SABER: SUBJECT-ADAPTIVE AND BOUNDARY-AWARE DEEP MULTI-MODAL CLUSTERING NETWORK FOR INFANT GROUP-LEVEL CORTICAL PARCELLATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Constructing a group-level multi-modal cortical parcellation map for infants during the first two postnatal years is essential for studying early brain development and identifying disease risks from structural and functional perspectives. However, infant data are noisy and high-dimensional, making machine learning prone to fragmented and unstable parcellations, which motivates the use of deep learning. Yet, deep learning remains underexplored for this task because jointly modeling associative and complementary structural-functional information is challenging, while existing unsupervised learning objectives insufficiently address group-level parcellation practical demands, including improving cross-subject adaptability and anatomical consistency while preserving spatial continuity. To this end, we propose **SABER**, the first **S**ubject-**A**daptive and **B**oundary-aware D**E**ep multi-modal cluste**R**ing network for infant group-level cortical parcellation. SABER employs a two-stage fusion strategy combining multi-layer graph mutual learning with vertex-level modality weighting to capture structural-functional associations and complementarity. It further introduces two tailored objectives: the Dynamic Spatial-Weighted Generalized Homogeneity Loss improves cross-subject adaptability by optimizing parcellation homogeneity over a large set of individual-level features, while incorporating dynamic spatial weighting mechanism to prevent fragmentation; the Local Gradient-Weighted Active Boundary Loss incorporates local gradient-based boundary evidence to enhance sensitivity to potential biological boundaries. Experiments on the BCP dataset show that SABER outperforms existing SOTA methods and published parcellations across evaluation metrics and downstream tasks. Spatial null model analysis further indicates that the resulting parcellations capture biologically meaningful organization rather than random spatial patterns. Ablation studies demonstrate the plug-and-play ability of both losses and show that improved cross-subject adaptability benefits downstream task performance. `https://anonymous.4open.science/r/SABER2027`.

1 INTRODUCTION

During the first two postnatal years, the human cerebral cortex undergoes rapid structural and functional development, shaping cognitive, sensory, and motor functions while representing a critical window of opportunity and vulnerability (Huang et al., 2022; Wang et al., 2023a). Constructing a group-level multi-modal cortical parcellation map for infants during this stage is essential for unraveling the mechanisms of brain development and identifying early disease risks from both structural and functional perspectives (Yin et al., 2025; Sun et al., 2025; Liang et al., 2025). However, existing infant parcellations rely almost exclusively on single-modality data, either structural (Li et al., 2015; Wu et al., 2019) or functional (Shi et al., 2018; Wang et al., 2023a), limiting insight into structure–function coupling. Moreover, available multi-modal parcellation maps are derived from adult or neonatal cohorts (Glasser et al., 2016; Li et al., 2023) and cannot be directly transferred to infants due to substantial organizational differences. These limitations highlight the need for a representative group-level multi-modal cortical parcellation map tailored to infancy, which would provide a parcellation foundation for subsequent developmental neuroimaging studies.

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.

Achieving this goal, however, poses several challenges. Infant neuroimaging data are inherently noisy and high-dimensional, making conventional machine learning methods prone to producing fragmented and unstable parcellations (Luo et al., 2020; Gao et al., 2025), which motivates the use of deep learning methods capable of joint representation learning. Nevertheless, despite its success in individual cortical parcellation (Li et al., 2024; Zhao et al., 2024; Ma et al., 2024), deep learning has rarely been explored for group-level parcellation (Zhu et al., 2025a), particularly in multi-modal infant settings. We attribute this gap to two main challenges. 1) Jointly modeling structural and functional modalities is inherently complex: They exhibit intrinsic associations (Baum et al., 2020; Fotiadis et al., 2024), where cross-modal interaction helps capture structure-function coupling and mitigate modality-specific noise, while also providing regionally heterogeneous complementary information (Glasser et al., 2016; Parisot et al., 2017). Therefore, how to jointly capture their associative and complementary information becomes critical. 2) Existing unsupervised learning objectives are poorly aligned with the practical demands of infant group-level parcellation. First, substantial inter-subject variability in infants (Li et al., 2017; Zhao et al., 2024) makes optimizing homogeneity solely on group-level features through contrastive learning or DEC-style annealing insufficient for constructing a representative infant group-level map (Zhu et al., 2025a; Liu et al., 2026b), as it lacks explicit adaptation to individual-level features and may compromise downstream analyses that rely on the group-level maps to construct individual brain networks due to severe group–individual mismatches. Moreover, optimizing only global homogeneity may overlook subtle local transitions associated with certain biologically meaningful cortical boundaries (Schaefer et al., 2018; Yan et al., 2023), reducing anatomical consistency. Finally, how to incorporate a mechanism into the optimization objective to prevent fragmented parcels, so as to address the effects of strong noise and long-range connections in infant data, also deserves careful consideration.

To address these challenges, we propose **SABER**, the first **S**ubject-**A**daptive and **B**oundary-aware D**E**ep multi-modal cluste**R**ing network for infant group-level cortical parcellation. Specifically, SABER exploits associative and complementary structural-functional information through two-stage fusion, combining early multi-layer graph mutual learning with late vertex-level modality weighting. The multi-layer graph enables intra- and inter-modal information exchange, yielding noise-robust latent embeddings enriched with cross-modal information, while vertex-level weighting integrates complementary modality information at each vertex to meet region-specific parcellation demands. For optimization, SABER designs two tailored losses: 1) the Dynamic Spatial-Weighted Generalized Homogeneity Loss, which optimizes homogeneity of group-level map over a large set of individual-level features to improve its cross-subject adaptability, while introducing a dynamic spatial weighting mechanism to discourage the grouping of spatially distant vertices and progressively anneal the spatial weights without inducing fragmented parcels, thereby accommodating the anatomical prior of elongated cortical areas; 2) and the Local Gradient-Weighted Active Boundary Loss, which incorporates boundary evidence from local gradient methods to enhance sensitivity to subtle local transitions associated with potential biological boundaries. Our contributions are:

- We propose SABER, the first deep multi-modal clustering network for infant group-level cortical parcellation, providing a key parcellation foundation for early infant studies.
- We design two novel tailored losses to better match the practical demands of infant group-level cortical parcellation, improving the cross-subject adaptability of the group-level parcellation and enhancing its sensitivity to local potential biological boundaries.
- We develop a two-stage fusion strategy combining multi-layer graph mutual learning and vertex-level modality weighting to fully use associative and complementary information.

Evaluation on the Baby Connectome Project (BCP) dataset (Howell et al., 2019) demonstrates that SABER outperforms existing SOTA methods and published parcellations across four complementary evaluation metrics and four standard downstream tasks. Spatial null model analysis further supports that the resulting parcellations reflect biologically meaningful organization rather than random spatial patterns. Finally, ablation studies verify the plug-and-play capability of both losses and show that improved cross-subject adaptability translates into better downstream performance.

2 METHODOLOGY

In this section, we present SABER in detail. Section 2.1 describes how SABER fully exploits the associative and complementary information between structural and functional modalities, while

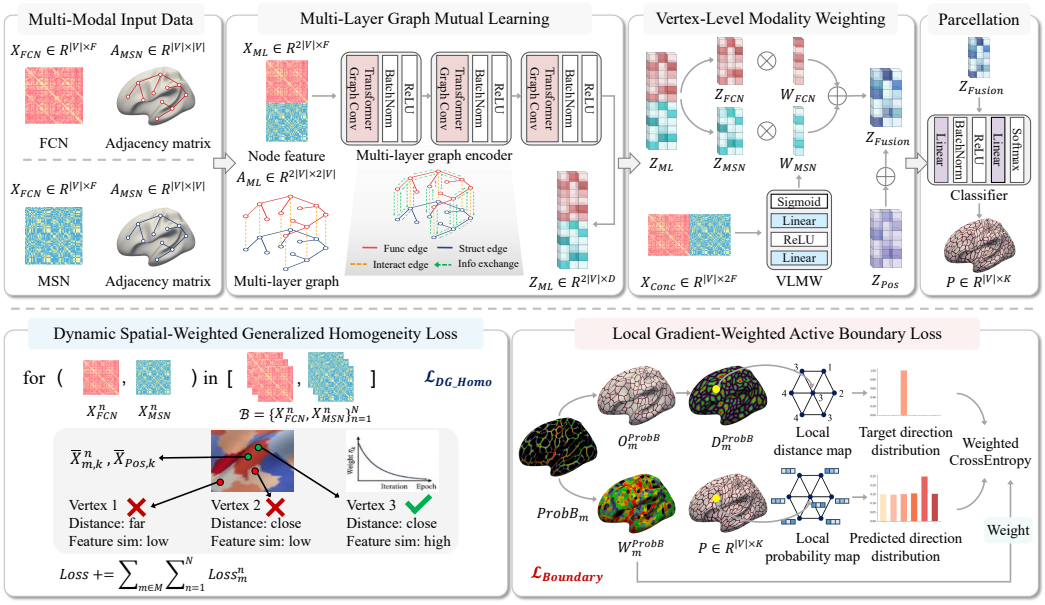


Figure 1: Overview of SABER. It employs a two-stage multi-modal integration: early mutual learning via a multi-layer graph, followed by late feature fusion via vertex-level adaptive weighting. Optimization is jointly driven by two tailored losses.

Section 2.2 shows how infant group-level cortical parcellation map is guided toward improved cross-subject adaptability and greater sensitivity to local potential biological boundaries. Here, structural and functional features are captured using morphological similarity networks (MSN) and functional connectivity networks (FCN), respectively. An overview of the framework is shown in Figure 1.

2.1 Network Architecture

Multi-Layer Graph Mutual Learning. Given infant group-level multi-modal feature matrices $\mathcal{X} = \{X_{\text{MSN}}, X_{\text{FCN}}\}$, where $X_{\text{MSN}}, X_{\text{FCN}} \in \mathbb{R}^{|V|\times F}$ denote structural and functional features averaged from individual-level data, respectively, with $|V|$ representing the number of cortical vertices and F the feature dimension. To mitigate modality-specific noise while propagating associative information across modalities, we construct a multi-layer graph $\mathcal{G} = \{V_{\text{ML}}, A_{\text{ML}}, X_{\text{ML}}\}$ in the early stage, enabling information exchange between neighboring vertices within each modality and identical vertices across modalities, thereby yielding noise-robust embeddings enriched with cross-modal information. Specifically, $V_{\text{ML}} = \binom{V_{\text{MSN}}}{V_{\text{FCN}}}$ and $X_{\text{ML}} = \binom{X_{\text{MSN}}}{X_{\text{FCN}}} \in \mathbb{R}^{2|V|\times F}$ aggregate vertices and features from both modalities. The adjacency matrix is defined as $A_{\text{ML}} = \begin{pmatrix} A_{\text{MSN}} & \text{diag}(I) \\ \text{diag}(I) & A_{\text{FCN}} \end{pmatrix} \in \mathbb{R}^{2|V|\times 2|V|}$, where A_{MSN} and $A_{\text{FCN}} \in \mathbb{R}^{|V|\times|V|}$ are constructed from Pearson correlations between 1-ring neighbouring vertices within each modality, and the diagonal identity matrix $\text{diag}(I)$ connects identical vertices across modalities. The resulting graph is processed by a multi-head Transformer GraphConv encoder (Shi et al., 2020) to obtain latent embeddings $\mathcal{Z} = \{Z_{\text{MSN}}, Z_{\text{FCN}}\}$, where the encoder consists of three stacked GraphConv-BatchNorm-ReLU blocks.

Vertex-Level Modality Weighting. Since modality contributions vary across cortical regions (Glasser et al., 2016; Parisot et al., 2017), we adopt a late-stage vertex-level weighted fusion strategy to adaptively integrate complementary information from both modalities. Specifically, the concatenated feature matrix $X_{\text{Conc}} = [X_{\text{MSN}}, X_{\text{FCN}}] \in \mathbb{R}^{|V|\times 2F}$ is fed into a Vertex-Level Modality Weighting (VLMW) module with Linear-ReLU-Linear-Sigmoid layers to obtain modality weights $W_{\text{MSN}}, W_{\text{FCN}} \in \mathbb{R}^{|V|\times 1}$. Meanwhile, an MLP maps spatial coordinates $X_{\text{Pos}} \in \mathbb{R}^{|V|\times 3}$ to positional embeddings Z_{Pos}. The fused embedding is then computed as:

$$Z_{\text{Fusion}} = W_{\text{MSN}} \odot Z_{\text{MSN}} + W_{\text{FCN}} \odot Z_{\text{FCN}} + Z_{\text{Pos}}. \tag{1}$$

Finally, Z_{Fusion} is passed through a Linear-BatchNorm-ReLU-Linear-Softmax classification head to obtain the final soft label assignment $P \in \mathbb{R}^{|V|\times K}$, where K is the number of parcels.

2.2 LOSS FUNCTION

Dynamic Spatial-Weighted Generalized Homogeneity Loss. Given the substantial inter-subject variability in infants (Li et al., 2017; Zhao et al., 2024), we aim to avoid group-level parcellations that adapt well to only a part of subjects while exhibiting poor homogeneity in others, thereby compromising downstream performance. Instead of optimizing homogeneity solely on group-level features, we explicitly optimize it over a large set of individual-level features to improve cross-subject adaptability. Meanwhile, to suppress fragmentation caused by noise and strong long-range connections, spatial distance is incorporated into inter-vertex similarity calculation, with each parcel-specific weight η_k progressively annealed when fragmentation is absent and increased otherwise, thereby preserving spatial continuity while accommodating the anatomical prior of elongated cortical areas. Here, fragmentation refers to the case where a parcel contains multiple connected components. The Dynamic Spatial-Weighted Generalized Homogeneity Loss $\mathcal{L}_{\text{DG}_{\text{Homo}}}$ is defined as:

$$\mathcal{L}_{\text{DG_Homo}} = \frac{1}{|M||\mathcal{B}|K} \sum_{m\in M} \sum_{n\in\mathcal{B}} \sum_{k=1}^{K} \left(\frac{\sum_{t\in\mathcal{N}(k)} Prr(\bar{X}^n_{m,t}, \bar{X}^n_{m,k})}{|\mathcal{N}(k)|} - \frac{\sum_{v\in V} P_{v,k}\, C^n_{m,v,k}}{\sum_{v\in V} P_{v,k}} \right) \quad (2)$$

$$C^n_{m,v,k} = Prr(X^n_{m,v}, \bar{X}^n_{m,k}) + \eta_k \cdot e^{-\frac{\| X_{\text{Pos},v} - \bar{X}_{\text{Pos},k} \|^2}{2\sigma^2}} \quad (3)$$

$$\bar{X}^n_{m,k} = \frac{\sum_{v\in V} P_{v,k}\, X^n_{m,v}}{\sum_{v\in V} P_{v,k}}, \quad \bar{X}_{\text{Pos},k} = \frac{\sum_{v\in V} P_{v,k}\, X_{\text{Pos},v}}{\sum_{v\in V} P_{v,k}} \quad (4)$$

where $M = \{\text{MSN}, \text{FCN}\}$ and $\mathcal{B}$ denotes a mini-batch of subjects. The first term in Eq. (2) minimizes similarity between each parcel and its neighbouring parcels, where $\mathcal{N}(k)$ denotes the neighbouring parcels of parcel k, and $\bar{X}^n_{m,t}$ and $\bar{X}^n_{m,k}$ are their mean feature profiles for subject n in modality m. $Prr(\cdot)$ denotes Pearson correlation. The second term maximizes intra-parcel homogeneity, where $X^n_{m,v}$ and $X_{\text{Pos},v}$ denote the feature profile and spatial coordinate of vertex v, $P_{v,k}$ is its assignment probability to parcel k. σ is a hyperparameter.

