SegCSR: Weakly-Supervised Cortical Surfaces Reconstruction from Brain Ribbon Segmentations

Anonymous Author(s)
Affiliation
Address
email

Abstract

Deep learning-based cortical surface reconstruction (CSR) approaches typically rely on supervision information provided by pseudo ground truth generated by conventional CSR methods, subject to errors associated with the supervision information and also increasing computational cost of training data preparation.We propose a new method to jointly reconstruct multiple cortical surfaces using *weak supervision* from brain MRI ribbon segmentation results. Our approach initializes a midthickness surface, which is then deformed inward and outward to form the inner (white matter) and outer (pial) cortical surfaces, respectively, by jointly learning diffeomorphic flows by minimizing loss functions to optimize the surfaces towards the boundaries of the cortical ribbon segmentation maps. Specifically, a boundary surface loss drives the initialization surface to the inner and outer boundaries, while an inter-surface normal consistency loss regularizes the pial surface in challenging deep cortical sulci regions. Additional regularization terms are utilized to enforce edge length uniformity and smoothness of the reconstructed surfaces. Our method has been evaluated on two large-scale adult brain MRI datasets and one infant brain MRI dataset, demonstrating comparable or superior performance in CSR in terms of accuracy and surface regularity compared to alternative supervised deep learning methods.

1 Introduction

Cortical surface reconstruction (CSR) is a crucial step for both qualitative visualization and quantitative characterization of cortical surfaces in imaging studies of brain morphology [15, 51], neurodegenerative diseases [6, 12, 43], and psychological disorders [42]. Well-established cortical analysis pipelines, such as BrainSuite [48], FreeSurfer [17], Connectome Workbench [18], and iBEAT V2.0 [52], have achieved significant success in reconstructing cortical surfaces from brain MRI data. However, these pipelines typically involve multiple processing steps, including iterative surface deformation and topology check and correction, resulting in lengthy processing time (e.g., $\sim$6h/subject). Moreover, each pipeline requires meticulously tuned parameters, posing challenges for generalization across diverse data domains, age groups, or acquisition protocols.

Deep learning (DL) approaches have significantly accelerated CSR, demonstrating orders of magnitude faster inference speeds while maintaining high accuracy and topology correctness [8, 11, 13, 22, 26, 30–32, 41, 47, 54]. One line of research predicts implicit surface representations, such as signed distance functions [13, 21] or level sets [41], from which 3D meshes are extracted using the Marching Cube (MC) algorithm [27] and refined with topology correction algorithms [4] to detect and rectify topology errors, ensuring that the reconstructed surface conforms to a sphere-like topology. Another line of research focuses on learning explicit surface deformations, using methods such as flow-based [8, 11, 22, 26, 47, 54] or NODE-based techniques [30, 31]), to deform an initial mesh towards target cortical surfaces. However, all these methods heavily rely on supervision information

Submitted to 38th Conference on Neural Information Processing Systems (NeurIPS 2024). Do not distribute.

provided by pseudo ground truth (pGT) of cortical surfaces generated by conventional CSR methods , regardless of whether they use implicit or explicit surface representations. The prolonged processing time for generating pGT surfaces limits the collection of sufficiently large datasets for training, and a general pipeline capable of extracting pGT surfaces across various data domains (e.g., age, modality) is currently lacking. Conversely, segmentation of brain structures is comparatively simpler, inspiring us to explore avenues to eliminate the need for supervised learning in CSR and to generalize *DL-based CSR* approaches to scenarios where ribbon segmentation results are readily available.

The key challenges for achieving accurate weakly supervised reconstruction of cortical surfaces span three primary aspects. ***First***, devising sub-voxel supervision signals presents a formidable hurdle. While existing approaches can produce precise segmentations [7, 20, 29, 45, 52], voxel-level representations may struggle to capture the intricate morphology of the cerebral cortex, especially its thin and highly-folded structure, due to the partial volume effect (PVE) inherent in brain MRI scans. This problem becomes particularly prominent in deep cortical sulci [17], where the two banks of grooves nearly converge, or in low-resolution images [52], such as under-sampled or infant MRIs. ***Second***, effectively modeling the interdependence between multiple surfaces is crucial. Incorporating this prior knowledge into the design of models and training algorithms can alleviate the complexity of reconstructing both the inner (white matter) and outer (pial) surfaces, ensuring the spherical topology of the reconstructed surfaces [8, 54]. However, in the absence of pGT, it becomes more challenging to forcibly deform surfaces and less stable to optimize multiple surfaces concurrently. ***Third***, maintaining optimal surface topology is paramount. Mesh uniformity, smoothness, and topology are susceptible to distortion during large deformations if networks are optimized based on randomly sampled vertices in 3D space for dense volumetric fields.

In this paper, we introduce ***SegCSR***, a novel weakly supervised DL framework aimed at reconstructing multiple cortical surfaces using ribbon segmentations derived from brain MRIs. We address the diffeomorphic deformation problem in a continues coordinate space, deforming the initialization midthickness surface towards the target inner and outer surfaces via innovative loss functions. Specifically, the boundary surface loss function based on the ribbon segmentations and the intensity gradient loss function based on the raw image facilitate sub-voxel-level surface movement. The inter-surface normal consistency loss function explicitly integrates the normal directions of the WM, midthickness, and pial surfaces, thereby regularizing the pial surface in challenging deep cortical sulci regions. Furthermore, we devise a customized edge length loss, in conjunction with the known normal consistency loss, to ensure surface uniformity and smoothness. Our main contributions can be summarized as follows:

- We propose a new weakly supervised paradigm for reconstructing multiple cortical surfaces, reducing the dependence on pGT cortical surfaces in training, unlike existing DL methods.
- We design two loss functions to optimize the surfaces towards the boundary of the cortical ribbon segmentation maps, along with regularization terms to enforce regularity of surfaces.
- We conduct extensive experiments on two large-scale adult brain MRI datasets and one infant brain MRI dataset. Our new method achieves comparable or superior performance compared to existing supervised DL-based CSR alternatives.

2 Related Works

Cortical Surface Reconstruction (CSR). (I) ***Traditional CSR methods*** typically rely on empirically defined automatic image/surface processing techniques to accomplish tissue segmentation (e.g., WM, GM, cerebrospinal fluid (CSF)), hemisphere separation, subcortical filling, topology correction, WM surface reconstruction, and pial surface reconstruction sequentially. Established pipelines such as FreeSurfer [17], BrainSuite [48], and HCP [18] are tailored for processing adult brain images, while dHCP [34] and iBEAT V2.0 [52] are designed for neonatal brain images, which exhibit distinct differences in intensity values, size, and shape compared to adult brains. Despite achieving sub-voxel accuracy and maintaining spherical topology, the iterative surface deformation and topology check and correction procedures lead to lengthy processing times. **(II)** ***DL-based CSR methods*** have significantly enhanced reconstruction speed while preserving high accuracy. Approaches like SegRecon [19] and DeepCSR [13] predict a signed distance map for implicit surface representation, embedding the target surface as the zero level-set and extracting it using MC algorithms. However, these methods require topology correction to eliminate artifacts and ensure spherical topology. Alternatively, PialNN [32], TopoFit [22], Vox2cortex [8], the CorticalFlow series [26, 47], SurfFlow [11], CortexODE [31], and CoCSR [54] leverage explicit representation to maintain good topology and overcome PVE by learning volumetric or vertex-wise diffeomorphic deformations and progressively deforming genus-0 template meshes. However, both implicit and explicit methods heavily rely on the supervision of pGT of cortical surfaces generated by traditional pipelines. Our proposed method is based on the explicit representation but differs significantly from them by utilizing ribbon segmentation maps for weakly supervising the model training process.

