

1	[APPENDIX] DCA: Graph-Guided Deep Embedding Clustering for Brain	
2	Atlases	
3	1 Cross-dataset generalization on CHCP	2
4	2 Atlas of the subcortex and white matter	3
5	3 Group atlas generation	4
6	4 AtlaScore: atlas evaluation platform	5
7	4.1 Distance-controlled boundary coefficient (DCBC)	5
8	4.2 Gender classification	6
9	4.3 Fluid, crystallized, and general intelligence level prediction	6
10	4.4 Cognitive task classification	6
11	4.5 Autism diagnosis and cross-site generalization	6
12	4.6 AD diagnosis	6
13	4.7 FC stability	7
14	4.8 Fingerprinting	7
15	4.9 Age group classification	7
16	5 Complete DCA performance on AtlaScore	8
17	5.1 Evaluation on different smoothing level	8
18	5.2 DCBC performance across Atlases	9
19	5.3 Downstream task performance across atlases	9
20	6 Main task ablation	12
21	6.1 Effect of DCA framework	12
22	6.2 Regularization and reconstruction loss	13
23	6.3 Clustering on different smoothing level	14
24	6.4 Choice of graph cut method	14
25	6.5 Choice of gray matter mask	14
26	7 Group atlas generation analysis	16
27	8 Spatiotemporal masking strategy	18
28	9 Data organization	19
29	9.1 DCA	19
30	9.2 AtlaScore	19

1 Cross-dataset generalization on CHCP

We further assess our model’s cross-dataset generalization by applying the Swin-UNETR encoder—pretrained on the HCP dataset—to individual atlas generation on the other independent dataset (Fig. 1 and Table 1). The Chinese Human Connectome Project (CHCP) dataset comprises high-resolution multimodal MRI—including structural, diffusion, and resting-state fMRI—from healthy Chinese adults, and uses the same acquisition parameters and HCP preprocessing pipeline [1]. Our results demonstrate that, without any additional fine-tuning, DCA produces coherent, spatially contiguous parcellations on CHCP dataset that surpass other atlases in both metric, underscoring the robustness of the learned voxel embeddings.

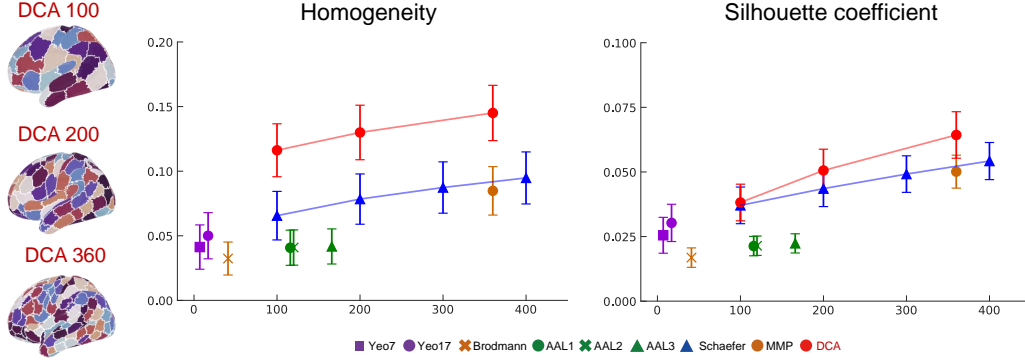


Figure 1: Homogeneity and Silhouette coefficients measured over 100 CHCP subjects at varying numbers of parcels.

Table 1: Evaluation of similarity metrics against DCA and other atlases on the CHCP dataset

Metrics \ Atlas	Yeo		Brodmann	Schaefer	DCA	AAL	
	7	17	41	100	100	116	120
Homogeneity ↑	0.0413 ±0.0171	0.0500 ±0.0179	0.0324 ±0.0127	0.6560 ±0.0188	0.1162 ± 0.0205	0.0407 ±0.0136	0.0409 ±0.0136
Silhouette ↑	0.0255 ±0.0069	0.0303 ±0.0072	0.0169 ±0.0038	0.0371 ±0.0071	0.0382 ± 0.0071	0.0213 ±0.0038	0.0215 ±0.0037

Metrics \ Atlas	AAL	Schaefer	DCA	Schaefer	MMP	DCA	Schaefer
	166	200	200	300	360	360	400
Homogeneity ↑	0.0418 ±0.0136	0.0784 ±0.0195	0.1300 ± 0.0211	0.0873 ±0.0199	0.0848 ±0.0187	0.1451 ± 0.0214	0.0948 ±0.0201
Silhouette ↑	0.0224 ±0.0037	0.0435 ±0.0069	0.0506 ± 0.0082	0.0492 ±0.0071	0.0501 ±0.0064	0.0643 ± 0.009	0.0542 ±0.0072

2 Atlas of the subcortex and white matter

Because our method learns voxel-level embeddings across the entire brain, it is applicable to any arbitrary brain structure. Given a region-of-interest (ROI) mask, our framework can generate parcellations at a specified resolution. Here, we demonstrate two applications: atlas construction for the subcortex and white matter (Fig. 2). The corresponding ROI masks are extracted from FreeSurfer’s `aparc+aseg.mgz` [2] and transformed into MNI152 space. We evaluate parcellations with $\{10, 20, 40, 50\}$ clusters for the subcortex and $\{50, 100, 200, 400\}$ clusters for the white matter. In future work, incorporating additional structural information, such as white matter fiber orientations, could further improve the quality of the parcellations.

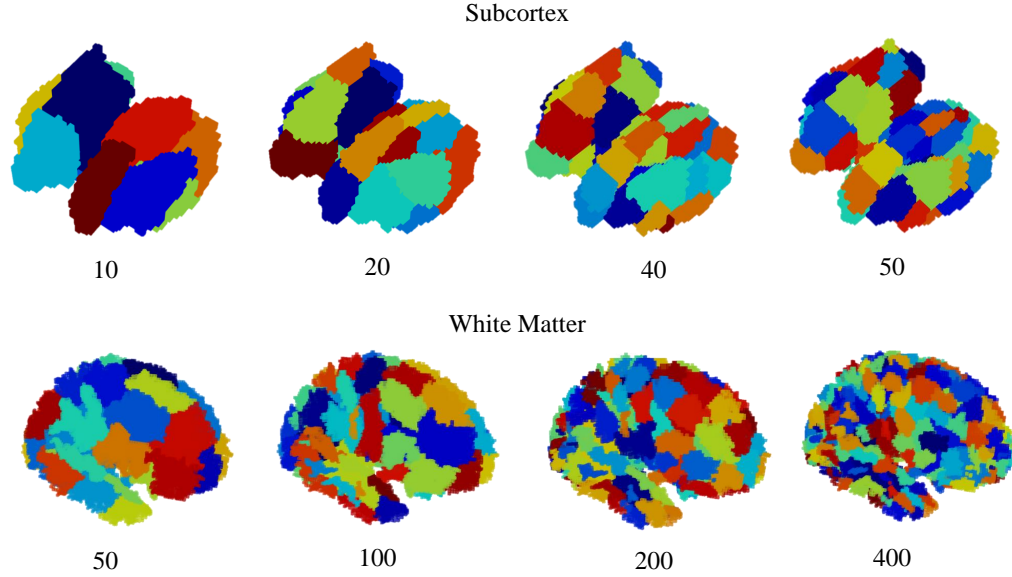


Figure 2: Voxel-wise parcellations of subcortical and white matter regions under varying granularity levels.

49 3 Group atlas generation

50 To facilitate downstream use and fair comparison, we developed a streamlined procedure to construct
 51 a group-level atlas from individual parcellations (Algorithm 1). Given N subject-specific atlases with
 52 a common label system of K parcels, we generate a spatially contiguous group-level atlas through
 53 three steps:

54 **Step 1: Template label selection.** We first identify a set of reliable voxels $\mathcal{V}_{\text{core}}$ that are absent
 55 in at most $\alpha \cdot N$ subjects (we use $\alpha = 0.2$). For each voxel $v \in \mathcal{V}_{\text{core}}$, we form a label vector
 56 $\mathbf{z}_v = [A^{(1)}(v), A^{(2)}(v), \dots, A^{(N)}(v)]$, where $A^{(i)}(v)$ denotes the label assigned to voxel v in subject
 57 i . We sort all such vectors by frequency of occurrence, and select the top K vectors that have pairwise
 58 Hamming distance less than $\beta \cdot N$ (we use $\beta = 0.2$) as the *template label vectors* $\{\mathbf{t}_1, \dots, \mathbf{t}_K\}$.