Local Gradient-Weighted Active Boundary Loss. Previous studies show that global similarity methods favor overall parcellation homogeneity but are relatively insensitive to subtle local transitions, whereas local gradient methods better capture transitions that may correspond to certain biologically meaningful boundaries (Schaefer et al., 2018; Gordon et al., 2016). To leverage this complementarity and improve boundary awareness, we introduce the boundary probability density map (ProbB) derived from local gradient methods (Wang et al., 2023a) as a supervised signal. Specifically, we extend the Active Boundary Loss (Wang et al., 2022) to the parcellation task, termed the Local Gradient-Weighted Active Boundary Loss $\mathcal{L}_{\text{Boundary}}$, which progressively aligns the predicted boundary PredB with high-probability boundaries in ProbB while relaxing constraints in low-probability regions. The loss involves three steps. First, for each vertex i on PredB, we select the neighboring vertex from its 1-ring neighborhood that is closest to ProbB_m of both modalities as the candidate target, and construct a target direction distribution $\mathcal{P}_i$ to indicate its next moving direction. This entails: 1) offline, for each modality, applying the watershed algorithm on ProbB_m to extract the boundary outline O^{ProbB}_m, computing the shortest geodesic distance to O^{ProbB}_m for each vertex to obtain the distance map D^{ProbB}_m, and recording the boundary probability at the closest point on O^{ProbB}_m to obtain the weight map W^{ProbB}_m; 2) identifying the neighbor-modality pair (j^*, m^*) with the minimum distance to O^{ProbB}_m among all 1-ring neighbors across both modalities:

$$(j^*, m^*) = \underset{j\in\mathcal{N}_1(i), m\in M}{\arg\min} D^{\text{ProbB}}_{m,j} \quad (5)$$

where $\mathcal{N}_1(i)$ denotes the 1-ring neighbors of vertex i; 3) encoding this direction as a one-hot vector $\mathcal{P}_i \in \mathbb{R}^{1\times|\mathcal{N}_1(i)|}$ (with the j-th entry as 1). Second, we construct a predicted direction distribution $\mathcal{Q}_i \in \mathbb{R}^{1\times|\mathcal{N}_1(i)|}$ via softmax over the KL divergences between the probability assignments of vertex i and its 1-ring neighbors:

$$\mathcal{Q}_i = \left\{ \frac{e^{KL(P_i||P_r)}}{\sum_{r\in\mathcal{N}_1(i)} e^{KL(P_i||P_r)}}, r \in \mathcal{N}_1(i) \right\} \quad (6)$$

Third, we compute the weighted cross-entropy between $\mathcal{Q}_i$ and $\mathcal{P}_i$ as the loss, encouraging alignment of PredB with high-probability boundaries. The loss is formulated as:

$$\mathcal{L}_{\text{Boundary}} = \frac{1}{|V_{\text{PredB}}|} \sum_{i\in V_{\text{PredB}}} W^{\text{ProbB}}_{m^*,j^*}\, D^{\text{ProbB}}_{m^*,j^*}\, CE(\mathcal{Q}_i, \mathcal{P}_i) \quad (7)$$

where $W^{\text{ProbB}}_{m^*,j^*}$ emphasizes vertices moving toward high-probability boundaries, $D^{\text{ProbB}}_{m^*,j^*}$ emphasizes vertices relatively far from the boundary, and V_{PredB} denotes the vertices located on PredB.

Balance Loss. To prevent the trivial solution of assigning most vertices to only a few parcels, the Balance Loss $\mathcal{L}_{\text{Balance}}$ is introduced and formulated as:

$$\mathcal{L}_{\text{Balance}} = \frac{1}{K}\sum_{k=1}^{K} T_k \log T_k, \quad T_k = \frac{1}{|V|}\sum_{v\in V} P_{v,k} \tag{8}$$

Consequently, the overall loss function is given by:

$$\mathcal{L} = \lambda_{\text{DG_Homo}}\,\mathcal{L}_{\text{DG_Homo}} + \lambda_{\text{Boundary}}\,\mathcal{L}_{\text{Boundary}} + \lambda_{\text{Balance}}\,\mathcal{L}_{\text{Balance}} \tag{9}$$

3 Experiments and Results

3.1 Datasets and Implementation Details

Datasets and Preprocessing. Focusing on the first two postnatal years, we constructed age-independent infant group-level multi-modal cortical parcellations to provide standardized localization for those subsequent cross-subject and cross-age neuroimaging studies, and conducted experiments on 370 sMRI and rs-fMRI scans from 181 infants aged 0–800 days in the BCP dataset (Howell et al., 2019). Detailed descriptions of the dataset, splits, and processing are provided in Appendix A.

Model Implementation Details. The model was implemented in PyTorch and trained on a single NVIDIA RTX 4090 GPU using AdamW optimizer with a learning rate of 0.001. The mini-batch size $\mathcal{B}$ was set to 30, where multi-modal individual-level features from 30 subjects were sampled in each iteration for optimizing group-level parcellation homogeneity. The model was trained for 250 epochs. Empirically, the hyperparameter σ was set to 10. The spatial weight η_k was initialized to 1 and fixed for the first 60 epochs for training stabilization, then updated every 30 epochs by multiplying by 0.5 when no fragmentation occurred and by 1.5 otherwise. The loss weights were set to $\lambda_{\text{DG_Homo}} = 1$ and $\lambda_{\text{Balance}} = 1$, while $\lambda_{\text{Boundary}}$ was linearly increased from 0 to 1 during training to prevent premature boundary constraints from hindering homogeneity optimization.

3.2 Method Evaluation

Evaluation Metrics. Assuming cortical parcels correspond to neurobiological units, desirable parcellations should exhibit internally coherent connectivity and align well with architectonic/anatomic boundaries (Schaefer et al., 2018; Gordon et al., 2016). To assess these two aspects, we specifically evaluated cross-modal and cross-subject parcellation homogeneity and measured anatomical consistency with Brodmann areas 1, 2, 3, 4, 6, 17, 18, 44, and 45 using Hausdorff distance (Van Essen et al., 2012). We further assessed parcellation quality from two additional perspectives: the silhouette coefficient jointly evaluates compactness and separability, while the fraction of voxels in each parcel's largest connected component measures spatial continuity. Detailed descriptions and formulas for these four complementary metrics are provided in Appendix B.

Spatial Null Model Analysis. In addition to comparisons with SOTA methods and published parcellations, we also compare SABER with spatial null models (Schaefer et al., 2018; Gordon et al., 2016). Only when the evaluation metrics significantly exceed those of random parcellations generated by randomly rotating the original parcellation while preserving its size, shape, and relative position can the group-level parcellation be considered to reflect genuine biological organization rather than spatial autocorrelation (Gordon et al., 2016). We repeated the spatial null model 1000 times and used z-scores to quantify how strongly the original parcellation outperforms the random parcellations. The z-score is computed as (original metric value − mean metric value of random parcellations) / standard deviation of the random parcellations' metric values.

Downstream Tasks. A key role of a group-level parcellation map is to reduce the dimensionality of the dense human connectome while preserving critical information that underlies identity recognition and complex cognitive processes (Zhu et al., 2025a; Li et al., 2024). We evaluated this capability using four standard downstream tasks: gender classification, age prediction, cognitive prediction, and longitudinal individual identification (Hu et al., 2022). Better performance on these tasks indicates that the group-level parcellation map better preserves such critical information. Detailed experimental settings are provided in Appendix B.

Table 1: Quantitative comparison on evaluation metrics. ▲ denotes machine learning, ■ denotes deep learning, ♣ denotes published parcellations, and ♠ denotes SABER. **Bold** denotes the best value in each column; * denotes that SABER significantly outperforms it by paired t-test ($p < 0.05$).

Methods	Parcellation homogeneity		Silhouette coefficient		Anatomical alignment	Spatial continuity
	Functional↑	Structural↑	Functional↑	Structural↑	Mean↓	Parcellation↑
▲ MV-Kmeans-400	0.7966 (0.0532)*	0.5206 (0.0364)*	0.4964 (0.0762)*	0.2848 (0.0242)*	4.5740 (1.6845)*	0.8517
▲ Mv-SC-400	0.8326 (0.0427)*	0.5362 (0.0350)*	0.5541 (0.0691)*	0.2984 (0.0235)*	2.4140 (0.3981)*	0.9968
▲ CGL-400	0.8266 (0.0448)*	0.5251 (0.0389)*	0.5475 (0.0704)*	0.2937 (0.0230)*	2.4470 (0.2620)*	0.9983
■ RML-400	0.8069 (0.0502)*	0.5127 (0.0386)*	0.4495 (0.0756)*	0.2292 (0.0197)*	2.4398 (0.5325)*	0.9536
■ SparseMVC-400	0.6788 (0.0604)*	0.4621 (0.0415)*	0.2905 (0.0813)*	0.1619 (0.0152)*	11.2135 (5.1806)*	0.6017
■ MSDIB-400	0.8061 (0.0501)*	0.5336 (0.0386)*	0.5084 (0.0756)*	0.2805 (0.0239)*	2.5958 (1.1235)*	0.9400
♠ **SABER-400**	**0.8408 (0.0460)**	**0.5415 (0.0368)**	**0.5614 (0.0712)**	**0.3220 (0.0285)**	**1.8990 (0.4094)**	**0.9995**
♣ Glasser-180	0.7569 (0.0589)*	0.4381 (0.0428)*	0.4405 (0.0589)*	0.2409 (0.0209)*	3.4751 (0.6215)	**1.0000**
♠ **SABER-180**	**0.7691 (0.0588)**	**0.4614 (0.0401)**	**0.5122 (0.0588)**	**0.2777 (0.0245)**	**3.4338 (1.2595)**	0.9957
♣ Li-105	0.7165 (0.0623)*	0.4227 (0.0411)*	0.4856 (0.0643)*	0.2691 (0.0310)*	4.6682 (1.9553)*	**1.0000**
♠ **SABER-105**	**0.7241 (0.0643)**	**0.4410 (0.0409)**	**0.4929 (0.0623)**	**0.2734 (0.0255)**	**4.3175 (2.4545)**	0.9968
♣ Wang-430	**0.8467 (0.0439)**	0.5256 (0.0400)*	0.5523 (0.0439)*	0.2747 (0.0201)*	2.5059 (0.5731)*	**1.0000**
♠ **SABER-430**	0.8438 (0.0454)	**0.5502 (0.0369)**	**0.5638 (0.0454)**	**0.3125 (0.0248)**	**1.9883 (0.7567)**	0.9975

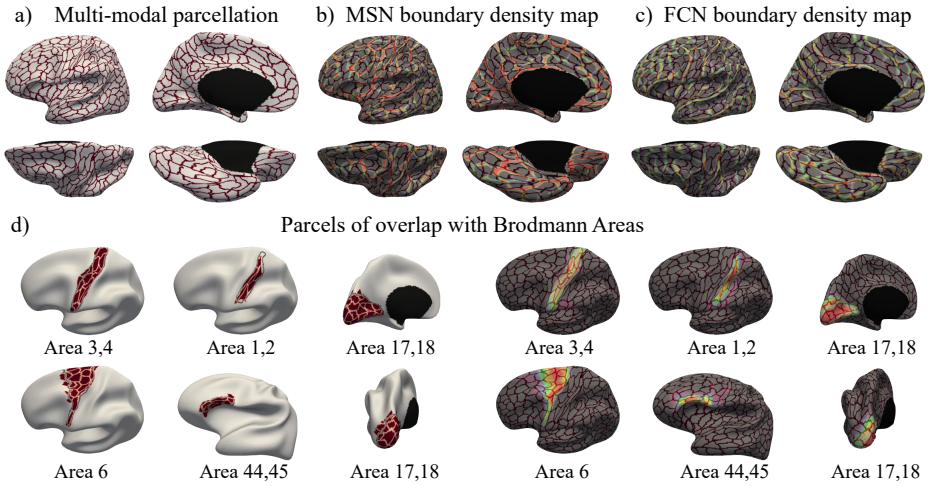


Figure 2: Qualitative visualization results for 400 parcels. a) Infant group-level multi-modal cortical parcellation map. b) and c) Boundary alignment between parcellation and boundary probability density maps. d) Overlap between parcellation and Brodmann areas.

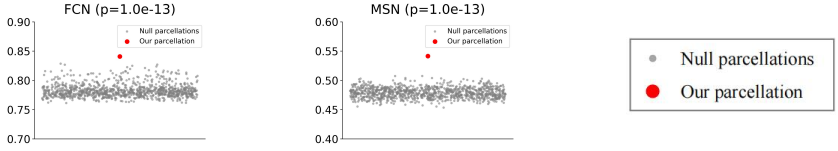
3.3 Comparison with SOTA Methods and Other Parcellations

Comparison Methods and Published Parcellations. To comprehensively evaluate SABER, we compared it with classical machine learning methods, SOTA deep learning methods, and widely used published group-level parcellations. Machine learning baselines include Multi-view Kmeans (Bickel et al., 2004), Multi-view Spectral Clustering (Kumar & Daumé, 2011), and CGL (Wen et al., 2024). As no deep cortical parcellation method exists for multi-modal infant data, we used recent general-purpose deep multi-modal clustering methods: RML (Xu et al., 2025), SparseMVC (Liu et al., 2026b), and MSDIB (Hu et al., 2025). Here, we exclude BDEC (Zhu et al., 2025a) from comparison because it is not multi-modal and lacks publicly available code, hindering faithful reproduction. For published parcellations, since no multi-modal group-level infant map exists for ages 0–2, we compared with the adult Glasser multi-modal map (Glasser et al., 2016) and neonatal Li multi-modal map (Li et al., 2023) to assess their adaptability to infant populations. We also included the Wang functional map (Wang et al., 2023a), built on the same 0–2-year-old cohort, to assess their adaptability to multi-modal data. All comparisons were performed in the UNC 4D infant surface template space (Wu et al., 2019) with identical resolution settings. Main-text results are at

Table 2: Quantitative comparison on downstream tasks. ▲ denotes machine learning, ■ denotes deep learning, ♣ denotes published parcellations, and ♠ denotes SABER. **Bold** denotes the best value in each column; * denotes that SABER significantly outperforms it by paired t-test ($p < 0.05$).

Methods	Gender classification	Age prediction	Cognitive prediction	Longitudinal individual identification	
	ACC↑	MAE↓	MAE↓	ACC (Functional)↑	ACC (Structural)↑
▲ MV-Kmeans-400	0.8595 (0.0107)*	81.7964 (1.8337)*	4.4196 (0.4477)*	0.3990 (0.0399)*	0.9326 (0.0187)*
▲ Mv-SC-400	0.8606 (0.0172)*	83.0008 (2.2617)*	4.2049 (0.4675)*	0.4435 (0.0275)	0.9594 (0.0126)*
▲ CGL-400	0.8592 (0.0140)*	80.9678 (1.4856)*	4.0938 (0.4125)*	0.4471 (0.0362)	0.9567 (0.0107)*
■ RML-400	0.8585 (0.0123)*	81.4590 (2.3630)*	4.1241 (0.3991)*	0.4278 (0.0083)*	0.9487 (0.0192)*
■ SparseMVC-400	0.8266 (0.0121)*	95.6629 (1.8053)*	4.5228 (0.4384)*	0.3677 (0.0314)*	0.9314 (0.0112)*
■ MSDIB-400	0.8473 (0.0139)*	82.3796 (1.6782)*	4.3192 (0.4163)*	0.4374 (0.0264)*	0.9326 (0.0143)*
♠ **SABER-400**	**0.8804 (0.0129)**	**78.0784 (1.3986)**	**4.0008 (0.4123)**	**0.4583 (0.0171)**	**0.9778 (0.0136)**
♣ Glasser-180	0.8322 (0.0126)*	82.8166 (1.8977)*	4.7060 (0.5565)*	0.3509 (0.0240)*	0.8903 (0.0246)
♠ **SABER-180**	**0.8494 (0.0090)**	**78.7473 (1.4819)**	**4.5949 (0.4634)**	**0.4134 (0.0187)**	**0.8963 (0.0247)**
♣ Li-105	0.8260 (0.0184)*	84.2968 (2.4434)*	5.4059 (0.5458)*	0.2548 (0.0336)*	0.8459 (0.0199)*
♠ **SABER-105**	**0.8422 (0.0125)**	**78.7703 (1.3363)**	**5.1820 (0.5459)**	**0.2980 (0.0267)**	**0.8742 (0.0250)**
♣ Wang-430	0.8572 (0.0179)*	78.9511 (1.7497)*	4.0009 (0.3856)	0.4423 (0.0231)*	0.9509 (0.0157)*
♠ **SABER-430**	**0.8757 (0.0108)**	**77.6459 (1.6533)**	**3.9467 (0.4047)**	**0.4651 (0.0213)**	**0.9791 (0.0116)**

a) Spatial null model analysis of parcellation homogeneity (Higher is better)



b) Spatial null model analysis of anatomical alignment (Lower is better)

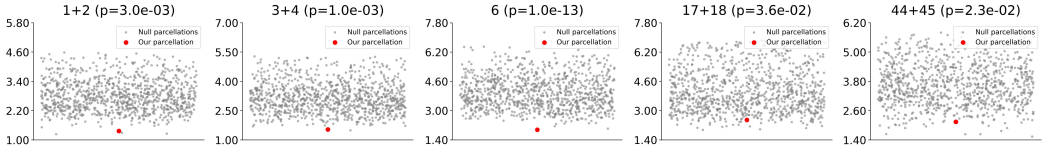


Figure 3: Spatial null model analysis. a) and b) compare our parcellation with 1,000 random parcellations in terms of parcellation homogeneity and anatomical alignment, respectively.

400 parcels, with additional resolutions (200, 300, 500, 105, 180, 430) and detailed methods and reproduction procedures in Appendix C.

