstruction Branch. The semantic branch combines two quantities that cannot be substituted for one another: the subject's deviation from the source healthy reference and the source-calibrated relation between the corresponding regions. A large deviation alone does not identify which ROI relations the prior emphasizes, and a high affinity alone does not describe the individual subject. Their interaction defines a deterministic auxiliary functional view, which regularizes the separate encoder E_{sem} alongside cross-modal reconstruction.

We first transform $\mathbf{S}$ into a bounded relative-affinity matrix:

$$\mathbf{C} = \mathrm{Sigmoid}\left(\frac{\mathbf{S} - \mu_{\mathrm{off}}(\mathbf{S})}{\sigma_{\mathrm{off}}(\mathbf{S})}\right), \tag{9}$$

where μ_{off} and σ_{off} denote the mean and nonzero standard deviation of off-diagonal entries. Standardization expresses each affinity relative to the other ROI pairs; the sigmoid then gives bounded weights. Thus $\mathbf{C}$ represents relative affinity rather than a calibrated biological probability. This

distinction matters when raw cosine similarities occupy a narrow range. The deviation of subject i from healthy reference is:

$$\mathbf{\Delta}_i = |\mathbf{X}_i^f - \mathbf{R}|. \tag{10}$$

A semantic-guided deviation modulation matrix G_i is then formulated as:

$$\mathbf{G}_i = \mathbf{1} + \tanh\left(\frac{\mathbf{\Delta}_i}{\tau}\right) \odot \mathbf{C}, \tag{11}$$

$$\widetilde{\mathbf{X}}_i^f = \mathbf{X}_i^f \odot \mathbf{G}_i, \tag{12}$$

where τ is a temperature parameter controlling sensitivity to functional deviations. Because $\mathbf{\Delta}_i$ is nonnegative and $\mathbf{C}$ is bounded, the multiplicative factor is at least one and below two. Connections with larger deviations and higher relative affinities receive stronger emphasis, while a zero deviation leaves the corresponding entry unchanged. The transformation preserves the signs and zero entries of FC; it does not create a nonzero entry from zero. However, the Top-k hypergraph $\widetilde{\mathbf{H}}_i^f = \mathcal{G}(\widetilde{\mathbf{X}}_i^f)$ can select different neighbors after this rescaling. Deviations from the source reference may reflect cohort effects as well as disease-related variation.

The original inputs and the modulated semantic view are encoded by E_{sem}:

$$\begin{aligned} \mathbf{U}_{i,s}^f &= E_{\text{sem}}(\mathbf{X}_i^f, \mathbf{H}_i^f), \qquad \mathbf{U}_{i,s}^s = E_{\text{sem}}(\mathbf{X}_i^s, \mathbf{H}_i^s), \\ \widetilde{\mathbf{U}}_{i,s}^f &= E_{\text{sem}}(\widetilde{\mathbf{X}}_i^f, \widetilde{\mathbf{H}}_i^f). \end{aligned} \tag{13}$$

While $\mathbf{U}_{i,s}^f$ and $\mathbf{U}_{i,s}^s$ optimize the standard cross-modal loss $\mathcal{L}_{\text{cm}}^s$ (in the same form as Eq. 8), the augmented view $\widetilde{\mathbf{U}}_{i,s}^f$ enforces semantically invariant feature learning via a Barlow Twins objective (Zbontar et al., 2021):

$$\mathcal{L}_{\text{sem}} = \text{BT}(\mathbf{U}_{i,s}^f, \widetilde{\mathbf{U}}_{i,s}^f). \tag{14}$$

The total loss for the semantic branch is:

$$\mathcal{L}_{E_{\text{sem}}} = \mathcal{L}_{\text{cm}}^s + \lambda(t)\mathcal{L}_{\text{sem}}. \tag{15}$$

Here, $\lambda(t)$ is an annealed weight that decreases from a large value to zero during training, providing strong semantic guidance at early stages while allowing intrinsic cross-modal learning in later stages. $\mathcal{L}_{\text{sem}}$ only regularizes E_{sem} and does not update the reconstruction decoders. Consequently, the auxiliary view constrains representation learning while the reconstruction targets remain the observed modalities. As $\lambda(t)$ decreases, this additional constraint weakens and the cross-modal objective continues to govern the later optimization. The prior therefore influences the training trajectory without replacing the subject's FC or SC reconstruction target.

3.4 Downstream Classification Task

The reconstruction stage learns representations before downstream classifier training. Once it is complete, all reconstruction decoders are discarded and both encoders (E_{base} and E_{sem}) are frozen. Classification then uses their representations of the original FC and SC inputs. The semantic branch's contribution is carried by its learned encoder, so synthesizing the auxiliary functional view is unnecessary at prediction time.