59 **Step 2: Voxel-to-template assignment.** We then identify $\mathcal{V}_{\text{assign}}$, the set of voxels absent in at
 60 most $\gamma \cdot N$ subjects (with $\gamma = 0.8$), and assign each $v \in \mathcal{V}_{\text{assign}}$ to the template k that maximizes the
 61 agreement:

$$L(v) = \arg \max_k \sum_{i=1}^N \mathbb{I}[A^{(i)}(v) = t_k^{(i)}]$$

62 This results in a K -label volumetric map \mathbf{L} that is not necessarily spatially contiguous.

63 **Step 3: Spatial contiguity enforcement.** For each parcel k , we retain only its largest 6-connected
 64 component, and mark all other voxels in k as unlabeled. We then iteratively reassign these dropped
 65 voxels as follows: for each unlabeled voxel v with at least one labeled 6-connected neighbor, we
 66 compute the Hamming distance between \mathbf{z}_v and the label vectors $\{\mathbf{z}_u\}$ of its labeled neighbors
 67 $u \in \mathcal{N}_6(v)$. The voxel v is then assigned the same label as the neighbor u^* with the smallest distance,
 68 i.e., $L(v) = L(u^*)$. This process repeats until all voxels are labeled, resulting in a group-level atlas
 69 with K spatially contiguous and functionally consistent regions.

Algorithm 1 Group-level atlas generation from individual parcellations

Require: Subject-wise atlases $\{\mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \dots, \mathbf{A}^{(N)}\}$, number of parcels K , thresholds α, β, γ

Ensure: Group-level atlas \mathbf{L}

```

1: Identify  $\mathcal{V}_{\text{core}} \leftarrow \{v : \text{voxel absent in } \leq \alpha N \text{ subjects}\}$ 
2: For each  $v \in \mathcal{V}_{\text{core}}$ , construct label vector  $\mathbf{z}_v = [A^{(1)}(v), \dots, A^{(N)}(v)]$ 
3: Count frequency of each  $\mathbf{z}_v$ ; sort descending
4: Select top  $K$  vectors  $\{\mathbf{t}_1, \dots, \mathbf{t}_K\}$  with pairwise Hamming distance  $\leq \beta N$ 
5: Identify  $\mathcal{V}_{\text{assign}} \leftarrow \{v : \text{voxel absent in } \leq \gamma N \text{ subjects}\}$ 
6: for all  $v \in \mathcal{V}_{\text{assign}}$  do
7:   Assign  $L(v) \leftarrow \arg \max_k \sum_{i=1}^N \mathbb{I}[A^{(i)}(v) = t_k^{(i)}]$ 
8: end for
9: for all label  $k = 1$  to  $K$  do
10:   Keep largest 6-connected component of label  $k$ 
11:   Mark all other voxels in  $k$  as unlabeled
12: end for
13: while any voxel is unlabeled do
14:   for all unlabeled voxel  $v$  do
15:     if  $v$  has at least one labeled 6-neighbor then
16:       Find  $u^* = \arg \min_{u \in \mathcal{N}_6(v), \text{ labeled}} \text{Hamming}(\mathbf{z}_v, \mathbf{z}_u)$ 
17:       Assign  $L(v) \leftarrow L(u^*)$ 
18:     end if
19:   end for
20: end while

```

70 4 AtlaScore: atlas evaluation platform

71 Due to space constraints, we have omitted the full set of results from the main text. In the Supple-
 72 mentary Material, we present detailed experimental designs and complete outcomes for 3 similarity
 73 evaluations and 12 downstream tasks (Table 6).

Table 2: Experiment index, name, and description.

#	Name	Description
1	Similarity–Homogeneity	Measure mean intra-cluster correlation
2	Similarity–Silhouette	Compute silhouette coefficient over voxels
3	Similarity–DCBC 4.1	Evaluating brain parcellations using the distance-controlled boundary coefficient [3]
4	Downstream–Gender classification 4.2	Predict biological sex using FC
5	Downstream–Fluid Intelligence 4.3	Predict fluid intelligence level using FC
6	Downstream–Cognitive task (7-way) 4.4	Classify 7 cognitive tasks using FC
7	Downstream–Cognitive task (24-way) 4.4	Classify 24 cognitive tasks using FC
8	Downstream–Autism diagnosis 4.5	Classify autism vs. healthy controls using FC
9	Downstream–AD diagnosis 4.6	Classify AD / MCI / CN using FC
10	Downstream–FC stability 4.7	Within-subject FC similarity
11	Downstream–Fingerprinting 4.8	Subject identification via FC matching
12	Downstream–Age group classification 4.9	Predict age group labels
13	Downstream–Crystallized intelligence 4.3	Predict crystallized intelligence level using FC
14	Downstream–General intelligence 4.3	Predict overall cognitive ability level using FC
15	Downstream–Autism cross-site 4.5	Cross-site classification of autism vs. healthy controls using FC

74 4.1 Distance-controlled boundary coefficient (DCBC)

75 In the main text, we have introduced Homogeneity and Silhouette in Section 3. Such conventional
 76 metrics overlook the intrinsic spatial smoothness of brain signals, often overestimating parcellation
 77 quality by conflating spatial proximity with functional similarity, especially in high-resolution
 78 cortical data, where false boundaries may emerge due to smoothness rather than genuine functional
 79 distinctions [3]. To mitigate this bias, DCBC groups vertex pairs based on their spatial separation and
 80 evaluates functional similarity differences between within- and between-parcel pairs at each distance
 81 level. The DCBC metric is formally defined as follows:

$$\text{DCBC} = \sum_{i=1}^N w_i d_i, \quad (1)$$

82 where per-bin correlation difference $d_i = \text{corr}_{\text{within}}(i)$ and $\text{corr}_{\text{between}}(i)$ are the mean functional
 83 correlations of within-parcel and between-parcel vertex pairs in the i -th spatial distance bin, respec-
 84 tively; The variance $\text{var}(d_i)$ reflects how vertex pair counts ($n_{w,i}$, $n_{b,i}$) affect the reliability of d_i
 85 in each spatial bin, while the precision weights w_i subsequently compensate for this uncertainty by
 86 assigning higher influence to bins with lower variance during DCBC computation. $\text{var}(d_i)$ and w_i
 87 take the following forms:

$$\text{var}(d_i) = \frac{n_{w,i} + n_{b,i}}{n_{w,i} n_{b,i}}, w_i = \frac{\frac{n_{w,i} n_{b,i}}{n_{w,i} + n_{b,i}}}{\sum_{j=1}^N \frac{n_{w,j} n_{b,j}}{n_{w,j} + n_{b,j}}}. \quad (2)$$

88 By controlling for spatial distance in this way, DCBC disentangles true functional boundaries from
 89 artifacts of spatial smoothness, providing a more reliable parcellation assessment. In our evaluation
 90 process, all parameters followed the settings specified in [3]. However, we note that DCBC was
 91 mainly developed for surface-based parcellations and becomes computationally prohibitive at the fine

voxel resolution employed by DCA. Therefore, we computed DCBC scores by projecting volumetric atlases onto the cortical surface (fsLR 32k template) [4], analyzing only data from the left hemisphere. This surface-based approach exceeds the scope of our native volumetric framework.

4.2 Gender classification

We evaluated the ability of each atlas to support gender classification based on resting-state functional connectivity (FC). We used data from 100 unrelated subjects in the Human Connectome Project (HCP) [5], each contributing multiple FC samples constructed from 300 consecutive TRs of resting-state fMRI. For each atlas, FC matrices were computed and their upper-triangular entries were used as features.

To ensure subject-level generalization, we performed 10-fold cross-validation across subjects: in each fold, 90 subjects were used for training and 10 for testing. A linear support vector classifier (SVC) was trained on the training set. If the FC dimensionality exceeded 100, we applied principal component analysis (PCA) to reduce dimensionality: features were projected onto the top 100 principal components computed from the training data, and test samples were projected into the same subspace. Classification accuracy on the test subjects was recorded for each fold and averaged to obtain final performance.