Evaluation Metrics. As shown in Table 1 and Appendix Tables 6, SABER consistently outperforms the compared methods and published parcellations across nearly all metrics. For cross-modal and cross-subject homogeneity and silhouette coefficient, SABER achieves superior performance at almost all resolutions, indicating stronger cross-subject adaptability to the highly variable 0–2-year-old infant population, which benefits downstream studies that derive individual brain networks from group-level maps. The only exception is the Wang functional map built on the same infant cohort, where SABER achieves higher structural but slightly lower functional homogeneity, reflecting the difficulty of jointly optimizing both modalities. For anatomical alignment, SABER shows greater consistency with anatomically defined boundaries across all resolutions. Figure 2 and Appendix Figures 8 visually illustrate the alignment with the boundary probability density map and Brodmann areas. It can be observed that high-probability boundaries inherently exhibit a good correspondence with certain biologically meaningful boundaries. Therefore, by incorporating such information into the optimization process, SABER further improves its anatomical alignment performance. Finally, SABER exhibits almost no fragmentation, as confirmed by Figure 2 and Appendix Figures 4, whereas other deep learning methods often produce fragmented parcels without explicit spatial continuity constraints, complicating neurobiological interpretation.

Spatial Null Model Analysis. As shown in Figure 3 and Appendix Figures 5, SABER significantly outperforms random parcellations at all resolutions, indicating biologically meaningful organiza-

Table 3: Quantitative ablation results on evaluation metrics. **Bold** denotes the best value in each column; ↑ or ↓ indicates whether higher or lower values are better.

Ablation Study						Parcellation homo		Silhouette coefficient		Anatomical alignment	Spatial continuity
$\mathcal{L}_{\text{Homo}}$	$\mathcal{L}_{\text{G_Homo}}$	$\mathcal{L}_{\text{DG_Homo}}$	$\mathcal{L}_{\text{Boundary}}$	ML	VL	Func↑	Struct↑	Func↑	Struct↑	Mean↓	Parcel↑
✓				✓	✓	0.8186 (0.0494)	0.5349 (0.0368)	0.5422 (0.0749)	0.3146 (0.0263)	2.2410 (0.1892)	0.9379
	✓			✓	✓	0.8357 (0.0468)	**0.5438 (0.0377)**	0.5503 (0.0717)	**0.3236 (0.0286)**	2.2225 (0.2163)	0.9823
		✓		✓	✓	**0.8476 (0.0446)**	0.5322 (0.0373)	**0.5713 (0.0704)**	0.3148 (0.0276)	2.1997 (0.4514)	**0.9998**
		✓	✓	✓	✓	0.8408 (0.0460)	0.5415 (0.0368)	0.5614 (0.0712)	0.3220 (0.0285)	**1.8990 (0.4094)**	0.9995
		✓	✓		✓	0.8383 (0.0468)	0.5335 (0.0370)	0.5558 (0.0718)	0.3073 (0.0260)	2.1161 (0.3771)	0.9993
		✓	✓	✓		0.8397 (0.0467)	0.5329 (0.0374)	0.5547 (0.0718)	0.3075 (0.0253)	2.0196 (0.5089)	0.9992
		✓	✓			0.8369 (0.0462)	0.5291 (0.0371)	0.5505 (0.0714)	0.3071 (0.0248)	2.2103 (0.6273)	0.9995

Table 4: Quantitative ablation results on the plug-and-play capability of $\mathcal{L}_{\text{DG_Homo}}$ and $\mathcal{L}_{\text{Boundary}}$. **Bold** denotes the best value in each column; ↑ / ↓ indicate higher / lower is better.

Ablation Study					Parcellation homo		Silhouette coefficient		Anatomical alignment	Spatial continuity
RML	SparseMVC	MSDIB	$\mathcal{L}_{\text{DG_Homo}}$	$\mathcal{L}_{\text{Boundary}}$	Func↑	Struct↑	Func↑	Struct↑	Mean↓	Parcel↑
✓					0.8069 (0.0502)	0.5127 (0.0386)	0.4495 (0.0756)	0.2292 (0.0197)	2.4398 (0.5325)	0.9536
✓			✓		**0.8377 (0.0464)**	0.5321 (0.0368)	**0.5597 (0.0715)**	0.3147 (0.0282)	2.2682 (0.5122)	**0.9914**
✓			✓	✓	0.8298 (0.0467)	**0.5382 (0.0371)**	0.5521 (0.0723)	**0.3204 (0.0280)**	**2.0241 (0.4912)**	0.9922
	✓				0.6788 (0.0604)	0.4621 (0.0415)	0.2905 (0.0813)	0.1619 (0.0152)	11.2135 (5.1806)	0.6017
	✓		✓		**0.7867 (0.0498)**	0.4724 (0.0360)	**0.4755 (0.0733)**	0.2673 (0.0208)	2.6016 (0.5962)	**0.9951**
	✓		✓	✓	0.7802 (0.0511)	**0.4792 (0.0364)**	0.4703 (0.0756)	**0.2692 (0.0223)**	**2.3169 (0.6362)**	0.9942
		✓			0.8061 (0.0501)	0.5336 (0.0386)	0.5084 (0.0756)	0.2805 (0.0239)	2.5958 (1.1235)	0.9400
		✓	✓		**0.8332 (0.0465)**	0.5388 (0.0394)	**0.5462 (0.0390)**	0.2991 (0.0225)	2.0753 (0.1829)	**0.9993**
		✓	✓	✓	0.8301 (0.05393)	**0.5393 (0.0398)**	0.5434 (0.0396)	**0.3013 (0.0232)**	**1.9433 (0.2214)**	0.9991

tion beyond random spatial patterns. Appendix Figures 6 further compare different methods and published parcellations against spatial null models. Most methods significantly outperform random parcellations, with exceptions such as Mv-SC and CGL showing lower functional homogeneity at 200 parcels, reflecting instability. Appendix Tables 7 report the corresponding z-scores. Except for Mv-Kmeans, SABER achieves higher z-scores than all other methods and parcellations, indicating more biologically meaningful solutions. The higher z-score of Mv-Kmeans may result from its fragmentation, as spatially discontinuous parcellations can yield inflated z-scores (Arslan et al., 2018; Zhu et al., 2025a), however, such discontinuity is difficult to interpret in fine-grained parcellations.

Downstream Tasks. As shown in Tables 2 and Appendix Tables 8, SABER outperforms other methods across the four downstream tasks at nearly all resolutions. We attribute this to the better cross-subject adaptability, which reduces the loss of identity- and cognition-related information when deriving individual brain networks from the group-level map due to severe mismatch.

3.4 Ablation Studies

Ablation Setup. First, we conduct ablation studies on $\mathcal{L}_{\text{DG_Homo}}$ by decomposing it into two variants: $\mathcal{L}_{\text{Homo}}$, which optimizes homogeneity using only group-level features, and $\mathcal{L}_{\text{G_Homo}}$, which uses individual-level features, neither includes dynamic spatial weighting. We also evaluate $\mathcal{L}_{\text{Boundary}}$

Table 5: Quantitative ablation results on downstream tasks. **Bold** denotes the best value in each column; * denotes that introducing $\mathcal{L}_{\text{DG_Homo}}$ significantly outperforms it by paired t-test ($p < 0.05$).

Ablation Study					Gender classification	Age prediction	Cognitive prediction
SABER	RML	SparseMVC	MSDIB	$\mathcal{L}_{\text{DG_Homo}}$	ACC↑	MAE↓	MAE↓
✓					0.8511 (0.0228)*	79.2748 (2.1286)	4.2431 (0.4494)*
✓				✓	**0.8804 (0.0129)**	**78.0784 (1.3986)**	**4.0008 (0.4123)**
	✓				0.8585 (0.0123)*	81.4590 (2.3630)*	4.1241 (0.3991)*
	✓			✓	**0.8739 (0.0104)**	**79.9404 (2.2068)**	**3.9755 (0.4043)**
		✓			0.8266 (0.0121)*	95.6629 (1.8053)*	4.5228 (0.4384)*
		✓		✓	**0.8669 (0.0162)**	**82.9186 (1.3692)**	**4.2197 (0.3546)**
			✓		0.8473 (0.0139)*	82.3796 (1.6782)*	4.3192 (0.4163)*
			✓	✓	**0.8654 (0.0151)**	**80.0321 (1.5681)**	**4.1147 (0.3783)**

by comparing models with and without it. Next, we replace multi-layer graph mutual learning (ML) with single-layer graph learning (SL), and vertex-level weighting (VL) with global weighting (GL), to assess their roles in capturing associative and complementary information. Furthermore, to examine the plug-and-play capability of $\mathcal{L}_{\text{DG_Homo}}$ and $\mathcal{L}_{\text{Boundary}}$, we integrate them into RML, SparseMVC, and MSDIB to evaluate generalization. Meanwhile, we assess whether $\mathcal{L}_{\text{DG_Homo}}$ improves downstream performance by enhancing cross-subject adaptability of group-level parcellation. Finally, we analyze the effects of progressively annealed spatial weights during training.

Evaluation Metrics. As shown in Table 3 and Appendix Tables 9, $\mathcal{L}_{\text{Homo}}$, which uses only group-level features, fails to ensure cross-subject adaptability, leading to marked declines in cross-subject and cross-modal homogeneity and silhouette coefficient. In contrast, $\mathcal{L}_{\text{G_Homo}}$ substantially improves cross-subject adaptability by optimizing homogeneity over individual-level features. Moreover, $\mathcal{L}_{\text{G_Homo}}$ also improves spatial continuity, likely because incorporating a large set of individual-level data makes vertex assignments more robust under broader constraints. However, fragmentation remains until dynamic spatial weighting is introduced in $\mathcal{L}_{\text{DG_Homo}}$, which nearly eliminates it. Adding $\mathcal{L}_{\text{Boundary}}$ further improves anatomical alignment. Replacing ML with SL or VL with GL causes slight overall performance drops, confirming their contributions to modeling associative and complementary information. As shown in Table 4 and Appendix Tables 10, integrating $\mathcal{L}_{\text{DG_Homo}}$ and $\mathcal{L}_{\text{Boundary}}$ into RML, SparseMVC, and MSDIB yields clear overall gains, demonstrating their plug-and-play capability and effectiveness for cortical parcellation. Moreover, SABER trains faster because it avoids the pretraining and more complex architectures required by these baselines.

Downstream Tasks. As shown in Table 5 and Appendix Tables 11, after incorporating $\mathcal{L}_{\text{DG_Homo}}$, all methods achieve significant performance improvements on the three downstream tasks. These results strongly indicate that explicitly optimizing the cross-subject adaptability of the group-level map is crucial for improving downstream task performance in infants with substantial inter-subject variability. If the group-level map is not well adapted to individual data, the quality of the derived individual brain networks deteriorates markedly, thereby compromising subsequent model fitting.

Dynamic Spatial Weight Analysis. As shown in Appendix Figure 7, as training proceeds, the spatial weights η_k of nearly all parcels approach zero, while structural and functional homogeneity gradually improves and converges, and spatial continuity remains well preserved with almost no fragmentation, demonstrating the effectiveness of progressive annealing of the spatial weights.

4 Conclusion

In this work, we propose SABER, the first subject-adaptive and boundary-aware deep multi-modal clustering network for infant group-level cortical parcellation. Extensive experiments on the BCP dataset demonstrate that explicitly improving the cross-subject adaptability of group-level parcellations and incorporating local boundary evidence can enhance downstream task performance and sensitivity to potential biological boundaries, which is valuable for subsequent analyses and neurobiological interpretation. In addition, a simple yet effective two-stage fusion strategy further improves parcellation quality by more fully exploiting associative and complementary structural-functional information. Collectively, these components contribute to the superior performance of SABER.

AI USE STATEMENT

In this work, we used generative AI tools, including ChatGPT, Claude, and DeepSeek, to assist with translation and method implementation. We did not use generative AI tools to generate synthetic data, develop theoretical models or conceptual frameworks, formulate or prove mathematical claims, propose or refine hypotheses, design research methodology or experiments, clean or reformat datasets, conduct qualitative or thematic data analysis, or interpret experimental results. Additionally, we used generative AI tools to improve the readability of the manuscript, assist in drafting parts of the paper, create or edit software code, summarize or analyze existing literature, and support brainstorming of research ideas. All AI-assisted content was reviewed by the authors. Specifically, manuscript text was checked sentence by sentence, translations were manually verified, generated or modified code was executed and tested for correctness, and the resulting experimental outputs were independently checked. We take full responsibility for the final content of this work, including all text, claims, code, and other artifacts produced with the assistance of generative AI.

ETHICS STATEMENT

This study exclusively uses publicly available data from the UNC/UMN Baby Connectome Project (BCP) and involves no new participant recruitment or data acquisition. The original BCP data collection followed the ethical approval and informed consent procedures described in the corresponding BCP work (Howell et al., 2019). Our work focuses solely on secondary study of the released neuroimaging data and does not involve subject re-identification, clinical decision-making, or the release of individual-level sensitive information. We therefore do not introduce additional ethical concerns beyond those associated with the original BCP data collection. There are no potential conflicts of interest or sponsorship relationships related to this work.

REPRODUCIBILITY STATEMENT

All MRI data analyzed in this study are from the publicly available UNC/UMN Baby Connectome Project (`https://nda.nih.gov/edit_collection.html?id=2848`), accessible via the NIMH Data Archive. The code and the resulting infant group-level multi-modal parcellation maps at multiple resolutions are provided in the SABER repository (`https://anonymous.4open.science/r/SABER2027`). Detailed experimental settings and procedures are described in the main text and appendix, while the code repository provides more specific annotations to facilitate understanding. These resources collectively support the reproducibility of our results.

REFERENCES

Chris L Adamson, Bonnie Alexander, Gareth Ball, Richard Beare, Jeanie LY Cheong, Alicia J Spittle, Lex W Doyle, Peter J Anderson, Marc L Seal, and Deanne K Thompson. Parcellation of the neonatal cortex using surface-based melbourne children's regional infant brain atlases (m-crib-s). *Scientific reports*, 10(1):4359, 2020.

Salim Arslan, Sofia Ira Ktena, Antonios Makropoulos, Emma C Robinson, Daniel Rueckert, and Sarah Parisot. Human brain mapping: A systematic comparison of parcellation methods for the human cerebral cortex. *NeuroImage*, 170:5–30, 2018.

Graham L Baum, Zaixu Cui, David R Roalf, Rastko Ciric, Richard F Betzel, Bart Larsen, Matthew Cieslak, Philip A Cook, Cedric H Xia, Tyler M Moore, et al. Development of structure–function coupling in human brain networks during youth. *Proceedings of the National Academy of Sciences*, 117(1):771–778, 2020.

Christian F Beckmann and Stephen M Smith. Probabilistic independent component analysis for functional magnetic resonance imaging. *IEEE transactions on medical imaging*, 23(2):137–152, 2004.

Steffen Bickel, Tobias Scheffer, et al. Multi-view clustering. In *Icdm*, pp. 19–26, 2004.

Yue Cui, Chengyi Li, Yuheng Lu, Liang Ma, Luqi Cheng, Long Cao, Shan Yu, and Tianzi Jiang. Multimodal connectivity-based individual parcellation and analysis for humans and rhesus monkeys. *IEEE Transactions on Medical Imaging*, 43(9):3343–3353, 2024.

Lingzhong Fan, Hai Li, Junjie Zhuo, Yu Zhang, Jiaojian Wang, Liangfu Chen, Zhengyi Yang, Congying Chu, Sangma Xie, Angela R Laird, et al. The human brainnetome atlas: a new brain atlas based on connectional architecture. *Cerebral cortex*, 26(8):3508–3526, 2016.

Emily S Finn, Xilin Shen, Dustin Scheinost, Monica D Rosenberg, Jessica Huang, Marvin M Chun, Xenophon Papademetris, and R Todd Constable. Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. *Nature neuroscience*, 18(11):1664–1671, 2015.

Bruce Fischl. Freesurfer. *Neuroimage*, 62(2):774–781, 2012.

Bruce Fischl, Martin I Sereno, and Anders M Dale. Cortical surface-based analysis: Ii: inflation, flattening, and a surface-based coordinate system. *Neuroimage*, 9(2):195–207, 1999.

Panagiotis Fotiadis, Linden Parkes, Kathryn A Davis, Theodore D Satterthwaite, Russell T Shinohara, and Dani S Bassett. Structure–function coupling in macroscale human brain networks. *Nature Reviews Neuroscience*, 25(10):688–704, 2024.

Jingjing Gao, Mingqi Liu, Maomin Qian, Heping Tang, Junyi Wang, Liang Ma, Yanling Li, Xin Dai, Zhengning Wang, Fengmei Lu, et al. Fine-scale striatal parcellation using diffusion mri tractography and graph neural networks. *Medical Image Analysis*, 101:103482, 2025.

Matthew F Glasser, Stamatios N Sotiropoulos, J Anthony Wilson, Timothy S Coalson, Bruce Fischl, Jesper L Andersson, Junqian Xu, Saad Jbabdi, Matthew Webster, Jonathan R Polimeni, et al. The minimal preprocessing pipelines for the human connectome project. *Neuroimage*, 80:105–124, 2013.

Matthew F Glasser, Timothy S Coalson, Emma C Robinson, Carl D Hacker, John Harwell, Essa Yacoub, Kamil Ugurbil, Jesper Andersson, Christian F Beckmann, Mark Jenkinson, et al. A multi-modal parcellation of human cerebral cortex. *Nature*, 536(7615):171–178, 2016.