For target subject i, representations are derived strictly from raw modal inputs $(\mathbf{X}_i^f, \mathbf{X}_i^s)$:

$$\begin{aligned} \mathbf{U}_{i,b}^f &= E_{\text{base}}(\mathbf{X}_i^f, \mathbf{H}_{i,b}^f), \qquad \mathbf{U}_{i,b}^s = E_{\text{base}}(\mathbf{X}_i^s, \mathbf{H}_i^s), \\ \mathbf{U}_{i,s}^f &= E_{\text{sem}}(\mathbf{X}_i^f, \mathbf{H}_{i,s}^f). \end{aligned} \tag{16}$$

The base and semantically regularized functional embeddings are fused via equal-weight averaging and concatenated with the structural representation:

$$\mathbf{U}_i^f = \frac{1}{2}\left(\mathbf{U}_{i,b}^f + \mathbf{U}_{i,s}^f\right), \qquad \mathbf{z}_i = \text{vec}\left(\mathbf{U}_i^f \,\|\, \mathbf{U}_{i,b}^s\right), \tag{17}$$

Equal weighting adds no learned fusion parameters and keeps the fused functional representation at the dimensionality of a single branch. Each branch can contribute a distinct representation of the same subject, while the structural representation is retained separately. This separates representation fusion from the earlier construction of the auxiliary semantic view.

Finally, the flattened multimodal representation is fed into an MLP classifier:

$$\widehat{\mathbf{y}}_i = \mathrm{Softmax}\left(\mathrm{MLP}(\mathbf{z}_i)\right), \qquad \mathcal{L}_{\mathrm{down}} = \mathrm{CE}(\widehat{\mathbf{y}}_i, y_i). \tag{18}$$

By decoupling semantic view synthesis from downstream inference, our model preserves cross-modal consistency without requiring auxiliary semantic views during inference.

4 Experiments

Our experiments address three questions: whether SCMT improves paired-cohort classification over the base reconstruction model, how the two functional representations should be combined, and how source-pretraining scale affects the target result. Main comparisons and ablations are reported below. Appendix C gives the full computational comparison, and Appendix D presents the ROI visualizations and prior audit.

4.1 Datasets and Preprocessing

We evaluate the proposed method on two paired multimodal neuroimaging cohorts, ABIDE I (Di Martino et al., 2014) and ADNI (Jack et al., 2008). Only subjects with both resting-state fMRI and DTI are included in the downstream experiments.

ABIDE. The paired ABIDE cohort contains 86 subjects (48 ASD, 38 NCs) collected from the NYU (50 subjects) and TCD (36 subjects) sites. Each brain is parcellated into 116 ROIs using the AAL atlas (Tzourio-Mazoyer et al., 2002). Functional connectivity matrices are constructed via Pearson correlation on regional fMRI time series (150–200 time points), while DTI connectivity matrices are binarized into structural adjacency matrices.

ADNI. The paired ADNI cohort is used for mild cognitive impairment (MCI) versus NC classification, comprising 247 subjects (118 MCI, 129 NCs). Following a similar preprocessing pipeline, regional FC and structural adjacency matrices are constructed with the AAL atlas.

Source pretraining and atlas correspondence. Source functional pretraining excludes all subjects from the NYU, TCD, and TRINITY sites, leaving 655 ABIDE subjects (310 ASD and 345 NC). These exclusions make source learning site-disjoint from the paired ABIDE evaluation cohort. Source diagnosis labels supervise the classification objective, while source healthy controls supply the mean FC reference $\mathbf{R}$. The ABIDE-derived affinity and reference are also used for ADNI, without ADNI-specific semantic pretraining. ABIDE uses AAL116, comprising 90 cerebral and 26 cerebellar ROIs, whereas ADNI uses AAL90. The 90 cerebral ROIs follow the same anatomical ordering, so the transferred matrices are restricted to those rows and columns. This establishes correspondence between region indices; it is not a cohort-harmonization procedure.

Target adaptation and evaluation. Target adaptation is label-blind and transductive: FC–SC reconstruction uses the complete target cohort before five-fold classification, including the connectivity of subjects subsequently held out for classification. Diagnosis labels are not used by this reconstruction stage. Encoders are then frozen, and a two-layer MLP is trained on the labeled training subjects of each fold. This evaluates classification within an already observed target cohort, not generalization to subjects unseen during encoder adaptation.

ROI text representations. ChatGPT-5 generates standardized AAL ROI descriptions offline. A frozen Llama-Encoder-1.0B from LLM2Vec (BehnamGhader et al., 2024) encodes them, and source calibration produces the cosine-affinity prior.

Table 1: Performance on ABIDE I and ADNI. Best/second-best means are bold/underlined. F1 denotes Macro-F1; row groups use fMRI, DTI, and multimodal inputs.