4.3 Fluid, crystallized, and general intelligence level prediction

We evaluated whether atlas-based FC features can predict individual differences in fluid, crystallized, and general intelligence. We used the corresponding HCP behavioral score (CogFluidComp_AgeAdj, CogCrystalComp_AgeAdj, and CogTotalComp_AgeAdj) to define three classes: low (<85), medium (85–115), and high (>115) intelligence. Each subject contributed multiple FC samples from resting-state fMRI (300 TRs per sample). 10-fold cross-validation was performed across subjects, using a linear SVC. When the number of FC features exceeded 100, PCA was applied to project the data onto the top 100 principal components computed from the training set.

4.4 Cognitive task classification

We assessed whether atlas-based FC can distinguish different cognitive states using task-fMRI data from 100 HCP subjects. For the 7-class classification, each subject completed seven tasks—working memory, gambling, motor, language, social, relational, and emotional—each contributing one FC matrix. For the 24-class classification, we further segmented each task into its constituent conditions (e.g., 0-back faces, math, fear), resulting in 24 fine-grained task labels. Each subject contributed one FC matrix per task condition.

We trained a linear SVC to predict the task label from FC features using 10-fold cross-validation across subjects. When the number of FC features exceeded 100, PCA was applied to project the data onto the top 100 principal components computed from the training set.

4.5 Autism diagnosis and cross-site generalization

We evaluated whether FC features can distinguish individuals with autism spectrum disorder (ASD) from healthy controls using resting-state fMRI data from the ABIDE dataset ($n = 871$) [6]. For each subject, an FC matrix was computed and used to train a linear SVC for binary classification. We considered two evaluation settings. In the first, we performed 10-fold cross-validation at the subject level to assess within-dataset classification performance. In the second, we tested cross-site generalization by holding out one acquisition site for testing while training on subjects from all other sites, repeating this procedure across all sites. When the number of FC features exceeded 100, PCA was applied to project the data onto the top 100 principal components computed from the training set.

4.6 AD diagnosis

We evaluated the ability of FC features to classify individuals into Alzheimer’s disease (AD), mild cognitive impairment (MCI), or cognitively normal (CN) groups. We used resting-state fMRI data from 267 subjects in the ADNI dataset [7], each labeled as AD, MCI, or CN. FC matrices were

139 computed for each subject and used to train a linear SVC for 3-class classification. 10-fold cross-
140 validation was performed across subjects. When the number of FC features exceeded 100, PCA was
141 applied to project the data onto the top 100 principal components computed from the training set.

142 **4.7 FC stability**

143 We evaluated the within-subject stability of FC to assess how consistently an atlas captures individual
144 functional architecture [8]. We used resting-state fMRI data from 100 HCP subjects. Each fMRI
145 scan was segmented into multiple non-overlapping windows of 300 TRs, and an FC matrix was
146 computed per window. For each subject, we calculated the Pearson correlation between the vectorized
147 upper triangles of all FC matrix pairs and averaged the resulting values to obtain a single stability
148 score. These scores were then aggregated across all subjects to evaluate the group-level FC stability
149 supported by each atlas.

150 **4.8 Fingerprinting**

151 To evaluate how well an atlas captures individual-specific features in FC, we conducted a subject
152 identification (fingerprinting) task [9]. We used resting-state fMRI data from 100 subjects, each
153 providing multiple FC matrices. For each subject, one FC was randomly selected as the reference.
154 The remaining FCs were matched to all reference FCs by computing the Pearson correlation between
155 the vectorized upper triangle of each pair. A prediction was considered correct if the most highly
156 correlated reference FC belonged to the same subject as the query FC. Each subject’s identification
157 accuracy was computed, and group-level performance was obtained by averaging across subjects.

158 **4.9 Age group classification**

159 We evaluated whether atlas-based FC features can predict individual differences in age group. We
160 used the HCP-provided Age variable to assign subjects into three age groups: 21–25, 26–30, and
161 31 years or older. Each subject contributed multiple FC samples from resting-state fMRI (300 TRs
162 per sample). 10-fold cross-validation was performed across subjects, using a linear SVC. When the
163 number of FC features exceeded 100, PCA was applied to project the data onto the top 100 principal
164 components computed from the training set.

5 Complete DCA performance on AtlaScore

5.1 Evaluation on different smoothing level

In the main text, we evaluated similarity-related metrics using data smoothed with a 3mm FWHM kernel. Here, we additionally report the results obtained with unsmoothed data and data smoothed with a 6mm FWHM kernel. The overall conclusions remain consistent across different smoothing levels (Fig. 3, Table 3, and Table 4).

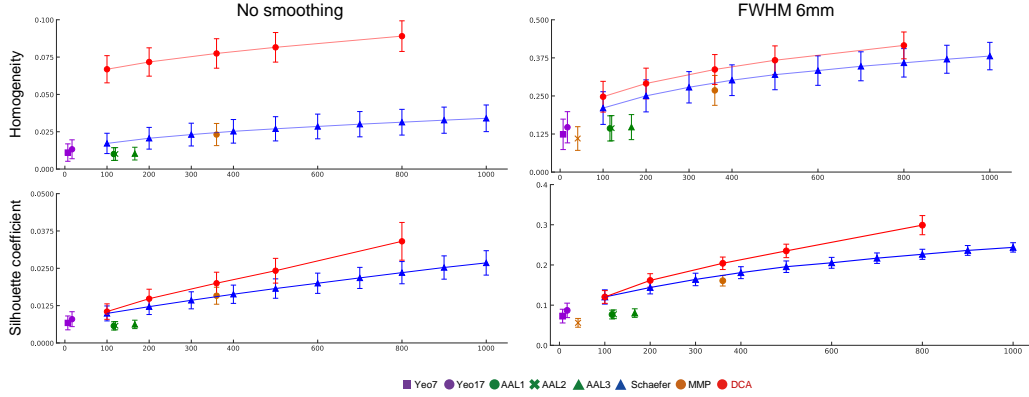


Figure 3: Homogeneity and silhouette measured over 100 HCP subjects on different smoothing level.

Table 3: Evaluation on no smoothing data.

Metrics \ Atlas	Yeo		Brodmann	Schaefer	DCA	AAL			Schaefer	DCA	Schaefer
	7	17	41	100	100	116	120	166	200	200	300
Homogeneity \uparrow	0.0110 ± 0.0058	0.0133 ± 0.0063	-	0.0172 ± 0.0068	0.0669 ± 0.0091	0.0100 ± 0.0043	0.0101 ± 0.0043	0.0103 ± 0.0043	0.0206 ± 0.0073	0.0717 ± 0.0095	0.0231 ± 0.0076
Silhouette \uparrow	0.0067 ± 0.0023	0.0080 ± 0.0025	-	0.0099 ± 0.0025	0.0105 ± 0.0026	0.0057 ± 0.0014	0.0058 ± 0.0014	0.0062 ± 0.0014	0.0122 ± 0.0026	0.0148 ± 0.0032	0.0143 ± 0.0029

Metrics \ Atlas	MMP	DCA	Schaefer		DCA	Schaefer			DCA	Schaefer	
	360	360	400	500	500	600	700	800	800	900	1000
Homogeneity \uparrow	0.0231 ± 0.0074	0.0774 ± 0.0099	0.0253 ± 0.0079	0.0270 ± 0.0081	0.0815 ± 0.0099	0.0285 ± 0.0083	0.0301 ± 0.0085	0.0314 ± 0.0086	0.0890 ± 0.0103	0.0327 ± 0.0088	0.0340 ± 0.0089
Silhouette \uparrow	0.0158 ± 0.0028	0.0200 ± 0.0037	0.0163 ± 0.0031	0.0182 ± 0.0032	0.0242 ± 0.0041	0.0200 ± 0.0034	0.0218 ± 0.0035	0.0235 ± 0.0037	0.0340 ± 0.0063	0.0253 ± 0.0039	0.0268 ± 0.0041

Table 4: Evaluation on 6mm FWHM smoothed data.