Evan M Gordon, Timothy O Laumann, Babatunde Adeyemo, Jeremy F Huckins, William M Kelley, and Steven E Petersen. Generation and evaluation of a cortical area parcellation from resting-state correlations. *Cerebral cortex*, 26(1):288–303, 2016.

Douglas N Greve and Bruce Fischl. Accurate and robust brain image alignment using boundary-based registration. *Neuroimage*, 48(1):63–72, 2009.

Seungbo Ha and Ilwoo Lyu. Spharm-net: Spherical harmonics-based convolution for cortical parcellation. *IEEE Transactions on Medical Imaging*, 41(10):2739–2751, 2022.

Brittany R Howell, Martin A Styner, Wei Gao, Pew-Thian Yap, Li Wang, Kristine Baluyot, Essa Yacoub, Geng Chen, Taylor Potts, Andrew Salzwedel, et al. The unc/umn baby connectome project (bcp): An overview of the study design and protocol development. *NeuroImage*, 185: 891–905, 2019.

Dan Hu, Fan Wang, Han Zhang, Zhengwang Wu, Zhen Zhou, Guoshi Li, Li Wang, Weili Lin, Gang Li, UNC/UMN Baby Connectome Project Consortium, et al. Existence of functional connectome fingerprint during infancy and its stability over months. *Journal of Neuroscience*, 42(3):377–389, 2022.

Shizhe Hu, Jiahao Fan, Guoliang Zou, and Yangdong Ye. Multi-aspect self-guided deep information bottleneck for multi-modal clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pp. 17314–17322, 2025.

Ying Huang, Zhengwang Wu, Fan Wang, Dan Hu, Tengfei Li, Lei Guo, Li Wang, Weili Lin, and Gang Li. Mapping developmental regionalization and patterns of cortical surface area from 29 post-menstrual weeks to 2 years of age. *Proceedings of the National Academy of Sciences*, 119 (33):e2121748119, 2022.

Ying Huang, Zhengwang Wu, Tengfei Li, Xifeng Wang, Ya Wang, Lei Xing, Hongtu Zhu, Weili Lin, Li Wang, Lei Guo, et al. Mapping genetic topography of cortical thickness and surface area in neonatal brains. *The Journal of Neuroscience*, 43(34):6010–6020, 2023.

Ying Huang, Zhengwang Wu, Ya Wang, Tengfei Li, Xifeng Wang, Hongtu Zhu, Weili Lin, Li Wang, Jun Feng, John H Gilmore, et al. Revealing fine-grained genetically informed cortical parcellation maps of neonates based on multi-view spectral clustering. In *2024 46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 1–4. IEEE, 2024.

Mark Jenkinson, Peter Bannister, Michael Brady, and Stephen Smith. Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage*, 17(2): 825–841, 2002.

Weixiong Jiang, Han Zhang, Li-Ming Hsu, Dan Hu, Guoshi Li, Ye Wu, and Dinggang Shen. Early development of infant brain complex network. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 832–840. Springer, 2019.

Tae-Eui Kam, Xuyun Wen, Bing Jin, Zhicheng Jiao, Li-Ming Hsu, Zhen Zhou, Yujie Liu, Koji Yamashita, Sheng-Che Hung, Weili Lin, et al. A deep learning framework for noise component detection from resting-state functional mri. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 754–762. Springer, 2019.

Abhishek Kumar and Hal Daumé. A co-training approach for multi-view spectral clustering. In *Proceedings of the 28th international conference on machine learning (ICML-11)*, pp. 393–400, 2011.

Chengyi Li, Yuheng Lu, Shan Yu, and Yue Cui. Ts-ai: A deep learning pipeline for multimodal subject-specific parcellation with task contrasts synthesis. *Medical Image Analysis*, 97:103297, 2024.

Gang Li, Jingxin Nie, Guorong Wu, Yaping Wang, Dinggang Shen, Alzheimer's Disease Neuroimaging Initiative, et al. Consistent reconstruction of cortical surfaces from longitudinal brain mr images. *Neuroimage*, 59(4):3805–3820, 2012.

Gang Li, Li Wang, Feng Shi, John H Gilmore, Weili Lin, and Dinggang Shen. Construction of 4d high-definition cortical surface atlases of infants: Methods and applications. *Medical image analysis*, 25(1):22–36, 2015.

Gang Li, Li Wang, Weili Lin, and Dinggang Shen. Developmental patterns based individualized parcellation of infant cortical surface. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 66–74. Springer, 2017.

Mingyang Li, Xinyi Xu, Zuozhen Cao, Ruike Chen, Ruoke Zhao, Zhiyong Zhao, Xixi Dang, Kenichi Oishi, and Dan Wu. Multi-modal multi-resolution atlas of the human neonatal cerebral cortex based on microstructural similarity. *NeuroImage*, 272:120071, 2023.

Xinyuan Liang, Lianglong Sun, Mingrui Xia, Tengda Zhao, Gaolang Gong, Qiongling Li, Xuhong Liao, Zaixu Cui, Dingna Duan, Chenxuan Pang, et al. Dissecting human cortical similarity networks across the lifespan. *Neuron*, 113(19):3275–3295, 2025.

Jiameng Liu, Feihong Liu, Kaicong Sun, Zhiming Cui, Tianyang Sun, Zehong Cao, Jiawei Huang, Shuwei Bai, Yulin Wang, Yulong Dou, et al. Brainparc: unified lifespan brain parcellation from structural magnetic resonance images. *Nature Computational Science*, 6(6):588–602, 2026a.

Ruimeng Liu, Xin Zou, Chang Tang, Xiao Zheng, Xingchen Hu, Kun Sun, and Xinwang Liu. Sparsemvc: probing cross-view sparsity variations for multi-view clustering. *Advances in Neural Information Processing Systems*, 38:154594–154622, 2026b.

Zhiguo Luo, Ling-Li Zeng, Jian Qin, Chenping Hou, Hui Shen, and Dewen Hu. Functional parcellation of human brain precuneus using density-based clustering. *Cerebral cortex*, 30(1):269–282, 2020.

Liang Ma, Yu Zhang, Hantian Zhang, Luqi Cheng, Zhengyi Yang, Yuheng Lu, Weiyang Shi, Wen Li, Junjie Zhuo, Jiaojian Wang, Lingzhong Fan, and Tianzi Jiang. Bai-net: Individualized anatomical cerebral cartography using graph neural network. *IEEE Transactions on Neural Networks and Learning Systems*, 35(6):7446–7457, 2024.

Sarah Parisot, Ben Glocker, Sofia Ira Ktena, Salim Arslan, Markus D Schirmer, and Daniel Rueckert. A flexible graphical model for multi-modal parcellation of the cortex. *Neuroimage*, 162:226–248, 2017.

Wenyuan Qiu, Liang Ma, Tianzi Jiang, and Yu Zhang. Unrevealing reliable cortical parcellation of individual brains using resting-state functional magnetic resonance imaging and masked graph convolutions. *Frontiers in neuroscience*, 16:838347, 2022.

Alexander Schaefer, Ru Kong, Evan M Gordon, Timothy O Laumann, Xi-Nian Zuo, Avram J Holmes, Simon B Eickhoff, and BT Thomas Yeo. Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity mri. *Cerebral cortex*, 28(9):3095–3114, 2018.

Feng Shi, Andrew P Salzwedel, Weili Lin, John H Gilmore, and Wei Gao. Functional brain parcellations of the infant brain and the associated developmental trends. *Cerebral Cortex*, 28(4): 1358–1368, 2018.

Yunsheng Shi, Zhengjie Huang, Shikun Feng, Hui Zhong, Wenjin Wang, and Yu Sun. Masked label prediction: Unified message passing model for semi-supervised classification. *arXiv preprint arXiv:2009.03509*, 2020.

John G Sled, Alex P Zijdenbos, and Alan C Evans. A nonparametric method for automatic correction of intensity nonuniformity in mri data. *IEEE transactions on medical imaging*, 17(1):87–97, 1998.

Stephen M Smith, Mark Jenkinson, Mark W Woolrich, Christian F Beckmann, Timothy EJ Behrens, Heidi Johansen-Berg, Peter R Bannister, Marilena De Luca, Ivana Drobnjak, David E Flitney, et al. Advances in functional and structural mr image analysis and implementation as fsl. *Neuroimage*, 23:S208–S219, 2004.

Liang Sun, Daoqiang Zhang, Chunfeng Lian, Li Wang, Zhengwang Wu, Wei Shao, Weili Lin, Dinggang Shen, Gang Li, UNC/UMN Baby Connectome Project Consortium, et al. Topological correction of infant white matter surfaces using anatomically constrained convolutional neural network. *NeuroImage*, 198:114–124, 2019.

Lianglong Sun, Tengda Zhao, Xinyuan Liang, Mingrui Xia, Qiongling Li, Xuhong Liao, Gaolang Gong, Qian Wang, Chenxuan Pang, Qian Yu, et al. Human lifespan changes in the brain's functional connectome. *Nature neuroscience*, 28(4):891–901, 2025.

David C Van Essen, Matthew F Glasser, Donna L Dierker, John Harwell, and Timothy Coalson. Parcellations and hemispheric asymmetries of human cerebral cortex analyzed on surface-based atlases. *Cerebral cortex*, 22(10):2241–2262, 2012.

Chendi Wang, Bernard Ng, and Rafeef Garbi. Multimodal brain parcellation based on functional and anatomical connectivity. *Brain Connectivity*, 8(10):604–617, 2018a.

Chi Wang, Yunke Zhang, Miaomiao Cui, Peiran Ren, Yin Yang, Xuansong Xie, Xian-Sheng Hua, Hujun Bao, and Weiwei Xu. Active boundary loss for semantic segmentation. In *Proceedings of the AAAI conference on artificial intelligence*, pp. 2397–2405, 2022.

Fan Wang, Chunfeng Lian, Zhengwang Wu, Han Zhang, Tengfei Li, Yu Meng, Li Wang, Weili Lin, Dinggang Shen, and Gang Li. Developmental topography of cortical thickness during infancy. *Proceedings of the National Academy of Sciences*, 116(32):15855–15860, 2019.

Fan Wang, Han Zhang, Zhengwang Wu, Dan Hu, Zhen Zhou, Jessica B Girault, Li Wang, Weili Lin, and Gang Li. Fine-grained functional parcellation maps of the infant cerebral cortex. *elife*, 12: e75401, 2023a.

Li Wang, Gang Li, Feng Shi, Xiaohuan Cao, Chunfeng Lian, Dong Nie, Mingxia Liu, Han Zhang, Guannan Li, Zhengwang Wu, et al. Volume-based analysis of 6-month-old infant brain mri for autism biomarker identification and early diagnosis. In *International conference on medical image computing and computer-assisted intervention*, pp. 411–419. Springer, 2018b.

Li Wang, Zhengwang Wu, Liangjun Chen, Yue Sun, Weili Lin, and Gang Li. ibeat v2. 0: a multisite-applicable, deep learning-based pipeline for infant cerebral cortical surface reconstruction. *Nature protocols*, 18(5):1488–1509, 2023b.

Qiushi Wang, Yuehua Xu, Tengda Zhao, Zhilei Xu, Yong He, and Xuhong Liao. Individual uniqueness in the neonatal functional connectome. *Cerebral Cortex*, 31(8):3701–3712, 2021.

Xuyun Wen, Yunxi Zhao, Geng Chen, Han Zhang, and Daoqiang Zhang. Constructing fine-grained spatiotemporal neonatal functional atlases with spectral functional network learning. *Human Brain Mapping*, 45(8):e26718, 2024.

Zhengwang Wu, Li Wang, Weili Lin, John H Gilmore, Gang Li, and Dinggang Shen. Construction of 4d infant cortical surface atlases with sharp folding patterns via spherical patch-based group-wise sparse representation. *Human brain mapping*, 40(13):3860–3880, 2019.

Jie Xu, Na Zhao, Gang Niu, Masashi Sugiyama, and Xiaofeng Zhu. Robust multi-view learning via representation fusion of sample-level attention and alignment of simulated perturbation. In *2025 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 4232–4241. IEEE, 2025.

Xiaoxuan Yan, Ru Kong, Aihuiping Xue, Qing Yang, Csaba Orban, Lijun An, Avram J Holmes, Xing Qian, Jianzhong Chen, Xi-Nian Zuo, et al. Homotopic local-global parcellation of the human cerebral cortex from resting-state functional connectivity. *NeuroImage*, 273:120010, 2023.

Ben Yang, Xuetao Zhang, Zhongheng Li, Feiping Nie, and Fei Wang. Efficient multi-view k-means clustering with multiple anchor graphs. *IEEE Transactions on Knowledge and Data Engineering*, 35(7):6887–6900, 2022.

BT Thomas Yeo, Mert R Sabuncu, Tom Vercauteren, Nicholas Ayache, Bruce Fischl, and Polina Golland. Spherical demons: fast diffeomorphic landmark-free surface registration. *IEEE transactions on medical imaging*, 29(3):650–668, 2009.

Weiyan Yin, Tengfei Li, Zhengwang Wu, Sheng-Che Hung, Dan Hu, Yiding Gui, Seoyoon Cho, Yue Sun, Mackenzie Allan Woodburn, Li Wang, et al. Charting brain functional development from birth to 6 years of age. *Nature Human Behaviour*, 9(6):1246–1259, 2025.

Jing Zhang, Jingguo Dai, Zhichao Wang, Shengfeng Liu, Haotian Qian, Botao Zhao, and Yu Zhang. Individualized brain parcellation using multimodal fusion and graph convolution networks. In *2024 IEEE International Symposium on Biomedical Imaging (ISBI)*, pp. 1–4. IEEE, 2024.

Qian Zhang, Li Wang, Xiaopeng Zong, Weili Lin, Gang Li, and Dinggang Shen. Frnet: Flattened residual network for infant mri skull stripping. In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pp. 999–1002. IEEE, 2019.

Fenqiang Zhao, Zhengwang Wu, Li Wang, Weili Lin, and Gang Li. Longitudinally consistent registration and parcellation of cortical surfaces using semi-supervised learning. *Medical image analysis*, 96:103193, 2024.

Zhen Zhou, Han Zhang, Li-Ming Hsu, Weili Lin, Gang Pan, Dinggang Shen, and UNC/UMN Baby Connectome Project Consortium. Multi-layer temporal network analysis reveals increasing temporal reachability and spreadability in the first two years of life. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 665–672. Springer, 2019.

Jianfei Zhu, Xiaoxiao Ma, Baichun Wei, Zhicai Zhong, Hui Zhou, Feng Jiang, Haiqi Zhu, and Chunzhi Yi. Bdec: brain deep embedded clustering model for resting state fmri group-level parcellation of the human cerebral cortex. *IEEE Transactions on Biomedical Engineering*, 73(2): 644–657, 2025a.

Yuanzhuo Zhu, Xianjun Li, Chen Niu, Fan Wang, and Jianhua Ma. Geometric deep learning with adaptive full-band spatial diffusion for accurate, efficient, and robust cortical parcellation. *Medical Image Analysis*, 101:103492, 2025b.

A Appendix A

A.1 Related works

Infant Group-Level Cortical Parcellation. Infant group-level cortical parcellation has gained increasing attention in developmental neuroimaging. Delineating distinct brain areas based on structural and functional properties is essential for uncovering early developmental and operational mechanisms, providing standardized spatial anchors for cross-subject and cross-age analyses, and offering an effective dimensionality reduction strategy that improves statistical power in dense infant brain network analyses (Glasser et al., 2016; Wang et al., 2023a; Huang et al., 2022). Prior work has established infant group-level cortical parcellations from structural folding patterns (e.g., the UNC 4D cortical surface (Wu et al., 2019) and surface-based Melbourne atlases (Adamson et al., 2020)), resting-state functional connectivity profiles (e.g., Shi et al. (Shi et al., 2018) and Wang et al. (Wang et al., 2023a)), morphological development (e.g., developmental maps of cortical thickness and surface area by Wang et al. (Wang et al., 2019) and Huang et al. (Huang et al., 2022), respectively), and genetic information about morphology (e.g., genetic maps of cortical thickness and surface area by Huang et al. Huang et al. (2023)). However, existing parcellations rely almost exclusively on single-modality data, limiting their utility across multi-modal neuroimaging studies. Thus, constructing a group-level multi-modal infant cortical parcellation is foundational for characterizing early neurodevelopment and identifying disease risks from structural and functional perspectives.