Method	ABIDE I (ASD vs. NC)			ADNI (MCI vs. NC)		
	ACC (%)	AUC	F1 (%)	ACC (%)	AUC	F1 (%)
BrainGNN	$65.26_{\pm 8.35}$	$0.5810_{\pm 0.0891}$	$61.17_{\pm 10.10}$	$52.77_{\pm 6.54}$	$0.5526_{\pm 0.0605}$	$58.73_{\pm 9.99}$
ALTER	$65.14_{\pm 4.36}$	$0.6234_{\pm 0.0453}$	$61.23_{\pm 5.64}$	$63.40_{\pm 6.31}$	$0.6076_{\pm 0.0738}$	$53.17_{\pm 7.66}$
BrainNPT	$64.25_{\pm 7.05}$	$0.6055_{\pm 0.0614}$	$59.54_{\pm 9.51}$	$64.29_{\pm 6.23}$	$0.6064_{\pm 0.0902}$	$55.24_{\pm 7.08}$
BrainMass	$64.73_{\pm 3.03}$	$0.6272_{\pm 0.0576}$	$52.28_{\pm 10.20}$	$64.32_{\pm 8.42}$	$0.6078_{\pm 0.0794}$	$57.64_{\pm 8.31}$
BrainNetCNN	$64.02_{\pm 5.31}$	$0.6118_{\pm 0.0420}$	$58.84_{\pm 8.50}$	$60.69_{\pm 12.30}$	$0.6125_{\pm 0.0879}$	$48.11_{\pm 16.60}$
GAT	$60.00_{\pm 5.41}$	$0.6040_{\pm 0.0573}$	$56.43_{\pm 10.00}$	$52.20_{\pm 11.40}$	$0.5395_{\pm 0.0560}$	$39.27_{\pm 11.90}$
BrainGNN	$60.11_{\pm 7.32}$	$0.5622_{\pm 0.0633}$	$56.78_{\pm 9.50}$	$53.08_{\pm 6.89}$	$0.5476_{\pm 0.0430}$	$64.27_{\pm 10.40}$
MME-GCN	$61.55_{\pm 8.32}$	$0.6024_{\pm 0.0834}$	$57.31_{\pm 9.75}$	$53.41_{\pm 5.82}$	$0.5703_{\pm 0.0592}$	$54.41_{\pm 9.28}$
Cross-GNN	$66.20_{\pm 5.50}$	$0.6298_{\pm 0.0487}$	$61.55_{\pm 9.76}$	$50.66_{\pm 9.09}$	$0.5313_{\pm 0.1502}$	$48.44_{\pm 10.60}$
MMBNF	$\underline{67.45_{\pm 6.52}}$	$0.6227_{\pm 0.0849}$	$62.04_{\pm 7.63}$	$62.21_{\pm 7.24}$	$0.6118_{\pm 0.0818}$	$60.47_{\pm 10.30}$
SCMT-Lite	$66.67_{\pm 6.75}$	$\underline{0.6496_{\pm 0.0639}}$	$\underline{63.97_{\pm 7.27}}$	$\underline{65.59_{\pm 3.73}}$	$\underline{0.6544_{\pm 0.0324}}$	$\underline{65.30_{\pm 3.82}}$
SCMT	$\mathbf{70.92}_{\pm 5.89}$	$\mathbf{0.6943}_{\pm 0.0647}$	$\mathbf{69.27}_{\pm 6.74}$	$\mathbf{66.19}_{\pm 3.06}$	$\mathbf{0.6556}_{\pm 0.0336}$	$\mathbf{65.65}_{\pm 3.27}$

4.2 Comparison Methods

We compared several unimodal and multimodal methods, including four unimodal supervised methods (BrainNetCNN (Kawahara et al., 2017), GAT (Veličković et al., 2018), BrainGNN (Li et al., 2021), ALTER (Yu et al., 2024)), two unimodal SSL methods (BrainNPT (Hu et al., 2024), BrainMass (Yang et al., 2024b)), and three multi-modal methods (MME-GCN (Liu et al., 2022), Cross-GNN (Yang et al., 2024a), MMBNF (Li et al., 2025)). These methods cover supervised connectome models, self-supervised functional encoders, and multimodal fusion architectures. All methods are evaluated on the same paired target cohorts using accuracy (ACC), area under the ROC curve (AUC), and Macro-F1. Table 1 groups the methods by their input modalities so that the unimodal and multimodal results can be read separately.

SCMT-Lite denotes our base multimodal reconstruction model without the source-guided branch. SCMT adds a separately trained target encoder regularized by the transferred affinity and reference. The two variants share target cohorts, outer folds, encoder seeds, downstream classifier, and model-selection protocol. Their comparison evaluates the complete extension, which changes both transferred information and encoder capacity; it does not isolate language semantics alone.

4.3 Implementation Details

All models are implemented in PyTorch and trained on two NVIDIA RTX 3090 Ti GPUs. Both the reconstruction backbone and semantic encoder are pretrained for 100 epochs using Adam with cosine learning-rate decay. During target adaptation, the backbone is optimized via cross-modal reconstruction, while the auxiliary semantic encoder is regularized by semantic consistency with a linearly annealed semantic loss. For downstream evaluation, encoder parameters remain frozen, functional and structural representations are fused, and a lightweight two-layer MLP is trained for classification. Unless otherwise specified, all hyperparameters are shared across compared methods. We set the hyperparameters β and τ to 0.1 and 0.5, respectively.

4.4 Experimental Results and Analysis

Table 1 reports the highest mean ACC, AUC, and Macro-F1 for SCMT among the evaluated methods on both datasets. The modality groups also show that access to both FC and SC is not sufficient by itself to ensure a strong result: performance varies across the multimodal architectures. SCMT-Lite already provides a competitive reconstruction-based baseline, making its comparison with the full model useful for assessing the added branch.