Weakly-/Un-supervised Mesh Reconstruction. Although geometric DL methods for general computer vision tasks have been extensively studied, research on mesh reconstruction from 3D images under weakly-/un-supervised settings is relatively underexplored. One approach involves constructing mesh-to-image rasterizer loss functions, as demonstrated in [36], where 2D projection views are extracted from predicted 3D meshes and compared with ground truth segmentations. Another line of research, exemplified by [39], focuses on learning the correspondence between a template image and a target image, which is then utilized to deform the template mesh to the target location. However, these methods have primarily been applied to biomedical tasks involving organs with relatively simple shapes, such as the liver and heart. But the cerebral cortex presents a highly-folded thin structure with a significantly complex shape, necessitating more advanced methods.

Diffeomorphic Deformation. Diffeomorphic deformation is a spatial transformation that guarantees both smoothness and invertibility in the mapping process [46]. It has been widely used in the modeling and analysis of brain morphometry, including image registration and surface reconstruction tasks. LDDMM [5] computes diffeomorphic deformation based on a time-dependent velocity vector field, while Arsigny et al. [2] employ a stationary velocity field (SVF) in conjunction with the scaling and squaring method to reduce computation complexity. Learning-based methods [3, 28, 38] improve the computation efficiency, with regularizations such as smoothness [3] and inverse-consistency [38] enchancing the diffeomorphic property of the deformation. In the CSR task, diffeomorphic deformation strategies have been adopted to solve an ordinary differential equation (ODE) modeling the trajectories of each vertex of a surface. For instance, CoticalFlow methods [26, 47] propose solving the ODE vertex-wise and derive a numerical condition to ensure homeomorphism of integration by training a *chain* of diffeomorphic deformation models in *sequential* stages. Recently, with the advances in neural ODE solver [10], CortexODE [31] parameterizes the trajectories of vertices on the surface as ODEs and proposes a pipeline to reconstruct WM and pial surfaces *sequentially*. Our method builds upon these works [31, 47, 54] and integrates multiple CSR tasks into a single framework, leveraging the efficiency and diffeomorphic properties of these strategies.

3 Methodology

Our proposed framework, depicted in Fig. 1, is designed to reconstruct multiple cortical surfaces simultaneously, eliminating the dependency on pGT generated by conventional and time-consuming CSR pipelines. We leverage as weak supervision the brain ribbon segmentation maps that are less accurate than pGT surfaces but more accessible. Section3.1 outlines the network structure that couples multiple cortical surfaces to reduce the learning difficulty. Section 3.2 describes the loss functions devised to supervise the network optimization, facilitating sub-voxel reconstruction accuracy and preserving optimal surface topology.

3.1 Coupled Cortical Surface Reconstruction

Existing supervised methods require pGT obtained from traditional CSR pipelines to provide precise sub-voxel supervision. They can effectively learn the deformation field, even from distant initial locations, to accurately align the initialization surface with the target surfaces [11, 26, 47]. However, brain ribbon segmentation maps are inherently discrete voxel grids, offering much coarser supervision. Consequently, the selection of the initialization surface becomes more critical. Moreover, given the intricate folded patterns of the cerebral cortex, the proximity of the two banks of grooves in deep cortical sulci often poses a considerable risk of generating topology errors (e.g., handles, holes) in the reconstructed surfaces. Conversely, voxels closer to the WM surface exhibit clearer contrast, enabling a distinct separation between sulci (Fig. 2 (b)). Thus, following [54], we opt for the midthickness layer, positioned midway between the WM and pial surfaces, to serve as a connection for coupling the reconstructions of both surfaces and achieve a balanced performance for both surfaces.

As illustrated in Fig. 1, SegCSR employs a neural network to jointly model three diffeomorphic flows: $F_\theta(I, \mathcal{S}_0) = (\mathbf{v}^m, \mathbf{v}^o, \mathbf{v}^i)$. Here, I represents a multi-channel input consisting of brain MRI, cortical

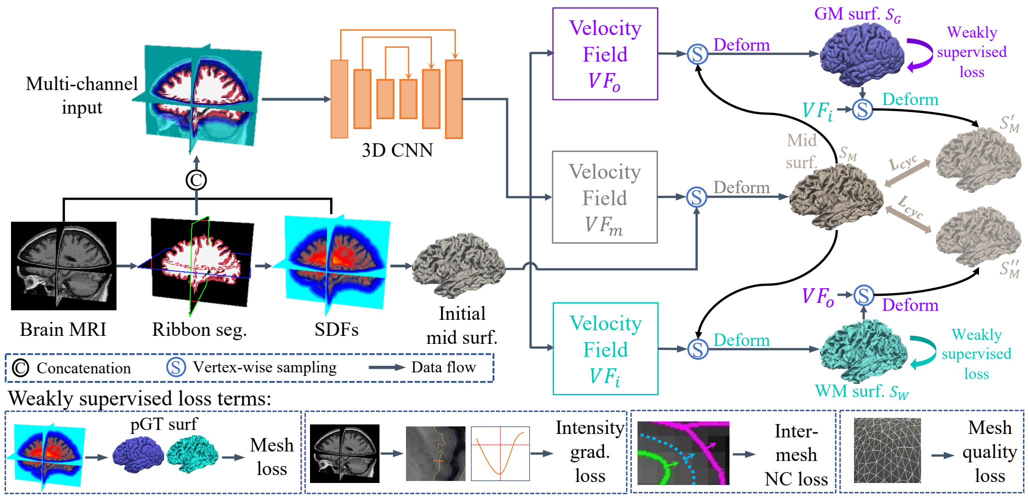


Figure 1: The SegCSR framework overview. SegCSR takes as input a brain MRI image, cortical ribbon segmentation maps, and signed distance maps of cortical surfaces, and simultaneously learns three diffeomorphic deformations to optimize the initial midthickness surface $\mathcal{S}_0$ to align with the target midthickness surface $\mathcal{S}_M$, and then deform $\mathcal{S}_M$ outwards and inwards to the pial surface $\mathcal{S}_G$ and the WM surface $\mathcal{S}_W$, respectively. The model is optimized using weakly supervised loss functions: the mesh loss guides the surfaces towards the boundaries of the cortical ribbon segmentation maps; the inter-surface normal consistency loss regularizes the pial surface in deep cortical sulci; the intensity gradient loss facilitates sub-voxel-level movement; and additional regularization terms control the deformation trajectories of multiple surfaces as well as the uniformity and smoothness of the surfaces.

ribbon masks, and signed distance functions (SDFs); $\mathcal{S}_0$ denotes the initialization midthickness surface; and $\mathbf{v}^m$, $\mathbf{v}^o$, $\mathbf{v}^i$ correspond to the velocity fields that drive $\mathcal{S}_0$ towards the true midthickness surface $\mathcal{S}_M$, outward to the pial surface $\mathcal{S}_G$, and inward to the WM surface $\mathcal{S}_W$, respectively. The SegCSR establishes an explicit *one-to-one mapping* between multiple surfaces and is trained by minimizing weakly supervised losses between the predicted mesh and the ribbon segmentations.