Multi-Modal Cortical Parcellation. To date, multi-modal cortical parcellation has mainly targeted adult and neonatal populations (Glasser et al., 2016; Parisot et al., 2017; Li et al., 2023). As substantial organizational differences preclude direct transfer to infants, we focus on how these frameworks integrate multi-modal information. Existing multi-modal integration strategies broadly fall into three paradigms. The first assigns spatially uniform global weights across modalities for multi-modal clustering (Huang et al., 2024), but neglects regionally heterogeneous modal contributions and blurs parcellation boundaries. The second uses hand-crafted vertex-level weighting to capture this heterogeneity via spatially varying weights derived from functional reliability (Wang et al., 2018a; Cui et al., 2024) or structural bias (Parisot et al., 2017), but lacks data-driven adaptivity, making parcellation quality susceptible to empirical bias. The third adopts neural networks or statistical models (e.g., GCNs (Zhang et al., 2024), spherical U-Nets (Li et al., 2024), or PCA (Li et al., 2023)) for adaptive multi-modal integration, but predominantly relies on naive feature concatenation, limiting fusion depth. Consequently, improving parcellation quality requires more flexible comprehensive strategies that fully exploit associative and complementary structural-functional information.

Deep Learning in Cortical Parcellation. Given the inherently high noise and high dimensionality of infant neuroimaging data, traditional machine learning often produces fragmented and unstable parcellations (Luo et al., 2020; Gao et al., 2025). The success of deep learning in computer vision has further motivated the use of deep learning with joint representation learning to construct group-level infant cortical parcellations. However, most existing studies target adult individual-level parcellation. They either train models with full or semi-supervision on large sets of manually annotated Desikan-Killiany maps (Zhu et al., 2025b; Ha & Lyu, 2022; Zhao et al., 2024; Liu et al., 2026a), or use other group-level maps (e.g., Glasser's (Glasser et al., 2016), Brainnetome (Fan et al., 2016), and Schaefer (Schaefer et al., 2018)) as reference priors to constrain training (Li et al., 2024; Ma et al., 2024; Qiu et al., 2022). In either case, these methods require an existing map and strong supervision to stabilize optimization. By contrast, group-level cortical parcellation itself constructs this "existing map", making it a standard unsupervised learning task. To date, few works apply deep learning to group-level cortical parcellation, with BDEC (Zhu et al., 2025a) as a representative example. BDEC takes group-level features as input, trains a specifically adjusted DEC model, and, after post-processing, yields adult functional maps that outperform previous ones. However, unsupervised objectives such as DEC-style annealing in BDEC (Zhu et al., 2025a) or contrastive learning in other novel deep clustering methods (Liu et al., 2026b; Xu et al., 2025) cannot adequately meet the demands of infant group-level cortical parcellation. First, they optimize group-level homogeneity using only group-level features. Given substantial infant inter-subject variability (Li et al., 2017; Zhao et al., 2024), the lack of explicit adaptation to individual-level features makes it difficult to construct a representative group-level map. Consequently, the resulting maps may fit only a part of subjects well, exhibit poor homogeneity on others due to severe group–individual mismatches, and compromise downstream tasks that rely on the group-level map to construct individual brain networks. Second, a purely clustering objective can easily yield high homogeneity while significantly

mismatching anatomical prior boundaries (Schaefer et al., 2018), complicating neurobiological interpretation. Therefore, designing targeted optimization objectives is both urgent and crucial.

A.2 Datasets and Preprocessing

Participants and Image Acquisition. This work focuses on infants during the first two postnatal years, a critical window characterized by rapid structural and functional development underlying the formation of cognitive, sensory, and motor functions (Huang et al., 2022; Wang et al., 2023a). Constructing group-level parcellation maps and conducting neuroimaging research during this developmental stage enhance our limited understanding of brain development and also provide valuable insights into neurodevelopmental and neurodegenerative disorders (Zhao et al., 2024). To this end, a subset of the Baby Connectome Project (BCP) (Howell et al., 2019) dataset was employed, comprising 370 longitudinal scans from 181 healthy term-born infants (83 males and 98 females) aged 0–800 days, with 46 infants contributing two scans, 26 contributing three, and 28 contributing four or more. The age distribution is relatively balanced: 3M (0–144 days): 54; 6M (145–223 days): 54; 9M (224–318 days): 59; 12M (319–410 days): 53; 18M (411–591 days): 82; and 24M (592–800 days): 68. Images were acquired at the University of North Carolina (UNC) at Chapel Hill and the University of Minnesota using a Siemens 3T Prisma MRI scanner equipped with a Siemens 32-channel head coil. Structural MRI (T1w and T2w) and resting-state functional MRI (rs-fMRI) data were used in this work. The T1w images (208 sagittal slices) were acquired with TR/TE = 2400/2.24 ms, flip angle = 8°, acquisition matrix = 320 × 320, and resolution = 0.8 × 0.8 × 0.8 mm^3. The T2w images (208 sagittal slices) were acquired with TR/TE = 3200/564 ms, flip angle = VAR, and resolution = 0.8 × 0.8 × 0.8 mm^3. For rs-fMRI, a blood oxygenation level-dependent (BOLD) contrast sensitive gradient echo echo-planar sequence was used with the following parameters: repetition time = 800 ms, echo time = 37 ms, flip angle = 80°, field of view = 208 × 208 mm, 72 axial slices per volume, resolution = 2 × 2 × 2 mm^3, and total volumes = 420. The rs-fMRI scans included both AP and PA phase encoding directions to correct geometric distortions.

The dataset is split at the subject level into a training set of 148 subjects with 299 longitudinal scans and a test set of 33 subjects with 71 longitudinal scans. We construct the group-level parcellation on the training set and compute the evaluation metrics on the test set.

Image Preprocessing. T1w and T2w images were processed using a structural pipeline specifically designed for infants (Wang et al., 2023b; Wu et al., 2019; Wang et al., 2018b). The preprocessing pipeline comprised the following major steps: 1) intensity inhomogeneity correction on both T1w and T2w images (Sled et al., 1998); 2) linear registration of each T2w image to its corresponding T1w image using FLIRT in FSL (Smith et al., 2004); 3) deep learning-based automatic skull stripping, removing the skull, brainstem, and cerebellum (Zhang et al., 2019); 4) deep learning-based segmentation of each brain into white matter (WM), gray matter (GM), and cerebrospinal fluid (CSF) using T1w and T2w images (Wang et al., 2018b); and 5) separation of the segmented brain into left and right hemispheres, with subcortical regions filled as white matter.

rs-fMRI data were processed using a functional pipeline tailored for infants (Jiang et al., 2019; Zhou et al., 2019). The preprocessing pipeline included the following major steps: 1) head motion correction using FSL (Jenkinson et al., 2002); 2) EPI geometric distortion correction for AP and PA scans using FSL topup (Smith et al., 2004), followed by visual inspection to verify the correction results; 3) boundary-based registration of rs-fMRI scans to the corresponding T1w image (Greve & Fischl, 2009), where all transformations and deformation fields were combined and applied via a one-step resampling strategy to resample the rs-fMRI data back to the native space; 4) conservative high-pass filtering with a sigma of 1000 s to remove linear trends, followed by individual independent component analysis (ICA) using FSL MELODIC (Beckmann & Smith, 2004) to decompose each preprocessed rs-fMRI dataset into 150 components; and 5) deep learning-based automatic noise component identification to detect and remove non-signal components for cleaning the rs-fMRI data (Kam et al., 2019).

Cortical Surface Reconstruction and Registration. Cortical surface reconstruction involved four major steps: 1) automatically identifying and correcting white matter topological defects with a learning-based approach (Sun et al., 2019); 2) triangulating the corrected white matter to reconstruct each hemisphere's inner cortical surface; 3) deforming the inner surface to the gray matter/cerebrospinal fluid interface while preventing mesh self-intersections, thereby constructing each

Table 6: Quantitative comparison on evaluation metrics. ▲ denotes machine learning, ■ denotes deep learning, and ♠ denotes the proposed method. **Bold** denotes the best value in each column; ↑ / ↓ indicate higher / lower is better; * denotes that SABER significantly outperforms it by paired t-test ($p < 0.05$).

Methods	Parcellation homogeneity		Silhouette coefficient		Anatomical alignment	Spatial continuity
	Functional↑	Structural↑	Functional↑	Structural↑	Mean↓	Parcellation↑
▲ MV-Kmeans-200	0.7229 (0.0619)*	0.4716 (0.0384)*	0.4730 (0.0821)*	0.2766 (0.0250)*	6.6177 (2.6774)*	0.8073
▲ Mv-SC-200	0.7637 (0.0556)*	0.4721 (0.0406)*	0.5027 (0.0818)*	0.2726 (0.0236)*	3.4634 (0.8113)*	0.9999
▲ CGL-200	0.7613 (0.0559)*	0.4639 (0.0418)*	0.5021 (0.0820)*	0.2672 (0.0226)*	3.1972 (0.8097)*	**0.9999**
■ RML-200	0.7532 (0.0602)*	0.4575 (0.0420)*	0.4285 (0.0812)*	0.2221 (0.0214)*	3.3530 (1.4183)*	0.9544
■ SparseMVC-200	0.6449 (0.0668)*	0.4347 (0.0423)*	0.3144 (0.0834)*	0.1945 (0.0221)*	7.3259 (2.6932)*	0.6778
■ MSDIB-200	0.7564 (0.0585)*	0.4820 (0.0409)*	0.4832 (0.0801)*	0.2706 (0.0250)*	4.4251 (3.7211)*	0.9494
♠ **SABER-200**	**0.7784 (0.0566)**	**0.4831 (0.0387)**	**0.5169 (0.0789)**	**0.2999 (0.0274)**	**2.8004 (0.4181)**	0.9978
▲ MV-Kmeans-300	0.7672 (0.0566)*	0.4972 (0.0368)*	0.4895 (0.0808)*	0.2828 (0.0248)*	5.8141 (1.7891)*	0.8296
▲ Mv-SC-300	0.8054 (0.0483)*	0.4983 (0.0384)*	0.5365 (0.0710)*	0.2869 (0.0228)*	2.6262 (0.3764)*	**0.9996**
▲ CGL-300	0.8033 (0.0490)*	0.5024 (0.0403)*	0.5295 (0.0712)*	0.2827 (0.0232)*	2.6596 (0.4014)*	0.9988
■ RML-300	0.7850 (0.0544)*	0.4855 (0.0396)*	0.4351 (0.0791)*	0.2210 (0.0206)*	3.1309 (1.7445)*	0.9554
■ SparseMVC-300	0.6605 (0.0633)*	0.4562 (0.0438)*	0.2942 (0.0824)*	0.1723 (0.0170)*	6.4107 (4.1715)*	0.6247
■ MSDIB-300	0.7912 (0.0537)*	0.5134 (0.0389)*	0.5038 (0.0787)*	0.2796 (0.0262)*	3.6663 (2.0994)*	0.9533
♠ **SABER-300**	**0.8157 (0.0509)**	**0.5157 (0.0380)**	**0.5414 (0.0759)**	**0.3124 (0.0288)**	**2.2685 (0.3797)**	0.9992
▲ MV-Kmeans-500	0.8177 (0.0490)*	0.5445 (0.0362)*	0.5150 (0.0745)*	0.2945 (0.0246)*	4.0386 (1.2921)*	0.8672
▲ Mv-SC-500	0.8421 (0.0341)*	0.5491 (0.0280)*	0.5550 (0.0632)*	0.3043 (0.0233)*	2.3581 (0.4511)*	0.9958
▲ CGL-500	0.8425 (0.0414)*	0.5470 (0.0373)*	0.5572 (0.0671)*	0.3060 (0.0231)*	2.3526 (0.3780)*	0.9968
■ RML-500	0.8219 (0.0475)*	0.5278 (0.0379)*	0.4503 (0.0717)*	0.2340 (0.0197)*	2.1052 (0.6817)*	0.9491
■ SparseMVC-500	0.7010 (0.0575)*	0.4860 (0.0389)*	0.2900 (0.0808)*	0.1694 (0.0152)*	7.1416 (5.3978)*	0.5797
■ MSDIB-500	0.8283 (0.0456)*	0.5584 (0.0367)*	0.5234 (0.0709)*	0.2859 (0.0237)*	2.1108 (0.7647)*	0.9408
♠ **SABER-500**	**0.8566 (0.0432)**	**0.5608 (0.0363)**	**0.5758 (0.0697)**	**0.3310 (0.0280)**	**1.9963 (0.1717)**	**0.9993**

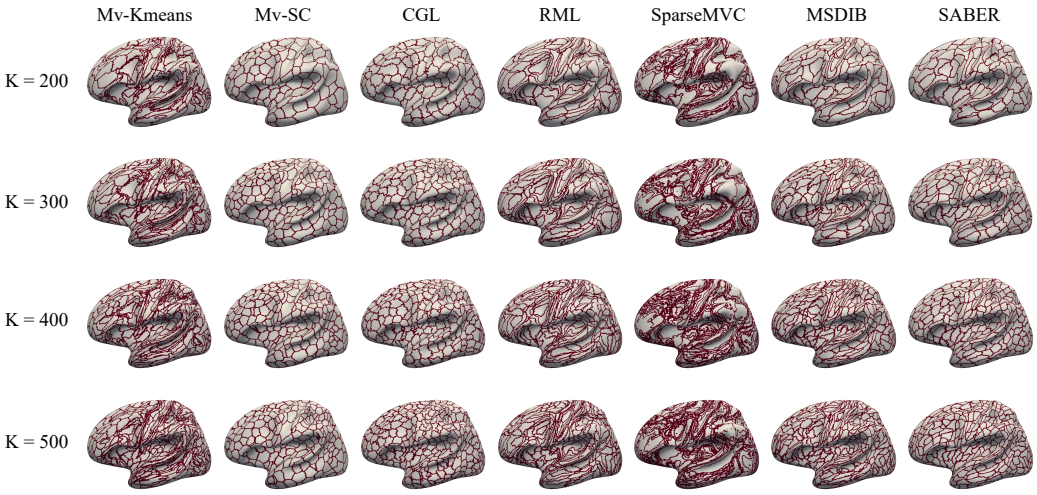


Figure 4: Qualitative visualization results of group-level parcellations from different methods at 200, 300, 400, and 500 parcels.

hemisphere's outer cortical surface (Li et al., 2012); and 4) further smoothing and inflating the inner surface (Fischl et al., 1999), mapping it to a standard sphere via FreeSurfer (Fischl, 2012), and extracting four cortical attributes: thickness, surface area, myelination, and curvature.

Cortical surface registration followed a longitudinal group-wise registration strategy (Li et al., 2015), comprising the following steps: 1) performing intra-subject co-registration of the longitudinal cortical surfaces for each subject; 2) averaging the longitudinal cortical attribute maps to obtain a subject-specific average surface map; 3) registering each subject's average surface map to the UNC 4D infant surface template (Wu et al., 2019) using Spherical Demons (Yeo et al., 2009); 4) concatenating the deformation fields from the individual surface to the subject-specific average surface, and from the

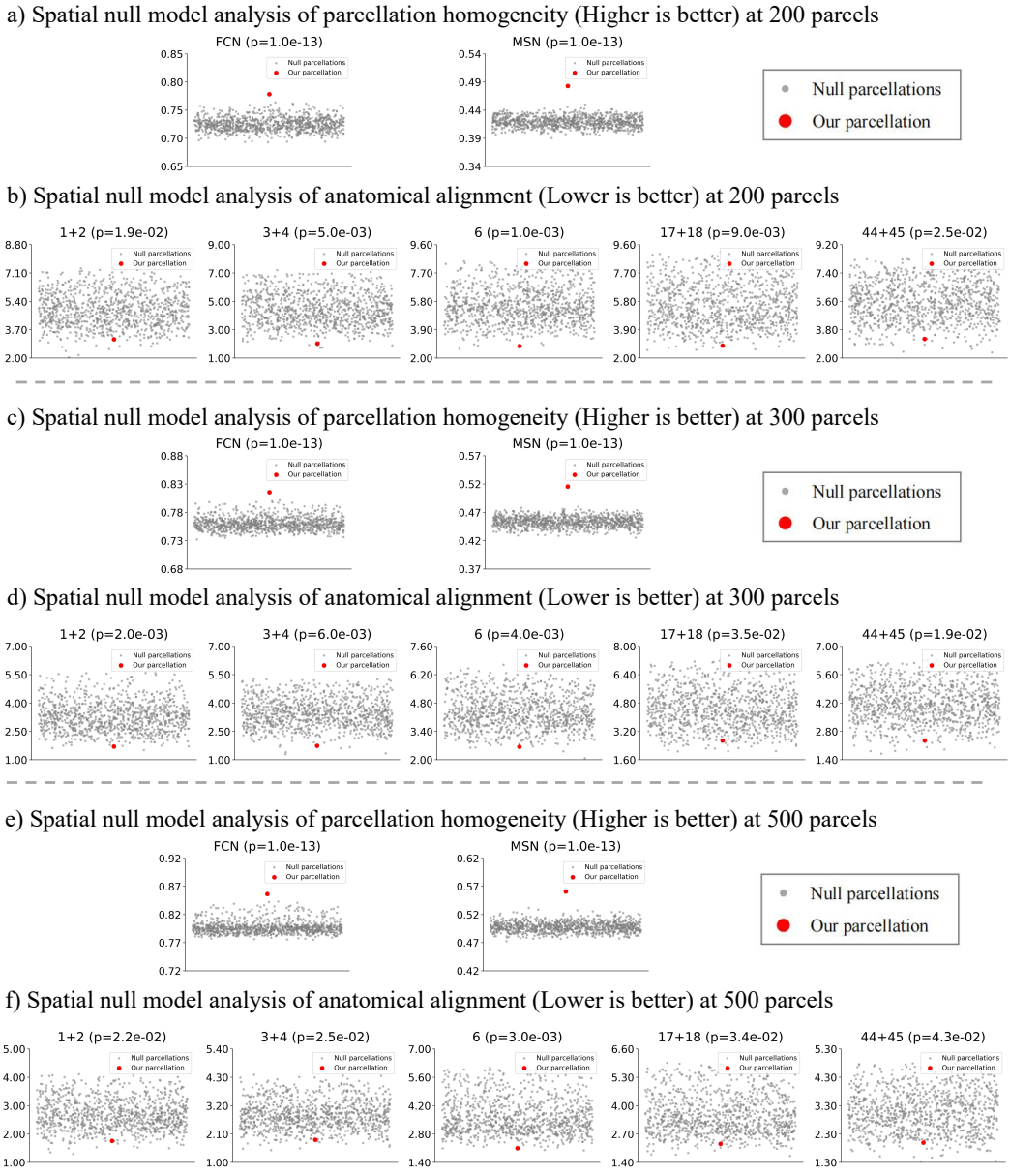


Figure 5: Spatial null model analysis at different resolutions. a), c), e) and b), d), f) compare our parcellation with 1,000 random parcellations in terms of parcellation homogeneity and anatomical alignment, respectively.

average surface to the UNC template surface, to align every longitudinal surface to the UNC template surface, thereby achieving cross-subject and cross-time-point alignment; and 5) resampling the rs-fMRI BOLD signals onto the middle cortical surface with 32,492 vertices using the inner and outer surfaces as constraints (Glasser et al., 2013), followed by spatial smoothing of the time courses on the middle cortical surface using a small Gaussian kernel ($\sigma = 2.55\text{mm}$). The four cortical attributes were also resampled to the 32k vertices.