Table 2: Component and fusion ablation: mean ± standard deviation over four encoder initializations and five outer folds. Best means are bold; F1 denotes Macro-F1.

Method	ABIDE I (ASD vs. NC)			ADNI (MCI vs. NC)		
	ACC (%)	AUC	F1 (%)	ACC (%)	AUC	F1 (%)
SCMT-Lite	$66.67_{\pm 6.75}$	$0.6496_{\pm 0.0639}$	$63.97_{\pm 7.27}$	$65.59_{\pm 3.73}$	$0.6544_{\pm 0.0324}$	$65.30_{\pm 3.82}$
Semantic only	$67.11_{\pm 5.50}$	$0.6527_{\pm 0.0583}$	$64.40_{\pm 7.46}$	$65.08_{\pm 2.28}$	$0.6471_{\pm 0.0343}$	$63.31_{\pm 3.31}$
Concatenation	$69.44_{\pm 7.71}$	$0.6765_{\pm 0.0888}$	$66.67_{\pm 10.89}$	$64.57_{\pm 3.90}$	$0.6396_{\pm 0.0432}$	$63.86_{\pm 4.22}$
Blend ($\alpha = 0.50$)	$\mathbf{70.92}_{\pm 5.89}$	$\mathbf{0.6943}_{\pm 0.0647}$	$\mathbf{69.27}_{\pm 6.74}$	$\mathbf{66.19}_{\pm 3.06}$	$\mathbf{0.6556}_{\pm 0.0336}$	$\mathbf{65.65}_{\pm 3.27}$

On ABIDE, SCMT increases ACC from 66.67% to 70.92%, AUC from 0.6496 to 0.6943, and Macro-F1 from 63.97% to 69.27% relative to SCMT-Lite. These correspond to 4.25 ACC points, 0.0447 AUC, and 5.30 Macro-F1 points. Improvements in both ACC and Macro-F1 accompany the AUC increase, so the observed change is not limited to a single reported metric. The result supports combining the base representation with the prior-regularized functional branch in this paired ABIDE setting.

On ADNI, ACC changes from 65.59% to 66.19%, AUC from 0.6544 to 0.6556, and Macro-F1 from 65.30% to 65.65%. The corresponding gains are 0.60 ACC points, 0.0012 AUC, and 0.35 Macro-F1 points. Here the source priors are learned from ABIDE, while the target cohort and classification task differ. The small positive changes show that the complete pipeline can be applied across these cohorts, but provide substantially weaker evidence of benefit than the ABIDE result. They do not establish disease-independent transfer or statistical significance.

Taken together, the two cohorts motivate retaining subject-specific reconstruction alongside transferred information. The complete model improves the reported means, but the magnitude depends on the target setting. The ablations below examine whether the prior-regularized representation should replace the base functional representation or be combined with it.

4.5 Ablation Study

We analyze the prior-regularized functional encoder and its fusion strategy. SCMT-Lite uses the base reconstruction representation. The semantic-branch variant replaces only the functional representation with the output of the semantic branch. Concatenation and Balanced Blend retain both functional representations; the structural representation and downstream classifier design are otherwise retained. These variants test how to use the additional functional pathway within the complete system.

Component contribution. Table 2 shows that the semantic branch alone gives only small gains on ABIDE and lower results on ADNI than SCMT-Lite. For example, the ADNI AUC decreases from 0.6544 to 0.6471 when the base functional representation is replaced. This makes the distinction between replacement and combination important: the source-guided pathway is not uniformly stronger than the base pathway on its own.

Balanced Blend achieves the highest reported mean in every metric on both cohorts. On ABIDE, concatenation also improves over SCMT-Lite, but its AUC of 0.6765 remains below the balanced average's 0.6943. On ADNI, concatenation reaches 0.6396 AUC, below both SCMT-Lite and Balanced Blend. Thus retaining both branches is useful only together with an appropriate fusion rule; increasing the feature dimension by concatenation does not consistently improve the result. These comparisons motivate the equal-weight average used in the complete model.

Fusion strategy. Figure 2 compares functional mixing weights $\alpha \in \{0, 0.25, 0.50, 0.75, 1.0\}$ and concatenation. The endpoints select the base or semantic functional representation, while intermediate weights retain both. Performance varies non-monotonically with α: giving the semantic branch more weight does not consistently improve classification. Balanced Blend therefore treats the prior-regularized features as a complementary contribution to the reconstruction features.

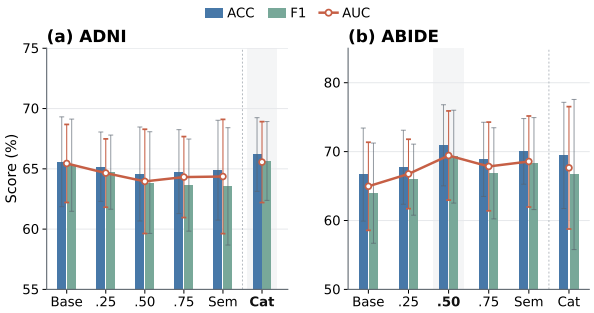


Figure 2: Fusion on ADNI/ABIDE I: ACC and Macro-F1 bars, AUC lines. Base/Sem select one branch; Cat concatenates both.