The diffeomorphic deformation between the initialization surface and the target surface can be computed as the integration of an ODE [1] based on the velocity field $\mathbf{v}$:

$$\frac{d\Phi(\mathbf{x},t)}{dt} = \mathbf{v}(\Phi(\mathbf{x},t),t) \text{ s.t. } \Phi(\mathbf{x},0) = \mathbf{x}^{(0)}, \text{and thus } \Phi(\mathbf{x},t) = \mathbf{x}^{(0)} + \int_o^t \mathbf{v}(\Phi(\mathbf{x},s),s)ds, \quad (1)$$

where $\Phi(\mathbf{x},t)$ defines a trajectory from the source position $\mathbf{x}^{(0)} = \Phi(\mathbf{x},0)$ to the target position $\mathbf{x}^{(1)} = \Phi(\mathbf{x},1)$. According to the *Cauchy-Lipschitz* theorem [50], if the velocity field is Lipschitz continuous, the resulting mapping Φ is bijective with continuous inverse (i.e., a diffeomorphism). To solve this initial value problem, we perform the integration on the predicted velocity fields using standard numerical integration techniques, such as the Euler method and the Runge-Kutta method [9]. Specifically, for each integration step $t \in [0,1]$, each vertex's coordinates can be updated by $\mathbf{x}^{(t+1)} = \mathbf{x}^{(t)} + h\mathbf{v}(\Phi(\mathbf{x},t),t)$, where $h = \frac{1}{T}$ is the step size and T is the total time steps, and the velocity vector $\mathbf{v}$ for a vertex is trilinearly interpolated from its neighboring velocity vectors [54].

3.2 Weak Supervision Loss Functions

Mesh Loss. Weak supervision for SegCSR is derived from cortical ribbon segmentation maps of WM and GM (see Fig. 1, the filled interior area of WM and pial surfaces), which can be obtained from existing segmentation approaches [7, 20, 29, 45, 52]. Although these ribbon segmentation maps do not perfectly represent the intricate pial surface, the WM surface is relatively easier to recognize due to its clear local intensity contrast, providing a better-separable boundary (see Fig. 2 (a-b)). Therefore, we use the boundary of the pGT WM segmentation to supervise the WM surface reconstruction. Inspired by [31, 54], we generate an SDF for the WM surface by using a distance transform algorithm, where voxels with values of zero represent the surface boundaries and voxels with negative or positive values encode their distances to the surface boundaries inward or outward,

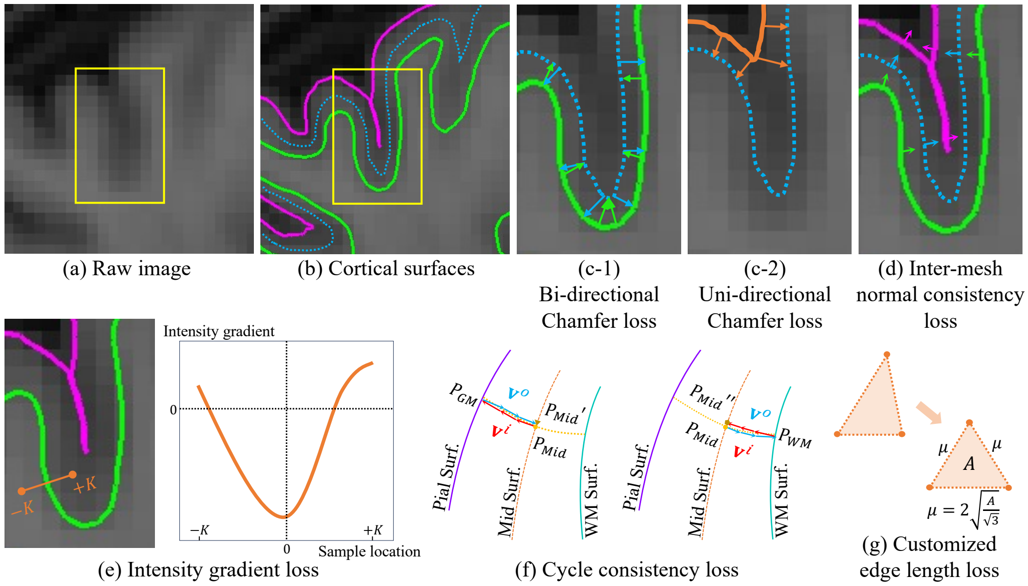


Figure 2: (a) A brain MRI region. (b)-(g) are illustration of loss terms. (b) WM, midthickness, pial surfaces in a deep sulcus region. (c-1) Bi-directional Chamfer loss for the WM surface; (c-2) Uni-directional Chamfer loss for the pGT pial surface generated from the GM segmentation. (d) Normal consistency between three reconstructed surfaces. (e) Intensity gradient along the normal direction of a vertex in the surface. (f) The symmetric deformation trajectory. $\mathbf{v}^o$ and $\mathbf{v}^i$ are outward and inward velocity fields respectively. (g) The customized edge length loss. A: area; μ: edge length.

respectively. We then apply a fast topology check and correction algorithm [4] to the SDF to ensure the surface maintains spherical topology. The WM surface $\mathcal{S}_{W_*}$ is extracted using the Marching Cubes algorithm [27]. The distance of the vertices between the predicted surface $\mathcal{S}_W$ and the pGT surface $\mathcal{S}_{W_*}$ is minimized using the bi-directional Chamfer distance [26]:

$$\mathcal{L}_{chW} = \frac{1}{|\mathcal{S}_W|}\sum_{\mathbf{p}\in\mathcal{S}_W}\min_{\mathbf{p}_*\in\mathcal{S}_{W_*}}\|\mathbf{p}-\mathbf{p}_*\|_2^2 + \frac{1}{|\mathcal{S}_{W_*}|}\sum_{\mathbf{p}_*\in\mathcal{S}_{W_*}}\min_{\mathbf{p}\in\mathcal{S}_W}\|\mathbf{p}_*-\mathbf{p}\|_2^2, \tag{2}$$

where $\mathbf{p}$ and $\mathbf{p}_*$ are the coordinates of vertices on meshes. See Fig. 2 (c-1) for illustration.

For the pial surface, GM segmentation may fail to delineate the boundary in deep cortical sulci. As shown in Fig. 2 (c-2), using a similar pGT surface generation protocol as the WM surface to generate the pial surface $\mathcal{S}_{G_*}$ fail to capture cortical folding accurately. Directly fitting to $\mathcal{S}_{G_*}$ with bi-directional Chamfer loss causes the model to predict similarly inaccurate cortical sulci. To address this issue, we propose the boundary surface loss, which uses a uni-directional Chamfer distance to compute the shortest distance from the pGT pial surface $\mathcal{S}_{G_*}$ to the predicted pial surface $\mathcal{S}_G$:

$$\mathcal{L}_{chG} = \frac{1}{|\mathcal{S}_G|}\sum_{\mathbf{p}\in\mathcal{S}_G}\min_{\mathbf{p}_*\in\mathcal{S}_{G_*}}\|\mathbf{p}-\mathbf{p}_*\|_2^2. \tag{3}$$

In this way, the deformed surface is not influenced by the inaccuracies of $\mathcal{S}_{G_*}$ and does not move outward from the deep sulci. The overall mesh loss is computed as $\mathcal{L}_{mesh} = \mathcal{L}_{chW} + \mathcal{L}_{chG}$.