Construction of MSN and FCN. Conventionally, MSN and FCN are derived from the Pearson correlation coefficients across vertices based on four cortical attributes and BOLD signals, respectively. However, explicitly constructing 32k $\times$ 32k feature matrices for model training would lead to memory overflow. To circumvent this issue, we adopt an anchor graph strategy (Yang et al., 2022), which computes the Pearson correlation between vertices and selected anchors to construct

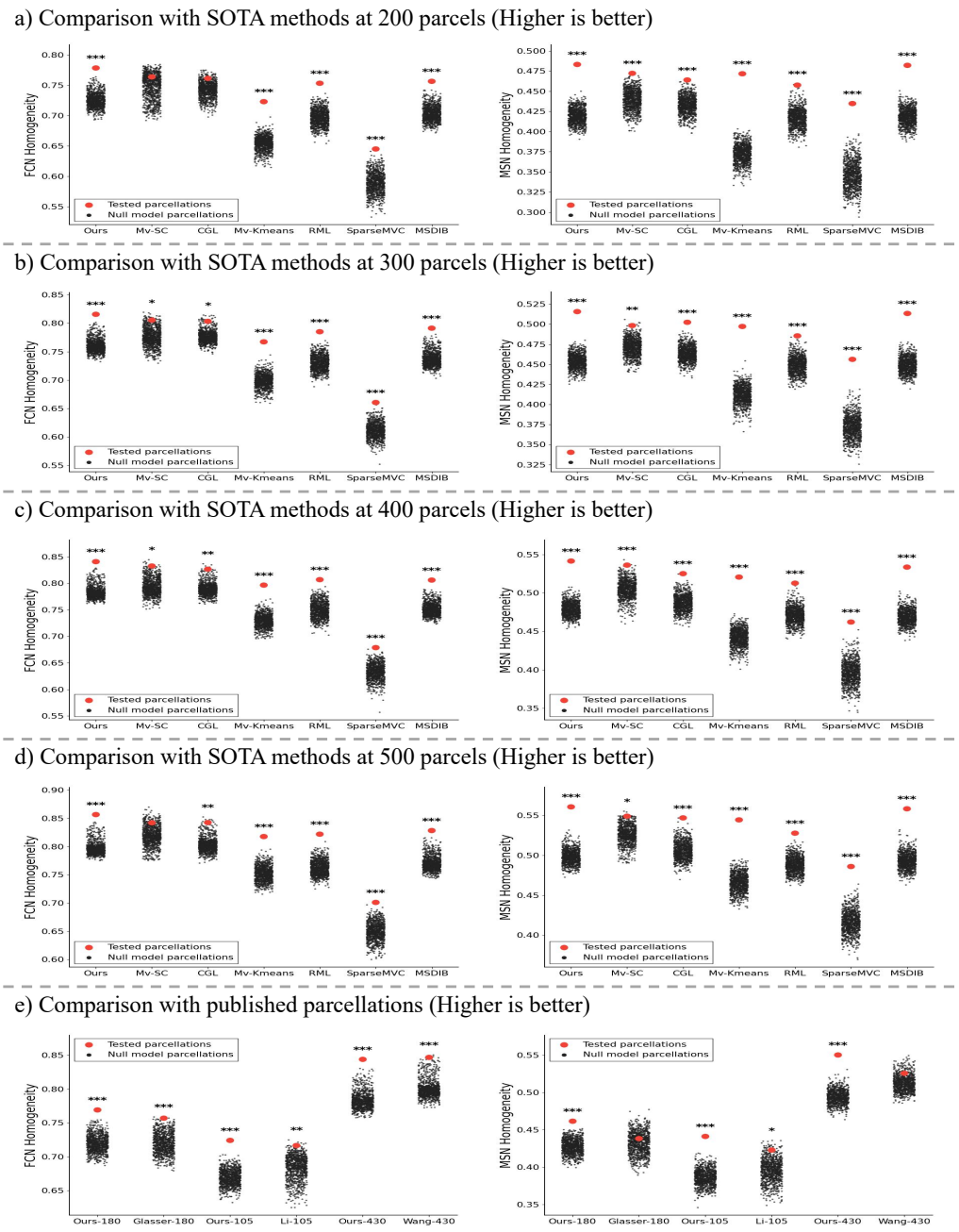


Figure 6: Comparison with spatial null models across methods and published parcellations. a), b), c), d) show comparisons of SABER and SOTA methods against spatial null models in terms of parcellation homogeneity. e) shows comparisons of SABER and published parcellations against spatial null models in terms of parcellation homogeneity.

dimensionally reduced MSN and FCN with minimal information loss. Specifically, we first compute the 32k $\times$ 32k MSN and FCN for all subjects. For the FCN, we calculate the Pearson correlation between the BOLD signals of all vertex pairs. For the MSN, we concatenate the four cortical attributes, apply Z-score normalization, and compute the Pearson correlation between all vertex pairs. After obtaining the MSN and FCN for all subjects, we average only the data from subjects in the training set to construct the group-averaged 32k $\times$ 32k FCN and MSN matrices. Spectral clustering is subsequently performed on these group-averaged matrices to generate 300 pre-parcellated parcels for MSN and FCN, respectively. Using these 300 parcels as anchors, for each subject in both the

Table 7: Quantitative comparison of parcellation homogeneity with null models. ▲ denotes machine learning, ■ denotes deep learning, and ♠ denotes the proposed method. **Bold** denotes the best value in each column; ↑ / ↓ indicate higher / lower is better.

Methods	Mean + SD homogeneity		Number of null model iterations with worse homogeneity		Z-score relative to null model	
	Functional↑	Structural↑	Functional↑	Structural↑	Functional↑	Structural↑
▲ MV-Kmeans-200	0.7229 (0.0619)	0.4716 (0.0384)	1000 ($p < 0.001$)	1000 ($p < 0.001$)	**5.9209**	**8.7089**
▲ Mv-SC-200	0.7637 (0.0556)	0.4721 (0.0406)	737 ($p = 0.263$)	1000 ($p < 0.001$)	0.7678	2.6705
▲ CGL-200	0.7613 (0.0559)	0.4639 (0.0418)	939 ($p = 0.061$)	1000 ($p < 0.001$)	1.4690	2.9484
■ RML-200	0.7532 (0.0602)	0.4575 (0.0420)	1000 ($p < 0.001$)	998 ($p = 0.002$)	4.2999	3.6494
■ SparseMVC-200	0.6449 (0.0668)	0.4347 (0.0423)	1000 ($p < 0.001$)	1000 ($p < 0.001$)	3.2596	5.3808
■ MSDIB-200	0.7564 (0.0585)	0.4820 (0.0409)	1000 ($p < 0.001$)	1000 ($p < 0.001$)	4.3030	6.2697
♠ **SABER-200**	**0.7784 (0.0566)**	**0.4831 (0.0387)**	**1000 ($p < 0.001$)**	**1000 ($p < 0.001$)**	4.5618	6.8591
▲ MV-Kmeans-300	0.7672 (0.0566)	0.4972 (0.0368)	1000 ($p < 0.001$)	1000 ($p < 0.001$)	**5.6631**	**7.5808**
▲ Mv-SC-300	0.8054 (0.0483)	0.4983 (0.0384)	966 ($p = 0.034$)	995 ($p = 0.005$)	1.8254	2.5485
▲ CGL-300	0.8033 (0.0490)	0.5024 (0.0403)	976 ($p = 0.024$)	1000 ($p < 0.001$)	2.5832	4.2965
■ RML-300	0.7850 (0.0544)	0.4855 (0.0396)	1000 ($p < 0.001$)	1000 ($p < 0.001$)	4.1296	3.4584
■ SparseMVC-300	0.6605 (0.0633)	0.4562 (0.0438)	1000 ($p < 0.001$)	1000 ($p < 0.001$)	3.6435	6.0857
■ MSDIB-300	0.7912 (0.0537)	0.5134 (0.0389)	1000 ($p < 0.001$)	1000 ($p < 0.001$)	4.4478	6.7846
♠ **SABER-300**	**0.8157 (0.0509)**	**0.5157 (0.0380)**	**1000 ($p < 0.001$)**	**1000 ($p < 0.001$)**	4.8720	6.9517
▲ MV-Kmeans-400	0.7966 (0.0532)	0.5206 (0.0364)	1000 ($p < 0.001$)	1000 ($p < 0.001$)	**5.6307**	7.1038
▲ Mv-SC-400	0.8326 (0.0427)	0.5362 (0.0350)	990 ($p = 0.010$)	998 ($p = 0.002$)	2.4250	2.5852
▲ CGL-400	0.8266 (0.0448)	0.5251 (0.0389)	995 ($p = 0.005$)	1000 ($p < 0.001$)	3.3267	3.7311
■ RML-400	0.8069 (0.0502)	0.5127 (0.0386)	1000 ($p < 0.001$)	1000 ($p < 0.001$)	3.7843	3.9868
■ SparseMVC-400	0.6788 (0.0604)	0.4621 (0.0415)	1000 ($p < 0.001$)	1000 ($p < 0.001$)	3.0290	4.4749
■ MSDIB-400	0.8061 (0.0501)	0.5336 (0.0386)	1000 ($p < 0.001$)	1000 ($p < 0.001$)	4.6779	6.9077
♠ **SABER-400**	**0.8408 (0.0460)**	**0.5415 (0.0368)**	**1000 ($p < 0.001$)**	**1000 ($p < 0.001$)**	5.2095	**7.1586**
▲ MV-Kmeans-500	0.8177 (0.0490)	0.5445 (0.0362)	1000 ($p < 0.001$)	1000 ($p < 0.001$)	4.9587	**7.0668**
▲ Mv-SC-500	0.8421 (0.0341)	0.5491 (0.0280)	905 ($p = 0.095$)	988 ($p = 0.012$)	1.3722	1.9218
▲ CGL-500	0.8425 (0.0414)	0.5470 (0.0373)	992 ($p = 0.008$)	1000 ($p < 0.001$)	3.2794	3.7500
■ RML-500	0.8219 (0.0475)	0.5278 (0.0379)	1000 ($p < 0.001$)	1000 ($p < 0.001$)	4.6955	4.0666
■ SparseMVC-500	0.7010 (0.0575)	0.4860 (0.0389)	1000 ($p < 0.001$)	1000 ($p < 0.001$)	3.2489	5.1108
■ MSDIB-500	0.8283 (0.0456)	0.5584 (0.0367)	1000 ($p < 0.001$)	1000 ($p < 0.001$)	4.7045	6.8002
♠ **SABER-500**	**0.8566 (0.0432)**	**0.5608 (0.0363)**	**1000 ($p < 0.001$)**	**1000 ($p < 0.001$)**	**5.4225**	7.0667
♣ Glasser-180	0.7569 (0.0589)	0.4381 (0.0428)	998 ($p = 0.002$)	633 ($p = 0.3670$)	2.4691	0.4334
♠ **SABER-180**	**0.7691 (0.0588)**	**0.4614 (0.0401)**	**1000 ($p < 0.001$)**	**1000 ($p < 0.001$)**	**4.0515**	**3.5216**
♣ Li-105	0.7165 (0.0623)	0.4227 (0.0411)	992 ($p = 0.008$)	963 ($p = 0.037$)	1.8349	1.7985
♠ **SABER-105**	**0.7241 (0.0643)**	**0.4410 (0.0409)**	**1000 ($p < 0.001$)**	**1000 ($p < 0.001$)**	**4.7199**	**5.1468**
♣ Wang-430	**0.8467 (0.0439)**	0.5256 (0.0400)	997 ($p = 0.003$)	891 ($p = 0.1090$)	3.4355	1.2374
♠ **SABER-430**	0.8438 (0.0454)	**0.5502 (0.0369)**	**1000 ($p < 0.001$)**	**1000 ($p < 0.001$)**	**4.9858**	**6.0914**

training and test sets, we compute the Pearson correlation between the vertex-level features and the average feature vector of each parcel, resulting in the individual 32k$\times$300 matrices X^n_{MSN} and X^n_{FCN}. The choice of 300 anchors is carefully selected to balance maximal dimensionality reduction with minimal information loss. Our empirical evaluations reveal that vertex-level similarity matrices derived from the 32k $\times$ 300 X^n_{MSN} and X^n_{FCN} are highly similar to those from the 32k $\times$ 32k matrices, with an average numerical difference of only about 0.021, confirming negligible information loss. Finally, the individual 32k $\times$ 300 X^n_{MSN} and X^n_{FCN} from the training set are averaged to obtain the group-level 32k $\times$ 300 raw feature matrices X_{MSN} and X_{FCN}, which are utilized as model inputs. Meanwhile, the individual 32k$\times$300 X^n_{MSN} and X^n_{FCN} from the training set are also used to optimize the cross-subject and cross-modality homogeneity of the group-level parcellation map.

B Appendix B

B.1 Evaluation Metrics

Parcellation Homogeneity. A representative group-level multi-modal cortical parcellation map should exhibit high intra-parcel uniformity of connectivity profiles across both subjects and modalities. To quantify this property, we use the group-level parcellation as a template and compute, for

Table 8: Quantitative comparison on downstream tasks. ▲ denotes machine learning, ■ denotes deep learning, and ♠ denotes the proposed method. **Bold** denotes the best value in each column; ↑ or ↓ indicates whether higher or lower values are better; * indicates that SABER significantly outperforms it by paired t-test ($p < 0.05$).