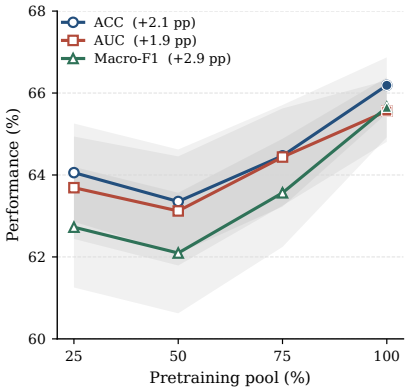


Figure 3: Source-data scaling on ADNI. Gray ribbons: ± 1 standard error.

Averaging preserves the functional feature dimension and introduces no learned mixing parameters, whereas concatenation retains both full vectors. This provides a practical reason to prefer the balanced average in addition to its reported results. Because adding a branch also increases encoder capacity, these ablations support the system design without separating the effects of the semantic prior from those of using two encoders.

4.6 Further Analysis

The initialization analysis repeats encoder pretraining with four seeds while holding the five-fold partition and classifier seed fixed. It measures initialization sensitivity under that partition, not robustness across new data splits. Figure 3 examines source-data scale with the ADNI target setting fixed: the full source pool gives the highest reported metrics, improving ACC by 2.1 points, AUC by 0.019, and Macro-F1 by 2.9 points relative to the 25% setting. The intermediate decline from 25% to 50% indicates that the trend is not uniformly monotonic. Appendix B provides the full protocol and interpretation of both analyses.

Additional analyses. The added branch increases parameters from 23.91M to 36.01M and downstream inference time from 4.69 to 6.98 ms per subject; the full cost breakdown is in Appendix C. Appendix D presents the discriminative ROI maps and the prior audit. The latter reports association with source mean FC and bilateral anatomical correspondence, providing descriptive context for the transferred matrix.

Limitations. The comparisons evaluate the complete framework and do not independently isolate language semantics, source diagnosis supervision, the healthy reference, and additional encoder capacity. Capacity-matched dual-base and non-semantic-prior controls are needed for that attribution. Common target cohorts alone do not establish equal source-data budgets or unlabeled-target access for every baseline. ADNI gains are small, the source healthy reference may retain cohort effects, and the transductive protocol does not evaluate entirely unseen target subjects. These retrospective experiments do not establish clinical utility.

5 Conclusion

SCMT transfers atlas-level ROI relations and a healthy-connectivity reference from functional source learning to paired FC–SC reconstruction. Its dual-branch design combines a base pathway with annealed consistency through an auxiliary functional view; frozen functional and structural features are fused for classification. Under the reported transductive protocol, gains are larger on ABIDE and small on ADNI. Fusion and source-scale analyses characterize transfer, while supplementary efficiency and ROI analyses describe its computational and anatomical context.

AI USE STATEMENT

AI tools were used to assist with language polishing.

REFERENCES

Parishad BehnamGhader, Vaibhav Adlakha, Marius Mosbach, Dzmitry Bahdanau, Nicolas Chapados, and Siva Reddy. LLM2Vec: Large language models are secretly powerful text encoders. In *First Conference on Language Modeling*, 2024. URL https://openreview.net/forum?id=IW1PR7vEBf.

A. Di Martino, C.-G. Yan, Q. Li, E. Denio, F. X. Castellanos, K. Alaerts, J. S. Anderson, M. Assaf, S. Y. Bookheimer, M. Dapretto, B. Deen, S. Delmonte, I. Dinstein, B. Ertl-Wagner, D. A. Fair, L. Gallagher, D. P. Kennedy, C. L. Keown, C. Keysers, J. E. Lainhart, C. Lord, B. Luna, V. Menon, N. J. Minshew, C. S. Monk, S. Mueller, R.-A. Müller, M. B. Nebel, J. T. Nigg, K. O'Hearn, K. A. Pelphrey, S. J. Peltier, J. D. Rudie, S. Sunaert, M. Thioux, J. M. Tyszka, L. Q. Uddin, J. S. Verhoeven, N. Wenderoth, J. L. Wiggins, S. H. Mostofsky, and M. P. Milham. The autism brain imaging data exchange: Towards a large-scale evaluation of the intrinsic brain architecture in autism. *Molecular Psychiatry*, 19(6):659–667, 2014. doi: 10.1038/mp.2013.78.

Jingxi Feng, Shaoyi Du, Heming Xu, Rundong Xue, Xiangmin Han, Dong Zhang, Jue Jiang, Yue Gao, and Juan Wang. Topology-aware multi-view hypergraph computation-based cross-modal brain network fusion for brain disease diagnosis. *Information Fusion*, 127:103751, 2026. doi: 10.1016/j.inffus.2025.103751.

Yifan Feng, Haoxuan You, Zizhao Zhang, Rongrong Ji, and Yue Gao. Hypergraph neural networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pp. 3558–3565, 2019. doi: 10.1609/aaai.v33i01.33013558.