Inter-Mesh Normal Consistency Loss. To further alleviate the difficulty of constraining the pial surface using the WM and midthickness surfaces, we propose leveraging the prior knowledge that the cerebral cortex has a sheet-like topology (i.e., the inner, middle, and outer surfaces are locally parallel to each other). As shown in Fig. 2 (d), this loss is defined to ensure that the deformation of the midthickness surface aligns with its normal direction, thereby maintaining similar normal directions on the target surfaces:

$$\mathcal{L}_{imnc} = \frac{1}{|\mathcal{S}_M|}\sum_{\mathbf{p}\in\mathcal{S}_M}(1-cos(\mathbf{n}_{\mathbf{p}_G},\mathbf{n}_{\mathbf{p}_W})), \tag{4}$$

where $\mathbf{n}_{\mathbf{p}_G}$ and $\mathbf{n}_{\mathbf{p}_W}$ are the normal vectors of the deformed vertex $\mathbf{p}$ on $\mathcal{S}_M$ and $\mathcal{S}_G$ respectively.

Intensity Gradient Loss. In addition to ribbon segmentaions, inspired by the fact that traditional methods utilize raw image intensity contrast to define and optimize the target surfaces, we propose to adjust the nuance between GT target surface and the pGT segmentation boundaries. By definition [17, 52], the WM (or pial) surface lies at the WM/GM (or GM/CSF) interface where image intensity change most drastically. We sample K points along the extended lines on each side of the normal direction at vertex $\mathbf{p}$, and compute the gradients of neighboring points: $\mathcal{L}_{grad} = \frac{1}{|\mathcal{S}_W|}\sum_{\mathbf{p}\in\mathcal{S}_M}\sum_{i=1}^{K} grad_i(\mathbf{p}) + \frac{1}{|\mathcal{S}_G|}\sum_{\mathbf{p}\in\mathcal{S}_G}\sum_{i=1}^{K} grad_i(\mathbf{p})$.

Cycle Consistency Loss. We utilize the midthickness layer to establish a correspondence between the inner and outer surfaces, thereby reducing the difficulty of learning large deformations. However, there is no true midthickness surface available for supervision, nor a definitive criterion for choosing between bi-directional or uni-directional approaches for different regions on the midthickness surface. Additionally, the learned velocity fields $\mathbf{v}^o$ and $\mathbf{v}^i$ could potentially cause non-inverse transformations at the midthickness surface. To address these issues, we propose a loss function that enforces the midthickness surface resides halfway between the WM and pial surfaces and maintains consistency along the entire trajectory:

$$\mathcal{L}_{cyc} = \frac{1}{|\mathcal{S}_M|}\sum_{\mathbf{p}\in\mathcal{S}_M} \|\mathbf{p}_{\Phi_W\circ\Phi_G} - \mathbf{p}\|_2^2 + \|\mathbf{p}_{\Phi_G\circ\Phi_W} - \mathbf{p}\|_2^2 + \|L_{Mid\to GM}(\mathbf{p}) - L_{Mid\to WM}(\mathbf{p})\|_2^2, \quad (5)$$

where $\mathbf{p}_{\Phi_b\circ\Phi_a}$ represents deforming a vertex $\mathbf{p}\in\mathcal{S}_M$ with velocity field $\mathbf{v}^a$ and $\mathbf{v}^b$ sequentially, and $L_{Mid\to GM}(\mathbf{p})$ is the accumulated trajectory length over T steps of deformation. For example, as shown in Fig. 2 (f), the deformations move a vertex $\mathbf{p}_{Mid}$ outward to $\mathbf{p}_{GM}$ using $\mathbf{v}^o$ and then inward to $\mathbf{p}'_{Mid}$ using $\mathbf{v}^i$, in which the two trajectories are aligned by minimizing the distance between $\mathbf{p}_{Mid}$ and $\mathbf{p}'_{Mid}$. Similarly, we enforce the consistency between $\mathbf{p}_{\Phi_G\circ\Phi_W}$ and $\mathbf{p}$. Furthermore, starting from the midthickness layer, the trajectory lengths of the vertex moving to the WM and pial surfaces should be equal, which is regularized by the third term in the equation above.

Mesh Quality Loss. First, the reconstructed surface should be composed of uniformally distributed triangles. To accommodate various sizes of brain volume and image resolution, we devise a *customized edge length loss* to constrain the size of triangles in the predicted meshes for each subject. Specifically, we assume an ideal prediction where the faces are equilateral and of the same area A and drive the edge length to the target edge length $\mu_{el} = 2\sqrt{\frac{A}{\sqrt{3}}}$ (see Fig. 2 (g)). Second, we employ a *normal consistency loss* to promote the surfaces' smoothness. The mesh quality loss is defined as:

$$\mathcal{L}_{qua} = \frac{1}{|S|}\left(\sum_{\mathbf{p}\in\mathcal{S}}\frac{1}{|\mathcal{N}(\mathbf{p})|}\sum_{\mathbf{k}\in\mathcal{N}(\mathbf{p})}(\mu_{el} - \|\mathbf{p}-\mathbf{k}\|_2)^2 + \sum_{e\in\mathcal{S}, f_0\cap f_1=e}(1 - cos(\mathbf{n}_{f_0}, \mathbf{n}_{f_1}))\right), \quad (6)$$

where $\mathcal{S}$ denotes the predicted mesh, $\mathcal{N}(\mathbf{p})$ are the neighbors of vertex $\mathbf{p}$, e is an edge, f_0 and f_1 are e's two neighboring faces with their unit normals $\mathbf{n}_{f_0}$ and $\mathbf{n}_{f_1}$.

In summary, we combine all the losses to jointly optimize our SegCSR model: $\mathcal{L} = \lambda_1\mathcal{L}_{mesh} + \lambda_2\mathcal{L}_{imnc} + \lambda_3\mathcal{L}_{grad} + \lambda_4\mathcal{L}_{cyc} + \lambda_5\mathcal{L}_{qua}$, where $\{\lambda_i\}_{i=1,\cdots,5}$ are weights to balance the loss terms.

4 Experiments

4.1 Experimental Setups

Datasets. We evaluate our method on two large-scale adult datasets and one infant dataset of low resolution. The ADNI-1 [24] dataset consists of 817 subjects aged 55 to 90. We randomly split it into subsets of 654, 50, and 113 subjects for training, validation, and testing, respectively. The OASIS-1 [35] dataset consists of 413 subjects aged 18 to 96. We randomly split it into subsets of 330, 25, and 58 subjects for training, validation, and testing, respectively. We followed a pre-processing protocol used in previous works [8, 13, 26, 31] for fair comparison. The T1-weighted MRI scans were aligned to the MNI152 template and clipped to the size of $192\times224\times192$ at $1mm^3$ isotropic resolution. The pseudo ground-truth (pGT) of ribbon segmentation and cortical surfaces were generated using FreeSurfer v7.2.0 [17]. The BCP [23] dataset consists of 121 subjects ranging in age from 2 weeks to 12 months. We randomly allocate 90, 12, and 19 subjects for training, validation, and testing,

Table 1: Quantitative analysis of cortical surface reconstruction on geometric accuracy and self-intersections. The Chamfer distance (CD), average symmetric surface distance (ASSD), Hausdorff distance (HD), and the ratio of the self-intersecting faces (SIF) were measured for WM and pial surfaces on three datasets. The mean value and standard deviation are reported. Lower scores indicate better results for all metrics. “S” denotes the use of pGT surfaces from conventional pipelines, while “W” represents weak supervision by pGT ribbon segmentations. In each supervision setting, the best results are in bold, and the second best results are underlined.