Metrics \ Atlas	Yeo		Brodmann	Schaefer	DCA	AAL			Schaefer	DCA	Schaefer
	7	17	41	100	100	116	120	166	200	200	300
Homogeneity \uparrow	0.1239 ± 0.0500	0.1472 ± 0.0512	0.1102 ± 0.0389	0.2102 ± 0.0537	0.2475 ± 0.0504	0.1434 ± 0.0413	0.1441 ± 0.0413	0.1476 ± 0.0411	0.2502 ± 0.0526	0.2909 ± 0.0503	0.2784 ± 0.0515
Silhouette \uparrow	0.0727 ± 0.0171	0.0870 ± 0.0180	0.0560 ± 0.0109	0.1204 ± 0.0177	0.1203 ± 0.0159	0.0766 ± 0.0109	0.0773 ± 0.0108	0.0803 ± 0.0106	0.1439 ± 0.0159	0.1615 ± 0.0168	0.1639 ± 0.0155

Metrics \ Atlas	MMP	DCA	Schaefer		DCA	Schaefer			DCA	Schaefer	
	360	360	400	500	500	600	700	800	800	900	1000
Homogeneity \uparrow	0.2682 ± 0.0491	0.3368 ± 0.0490	0.3016 ± 0.0504	0.3198 ± 0.0494	0.3671 ± 0.0472	0.3332 ± 0.0484	0.3473 ± 0.0476	0.3587 ± 0.0467	0.4159 ± 0.0441	0.3705 ± 0.0460	0.3809 ± 0.0451
Silhouette \uparrow	0.1608 ± 0.0132	0.2041 ± 0.0155	0.1808 ± 0.0149	0.1956 ± 0.0143	0.2350 ± 0.0168	0.2054 ± 0.0135	0.2169 ± 0.0130	0.2265 ± 0.0126	0.2990 ± 0.0238	0.2361 ± 0.0122	0.2437 ± 0.0117

5.2 DCBC performance across Atlases

DCBC was mainly developed for surface-based parcellations and becomes computationally prohibitive at the fine voxel resolution employed by DCA. Therefore, we computed DCBC scores by projecting volumetric atlases onto the cortical surface (fsLR 32k template [4]), analyzing only data from the left hemisphere. This surface-based approach exceeds the scope of our native volumetric framework (Fig. 4 and Table 5).

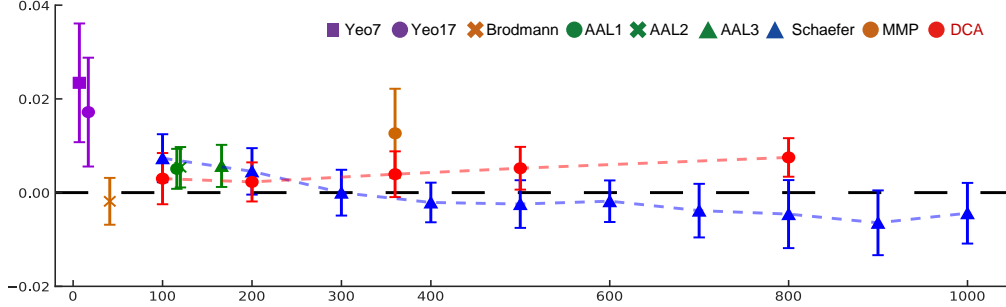


Figure 4: DCBC measured over 100 HCP subjects at varying numbers of parcels.

Table 5: DCBC values and standard deviations for various atlases.

Atlas	nParcels	Value	ValueStd
Yeo7	7	0.0234	0.0127
Yeo17	17	0.0172	0.0116
Brodmann	41	-0.0019	0.0050
Schaefer	100	0.0073	0.0051
DCA	100	0.0030	0.0055
AAL1	116	0.0051	0.0043
AAL2	120	0.0054	0.0043
AAL3	166	0.0057	0.0045
Schaefer	200	0.0045	0.0050
DCA	200	0.0023	0.0042
Schaefer	300	-0.0000	0.0049
MMP	360	0.0127	0.0095
DCA	360	0.0039	0.0049
Schaefer	400	-0.0021	0.0042
Schaefer	500	-0.0024	0.0051
DCA	500	0.0052	0.0046
Schaefer	600	-0.0018	0.0044
Schaefer	700	-0.0038	0.0057
Schaefer	800	-0.0046	0.0073
DCA	800	0.0075	0.0041
Schaefer	900	-0.0064	0.0069
Schaefer	1000	-0.0044	0.0065

5.3 Downstream task performance across atlases

In the main paper, we visualized the performance of 16 atlases on 6 downstream tasks using radar plots. However, the results for the 24-way cognitive task classification may have underestimated the differences between atlases and failed to highlight the strengths of certain methods. Here, we refine our evaluation and provide an updated visualization in Fig. 5. As shown, our proposed DCA atlas achieves consistently strong performance across all resolution levels, while traditional atlases often underperform by comparison.

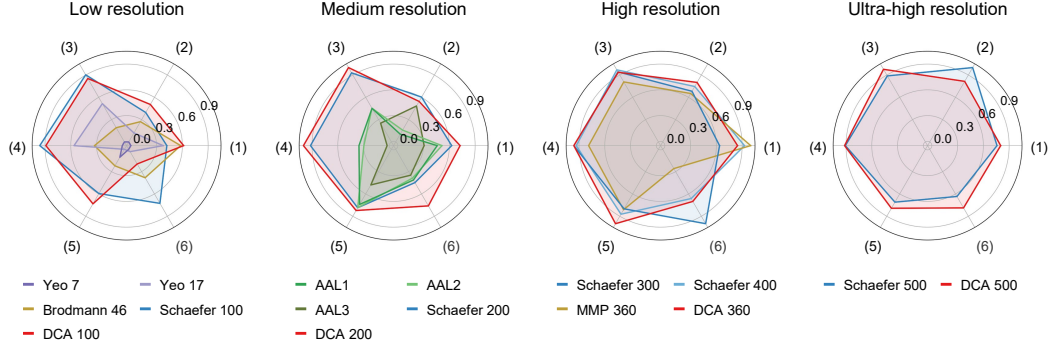


Figure 5: Performance of DCA and baseline atlases across different spatial resolutions on six downstream tasks: (1) gender classification from resting-state FC (HCP), (2) fluid intelligence level prediction from resting-state FC (HCP), (3) classification of 7 cognitive tasks from task-based FC (HCP), (4) classification of 24 cognitive tasks from task-based FC (HCP), (5) autism diagnosis from resting-state FC (ABIDE), (6) AD diagnosis from resting-state FC (ADNI). DCA achieves competitive or superior performance at each resolution level. Values are linearly scaled per task with 0 and 1 corresponding to the lowest and highest performing atlases, respectively.

184 Due to space constraints, only a subset of results was included in the main text. Here, we provide
 185 the complete evaluation on all 12 downstream tasks across 16 atlases (Table 6). Results are reported
 186 as mean \pm standard deviation, averaged over 10-fold cross-validation or subject-level evaluation.
 187 Within each resolution group, the best-performing atlas for each task is highlighted in bold.

Table 6: Evaluation of downstream task performance across atlases. Values are shown as mean \pm standard deviation.

Task	Atlas				Yeo				Brodmann				Schaefer				DCA				AAL				Schaefer				DCA				Schaefer				MMP				DCA				Schaefer				Schaefer				DCA																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
	7	17	41		0.547 ±0.077	0.620 ±0.063	0.659 ±0.078		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053		0.628 ±0.053	0.628 ±0.053	0.628 ±0.053	

6 Main task ablation

6.1 Effect of DCA framework

In the main text, we report ablation results on 20 subjects. Here, we extend the same analysis to the full cohort of 100 subjects used in our primary experiments, and find that the conclusions remain fully consistent with those presented earlier.

First, we applied (1) K-Means clustering and (2) the same graph construction and graph-cut pipeline from our method directly to the raw fMRI time series. This baseline highlights the necessity of our pretrained encoder for extracting informative features. Second, we ran both (3) K-Means and (4) the same graph cut pipeline on the encoded embeddings **without** the KL-guided joint optimization of encoder parameters and centroids, isolating the impact of our graph-guided deep clustering mechanism. In both ablations, all similarity metrics (homogeneity and silhouette) drop noticeably below those of (5) the full DCA model (Fig. 6 and Table 7). Moreover, K-Means fails to produce spatially contiguous parcels. And graph-cut improves continuity, but it still yields isolated regions. In contrast, DCA’s iterative KL-driven refinement—which alternates updating graph weights, encoder parameters, and cluster centers—produces brain atlases that are highly homogeneous. Across five experimental settings, we measured the average number of connected components per parcel as follows: (1) yielded 447.90 components, (2) yielded 8.92, (3) yielded 322.32, (4) yielded 4.94, and (5) our DCA method yielded just 1.0052 components per parcel. These results demonstrate that DCA produces nearly fully contiguous parcels in contrast to the highly fragmented outputs of the baselines (Fig. 6). The few residual disconnections occur primarily along deep cortical sulci and could be readily eliminated via lightweight post-processing or by enforcing connectivity in the group-atlas construction.