Jinlong Hu, Yangmin Huang, Nan Wang, and Shoubin Dong. BrainNPT: Pre-training transformer networks for brain network classification. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 32:2727–2736, 2024. doi: 10.1109/tnsre.2024.3434343.

Clifford R. Jack, Jr., Matt A. Bernstein, Nick C. Fox, Paul Thompson, Gene Alexander, Danielle Harvey, Bret Borowski, Paula J. Britson, Jennifer L. Whitwell, Chadwick Ward, Anders M. Dale, Joel P. Felmlee, Jeffrey L. Gunter, Derek L.G. Hill, Ron Killiany, Norbert Schuff, Sabrina Fox-Bosetti, Chen Lin, Colin Studholme, Charles S. DeCarli, Gunnar Krueger, Heidi A. Ward, Gregory J. Metzger, Katherine T. Scott, Richard Mallozzi, Daniel Blezek, Joshua Levy, Josef P. Debbins, Adam S. Fleisher, Marilyn Albert, Robert Green, George Bartzokis, Gary Glover, John Mugler, and Michael W. Weiner. The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. *Journal of Magnetic Resonance Imaging*, 27(4):685–691, 2008. doi: 10.1002/jmri.21049.

Xuan Kan, Wei Dai, Hejie Cui, Zilong Zhang, Ying Guo, and Carl Yang. Brain network transformer. In *Advances in Neural Information Processing Systems*, volume 35, pp. 25586–25599, 2022. doi: 10.52202/068431-1855.

Jeremy Kawahara, Colin J. Brown, Steven P. Miller, Brian G. Booth, Vann Chau, Ruth E. Grunau, Jill G. Zwicker, and Ghassan Hamarneh. BrainNetCNN: Convolutional neural networks for brain networks; towards predicting neurodevelopment. *NeuroImage*, 146:1038–1049, 2017. doi: 10.1016/j.neuroimage.2016.09.046.

Shengrong Li, Qi Zhu, Liang Sun, Kai Ma, Yixin Ji, Shile Qi, and Daoqiang Zhang. Multi-modal brain network fusion for intelligent diagnostic devices. *IEEE Transactions on Consumer Electronics*, 71(2):3654–3666, 2025. doi: 10.1109/tce.2025.3563691.

Xiaoxiao Li, Yuan Zhou, Nicha Dvornek, Muhan Zhang, Siyuan Gao, Juntang Zhuang, Dustin Scheinost, Lawrence H. Staib, Pamela Ventola, and James S. Duncan. BrainGNN: Interpretable brain graph neural network for fMRI analysis. *Medical Image Analysis*, 74:102233, 2021. doi: 10.1016/j.media.2021.102233.

Liangliang Liu, Yu-Ping Wang, Yi Wang, Pei Zhang, and Shufeng Xiong. An enhanced multi-modal brain graph network for classifying neuropsychiatric disorders. *Medical Image Analysis*, 81:102550, 2022. doi: 10.1016/j.media.2022.102550.

Hae-Jeong Park and Karl Friston. Structural and functional brain networks: from connections to cognition. *Science*, 342(6158):1238411, 2013. doi: 10.1126/science.1238411.

Olaf Sporns. Structure and function of complex brain networks. *Dialogues in Clinical Neuroscience*, 15(3):247–262, 2013. doi: 10.31887/dcns.2013.15.3/osporns.

Nathalie Tzourio-Mazoyer, Brigitte Landeau, Dimitri Papathanassiou, Fabrice Crivello, Octave Etard, Nicolas Delcroix, Bernard Mazoyer, and Marc Joliot. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage*, 15(1):273–289, 2002. doi: 10.1006/nimg.2001.0978.

Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. Graph attention networks. In *International Conference on Learning Representations*, 2018. URL https://openreview.net/forum?id=rJXMpikCZ.

Guangqi Wen, Peng Cao, Lingwen Liu, Jinzhu Yang, Xizhe Zhang, Fei Wang, and Osmar R. Zaiane. Graph self-supervised learning with application to brain networks analysis. *IEEE Journal of Biomedical and Health Informatics*, 27(8):4154–4165, 2023. doi: 10.1109/jbhi.2023.3274531.

Jiaxing Xu, Kai He, Yue Tang, Wei Li, Mengcheng Lan, Xia Dong, Yiping Ke, and Mengling Feng. BrainPrompt: Multi-level brain prompt enhancement for neurological condition identification. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2025*, volume 15971 of *Lecture Notes in Computer Science*, pp. 172–182. Springer, 2025. doi: 10.1007/978-3-032-05162-2_17.

Yidan Xu, Xiangmin Han, Rundong Xue, and Huihui Ye. SABER: A semantic-aligned brain network analysis framework via multi-scale hypergraphs. In *2026 IEEE International Conference on Multimedia and Expo (ICME)*, 2026. URL https://arxiv.org/abs/2607.01901.

Yanwu Yang, Chenfei Ye, Xutao Guo, Tao Wu, Yang Xiang, and Ting Ma. Mapping multi-modal brain connectome for brain disorder diagnosis via cross-modal mutual learning. *IEEE Transactions on Medical Imaging*, 43(1):108–121, 2024a. doi: 10.1109/tmi.2023.3294967.