Methods	Gender classification	Age prediction	Cognitive prediction	Longitudinal individual identification	
	ACC↑	MAE↓	MAE↓	ACC (Functional)↑	ACC (Structural)↑
▲ MV-Kmeans-200	0.8443 (0.0171)*	78.8738 (1.3655)*	4.7265 (0.6062)*	0.2980 (0.0339)*	0.8855 (0.0276)*
▲ Mv-SC-200	0.8308 (0.0149)*	80.4340 (1.6402)*	4.6460 (0.4954)	0.3771 (0.0251)*	0.9086 (0.0211)
▲ CGL-200	0.8323 (0.0171)*	79.2412 (2.0897)*	4.6397 (0.4809)*	0.3621 (0.0314)*	0.8884 (0.0207)*
■ RML-200	0.8250 (0.0132)*	80.0858 (1.7856)*	4.7476 (0.4414)*	0.3677 (0.0243)*	0.8913 (0.0172)*
■ SparseMVC-200	0.8453 (0.0188)*	89.4389 (1.8987)*	4.9228 (0.5213)*	0.3064 (0.0298)*	0.8730 (0.0214)*
■ MSDIB-200	0.8271 (0.0129)*	77.4652 (2.1180)*	4.7311 (0.4747)*	0.3834 (0.0301)*	0.9001 (0.0183)*
♠ **SABER-200**	**0.8750 (0.0126)**	**75.1561 (1.3272)**	**4.5062 (0.5335)**	**0.4102 (0.0297)**	**0.9115 (0.0205)**
▲ MV-Kmeans-300	0.8456 (0.0176)*	77.5387 (2.0433)*	4.4828 (0.4596)*	0.3653 (0.0280)*	0.9173 (0.0167)*
▲ Mv-SC-300	0.8446 (0.0190)*	79.8670 (1.2571)*	4.3945 (0.4175)*	0.4218 (0.0334)	0.9388 (0.0122)
▲ CGL-300	0.8571 (0.0172)*	79.3930 (1.5339)*	4.3586 (0.4717)	0.4262 (0.0345)	0.9336 (0.0145)*
■ RML-300	0.8266 (0.0195)*	78.0493 (1.0151)*	4.4840 (0.4558)*	0.3978 (0.0280)*	0.9317 (0.0157)*
■ SparseMVC-300	0.8191 (0.0182)*	88.8518 (1.2082)*	4.7803 (0.5705)*	0.3329 (0.0117)*	0.8971 (0.0210)*
■ MSDIB-300	0.8275 (0.0126)*	78.0419 (1.4073)*	4.5537 (0.4933)*	0.3930 (0.0351)*	0.9182 (0.0162)*
♠ **SABER-300**	**0.8767 (0.0156)**	**76.1410 (1.6405)**	**4.2854 (0.4374)**	**0.4417 (0.0243)**	**0.9484 (0.0115)**
▲ MV-Kmeans-500	0.8728 (0.0164)	79.4299 (1.9093)*	3.9191 (0.3904)*	0.4076 (0.0232)*	0.9644 (0.0122)*
▲ Mv-SC-500	0.8678 (0.0198)*	85.5745 (2.1073)*	3.9906 (0.4076)*	**0.4662 (0.0165)**	0.9764 (0.0151)*
▲ CGL-500	0.8558 (0.0229)*	83.3472 (1.6910)*	3.9614 (0.4216)*	0.4546 (0.0276)*	0.9750 (0.0129)*
■ RML-500	0.8570 (0.0162)*	80.1032 (0.7244)*	4.1877 (0.3686)*	0.4076 (0.0372)*	0.9605 (0.0145)*
■ SparseMVC-500	0.8514 (0.0133)*	89.9762 (1.5970)*	4.2336 (0.4504)*	0.3658 (0.0310)*	0.9288 (0.0215)*
■ MSDIB-500	0.8503 (0.0182)*	78.4164 (1.6902)*	4.2068 (0.3835)*	0.4336 (0.0256)*	0.9509 (0.0151)*
♠ **SABER-500**	**0.8825 (0.0128)**	**77.2532 (1.4079)**	**3.9438 (0.3833)**	0.4648 (0.0384)	**0.9894 (0.0103)**

each subject and modality, the mean Pearson correlation between the connectivity profiles of all vertex pairs within each parcel (Schaefer et al., 2018). The resulting values are then averaged across all parcels to obtain the parcellation homogeneity score, formulated as:

$$Homo_m^n = \sum_{k=1}^{K} \frac{\sum_{i,j \in V, i \neq j} Y_{i,k}\, Y_{j,k} Prr\left(X_{m,i}^n, X_{m,j}^n\right)}{K \sum_{i,j \in V, i \neq j} Y_{i,k}\, Y_{j,k}} \tag{10}$$

where $m = \{\text{MSN}, \text{FCN}\}$; K denotes the number of parcels; V denotes cortical vertices; $Y_{i,k} \in \{0, 1\}$ indicates whether vertex i is assigned to parcel k in the group-level parcellation map; $X_{m,i}^n$ denotes the connectivity profile of vertex i for subject n under modality m; and $Prr(\cdot)$ denotes the Pearson correlation calculation. Finally, we report the mean and standard deviation of cross-subject and cross-modality parcellation homogeneity on the test set. Higher scores indicate that the group-level parcellation consistently exhibits strong intra-parcel uniformity across subjects, reflecting better cross-subject adaptability.

Silhouette Coefficient. The silhouette coefficient is commonly used to jointly assess the compactness and separability of a parcellation, reflecting how well each vertex is assigned to its corresponding parcel (Zhu et al., 2025a). Notably, following Zhu et al. (2025a), we define the inter-parcel distance in the silhouette coefficient as the average distance from a vertex to its neighboring parcels rather than the minimum distance to all other parcels. This formulation reflects whether assigning a vertex to its current parcel or neighboring parcels provides a better fit. We avoid using the minimum distance to all other parcels because it can result in systematically lower silhouette coefficients due to strong long-range connections, thereby reducing their discriminative power for comparing parcellations. It is defined as follows:

$$SC_m^n = \frac{1}{|V|} \sum_{i \in V} \frac{b_{m,i}^n - a_{m,i}^n}{max\{a_{m,i}^n, b_{m,i}^n\}} \tag{11}$$

Given a parcellation $U = \{U_1, ..., U_K\}$, where U_K is parcel K, $a_{m,i}^n$ and $b_{m,i}^n$ denote the intra- and inter-parcel distances for vertex i of subject n under modality m, with $i \in U_k$:

$$a_{m,i}^n = \frac{1}{n_k - 1} \sum_{j \in U_k, i \neq j} (1 - Prr(X_{m,i}^n, X_{m,j}^n)) \tag{12}$$

Table 9: Quantitative ablation results on evaluation metrics. **Bold** denotes the best value in each column; ↑ or ↓ indicates whether higher or lower values are better.

Ablation Study						Parcellation homo		Silhouette coefficient		Anatomical alignment	Spatial continuity
$\mathcal{L}_{\text{Homo}}$	$\mathcal{L}_{\text{G_Homo}}$	$\mathcal{L}_{\text{DG_Homo}}$	$\mathcal{L}_{\text{Boundary}}$	ML	VL	Func↑	Struct↑	Func↑	Struct↑	Mean↓	Parcel↑
Number of Clusters Set to 200											
✓				✓	✓	0.7478 (0.0599)	**0.4836 (0.0415)**	0.4958 (0.0812)	0.2972 (0.0274)	3.2956 (0.3677)	0.9291
	✓			✓	✓	0.7792 (0.0571)	0.4824 (0.0415)	0.5093 (0.0793)	0.2939 (0.0286)	2.9700 (0.6962)	0.9859
		✓		✓	✓	**0.7839 (0.0561)**	0.4712 (0.0414)	**0.5209 (0.0781)**	0.2869 (0.0270)	2.9906 (0.8494)	0.9965
		✓	✓	✓	✓	0.7784 (0.0566)	0.4831 (0.0387)	0.5169 (0.0789)	**0.2999 (0.0274)**	**2.8004 (0.4181)**	0.9978
		✓	✓		✓	0.7702 (0.0557)	0.4735 (0.0403)	0.5078 (0.0778)	0.2888 (0.0258)	2.9647 (0.6657)	**0.9989**
		✓	✓	✓		0.7694 (0.0568)	0.4727 (0.0411)	0.5060 (0.0789)	0.2885 (0.0260)	2.9716 (0.8796)	0.9976
		✓	✓			0.7689 (0.0567)	0.4693 (0.0411)	0.5052 (0.0788)	0.2867 (0.0259)	3.0832 (0.7861)	0.9973
Number of Clusters Set to 300											
✓				✓	✓	0.7969 (0.0534)	0.5143 (0.0383)	0.5190 (0.0768)	0.3024 (0.0270)	2.9047 (0.8557)	0.9535
	✓			✓	✓	0.8159 (0.0512)	**0.5165 (0.0383)**	0.5375 (0.0760)	**0.3156 (0.0270)**	2.4019 (0.3768)	0.9897
		✓		✓	✓	**0.8220 (0.0491)**	0.5078 (0.0383)	**0.5491 (0.0752)**	0.3035 (0.0280)	2.5359 (0.3394)	**0.9998**
		✓	✓	✓	✓	0.8157 (0.0509)	0.5157 (0.0380)	0.5414 (0.0759)	0.3124 (0.0288)	**2.2685 (0.3797)**	0.9992
		✓	✓		✓	0.8172 (0.0512)	0.5067 (0.0387)	0.5392 (0.0760)	0.2958 (0.0243)	2.4981 (0.2544)	0.9991
		✓	✓	✓		0.8159 (0.0506)	0.5051 (0.0378)	0.5362 (0.0758)	0.2977 (0.0255)	2.5720 (0.6737)	0.9994
		✓	✓			0.8151 (0.0507)	0.5008 (0.0380)	0.5354 (0.0758)	0.2960 (0.0248)	2.6007 (0.9793)	0.9997
Number of Clusters Set to 500											
✓				✓	✓	0.8278 (0.0503)	0.5502 (0.0376)	0.5508 (0.0751)	0.3217 (0.0286)	2.3561 (0.4729)	0.9309
	✓			✓	✓	0.8562 (0.0457)	**0.5611 (0.0366)**	0.5756 (0.0709)	**0.3316 (0.0291)**	2.2342 (0.3223)	0.9872
		✓		✓	✓	**0.8593 (0.0468)**	0.5523 (0.0357)	**0.5771 (0.0711)**	0.3244 (0.0281)	2.2169 (0.4334)	**0.9996**
		✓	✓	✓	✓	0.8566 (0.0432)	0.5608 (0.0363)	0.5758 (0.0697)	0.3310 (0.0280)	**1.9963 (0.1717)**	0.9993
		✓	✓		✓	0.8508 (0.0442)	0.5521 (0.0356)	0.5682 (0.0703)	0.3241 (0.0280)	2.1893 (0.3452)	0.9994
		✓	✓	✓		0.8511 (0.0473)	0.5528 (0.0364)	0.5686 (0.0735)	0.3251 (0.0287)	2.1321 (0.2866)	0.9996
		✓	✓			0.8496 (0.0472)	0.5509 (0.0358)	0.5676 (0.0736)	0.3230 (0.0277)	2.2873 (0.3315)	0.9989

$$b_{m,i}^{n} = \frac{1}{R} \sum_{j \in \mathcal{N}(U_k)} \left(1 - \Pr\left(X_{m,i}^{n}, X_{m,j}^{n}\right)\right), \tag{13}$$

where n_k denotes the number of vertices in U_k; $\mathcal{N}(U_k)$ denotes the set of neighboring parcels of U_k and R denotes the total number of vertices in these neighboring parcels. Finally, we report the mean and standard deviation of cross-subject and cross-modality silhouette coefficient on the test set. Higher scores indicate that the group-level parcellation consistently exhibits better compactness and separability across subjects, reflecting the validity of the group-level parcellation.

Anatomical Alignment. Anatomical alignment is evaluated using the Hausdorff distance between architectonic boundaries of Brodmann areas and predicted group-level parcellation boundaries. Briefly, for each boundary vertex of each architectonic area, the geodesic distance to the closest parcellation boundary was computed. These distances were averaged across all boundary vertices of an architectonic area, yielding a Hausdorff distance per architectonic area (Schaefer et al., 2018). Finally, we report the mean and standard deviation of the Hausdorff distances across all Brodmann areas. Smaller scores indicate better agreement between parcellation and anatomical boundaries.

Table 10: Quantitative ablation results on the plug-and-play capability of $\mathcal{L}_{\text{DG_Homo}}$ and $\mathcal{L}_{\text{Boundary}}$. **Bold** denotes the best value in each column; ↑ / ↓ indicate higher / lower is better.

Ablation Study					Parcellation homo		Silhouette coefficient		Anatomical alignment	Spatial continuity
RML	SparseMVC	MSDIB	$\mathcal{L}_{\text{DG_Homo}}$	$\mathcal{L}_{\text{Boundary}}$	Func↑	Struct↑	Func↑	Struct↑	Mean↓	Parcel↑
Number of Clusters Set to 200										
✓					0.7532 (0.0602)	0.4575 (0.0420)	0.4285 (0.0812)	0.2221 (0.0214)	3.3530 (1.4183)	0.9544
✓			✓		**0.7786 (0.0566)**	0.4812 (0.0389)	**0.5148 (0.0787)**	0.2986 (0.0277)	2.9927 (0.6019)	**0.9973**
✓			✓	✓	0.7711 (0.0571)	**0.4823 (0.0386)**	0.5112 (0.0792)	**0.2991 (0.0274)**	**2.8362 (0.6324)**	0.9967
	✓				0.6449 (0.0668)	0.4347 (0.0423)	0.3144 (0.0834)	0.1945 (0.0221)	7.3259 (2.6932)	0.6778
	✓		✓		**0.7461 (0.0534)**	0.4434 (0.0436)	**0.4621 (0.0781)**	0.2458 (0.0241)	3.4856 (0.9749)	0.9924
	✓		✓	✓	0.7397 (0.0562)	**0.4502 (0.0481)**	0.4604 (0.0793)	**0.2465 (0.0247)**	**3.1947 (1.0227)**	**0.9962**
		✓			0.7564 (0.0585)	**0.4820 (0.0409)**	0.4832 (0.0801)	0.2706 (0.0250)	4.4251 (3.7211)	0.9494
		✓	✓		**0.7723 (0.0567)**	0.4781 (0.0411)	**0.5020 (0.0452)**	0.2737 (0.0234)	3.1658 (0.6108)	**0.9978**
		✓	✓	✓	0.7696 (0.5731)	0.4813 (0.0412)	0.4998 (0.0461)	**0.2745 (0.0238)**	**2.9453 (0.5483)**	0.9976
Number of Clusters Set to 300										
✓					0.7850 (0.0544)	0.4855 (0.0396)	0.4351 (0.0791)	0.2210 (0.0206)	3.1309 (1.7445)	0.9554
✓			✓		**0.8144 (0.0507)**	0.5117 (0.0380)	**0.5335 (0.0757)**	0.2991 (0.0285)	2.6612 (0.6577)	0.9931
✓			✓	✓	0.8108 (0.0514)	**0.5154 (0.0386)**	0.5243 (0.0763)	**0.3046 (0.0287)**	**2.4451 (0.5864)**	**0.9942**
	✓				0.6605 (0.0633)	0.4562 (0.0438)	0.2942 (0.0824)	0.1723 (0.0170)	6.4107 (4.1715)	0.6247
	✓		✓		**0.7678 (0.0539)**	0.4622 (0.0369)	**0.4664 (0.0753)**	0.2589 (0.0212)	2.6602 (0.8005)	0.9937
	✓		✓	✓	0.7598 (0.0528)	**0.4693 (0.0381)**	0.4621 (0.0724)	**0.2631 (0.0243)**	**2.4189 (0.9325)**	0.9943
		✓			0.7912 (0.0537)	**0.5134 (0.0389)**	0.5038 (0.0787)	0.2796 (0.0262)	3.6663 (2.0994)	0.9533
		✓	✓		**0.8085 (0.0510)**	0.5100 (0.0399)	**0.5284 (0.0426)**	0.2878 (0.0232)	2.5291 (0.4067)	0.9982
		✓	✓	✓	0.8038 (0.0524)	0.5128 (0.0406)	0.5231 (0.0441)	**0.2913 (0.0246)**	**2.3867 (0.4963)**	**0.9989**
Number of Clusters Set to 500										
✓					0.8219 (0.0475)	0.5278 (0.0379)	0.4503 (0.0717)	0.2340 (0.0197)	2.1052 (0.6817)	0.9491
✓			✓		**0.8583 (0.0426)**	0.5457 (0.0361)	**0.5736 (0.0683)**	0.3172 (0.0279)	2.1243 (0.5326)	**0.9916**
✓			✓	✓	0.8524 (0.0418)	**0.5571 (0.0358)**	0.5678 (0.0672)	**0.3230 (0.0271)**	**1.9973 (0.3208)**	0.9911
	✓				0.7010 (0.0575)	0.4860 (0.0389)	0.2900 (0.0808)	0.1694 (0.0152)	7.1416 (5.3978)	0.5797
	✓		✓		**0.7996 (0.0471)**	0.4886 (0.0352)	**0.4796 (0.0711)**	0.2787 (0.0205)	2.3628 (0.2821)	0.9928
	✓		✓	✓	0.7945 (0.0486)	**0.4932 (0.0364)**	0.4712 (0.0721)	**0.2806 (0.0213)**	**2.1457 (0.4221)**	**0.9943**
		✓			0.8283 (0.0456)	**0.5584 (0.0367)**	0.5234 (0.0709)	0.2859 (0.0237)	2.1108 (0.7647)	0.9408
		✓	✓		**0.8534 (0.0429)**	0.5596 (0.0375)	**0.5695 (0.0366)**	0.3050 (0.0232)	2.0387 (0.4541)	0.9990
		✓	✓	✓	0.8503 (0.0426)	**0.5611 (0.0368)**	0.5652 (0.0357)	**0.3107 (0.0227)**	**1.9686 (0.4489)**	**0.9993**

Spatial Continuity. This metric evaluates the ratio of the largest connected component (LCC) to the total size of each parcel in the group-level parcellation $U = \{U_1, ..., U_K\}$, defined as:

$$Spa_cont = \frac{1}{K} \sum_{k=1}^{K} \frac{|LCC(U_k)|}{|U_k|} \tag{14}$$

Table 11: Quantitative ablation results on downstream tasks. **Bold** denotes the best value in each column; ↑ / ↓ indicate higher / lower is better; * denotes that introducing $\mathcal{L}_{\text{DG_Homo}}$ significantly outperforms it by paired t-test ($p < 0.05$).