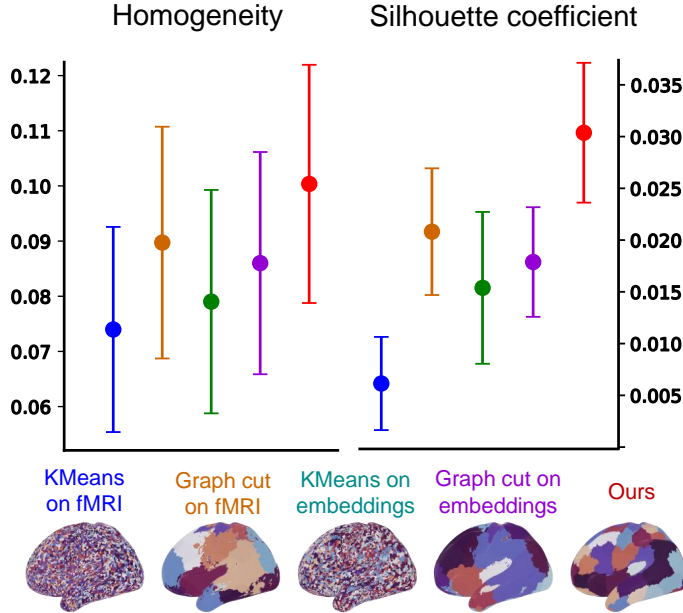


Figure 6: Directly applying K-means or graph-cut method to the raw fMRI time-series or embeddings produces parcellations with substantially lower homogeneity and silhouette scores than our iteratively optimized method. Moreover, K-means on the unprocessed signals cannot guarantee spatially contiguous regions, and while graph-cut method can partially enforce contiguity, it still fails to produce fully continuous parcels. Our method yields brain atlases that are both highly homogeneous and spatially contiguous.

Table 7: Evaluation for fMRI and Embedding-based Clustering with KMeans, Graph Cut, and DCA.

		Homogeneity	Silhouette	connected components per parcel
fMRI	KMeans	0.0740±0.0186	0.0061±0.0045	447.90
	Graph Cut	0.0860±0.0201	0.0179±0.0053	8.92
Embedding	KMeans	0.0790±0.0203	0.0154±0.0073	322.32
	Graph Cut	0.0897±0.0210	0.0208±0.0061	4.94
DCA		0.1004±0.0216	0.0304±0.0068	1.0052

210 6.2 Regularization and reconstruction loss

211 We investigated whether incorporating an orthogonality regularizer or a reconstruction loss would
 212 further improve parcellation quality (Table 8). Applying these augmented objectives to the same 100
 213 subjects used in our main experiments yielded no statistically significant gains in homogeneity or
 214 silhouette coefficient for 100 parcels, indicating that the core KL-based clustering loss is sufficient to
 215 drive optimal voxel-level atlas generation.

216 **orthogonal loss** The orthogonality regularizer is defined on the centroid matrix $\mathbf{D} \in \mathbb{R}^{K \times d}$ (with
 217 unit-norm rows) as follows. Let

$$\mathbf{G} = \mathbf{D} \mathbf{D}^\top \quad (\in \mathbb{R}^{K \times K})$$

218 be the Gram matrix of pairwise inner products. We zero out the diagonal to isolate off-diagonal
 219 similarities:

$$\mathbf{G}_{\text{off}} = \mathbf{G} - \mathbf{I}_K.$$

220 The orthogonality loss then penalizes the mean absolute off-diagonal entry via

$$\mathcal{R}_\perp(\mathbf{D}) = \sqrt{\frac{1}{K(K-1)} \sum_{i \neq j} (\mathbf{G}_{\text{off}})_{ij}}.$$

221 Minimizing \mathcal{R}_\perp encourages the rows of \mathbf{D} to remain mutually orthogonal.

222 **masked reconstructed loss** We only consider the reconstruction of non-background voxels.

$$\mathcal{L}_{\text{masked_MSE}} = \begin{cases} \frac{\sum_{i=1}^N (1 - m_i) (\hat{x}_i - x_i)^2}{\sum_{i=1}^N (1 - m_i)}, & \text{if } \sum_{i=1}^N (1 - m_i) > 0, \\ 0, & \text{otherwise,} \end{cases}$$

223 where \hat{x}_i and x_i are the predicted and target values at voxel i , respectively, and $m_i \in \{0, 1\}$ is the
 224 binary mask indicating background ($m_i = 1$) or foreground ($m_i = 0$).

Table 8: Ablation study on orthogonality and masked reconstruction loss components.

KL	Loss		Homogeneity	Silhouette
	Orthogonality	Reconstruction		
✓			0.1002±0.0214	0.0301±0.0066
✓		✓	0.0890±0.0091	0.0267±0.0045
✓	✓		0.1004±0.0215	0.0304±0.0067

6.3 Clustering on different smoothing level

To improve signal quality and spatial coherence, we applied spatial smoothing using AFNI’s 3dBlurToFWHM [10], targeting a 3 mm full width at half maximum (FWHM). The preprocessed volumetric images were resampled to 2 mm isotropic resolution. This follows the common practice of setting the smoothing kernel to approximately 1.5 times the image resolution. For comparison, we also present results under two additional conditions: no smoothing and 6 mm FWHM smoothing. And there is no significant difference (Table 9).

Table 9: Evaluation of similarity metrics across smoothing levels and parcel resolutions.

Raw		100	200	360	500	800
Metrics						
Homogeneity \uparrow		0.1004 ± 0.0216	0.1127 ± 0.0225	0.1266 ± 0.0229	0.1363 ± 0.0228	0.1535 ± 0.0226
Silhouette \uparrow		0.0305 ± 0.0068	0.0422 ± 0.0094	0.0554 ± 0.0085	0.0647 ± 0.0090	0.0883 ± 0.0114
FWHM 3mm		100	200	360	500	800
Metrics						
Homogeneity \uparrow		0.1004 ± 0.0216	0.1127 ± 0.0225	0.1266 ± 0.0230	0.1364 ± 0.0229	0.1536 ± 0.0227
Silhouette \uparrow		0.0304 ± 0.0068	0.0417 ± 0.0078	0.0545 ± 0.0080	0.0644 ± 0.0086	0.0866 ± 0.0114
FWHM 6mm		100	200	360	500	800
Metrics						
Homogeneity \uparrow		0.1005 ± 0.0217	0.1128 ± 0.0225	0.1267 ± 0.0229	0.1364 ± 0.0229	0.1536 ± 0.0227
Silhouette \uparrow		0.0306 ± 0.0069	0.0419 ± 0.0078	0.0546 ± 0.0081	0.0642 ± 0.0083	0.0868 ± 0.0102

6.4 Choice of graph cut method

In addition to spectral clustering, we evaluated several graph-cut algorithms (Fig. 7 and Table 10). However, most failed to guarantee that each resulting parcel forms a single connected subgraph, leading to fragmented regions. Here, we use a breadth-first search (BFS) based algorithm. The weighted BFS-connected clustering algorithm begins by converting the input edge list and weights into an undirected adjacency list, then randomly seeds k initial clusters by assigning one unique node to each cluster. Each cluster maintains a max-heap of its unassigned neighboring nodes, prioritized by edge weight. Clusters then grow in parallel: at each step, a cluster pops the highest-weight neighbor from its heap, claims that node (if unassigned), and pushes all of its unassigned neighbors onto the heap. To enforce roughly equal cluster sizes, each cluster stops growing once it reaches $\lceil N/k \rceil$. If any nodes remain unassigned after this frontier-driven expansion, they are absorbed into the smallest adjacent cluster. By always selecting the strongest edges first and only adding connected nodes, this method produces contiguous clusters that respect the underlying graph’s weighted connectivity.