Yanwu Yang, Chenfei Ye, Guinan Su, Ziyao Zhang, Zhikai Chang, Hairui Chen, Piu Chan, Yue Yu, and Ting Ma. BrainMass: Advancing brain network analysis for diagnosis with large-scale self-supervised learning. *IEEE Transactions on Medical Imaging*, 43(11):4004–4016, 2024b. doi: 10.1109/tmi.2024.3414476.

Yaokai Yang, Nan Zhao, Zeyu Zhang, Jing Zhang, and Lu Zhang. Brain-CLIP: An individual-preserving contrastive framework for bidirectional cross-modal brain connectome translation. In *2026 IEEE 23rd International Symposium on Biomedical Imaging (ISBI)*, pp. 1–5, 2026. doi: 10.1109/isbi61048.2026.11515388.

Shuo Yu, Shan Jin, Ming Li, Tabinda Sarwar, and Feng Xia. Long-range brain graph transformer. In *Advances in Neural Information Processing Systems*, volume 37, pp. 24472–24495, 2024. doi: 10.52202/079017-0771.

Jure Zbontar, Li Jing, Ishan Misra, Yann LeCun, and Stéphane Deny. Barlow Twins: Self-supervised learning via redundancy reduction. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 12310–12320. PMLR, 2021. URL https://proceedings.mlr.press/v139/zbontar21a.html.

A Detailed Related Work

A.1 Multimodal Brain Network Learning

Brain network analysis typically represents brain regions as nodes and connectivity patterns as graphs or connectivity matrices. BrainNetCNN introduced specialized edge-to-edge and edge-to-node operations for connectome analysis (Kawahara et al., 2017), while BrainGNN learned ROI-aware graph representations with interpretability considerations (Li et al., 2021). Recent multimodal approaches integrate functional and structural connectomes through graph fusion, cross-modal interaction, or mutual learning (Liu et al., 2022; Yang et al., 2024a; Li et al., 2025; Feng et al., 2026; Yang et al., 2026). These methods exploit complementary modalities, but they mainly rely on paired multimodal cohorts and learn cross-modal correspondence from limited subject-level observations.

A.2 Pretraining and Transfer for Neuroimaging

Large-scale pretraining has recently been explored for brain representation learning through self-supervised graph augmentation, masked connectivity modeling, and Transformer-based objectives (Wen et al., 2023; Hu et al., 2024; Yang et al., 2024b). These approaches improve the quality and robustness of unimodal representations, but transferring knowledge from large-scale functional datasets to multimodal FC-SC learning remains challenging due to the mismatch between source and target settings. Meanwhile, graph Transformers and brain network Transformers further enhance representation capacity (Yu et al., 2024; Kan et al., 2022), but their effectiveness can be limited by the small scale of paired neuroimaging datasets.

A.3 Knowledge-Guided Graph Learning

Hypergraph neural networks extend conventional graph learning by capturing higher-order relationships among multiple nodes (Feng et al., 2019). Recent studies have further explored external knowledge, including language-derived semantic information, to enrich brain network representations (Xu et al., 2025; 2026). SCMT uses language-derived ROI representations as an initial source of regional information, calibrates them through source functional learning, and then transfers their pairwise affinities to a separate target model. The affinity consequently combines language, source connectivity, and source-task supervision. Its role is to regularize target representations through a subject-dependent auxiliary view, with an annealed weight balancing that guidance against paired FC–SC reconstruction.

B Additional Experimental Analyses

B.1 Sensitivity to Encoder Initialization

We repeat encoder pretraining with four initialization seeds (0, 42, 2022, and 2020), while keeping the five-fold partition and downstream classifier seed (2020) fixed. The ablation table reports means and standard deviations across the four initializations and five outer folds. This setup examines the dependence on encoder initialization within a fixed partition, rather than adding new train–test splits. The reported variability should therefore be interpreted as initialization/fold variation under this protocol, not as a repeated-split validation or significance test.

B.2 Effect of Pretraining Scale

Figure 3 varies the source-pretraining pool while keeping the ADNI target cohort and downstream training configuration fixed. There is an intermediate decline when the source fraction increases from 25% to 50%, followed by higher results at larger fractions. The full source pool gives the highest reported metrics. Relative to the 25% setting, it improves ACC by 2.1 points, AUC by 0.019, and Macro-F1 by 2.9 points. The uncertainty bands also narrow at full scale, particularly for ACC and Macro-F1.

The trend is consistent with a benefit from additional source information, while its non-monotonicity indicates remaining variability in this limited target cohort. Since changing the source pool jointly

changes the connectivity and supervision available for calibration, this analysis assesses source-data scale for the complete pretraining stage. It does not separately quantify the contribution of language-derived descriptions.

C Computational Efficiency

Table 3: Efficiency comparison between the baseline and the proposed model. Training time refers to the reported downstream stage; it excludes the full pretraining pipeline.