Data	Sup.	Method	L-Pial Surface				L-WM Surface			
			CD (mm)	ASSD (mm)	HD (mm)	SIF (%)	CD (mm)	ASSD (mm)	HD (mm)	SIF (%)
ADNI	S	CorticalFlow++ [47]	0.545±0.036	0.410±0.033	0.886±0.069	0.098±0.067	0.544±0.034	0.401±0.030	0.878±0.066	0.069±0.042
		cortexODE [31]	0.476±0.017	0.214±0.020	0.455±0.058	0.022±0.012	0.458±0.016	0.192±0.015	0.436±0.014	0.015±0.011
		Vox2Cortex [8]	0.582±0.028	0.370±0.025	0.746±0.057	0.059±0.039	0.577±0.027	0.353±0.022	0.722±0.055	0.043±0.023
		CoCSR [54]	**0.322**±0.021	**0.123**±0.010	**0.267**±0.022	**0.013**±0.011	**0.303**±0.018	**0.117**±0.010	**0.254**±0.021	**0.005**±0.002
	W	DeepCSR [13]	0.945±0.078	0.593±0.065	1.149±0.203	\	0.938±0.076	0.587±0.064	1.137±0.193	\
		3D U-Net [44]	0.598±0.049	0.341±0.037	0.782±0.163	\	0.473±0.013	0.265±0.015	0.558±0.028	\
		SegCSR (Ours)	**0.578**±0.019	**0.324**±0.019	**0.749**±0.049	0.008±0.009	**0.467**±0.014	**0.258**±0.019	**0.545**±0.036	0.009±0.009
OASIS	S	CorticalFlow++ [47]	0.531±0.035	0.399±0.030	0.812±0.057	0.088±0.045	0.529±0.033	0.398±0.030	0.810±0.055	0.086±0.042
		cortexODE [31]	0.481±0.019	0.218±0.021	0.461±0.062	0.026±0.015	0.463±0.018	0.207±0.017	0.435±0.015	0.018±0.010
		Vox2Cortex [8]	0.588±0.032	0.381±0.030	0.750±0.063	0.061±0.037	0.581±0.028	0.375±0.027	0.731±0.059	0.046±0.027
		CoCSR [54]	**0.410**±0.034	**0.142**±0.016	**0.281**±0.024	**0.016**±0.012	**0.349**±0.024	**0.128**±0.019	**0.266**±0.022	**0.007**±0.002
	W	DeepCSR [13]	0.986±0.085	0.617±0.070	1.331±0.212	\	0.975±0.081	0.594±0.067	1.151±0.197	\
		3D U-Net [44]	0.611±0.069	0.332±0.050	0.774±0.267	\	0.454±0.013	0.245±0.017	0.489±0.031	\
		SegCSR (Ours)	**0.581**±0.016	**0.321**±0.018	**0.725**±0.040	0.010±0.010	**0.449**±0.011	**0.223**±0.016	**0.461**±0.027	0.010±0.009
BCP	S	CorticalFlow++ [47]	0.927±0.271	0.731±0.036	1.943±0.175	1.114±0.385	0.895±0.242	0.722±0.034	1.880±0.151	0.533±0.107
		cortexODE [31]	0.759±0.082	0.396±0.032	0.823±0.103	0.124±0.061	0.678±0.071	0.349±0.031	0.816±0.099	0.101±0.034
		CoCSR [54]	**0.576**±0.041	**0.216**±0.023	**0.468**±0.063	**0.064**±0.040	**0.544**±0.038	**0.199**±0.020	**0.447**±0.049	**0.058**±0.033
	W	DeepCSR [13]	2.673±1.131	1.224±0.215	3.112±1.218	\	1.440±0.521	0.428±0.051	0.933±0.118	\
		3D U-Net [44]	1.175±0.314	0.793±0.059	2.140±1.021	\	0.688±0.120	0.377±0.041	0.791±0.064	\
		SegCSR (Ours)	**0.927**±0.070	**0.497**±0.061	**1.287**±0.144	0.061±0.058	**0.876**±0.067	**0.478**±0.052	**1.206**±0.132	0.055±0.057

respectively. Rigid registration was applied to the T1w and T2w image pairs. The pGT of ribbon segmentation and cortical surfaces were generated by the iBEAT v2.0 [52]. The intensity values of MRI scans, ribbon segmentation maps, and SDFs were normalized to $[0, 1]$ and the coordinates of the vertices were normalized to $[-1, 1]$. All the models were trained on the training set until they reached a loss plateau on the validation set and evaluated on the test set.

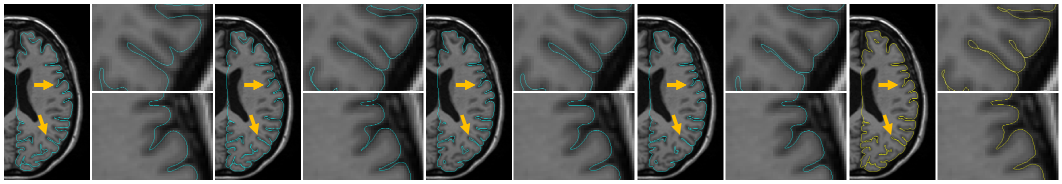
Implementation Details Our framework was implemented in PyTorch [40] and trained on a workstation with 12 GB NVIDIA P100 GPU. The 3D U-Net [44] for segmentation of ribbons was trained for 200 epochs using Adam [25] optimization and achieved an average Dice index of 0.96 on the testing set. The SegCSR model utilized $T = 5$ steps (i.e., step size is 0.2) in Euler solver. We trained our SegCSR model using Adam optimizer ($\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon = 1e^{-10}$, learning rate $1e^{-4}$) for 400 epochs to reconstruct both WM, midthickness, and pial surfaces of both brain hemispheres. We set $\lambda_1 = \lambda_4 = 1$ and $\lambda_2 = \lambda_3 = \lambda_5 = 0.1$. The surface meshes had $\sim 130k$ vertices. More details can be found in the Supplementary Materials.

Evaluation Metrics We utilized three distance-based metrics to measure the CSR accuracy: Chamfer distance (CD), average symmetric surface distance (ASSD), and 90th-percentile Hausdorff distance (HD). CD [16, 53] measures the mean distance between two sets of vertices. ASSD [13] and HD [13, 49] measure the average and maximum distance between two surfaces. They were computed bidirectionally over $\sim 130k$ points uniformly sampled from the predicted and target surfaces. A lower distance means a better result. Since topology is also important in CSR, we utilized the ratio of self-intersection faces (SIF) [13, 14, 31, 54] to measure reconstructed surface quality.

4.2 Comparison with Related Works

We compare SegCSR with both implicit and explicit learning-based cortical surface reconstruction approaches described in Section 1 and summarize the experimental results in Table 5.