6.5 Choice of gray matter mask

In the main text, the corresponding ROI masks are extracted from FreeSurfer’s aparc+aseg.mgz [2]. Here, we show the result (Table 11) by defining the ROI with the same mask to MMP [11] and Schaefer al [12]. DCA still shows higher homogeneity and silhouette coefficients compared to corresponding atlas.

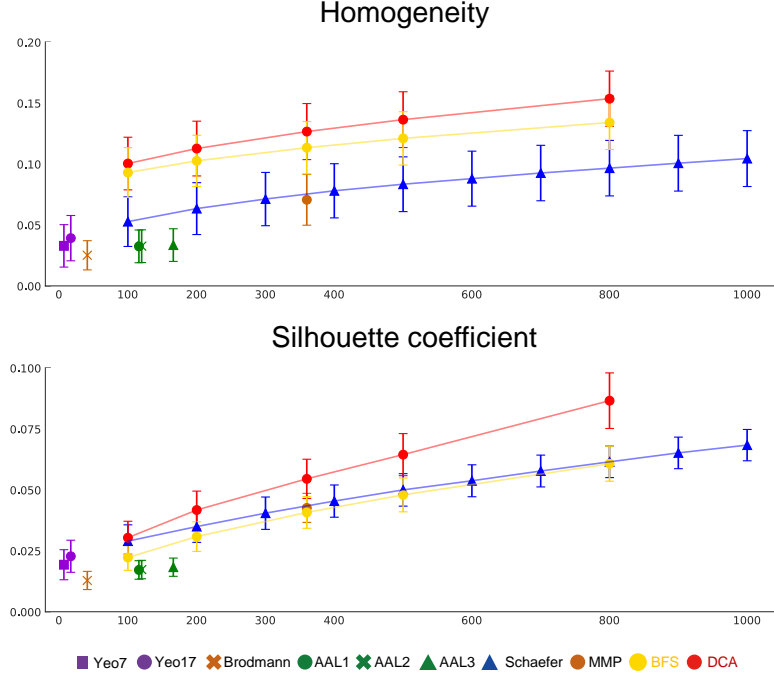


Figure 7: Homogeneity and Silhouette coefficients for weighted BFS-connected clustering and baselines, measured over 100 HCP subjects at varying numbers of parcels.

Table 10: Evaluation of similarity metrics across graph cut methods and parcel resolutions.

graph		100	200	360	500	800
Metrics						
Homogeneity ↑		0.1004 ± 0.0216	0.1127 ± 0.0225	0.1266 ± 0.0230	0.1364 ± 0.0229	0.1536 ± 0.0227
Silhouette ↑		0.0304 ± 0.0068	0.0417 ± 0.0078	0.0545 ± 0.0080	0.0644 ± 0.0086	0.0866 ± 0.0114

mst		100	200	360	500	800
Metrics						
Homogeneity ↑		0.0930 ± 0.0204	0.1026 ± 0.0210	0.1134 ± 0.0215	0.1210 ± 0.0219	0.1341 ± 0.0222
Silhouette ↑		0.0223 ± 0.0052	0.0308 ± 0.0061	0.0407 ± 0.0065	0.0479 ± 0.0069	0.0607 ± 0.0072

Table 11: Evaluation of Similarity Metrics on Schaefer, DCA, MMP, and DCA360 with Gray Matter Masks

	Schaefer100	DCA100	MMP	DCA360
Homogeneity	0.0885±0.0212	0.1004±0.0216	0.1212±0.0211	0.1266±0.0230
Silhouette	0.0191±0.0076	0.0304±0.0068	0.0470±0.0066	0.0545±0.0080

7 Group atlas generation analysis

We systematically examined how three key hyperparameters in group-level atlas generation—voxel reliability threshold α , template distinctiveness threshold β , and voxel inclusion threshold γ —affect downstream task performance. The following four tables report performance across 12 tasks using atlases of increasing resolution (100 to 500 parcels), under various parameter settings.

Table 12: Effect of group atlas generation parameters on performance (100 regions). Values are shown as mean \pm standard deviation.

	0.2			0.3		
	0.7	0.8	0.9	0.7	0.8	0.9
α						
β						
γ						
Gender classification	0.639 \pm 0.076	0.666 \pm 0.080	0.648 \pm 0.074	0.676 \pm 0.066	0.674 \pm 0.050	0.684\pm0.057
Fluid intelligence	0.477 \pm 0.080	0.491 \pm 0.082	0.486 \pm 0.084	0.508 \pm 0.089	0.508 \pm 0.078	0.516 \pm 0.066
Cognitive task (7-way)	0.863 \pm 0.056	0.869 \pm 0.062	0.867 \pm 0.060	0.872 \pm 0.050	0.866 \pm 0.038	0.864 \pm 0.053
Cognitive task (24-way)	0.463\pm0.027	0.452 \pm 0.030	0.440 \pm 0.033	0.458 \pm 0.020	0.450 \pm 0.024	0.459 \pm 0.024
Autism diagnosis	0.681\pm0.048	0.655 \pm 0.054	0.653 \pm 0.024	0.658 \pm 0.032	0.658 \pm 0.026	0.665 \pm 0.046
AD diagnosis	0.387 \pm 0.093	0.387 \pm 0.077	0.387 \pm 0.099	0.432 \pm 0.064	0.428 \pm 0.073	0.451\pm0.054
FC stability	0.650 \pm 0.045	0.650 \pm 0.045	0.649 \pm 0.045	0.656 \pm 0.045	0.656 \pm 0.045	0.654 \pm 0.045
Fingerprinting	0.700 \pm 0.202	0.696 \pm 0.201	0.694 \pm 0.202	0.691 \pm 0.204	0.689 \pm 0.205	0.690 \pm 0.210
Age group classification	0.447 \pm 0.110	0.452\pm0.136	0.446 \pm 0.118	0.419 \pm 0.127	0.425 \pm 0.113	0.424 \pm 0.100
Crystallized intelligence	0.479 \pm 0.083	0.472 \pm 0.095	0.489 \pm 0.082	0.511 \pm 0.055	0.505 \pm 0.060	0.521\pm0.061
General intelligence	0.432 \pm 0.108	0.442 \pm 0.104	0.438 \pm 0.097	0.474 \pm 0.094	0.473 \pm 0.101	0.498\pm0.104
Autism cross-site	0.652 \pm 0.075	0.662 \pm 0.068	0.671 \pm 0.093	0.675 \pm 0.070	0.678\pm0.077	0.659 \pm 0.091

Table 13: Effect of group atlas generation parameters on performance (200 regions). Values are shown as mean \pm standard deviation.

	0.2			0.3		
	0.7	0.8	0.9	0.7	0.8	0.9
α						
β						
γ						
Gender classification	0.703\pm0.063	0.687 \pm 0.073	0.676 \pm 0.061	0.678 \pm 0.042	0.682 \pm 0.044	0.686 \pm 0.043
Fluid intelligence	0.524 \pm 0.065	0.497 \pm 0.074	0.518 \pm 0.063	0.521 \pm 0.083	0.525 \pm 0.078	0.502 \pm 0.067
Cognitive task (7-way)	0.897 \pm 0.041	0.900\pm0.044	0.895 \pm 0.043	0.888 \pm 0.045	0.877 \pm 0.046	0.879 \pm 0.050
Cognitive task (24-way)	0.480\pm0.022	0.479 \pm 0.031	0.469 \pm 0.032	0.466 \pm 0.043	0.465 \pm 0.045	0.461 \pm 0.033
Autism diagnosis	0.672\pm0.057	0.663 \pm 0.040	0.654 \pm 0.035	0.657 \pm 0.036	0.650 \pm 0.035	0.646 \pm 0.046
AD diagnosis	0.448 \pm 0.119	0.456 \pm 0.107	0.478\pm0.119	0.378 \pm 0.148	0.407 \pm 0.124	0.407 \pm 0.132
FC stability	0.644 \pm 0.043	0.644 \pm 0.043	0.642 \pm 0.043	0.651 \pm 0.044	0.652 \pm 0.044	0.652 \pm 0.043
Fingerprinting	0.777\pm0.171	0.776 \pm 0.172	0.767 \pm 0.177	0.750 \pm 0.188	0.743 \pm 0.193	0.739 \pm 0.199
Age group classification	0.474\pm0.065	0.473 \pm 0.048	0.450 \pm 0.065	0.455 \pm 0.098	0.442 \pm 0.097	0.472 \pm 0.099
Crystallized intelligence	0.509 \pm 0.097	0.505 \pm 0.082	0.521 \pm 0.083	0.524 \pm 0.084	0.532 \pm 0.087	0.506 \pm 0.084
General intelligence	0.453 \pm 0.114	0.461 \pm 0.108	0.460 \pm 0.112	0.443 \pm 0.105	0.423 \pm 0.120	0.426 \pm 0.109
Autism cross-site	0.636 \pm 0.139	0.635 \pm 0.091	0.667 \pm 0.111	0.656 \pm 0.125	0.660 \pm 0.115	0.657 \pm 0.064

Table 14: Effect of group atlas generation parameters on performance (360 regions). Values are shown as mean \pm standard deviation.