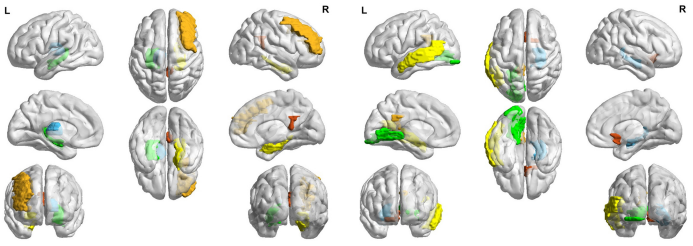
Method	**Params** (M)	**Training** (s↓)	**Inference** (ms/sample↓)	**Memory** (MiB↓)
SCMT-Lite(Ours)	23.91	**40.59**	**4.69**	**857.6**
SCMT(Ours)	36.01	45.39	6.98	1280.4
Overhead	+50.6%	+11.8%	+48.8%	+49.3%

Table 3 reports the computational cost of SCMT-Lite and SCMT. The complete model adds a separate semantic encoder, increasing the parameter count from 23.91M to 36.01M (50.6%). The reported downstream training time rises from 40.59 to 45.39 s (11.8%), inference latency from 4.69 to 6.98 ms per subject (48.8%), and GPU memory from 857.6 to 1280.4 MiB (49.3%). These quantities describe different costs and should be considered separately: the smaller training-time increase does not remove the larger parameter and inference overhead.

The auxiliary FC view and reconstruction decoders are used during representation learning. At downstream prediction, both functional encoders remain active, together with the base structural pathway and classifier, even though the auxiliary view is no longer synthesized. Source semantic-pretraining costs are additional to the reported downstream timing. The measurements therefore characterize the downstream accuracy–cost tradeoff rather than the end-to-end cost of the complete source-and-target pipeline.

D Visualization and Semantic Prior Analysis

D.1 Discriminative ROI Visualization



(a) ABIDE (ASD/NC) (b) ADNI (MCI/NC)

Figure 4: Visualization of discriminative ROIs.

Figure 4 shows the highlighted ROIs in the paired ABIDE and ADNI cohorts. On ABIDE, the visualization emphasizes the hippocampus and middle frontal regions; on ADNI, it highlights regions around the entorhinal and posterior cingulate areas. These maps complement the classification tables by showing the anatomical regions emphasized in the learned target representations.

The maps are descriptive: they do not independently establish disease-specific biomarkers, a causal disease mechanism, or the contribution of language-derived information. In particular, the visualizations of target representations should be distinguished from the source-level affinity audit below. They concern different objects and do not provide interchangeable evidence about the model.

D.2 ROI-Level Prior Audit

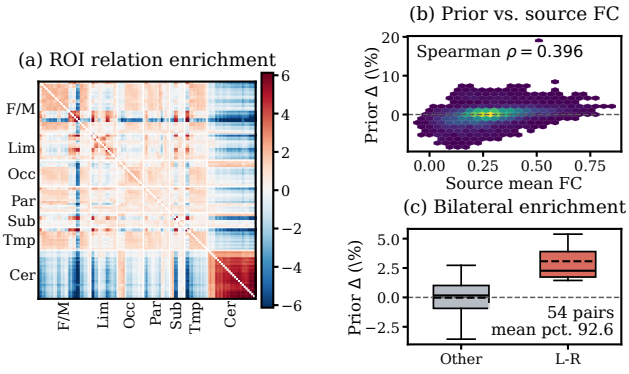


Figure 5: ROI-level prior audit: (a) off-diagonal enrichment grouped anatomically; (b) enrichment versus source mean FC; (c) left–right homologous pairs versus other pairs. This describes population structure, not disease-specific saliency.

Figure 5 examines the source-calibrated ROI affinity from three perspectives. The raw weights occupy a narrow dynamic range, so panel (a) displays percentage deviations from the mean off-diagonal weight, grouped by anatomical region. This visualization exposes relative structure that would be difficult to see on the raw scale. It does not imply that the absolute differences between affinities are large.

Panel (b) compares pairwise prior weights with source-cohort mean FC and reports a Spearman correlation of 0.396. Panel (c) contrasts the 54 left–right homologous ROI pairs with other off-diagonal pairs: homologous pairs receive 3.10% higher weights on average and rank at the 92.6th percentile. Correlation with mean FC describes association with source functional organization, whereas bilateral enrichment describes correspondence with anatomical pairing. Together they characterize the transferred population structure rather than disease-specific saliency.

These associations are compatible with anatomically organized relations, but source connectivity also participates in calibrating the prior. The audit therefore does not isolate a language-specific contribution or establish that the prior improves classification for an individual subject. The current experiments do not include shuffled or non-language prior controls that would resolve this attribution.

D.3 Interpreting the Narrow Affinity Range

The standardization used to construct $\mathbf{C}$ rescales relative differences between ROI affinities. This can make a narrow raw range influential in the auxiliary-view construction, but can also amplify small perturbations. Consequently, near-uniform raw similarities do not by themselves imply either a useless modulation or a biologically validated one. Both the transformation and the downstream use of the prior matter.

Cosine affinities are not probability distributions, and a row-entropy calculation requires an explicit probability normalization. We therefore use the correlation and bilateral-enrichment statistics above to describe the prior, without treating near-uniform affinity or an unspecified row-entropy calculation as evidence of semantic validity.