On Adult Datasets. ***(I) Comparison with Implicit Approaches***. We compare SegCSR with two representative implicit representation approaches on the ADNI and OASIS datasets. As shown in Table 5, SegCSR achieves superior geometric accuracy. Note that both DeepCSR [13] and 3D U-Net [44] require post-processing to correct topology and extract a mesh, resulting in SIFs of 0. Without post-processing, the SIFs for 3D U-Net's WM and pial surfaces range from 3% to 15%. SegCSR produces a negligible number of self-intersecting faces, $\sim 0.3\%$ on average for both white and pial surfaces. Fig. 3 shows that SegCSR effectively deforms the pial surface into deep sulci, while the baseline approaches exhibit large geometric errors due to the PVE problem of brain MRI.



(a) DeepCSR (b) CortexODE (c) SegCSR (S0 setting) (d) SegCSR (S0* setting) (e) pGT from FreeSurfer

Figure 3: Visualization of reconstructed pial surfaces compared to DeepCSR and CortexODE. CortexODE is trained with pGT from FreeSurfer; DeepCSR and ours are trained with pGT ribbon segmentations.

Additionally, SegCSR requires only 0.37s of runtime per brain hemisphere, orders of magnitude faster than traditional FreeSurfer pipelines. ***(II) Comparison with Explicit Approaches***. We compare SegCSR with explicit learning-based approaches, including CorticalFlow++ [47], Vox2Cortex [8], CortexODE [31], and CoCSR [54]. These methods are trained with pGT surfaces generated by conventional pipelines, providing more accurate supervision than pGT ribbon segmentations. For a fair comparison, we employ the same network structure for the current best CoCSR [54] and our SegCSR, with CoCSR serving as an upper-bound performance benchmark for our weakly supervised SegCSR. As shown in Table 5, SegCSR surprisingly surpasses some supervised baselines in terms of both geometric and morphological accuracy, demonstrating its potential to replace existing methods when accurate surface supervision is not available.

On Infant Dataset. Infant brain MRIs present additional challenges due to the smaller size of fetal brains, limited image resolution, and lower image contrast, which together make the reconstruction task more difficult. Consequently, overall performance is inferior compared to adult datasets. We compare SegCSR with both implicit and explicit representation approaches. The results in Table 5 show that SegCSR achieves superior performance than the implicit DeepCSR and 3D U-Net methods, and comparable performance to explicit methods like CorticalFlow++, CortexODE, and CoCSR.

4.3 Ablation Studies

Table 2: Ablation studies on the ADNI dataset. The setting S0 refers to our complete setting (cf. Table 5). Top: The impact of loss functions. Bottom: The impact of initialization surface location.

Setting	Loss					L-Pial Surface				L-WM Surface			
	$\mathcal{L}_{mesh}$	$\mathcal{L}_{imnc}$	$\mathcal{L}_{grad}$	$\mathcal{L}_{cyc}$	$\mathcal{L}_{qua}$	CD (mm)	ASSD (mm)	HD (mm)	SIF (%)	CD (mm)	ASSD (mm)	HD (mm)	SIF (%)
S0	✓	✓	✓	✓	✓	0.578±0.019	0.324±0.019	0.749±0.049	0.008±0.009	0.467±0.014	0.258±0.019	0.545±0.036	0.009±0.009
S1	✓	✓	✓	✓		0.576±0.019	0.323±0.019	0.747±0.046	0.012±0.011	0.467±0.015	0.257±0.020	0.542±0.036	0.011±0.011
S2	✓	✓	✓			0.579±0.019	0.325±0.019	0.748±0.047	0.014±0.013	0.469±0.016	0.248±0.019	0.544±0.042	0.015±0.014
S3	✓	✓				0.579±0.020	0.325±0.021	0.749±0.050	0.018±0.014	0.473±0.013	0.249±0.018	0.544±0.039	0.017±0.013
S4	✓					0.589±0.034	0.356±0.039	0.764±0.067	0.015±0.012	0.473±0.012	0.256±0.020	0.564±0.042	0.014±0.013
S0*	✓*	✓	✓	✓	✓	0.607±0.034	0.327±0.024	0.752±0.077	0.026±0.016	0.469±0.015	0.258±0.020	0.547±0.038	0.020±0.015
S4*	✓*					0.626±0.053	0.321±0.039	0.773±0.168	0.034±0.025	0.476±0.013	0.256±0.018	0.562±0.034	0.031±0.017

Init. Surface Location	L-Pial Surface				L-WM Surface			
	CD (mm)	ASSD (mm)	HD (mm)	SIF(%)	CD (mm)	ASSD (mm)	HD (mm)	SIF(%)
WM	0.878±0.077	0.587±0.060	1.084±0.097	0.012±0.011	0.439±0.011	0.211±0.013	0.430±0.028	0.007±0.008
Mid	0.578±0.019	0.324±0.019	0.749±0.049	0.008±0.009	0.467±0.014	0.258±0.019	0.545±0.036	0.009±0.009
GM	0.489±0.016	0.317±0.018	0.567±0.044	0.008±0.008	0.889±0.085	0.597±0.059	1.211±0.104	0.020±0.018

Loss Functions. We evaluated the contribution of different losses of our method to the surface reconstruction performance in terms of both accuracy (CD, ASSD, HD) and topological correctness (SIF). The results are summarized in Table 2 (Top). The setting S4 represents using our proposed Chamfer loss (i.e., uni-directional for the pial surface) alone, while S4* referes to using existing bi-directional Chamfer loss for both WM and pial surfaces. The results of S4 and S4* indicated that the model using bi-directional Chamfer loss overfitted to the pGT segmentation boundary and failed to fit the deep cortical sulci. Another pair of comparison, S0 and S0*, showed a similar phenomenon. Enforcing the inter-mesh normal consistency of the WM and pial surfaces (S3, $\mathcal{L}_{imnc}$) improved geometric accuracy by explicitly constraining the nromal direction of two surfaces but slightly worsened the topology, which might be caused by the discrepancy between the midthickness and the WM (and pial) surface. The proposed intensity gradient loss (S2, $\mathcal{L}_{grad}$) helped adjust the deformed surfaces locally, leading to slightly improved geometric accuracy and reduced topology error. Enforcing equality of the trajectories from the midthickness surface to the WM and pial surfaces and symmetric cycle consistency of two trajectories (S1, $\mathcal{L}_{cyc}$) helped optimize the midthickness surface and promoted the invertibility of deformations. Moreover, the inclusion of regularization terms on the uniformity and smoothness of the reconstructed surfaces (S0, $\mathcal{L}_{qua}$) enhanced the surface quality and significantly reduce the self-intersection face ratio. Overall, our proposed method struck a balance between geometric accuracy and topology quality, with each component playing a complementary role.

Initialization Surface Location. Table 2 (Bottom) shows the impact of the initialization surface location. Starting from either the WM or midthickness surfaces leads to satisfactory results. Conversely, initializing from the GM surface introduced more difficulty in learning large deformations into deep sulci due to the severe partial volume effect, resulting in worse average geometric accuracy for both surfaces. The results also indicated that the closer the initial surface was to its target surface, the higher the reconstruction accuracy achieved. Therefore, starting from the midthickness surface strikes a balance between WM and pial surface reconstruction outcomes.