Ablation Study					Gender classification	Age prediction	Cognitive prediction
SABER	RML	SparseMVC	MSDIB	$\mathcal{L}_{\text{DG_Homo}}$	ACC↑	MAE↓	MAE↓
Number of Clusters Set to 200							
✓					0.8448 (0.0102)*	77.3210 (1.2489)*	4.6363 (0.4456)
✓				✓	**0.8750 (0.0126)**	**75.1561 (1.3272)**	**4.5062 (0.5335)**
	✓				0.8250 (0.0132)*	80.0858 (1.7856)*	4.7476 (0.4414)*
	✓			✓	**0.8507 (0.0161)**	**77.9930 (2.2851)**	**4.6218 (0.4344)**
		✓			0.8453 (0.0188)*	89.4389 (1.8987)*	4.9228 (0.5213)*
		✓		✓	**0.8576 (0.0175)**	**81.8641 (1.5390)**	**4.6428 (0.4857)**
			✓		0.8271 (0.0129)*	77.4652 (2.1180)	4.7311 (0.4747)
			✓	✓	**0.8537 (0.0112)**	**77.1216 (1.6135)**	**4.6327 (0.5380)**
Number of Clusters Set to 300							
✓					0.8583 (0.0103)*	77.2072 (1.8956)*	4.3336 (0.4190)
✓				✓	**0.8767 (0.0156)**	**76.1410 (1.6405)**	**4.2854 (0.4374)**
	✓				0.8266 (0.0195)*	78.0493 (1.0151)*	4.4840 (0.4558)*
	✓			✓	**0.8562 (0.0153)**	**76.8905 (1.4372)**	**4.2921 (0.5042)**
		✓			0.8191 (0.0182)*	88.8518 (1.2082)*	4.7803 (0.5705)*
		✓		✓	**0.8549 (0.0164)**	**83.1629 (1.4710)**	**4.3237 (0.4433)**
			✓		0.8275 (0.0126)*	78.0419 (1.4073)*	4.5537 (0.4933)*
			✓	✓	**0.8517 (0.0103)**	**76.9882 (1.4133)**	**4.3235 (0.4573)**
Number of Clusters Set to 500							
✓					0.8760 (0.0086)*	79.0367 (1.1255)*	4.1564 (0.3656)*
✓				✓	**0.8825 (0.0128)**	**77.2532 (1.4079)**	**3.9438 (0.3833)**
	✓				0.8570 (0.0162)*	80.7032 (1.4244)	4.1877 (0.3686)*
	✓			✓	**0.8644 (0.0142)**	**79.7560 (1.3864)**	**3.9755 (0.4043)**
		✓			0.8514 (0.0133)*	89.9762 (1.5970)*	4.2336 (0.4504)*
		✓		✓	**0.8698 (0.0158)**	**83.6178 (1.3105)**	**4.0732 (0.4404)**
			✓		0.8503 (0.0182)	78.4164 (1.6902)	4.2068 (0.3835)*
			✓	✓	**0.8560 (0.0152)**	**78.2364 (1.5863)**	**3.9761 (0.4128)**

where $|LCC(U_k)|$ denotes the number of vertices in the largest connected component of U_k, and $|U_k|$ denotes the number of vertices in U_k. Higher spatial continuity scores indicate that parcels are more continuous rather than overly fragmented.

B.2 Downstream Tasks

Gender Classification, Age and Cognitive Prediction. To assess whether individual brain networks constructed using the group-level parcellation as a template effectively preserve identity- and cognition-related information after dimensionality reduction, we evaluate them on three standard downstream tasks: gender classification, age prediction, and cognitive prediction. Specifically, for each subject in both the training and test sets, we first construct region-wise FCN and MSN based on group-level parcellation. For instance, when the parcellation resolution is set to 400 parcels, the resulting FCN and MSN are both of size 400×400. Subsequently, we extract the upper triangular elements of the FCN and MSN, flatten them into vectors, and finally concatenate them to form the input multi-modal feature vector for each individual. For the prediction ground truth, we used postnatal days as the target for age prediction and adopted the Early Learning Composite from the Mullen Scales of Early Learning as the reference standard for cognitive prediction. On the training set, ridge regression is used for feature selection, while SVM and SVR are employed for classification and prediction, respectively. Performance is evaluated on the test set using accuracy (ACC) for classification and mean absolute error (MAE) for prediction. We repeat the experiment 10 times with different random seeds and report the mean and standard deviation across runs.

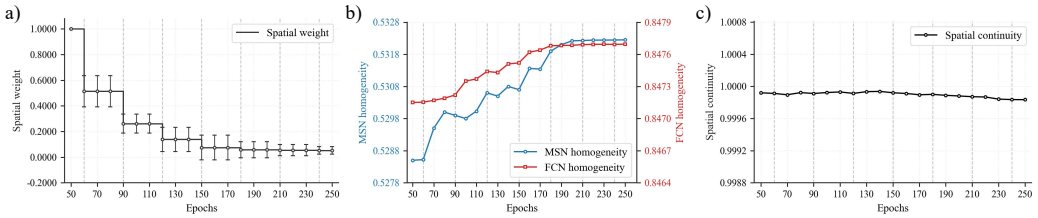


Figure 7: Effect of dynamic spatial-weight annealing. a) shows how the spatial weights progressively decrease over training epochs. The mean and standard deviation across all parcels are reported, showing that nearly all weights approach zero in later epochs. b) shows the changes in structural and functional homogeneity during spatial-weight annealing. Only the mean is reported to avoid obscuring the overall trends. c) shows the changes in spatial continuity, which remains consistently high throughout the annealing process.

Longitudinal Individual Identification. Individual identification based on morphological, structural, and functional brain networks has long been a crucial topic in neuroscience (Hu et al., 2022; Finn et al., 2015; Wang et al., 2021), facilitating the understanding of individual variations in cognition, emotion, and behavior. Recent infant studies further demonstrate that individual-specific brain network patterns are established as early as the first two postnatal years and remain stable across several months during this rapid developmental stage, enabling relatively accurate identification of the same individual through longitudinal scans (Hu et al., 2022). Therefore, to evaluate whether individual brain networks driven by group-level parcellations can effectively preserve individual specificity, we conduct longitudinal individual identification experiments and use identification accuracy as the evaluation metric. Higher scores indicate a stronger ability to capture unique individual brain network patterns and their dynamic developmental evolution.

The experiment was conducted following previous studies (Hu et al., 2022; Finn et al., 2015). Specifically, for longitudinal individual identification, we selected 100 infant subjects with multiple longitudinal scans, including 46 infants with two time points, 26 with three, and 28 with four or more. For each infant, two scans from distinct time points were randomly selected to construct the target and base sets. The target set contained one scan per infant as the query sample, while the base set contained the other scan from the same infant as the reference sample. Based on group-level parcellations, functional connectivity networks and morphological similarity networks were constructed for both sets. The upper triangular elements of these networks were extracted and flattened into one-dimensional vectors as identification features. For each query sample, Pearson correlation was calculated between its features and those of all reference samples, and the identity corresponding to the reference sample with the highest correlation was assigned as the prediction. Identification was considered successful if the predicted identity matched the true identity. The identification accuracy was measured as the percentage of subjects whose identity was correctly predicted among the total number of subjects. We repeated this experiment 10 times and reported the mean and standard deviation.

C Appendix C

C.1 Comparison Methods

RML. RML (Xu et al., 2025) is an advanced robust multi-modal representation learning method that converts heterogeneous multi-modal data into homogeneous word embeddings and fuses them via sample-level attention, yielding a unified and discriminative latent embeddings. To handle imperfect multi-modal data, RML introduces simulated noise and unavailable-data perturbations, and aligns the fused embeddings through contrastive learning. After obtaining the fused embeddings, Kmeans is applied for clustering.

For reproduction, we use the official code repository provided by RML, replacing the input data with $32\text{k} \times 300$ X_{MSN} and X_{FCN} raw feature matrices while keeping the parameter settings consistent with the paper. After training, following the paper, we directly perform Kmeans on the obtained fused embeddings to obtain the final group-level parcellation map.

To validate the plug-and-play capability of $\mathcal{L}_{\text{DG_Homo}}$ and $\mathcal{L}_{\text{Boundary}}$, we first attach a classification head after the fused embeddings and then incorporate $\mathcal{L}_{\text{DG_Homo}}$ and $\mathcal{L}_{\text{Boundary}}$ into the network for joint training. Notably, we still train from scratch, in which case RML can be viewed as a regularizer within the clustering task (Xu et al., 2025).

SparseMVC. SparseMVC (Liu et al., 2026b) is an advanced multi-modal clustering method that explicitly probes cross-modal sparsity variations and adapts the encoding process to each modality's sparsity profile. It employs sparse autoencoders with adaptive constraints, in which an entropy-matching sparsity loss is dynamically scaled by the probed modality sparsity ratio, and it introduces a correlation-informed sample reweighting module that computes sample-level fusion weights based on correlations between early-fused global embeddings and modality-specific embeddings. A cross-modal distribution alignment module further aligns the late-fused global embeddings with each modality's compressed embeddings via contrastive learning, producing unified and discriminative latent embeddings. After obtaining the late-fused global embeddings, Kmeans is applied for clustering.

For reproduction, we use the official code repository provided by SparseMVC, replace the input data with 32k $\times$ 300 X_{MSN} and X_{FCN} raw feature matrices, and follow the same training procedure and parameter settings as in the paper. After training, we run kmeans to obtain the final group-level parcellation map.

To validate the plug-and-play capability of $\mathcal{L}_{\text{DG_Homo}}$ and $\mathcal{L}_{\text{Boundary}}$, we adapt the SparseMVC framework as follows. We remove the modality-specific compression layers and retain only the global compression layer to obtain the late-fused global embeddings, after which a classification head is attached to produce soft label assignments. For optimization, we discard the contrastive learning loss and instead train with $\mathcal{L}_{\text{DG_Homo}}$ and $\mathcal{L}_{\text{Boundary}}$, while keeping others unchanged.

MSDIB. MSDIB (Hu et al., 2025) is an novel multi-modal clustering method that performs multi-aspect self-guided deep information bottleneck learning to extract cluster-friendly information from heterogeneous modalities. It jointly conducts information compression and information preservation: the former learns compact modality-private embeddings and minimizes mutual information within and between modalities to remove redundancy, while the latter maximizes mutual information between modality-private and modality-shared embeddings and between local and global cluster assignments to preserve shared and clustering-relevant information. Finally, a deep divergence-based clustering module takes the fused embeddings as input to obtain the global cluster assignment.

For reproduction, we use the official code repository provided by MSDIB, replace only the input data while keeping all other settings at their defaults, and train the model to obtain the final group-level parcellation.

To validate the plug-and-play capability of $\mathcal{L}_{\text{DG_Homo}}$ and $\mathcal{L}_{\text{Boundary}}$, we adapt the MSDIB framework with minimal modifications: modality-specific private embeddings no longer pass through the cluster assignment module. Instead, only the fused embeddings are fed into this module to produce the global cluster assignment. During optimization, we discard both the mutual information maximization between local and global cluster assignments and the deep divergence clustering objective, employing $\mathcal{L}_{\text{DG_Homo}}$ and $\mathcal{L}_{\text{Boundary}}$ to optimize the clustering process, while keeping others unchanged.

BDEC. BDEC (Zhu et al., 2025a) is a deep embedded clustering model for group-level resting-state fMRI cortical parcellation, designed to reduce the model assumptions of previous approaches. It employs an encoder-decoder architecture to learn compact representations of functional activation, incorporates a soft location embedding, and optimizes a loss combining reconstruction, intra-class similarity via distribution projection, and inter-class separation between cluster centers. A post-processing step further enforces spatial contiguity of parcels. Compared with ten widely used parcellation methods, BDEC achieves improved functional homogeneity, validity, downstream task performance, and generalization on independent datasets.

We initially intended to compare against BDEC. However, BDEC does not provide an official code repository, making it difficult to faithfully reproduce the authors' implementation. Moreover, BDEC is not designed for multi-modal data, so even with official code, additional modifications would be required, which could introduce further performance degradation. Nevertheless, we attempted to manually reconstruct BDEC based on the methodological descriptions in the paper and performed

single-modal group-level parcellation using only the group-level X_{FCN}. Unfortunately, we encountered severe and unstable empty-cluster issues. Even after introducing a balance loss $\mathcal{L}_{\mathrm{Balance}}$, only a part of clusters were assigned vertices, while the other clusters remained empty, making a meaningful comparison infeasible. Overall, the lack of an official code repository substantially limits faithful reproduction of BDEC, which is why we do not directly compare against it in the main text.

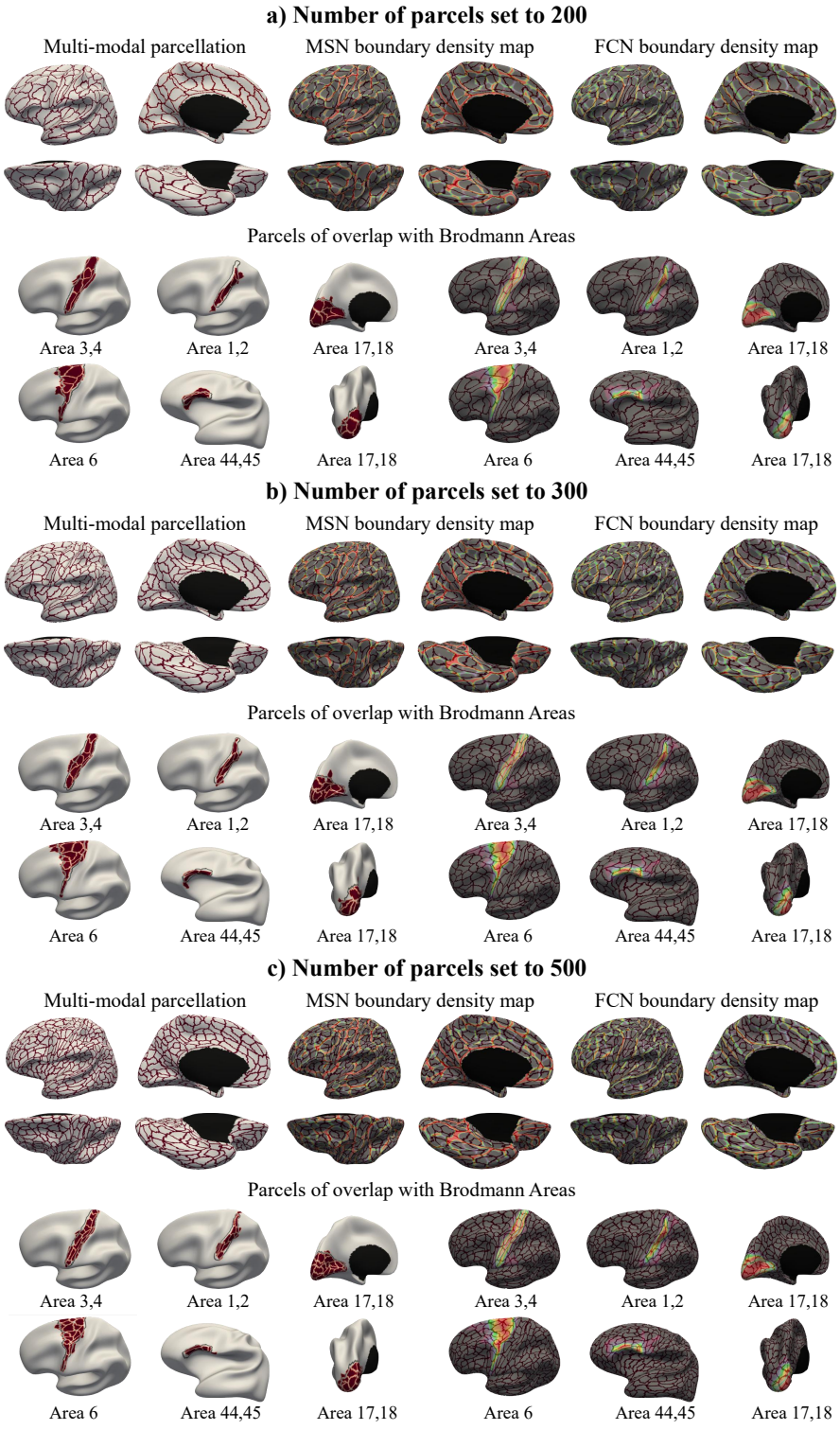


Figure 8: Qualitative visualization results of SABER parcellations at different resolutions. a), b) and c) show the results for 200, 300, and 500 parcels, respectively, including the Infant group-level multi-modal cortical parcellation map, boundary alignment between parcellation and boundary probability density maps, and overlap between parcellation and Brodmann areas.