	0.2			0.3			0.2			0.3		
	α	β	γ	α	β	γ	α	β	γ	α	β	γ
Gender classification	0.705 \pm 0.050	0.710 \pm 0.059	0.715\pm0.064	0.692 \pm 0.064	0.701 \pm 0.060	0.681 \pm 0.054	0.683 \pm 0.055	0.683 \pm 0.058	0.683 \pm 0.049	0.682 \pm 0.084	0.691 \pm 0.087	0.691 \pm 0.087
Fluid intelligence	0.521 \pm 0.068	0.535 \pm 0.084	0.505 \pm 0.094	0.517 \pm 0.086	0.528 \pm 0.077	0.528 \pm 0.076	0.532 \pm 0.096	0.554 \pm 0.083	0.539 \pm 0.093	0.555 \pm 0.078	0.562\pm0.090	0.562\pm0.090
Cognitive task (7-way)	0.888 \pm 0.045	0.887 \pm 0.042	0.879 \pm 0.043	0.900 \pm 0.040	0.901 \pm 0.037	0.898 \pm 0.046	0.902 \pm 0.049	0.905\pm0.049	0.899 \pm 0.051	0.898 \pm 0.054	0.898 \pm 0.049	0.893 \pm 0.052
Cognitive task (24-way)	0.462 \pm 0.023	0.469 \pm 0.037	0.460 \pm 0.039	0.474 \pm 0.031	0.473 \pm 0.024	0.458 \pm 0.036	0.477 \pm 0.035	0.478 \pm 0.031	0.473 \pm 0.030	0.480\pm0.036	0.472 \pm 0.034	0.472 \pm 0.034
Autism diagnosis	0.673 \pm 0.043	0.680 \pm 0.044	0.658 \pm 0.058	0.688\pm0.025	0.677 \pm 0.056	0.685 \pm 0.055	0.654 \pm 0.049	0.658 \pm 0.058	0.653 \pm 0.046	0.660 \pm 0.065	0.661 \pm 0.050	0.670 \pm 0.059
AD diagnosis	0.463 \pm 0.118	0.448 \pm 0.131	0.463 \pm 0.117	0.458 \pm 0.083	0.462 \pm 0.080	0.432 \pm 0.082	0.432 \pm 0.121	0.402 \pm 0.119	0.452 \pm 0.113	0.474\pm0.133	0.459 \pm 0.110	0.455 \pm 0.104
FC stability	0.615 \pm 0.043	0.615 \pm 0.043	0.614 \pm 0.043	0.615 \pm 0.043	0.615 \pm 0.043	0.614 \pm 0.043	0.615 \pm 0.043	0.615 \pm 0.043	0.614 \pm 0.043	0.618 \pm 0.044	0.619\pm0.043	0.618 \pm 0.043
Fingerprinting	0.853 \pm 0.161	0.852 \pm 0.164	0.851 \pm 0.161	0.857 \pm 0.159	0.854 \pm 0.166	0.852 \pm 0.167	0.863 \pm 0.155	0.861 \pm 0.157	0.864 \pm 0.154	0.870\pm0.144	0.866 \pm 0.147	0.866 \pm 0.149
Age group classification	0.448 \pm 0.071	0.433 \pm 0.079	0.448 \pm 0.090	0.471 \pm 0.080	0.479\pm0.073	0.465 \pm 0.067	0.471 \pm 0.084	0.477 \pm 0.099	0.467 \pm 0.075	0.449 \pm 0.067	0.458 \pm 0.062	0.449 \pm 0.070
Crystallized intelligence	0.510 \pm 0.115	0.516 \pm 0.117	0.496 \pm 0.122	0.527 \pm 0.103	0.550\pm0.114	0.520 \pm 0.110	0.532 \pm 0.089	0.528 \pm 0.074	0.537 \pm 0.098	0.481 \pm 0.085	0.505 \pm 0.086	0.502 \pm 0.076
General intelligence	0.452 \pm 0.092	0.446 \pm 0.085	0.455 \pm 0.087	0.441 \pm 0.098	0.453 \pm 0.089	0.476 \pm 0.114	0.462 \pm 0.117	0.478\pm0.101	0.461 \pm 0.107	0.425 \pm 0.059	0.432 \pm 0.065	0.454 \pm 0.053
Autism cross-site	0.672 \pm 0.114	0.696\pm0.136	0.663 \pm 0.126	0.639 \pm 0.177	0.672 \pm 0.111	0.668 \pm 0.082	0.672 \pm 0.109	0.686 \pm 0.150	0.665 \pm 0.133	0.646 \pm 0.211	0.639 \pm 0.203	0.646 \pm 0.205

Table 15: Effect of group atlas generation parameters on performance (500 regions). Values are shown as mean \pm standard deviation.

	0.2			0.3			0.2			0.3		
	α	β	γ	α	β	γ	α	β	γ	α	β	γ
Gender classification	0.711 \pm 0.081	0.702 \pm 0.079	0.707 \pm 0.056	0.708 \pm 0.057	0.710 \pm 0.057	0.719 \pm 0.041	0.742\pm0.056	0.728 \pm 0.067	0.714 \pm 0.069	0.703 \pm 0.051	0.702 \pm 0.060	0.700 \pm 0.062
Fluid intelligence	0.548 \pm 0.083	0.537 \pm 0.088	0.548 \pm 0.094	0.539 \pm 0.082	0.544 \pm 0.084	0.555\pm0.085	0.537 \pm 0.068	0.545 \pm 0.083	0.548 \pm 0.089	0.505 \pm 0.069	0.505 \pm 0.071	0.507 \pm 0.066
Cognitive task (7-way)	0.892 \pm 0.058	0.895 \pm 0.054	0.886 \pm 0.061	0.880 \pm 0.057	0.873 \pm 0.049	0.859 \pm 0.049	0.898\pm0.052	0.897 \pm 0.052	0.893 \pm 0.052	0.886 \pm 0.055	0.885 \pm 0.051	0.879 \pm 0.050
Cognitive task (24-way)	0.462 \pm 0.035	0.459 \pm 0.025	0.445 \pm 0.030	0.444 \pm 0.027	0.434 \pm 0.030	0.432 \pm 0.035	0.468\pm0.036	0.468\pm0.034	0.456 \pm 0.037	0.452 \pm 0.032	0.457 \pm 0.035	0.459 \pm 0.018
Autism diagnosis	0.649 \pm 0.070	0.661 \pm 0.060	0.647 \pm 0.047	0.633 \pm 0.089	0.590 \pm 0.051	0.635 \pm 0.050	0.655 \pm 0.059	0.667\pm0.040	0.640 \pm 0.063	0.575 \pm 0.070	0.580 \pm 0.109	0.572 \pm 0.085
AD diagnosis	0.448 \pm 0.092	0.459 \pm 0.086	0.455 \pm 0.094	0.455 \pm 0.095	0.429 \pm 0.092	0.474 \pm 0.093	0.504\pm0.105	0.489 \pm 0.113	0.497 \pm 0.114	0.448 \pm 0.101	0.471 \pm 0.107	0.478 \pm 0.077
FC stability	0.603 \pm 0.044	0.603 \pm 0.044	0.602 \pm 0.043	0.603 \pm 0.045	0.603 \pm 0.044	0.602 \pm 0.044	0.602 \pm 0.044	0.602 \pm 0.044	0.601 \pm 0.044	0.607\pm0.044	0.607\pm0.044	0.607\pm0.044
Fingerprinting	0.883 \pm 0.152	0.884 \pm 0.148	0.880 \pm 0.150	0.864 \pm 0.163	0.864 \pm 0.167	0.865 \pm 0.161	0.885\pm0.146	0.885\pm0.143	0.884 \pm 0.144	0.878 \pm 0.144	0.879 \pm 0.142	0.876 \pm 0.142
Age group classification	0.475 \pm 0.089	0.475 \pm 0.096	0.494\pm0.101	0.476 \pm 0.078	0.480 \pm 0.102	0.479 \pm 0.082	0.477 \pm 0.105	0.473 \pm 0.116	0.491 \pm 0.118	0.469 \pm 0.087	0.467 \pm 0.083	0.482 \pm 0.092
Crystallized intelligence	0.530 \pm 0.085	0.515 \pm 0.114	0.518 \pm 0.097	0.538 \pm 0.100	0.505 \pm 0.088	0.501 \pm 0.079	0.527 \pm 0.097	0.520 \pm 0.105	0.493 \pm 0.091	0.545\pm0.080	0.539 \pm 0.071	0.539 \pm 0.073
General intelligence	0.470 \pm 0.110	0.459 \pm 0.112	0.463 \pm 0.097	0.503 \pm 0.133	0.504 \pm 0.096	0.506\pm0.099	0.446 \pm 0.117	0.460 \pm 0.131	0.453 \pm 0.117	0.438 \pm 0.119	0.430 \pm 0.096	0.445 \pm 0.082
Autism cross-site	0.636 \pm 0.175	0.638 \pm 0.166	0.659\pm0.152	0.560 \pm 0.196	0.565 \pm 0.170	0.559 \pm 0.196	0.644 \pm 0.093	0.655 \pm 0.125	0.612 \pm 0.120	0.531 \pm 0.142	0.598 \pm 0.131	0.568 \pm 0.164