4.4 Reproducibility

We conducted an experiment on the Test-Retest dataset [33], which comprises 40 MRIs collected within a short period for each of the 3 subjects. The cortical surfaces of the same subject should be nearly identical. Following the experimental setup outlined in [8, 13, 31, 54], we utilized the iterative closest-point algorithm to align image pairs and computed the geometric distance between surfaces. The results for the left hemisphere are presented in Table 3, showing that SegCSR obtained superior reproducibility compared with DeepCSR (implicit representation; weakly supervised) and was comparable to the conventional FreeSurfer pipeline and supervised DL-based CSR methods. This implied that the results generated by SegCSR can be reliably used for downstream analyses, such as investigating cortical thickness changes in patients.

Table 3: Reproducibility analysis.

Method	L-WM Surface		
	CD (mm)	ASSD (mm)	HD (mm)
SegCSR (Ours)	0.473±0.016	0.254±0.024	0.520±0.062
DeepCSR	0.505±0.047	0.297±0.053	0.610±0.100
CoCSR	0.451±0.019	0.235±0.030	0.492±0.059
CortexODE	0.457±0.021	0.238±0.031	0.504±0.071
FreeSurfer	0.476±0.015	0.253±0.022	0.519±0.048

Method	L-Pial Surface		
	CD (mm)	ASSD (mm)	HD (mm)
SegCSR (Ours)	0.529±0.023	0.285±0.033	0.622±0.066
DeepCSR	0.560±0.055	0.341±0.060	0.668±0.118
CoCSR	0.493±0.024	0.276±0.036	0.573±0.070
CortexODE	0.506±0.029	0.272±0.034	0.581±0.079
FreeSurfer	0.526±0.021	0.283±0.032	0.595±0.068

5 Conclusions

We introduce SegCSR, a novel approach to jointly reconstruct multiple cortical surfaces using weak supervision from ribbon segmentations derived from brain MRIs. Our method initializes a midthickness surface and then deforms it inward and outward to the inner and outer cortical surfaces by jointly learning diffeomorphic flows. The new boundary loss function optimizes the surfaces toward the boundaries of the cortical ribbon segmentation maps while the inter-surface normal consistency loss regularizes the pial surface in complex and challenging cortical sulci regions. Additional regularization terms are incorporated to enforce reconstructed surfaces' uniformity, smoothness, and topology. Extensive experiments conducted on large-scale adult and infant brain MRI datasets demonstrate superior performance in terms of accuracy and surface regularity compared to existing supervised DL-based alternatives.

Limitations and Future Directions. The efficacy of SegCSR is influenced by the quality of pGT segmentations. Also, We can utilize brain tissue segmentation as auxiliary functions to supervise the model training. SegCSR constrains the inter-mesh consistency of the deformation on the midthickness surface, potentially affecting anatomical fidelity of pial surfaces. The method should be tested on more diverse cohorts of subjects to demonstrate its efficacy on real world neuroimage analysis tasks.

Societal Impact. Our proposed method has been rigorously evaluated on four real-world brain MRI datasets, showcasing its capacity to assist doctors and scientists in both quantitative and qualitative analyses of the cerebral cortex. Nonetheless it is imperative to conduct more thorough evaluation on a larger cohort of subjects and across various imaging qualities. And the deployment of the model in clinical settings should be approached with caution and under human supervision.

References

[1] V. I. Arnold. *Ordinary differential equations*. Springer Science & Business Media, 1992.

[2] V. Arsigny, O. Commowick, X. Pennec, and N. Ayache. A log-euclidean framework for statistics on diffeomorphisms. In *International Conference on Medical Image Computing and Computer Assisted Intervention*, pages 924–931. Springer, 2006.

[3] G. Balakrishnan, A. Zhao, M. R. Sabuncu, J. Guttag, and A. V. Dalca. Voxelmorph: a learning framework for deformable medical image registration. *IEEE Transactions on Medical Imaging*, 38(8):1788–1800, 2019.

[4] P.-L. Bazin and D. L. Pham. Topology correction of segmented medical images using a fast marching algorithm. *Computer Methods and Programs in Biomedicine*, 88(2):182–190, 2007.

[5] M. F. Beg, M. I. Miller, A. Trouvé, and L. Younes. Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *International Journal of Computer Vision*, 61:139–157, 2005.

[6] M. Bertoux, J. Lagarde, F. Corlier, L. Hamelin, J.-F. Mangin, O. Colliot, M. Chupin, M. N. Braskie, P. M. Thompson, M. Bottlaender, et al. Sulcal morphology in alzheimer's disease: an effective marker of diagnosis and cognition. *Neurobiology of Aging*, 84:41–49, 2019.

[7] B. Billot, D. N. Greve, O. Puonti, A. Thielscher, K. Van Leemput, B. Fischl, A. V. Dalca, J. E. Iglesias, et al. Synthseg: Segmentation of brain mri scans of any contrast and resolution without retraining. *Medical image analysis*, 86:102789, 2023.

[8] F. Bongratz, A.-M. Rickmann, S. Pölsterl, and C. Wachinger. Vox2Cortex: Fast explicit reconstruction of cortical surfaces from 3D MRI scans with geometric deep neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 20773–20783, 2022.

[9] R. L. Burden, J. D. Faires, and A. M. Burden. *Numerical analysis*. Cengage learning, 2015.

[10] R. T. Chen, Y. Rubanova, J. Bettencourt, and D. K. Duvenaud. Neural ordinary differential equations. *Advances in Neural Information Processing Systems*, 31:6572–6583, 2018.

[11] X. Chen, J. Zhao, S. Liu, S. Ahmad, and P.-T. Yap. SurfFlow: A flow-based approach for rapid and accurate cortical surface reconstruction from infant brain mri. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 380–388. Springer, 2023.

[12] S. J. Crutch, M. Lehmann, J. M. Schott, G. D. Rabinovici, M. N. Rossor, and N. C. Fox. Posterior cortical atrophy. *The Lancet Neurology*, 11(2):170–178, 2012.

[13] R. S. Cruz, L. Lebrat, P. Bourgeat, C. Fookes, J. Fripp, and O. Salvado. DeepCSR: A 3D deep learning approach for cortical surface reconstruction. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 806–815, 2021.

[14] R. Dahnke, R. A. Yotter, and C. Gaser. Cortical thickness and central surface estimation. *Neuroimage*, 65:336–348, 2013.

[15] A. M. Dale, B. Fischl, and M. I. Sereno. Cortical surface-based analysis: I. segmentation and surface reconstruction. *Neuroimage*, 9(2):179–194, 1999.

[16] H. Fan, H. Su, and L. J. Guibas. A point set generation network for 3d object reconstruction from a single image. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 605–613, 2017.

[17] B. Fischl. Freesurfer. *Neuroimage*, 62(2):774–781, 2012.

[18] M. F. Glasser, S. N. Sotiropoulos, J. A. Wilson, T. S. Coalson, B. Fischl, J. L. Andersson, J. Xu, S. Jbabdi, M. Webster, J. R. Polimeni, et al. The minimal preprocessing pipelines for the human connectome project. *Neuroimage*, 80:105–124, 2013.

[19] K. Gopinath, C. Desrosiers, and H. Lombaert. SegRecon: Learning joint brain surface reconstruction and segmentation from images. In *International Conference on Medical Image Computing and Computer Assisted Intervention*, pages 650–659. Springer, 2021.

[20] L. Henschel, S. Conjeti, S. Estrada, K. Diers, B. Fischl, and M. Reuter. Fastsurfer-a fast and accurate deep learning based neuroimaging pipeline. *NeuroImage*, 219:117012, 2020.