Our results reveal several trends. When the number of parcels is small (e.g., 100), using $\alpha = 0.2$ yields better performance, while $\alpha = 0.3$ becomes more advantageous as the resolution increases, possibly because finer parcellations require more voxels to capture individual variability—even if some voxels are less reliable. For β , a lower threshold ($\beta = 0.2$) tends to offer a marginal performance benefit, possibly due to increased distinctiveness between parcels. Although $\gamma = 0.7$ yields slightly better performance overall, we choose $\gamma = 0.8$ for group-level atlas generation to match the voxel coverage of MMP [11] and ensure comparability. Consequently, our final group atlas is generated using hyperparameters $\alpha = 0.2$, $\beta = 0.2$, and $\gamma = 0.8$.

259 8 Spatiotemporal masking strategy

260 To ensure that masked regions form contiguous blocks in the high-resolution volume, we first spatially
 261 downsample the original $96 \times 96 \times 96$ grid by a factor of 16 along each axis, yielding a coarse
 262 $6 \times 6 \times 6$ volume (time dimension T unchanged). We then sample two independent binary masks on
 263 this downsampled grid: a spatial mask with fraction x_s of voxels set to zero, and a temporal mask
 264 with fraction x_t of frames set to zero. We choose x_s and x_t so that the overall masking ratio satisfies

$$(1 - x_s)(1 - x_t) = 1 - r, \quad (3)$$

265 where r is the desired fraction of masked spatiotemporal volume (e.g. $r = 0.8$ for 80% masking [13]).
 266 Finally, we upsample these binary masks back to the original resolution by expanding each down-
 267 sampled voxel mask to a 16^3 block in space and each temporal mask entry to the corresponding
 268 contiguous frames. Applying the resulting mask to the full-resolution data yields large, continuous
 269 spatiotemporal occlusions, encouraging the encoder to reconstruct missing patches using both local
 270 and long-range context.

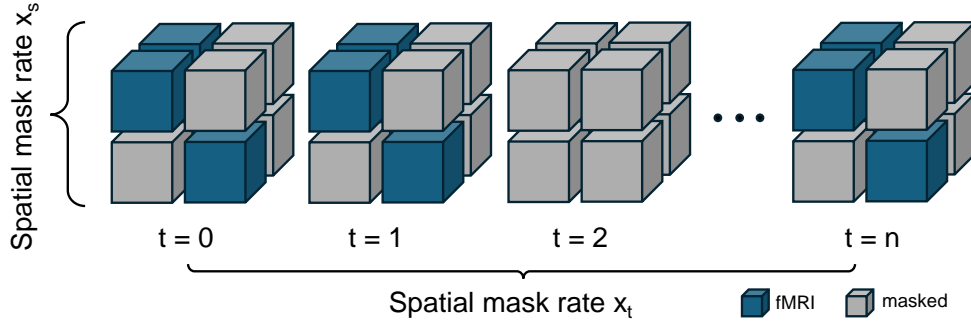


Figure 8: Illustration of the spatiotemporal masking scheme. We first select a fixed subset comprising x_s of spatial voxels and mask them to zero across all time steps; additionally, we mask all voxels at a fraction x_t of temporal frames to zero.

271 9 Data organization

272 There are two parts in our repository, DCA and ATlaScore. Please follow <https://anonymous.4open.science/r/DCA-03B1> for details.

274 9.1 DCA

275 The codes for pretraining and personalized clustering. One example subject is provided.

```
DCA/
├── data/
│   ├── fmri/
│   ├── mask/
│   ├── sub_test.txt
│   └── swin_model_epoch_30.pth
├── results/
│   └── demo/
├── ablation_fmri.py
├── main.py
├── utils.py
└── swin_unetr.py
```

276 9.2 AtlaScore

277 The codes for evaluating atlases. The volumetric Homogeneity and Silhouette evaluations, as well
278 as the surface-based DCBC evaluation, are provided. All 12 downstream tasks are implemented in
279 downstream/, with FC features for DCA100 provided as an example.

```
AtlaScore/
├── similarity/
│   ├── compute_adj.py
│   ├── eva.py
│   └── eva_DCBC.py
├── downstream/
│   ├── docs/
│   ├── fc_data/
│   ├── nii_data/
│   ├── downstream.py
│   └── demo.ipynb
```

References

- [1] Jianqiao Ge, Guoyuan Yang, Meizhen Han, Sizhong Zhou, Weiwei Men, Lang Qin, Bingjiang Lyu, Hai Li, Haobo Wang, Hengyi Rao, et al. Increasing diversity in connectomics with the chinese human connectome project. *Nature Neuroscience*, 26(1):163–172, 2023.
- [2] Bruce Fischl. Freesurfer. *Neuroimage*, 62(2):774–781, 2012.
- [3] Da Zhi, Maedbh King, Carlos R Hernandez-Castillo, and Jörn Diedrichsen. Evaluating brain parcellations using the distance-controlled boundary coefficient. *Human Brain Mapping*, 43(12):3706–3720, 2022.
- [4] 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.
- [5] David C Van Essen, Stephen M Smith, Deanna M Barch, Timothy EJ Behrens, Essa Yacoub, Kamil Ugurbil, Wu-Minn HCP Consortium, et al. The wu-minn human connectome project: an overview. *Neuroimage*, 80:62–79, 2013.
- [6] Adriana Di Martino, Chao-Gan Yan, Qingyang Li, Erin Denio, Francisco X Castellanos, Kaat Alaerts, Jeffrey S Anderson, Michal Assaf, Susan Y Bookheimer, Mirella Dapretto, et al. The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. *Molecular psychiatry*, 19(6):659–667, 2014.
- [7] Clifford R Jack Jr, Matt A Bernstein, Nick C Fox, Paul Thompson, Gene Alexander, Danielle Harvey, Bret Borowski, Paula J Britson, Jennifer L. Whitwell, Chadwick Ward, et al. The alzheimer’s disease neuroimaging initiative (adni): Mri methods. *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 27(4):685–691, 2008.
- [8] Stephanie Noble, Marisa N Spann, Fuyuze Tokoglu, Xilin Shen, R Todd Constable, and Dustin Scheinost. Influences on the test–retest reliability of functional connectivity mri and its relationship with behavioral utility. *Cerebral cortex*, 27(11):5415–5429, 2017.
- [9] 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.
- [10] Robert W Cox. Afni: software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical research*, 29(3):162–173, 1996.
- [11] 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.
- [12] 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.
- [13] Zhenda Xie, Zheng Zhang, Yue Cao, Yutong Lin, Jianmin Bao, Zhuliang Yao, Qi Dai, and Han Hu. Simmim: a simple framework for masked image modeling. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9643–9653, 2021.