[21] Y. Hong, S. Ahmad, Y. Wu, S. Liu, and P.-T. Yap. Vox2Surf: Implicit surface reconstruction from volumetric data. In *Intl. Workshop on Machine Learning in Medical Imaging*, pages 644–653. Springer, 2021.

[22] A. Hoopes, J. E. Iglesias, B. Fischl, D. Greve, and A. V. Dalca. Topofit: Rapid reconstruction of topologically-correct cortical surfaces. In *Medical Imaging with Deep Learning*, 2022.

[23] B. R. Howell, M. A. Styner, W. Gao, P.-T. Yap, L. Wang, K. Baluyot, E. Yacoub, G. Chen, T. Potts, A. Salzwedel, et al. The unc/umn baby connectome project (bcp): An overview of the study design and protocol development. *NeuroImage*, 185:891–905, 2019.

[24] C. R. Jack Jr, M. A. Bernstein, N. C. Fox, P. Thompson, G. Alexander, D. Harvey, B. Borowski, P. J. Britson, J. L. Whitwell, C. 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.

[25] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. In *International Conference on Learning Representations*, 2015.

[26] L. Lebrat, R. Santa Cruz, F. de Gournay, D. Fu, P. Bourgeat, J. Fripp, C. Fookes, and O. Salvado. CorticalFlow: A diffeomorphic mesh transformer network for cortical surface reconstruction. *Advances in Neural Information Processing Systems*, 34:29491–29505, 2021.

[27] T. Lewiner, H. Lopes, A. W. Vieira, and G. Tavares. Efficient implementation of marching cubes' cases with topological guarantees. *Journal of Graphics Tools*, 8(2):1–15, 2003.

[28] H. Li, Y. Fan, and A. D. N. Initiative. MDReg-Net: Multi-resolution diffeomorphic image registration using fully convolutional networks with deep self-supervision. *Human Brain Mapping*, 43(7):2218–2231, 2022.

[29] Y. Li, H. Li, and Y. Fan. ACEnet: Anatomical context-encoding network for neuroanatomy segmentation. *Medical image analysis*, 70:101991, 2021.

[30] Q. Ma, L. Li, V. Kyriakopoulou, J. V. Hajnal, E. C. Robinson, B. Kainz, and D. Rueckert. Conditional temporal attention networks for neonatal cortical surface reconstruction. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 312–322. Springer, 2023.

[31] Q. Ma, L. Li, E. C. Robinson, B. Kainz, D. Rueckert, and A. Alansary. CortexODE: Learning cortical surface reconstruction by neural ODEs. *IEEE Trans. on Medical Imaging*, 42(2):430–443, 2022.

[32] Q. Ma, E. C. Robinson, B. Kainz, D. Rueckert, and A. Alansary. PialNN: A fast deep learning framework for cortical pial surface reconstruction. In *International Workshop on Machine Learning in Clinical Neuroimaging*, pages 73–81. Springer, 2021.

[33] J. Maclaren, Z. Han, S. B. Vos, N. Fischbein, and R. Bammer. Reliability of brain volume measurements: a test-retest dataset. *Scientific Data*, 1(1):1–9, 2014.

[34] A. Makropoulos, E. C. Robinson, A. Schuh, R. Wright, S. Fitzgibbon, J. Bozek, S. J. Counsell, J. Steinweg, K. Vecchiato, J. Passerat-Palmbach, et al. The developing human connectome project: A minimal processing pipeline for neonatal cortical surface reconstruction. *Neuroimage*, 173:88–112, 2018.

[35] D. S. Marcus, T. H. Wang, J. Parker, J. G. Csernansky, J. C. Morris, and R. L. Buckner. Open access series of imaging studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. *Journal of Cognitive Neuroscience*, 19(9):1498–1507, 2007.

[36] Q. Meng, W. Bai, D. P. O'Regan, and D. Rueckert. Deepmesh: Mesh-based cardiac motion tracking using deep learning. *IEEE Transactions on Medical Imaging*, 2023.

[37] M. Modat, D. M. Cash, P. Daga, G. P. Winston, J. S. Duncan, and S. Ourselin. Global image registration using a symmetric block-matching approach. *Journal of medical imaging*, 1(2):024003–024003, 2014.

[38] T. C. Mok and A. Chung. Fast symmetric diffeomorphic image registration with convolutional neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4644–4653, 2020.

[39] D. H. Pak, M. Liu, S. S. Ahn, A. Caballero, J. A. Onofrey, L. Liang, W. Sun, and J. S. Duncan. Weakly supervised deep learning for aortic valve finite element mesh generation from 3D CT images. In *International Conference on Information Processing in Medical Imaging*, pages 637–648. Springer, 2021.

[40] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, et al. PyTorch: An imperative style, high-performance deep learning library. *Advances in Neural Information Processing Systems*, 32:8026–8037, 2019.

[41] J. Ren, Q. Hu, W. Wang, W. Zhang, C. S. Hubbard, P. Zhang, N. An, Y. Zhou, L. Dahmani, D. Wang, et al. Fast cortical surface reconstruction from MRI using deep learning. *Brain Informatics*, 9(1):1–16, 2022.

[42] L. M. Rimol, R. Nesvåg, D. J. Hagler Jr, Ø. Bergmann, C. Fennema-Notestine, C. B. Hartberg, U. K. Haukvik, E. Lange, C. J. Pung, A. Server, et al. Cortical volume, surface area, and thickness in schizophrenia and bipolar disorder. *Biological Psychiatry*, 71(6):552–560, 2012.

[43] J. M. Roe, D. Vidal-Piñeiro, Ø. Sørensen, A. M. Brandmaier, S. Düzel, H. A. Gonzalez, R. A. Kievit, E. Knights, S. Kühn, U. Lindenberger, et al. Asymmetric thinning of the cerebral cortex across the adult lifespan is accelerated in alzheimer's disease. *Nature Communications*, 12(1):721, 2021.

[44] O. Ronneberger, P. Fischer, and T. Brox. U-Net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 234–241, 2015.

[45] A. G. Roy, S. Conjeti, N. Navab, C. Wachinger, A. D. N. Initiative, et al. Quicknat: A fully convolutional network for quick and accurate segmentation of neuroanatomy. *NeuroImage*, 186:713–727, 2019.

[46] D. Ruelle and D. Sullivan. Currents, flows and diffeomorphisms. *Topology*, 14(4):319–327, 1975.

[47] R. Santa Cruz, L. Lebrat, D. Fu, P. Bourgeat, J. Fripp, C. Fookes, and O. Salvado. CorticalFlow++: Boosting cortical surface reconstruction accuracy, regularity, and interoperability. In *International Conferences on Medical Image Computing and Computer Assisted Intervention*, pages 496–505. Springer, 2022.

[48] D. W. Shattuck and R. M. Leahy. Brainsuite: an automated cortical surface identification tool. *Medical Image Analysis*, 6(2):129–142, 2002.

[49] A. A. Taha and A. Hanbury. Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. *BMC Medical Imaging*, 15(1):1–28, 2015.

[50] G. Teschl. *Ordinary differential equations and dynamical systems*, volume 140. American Mathematical Soc., 2012.

[51] D. C. Van Essen, H. A. Drury, S. Joshi, and M. I. Miller. Functional and structural mapping of human cerebral cortex: solutions are in the surfaces. *Proceedings of the National Academy of Sciences*, 95(3):788–795, 1998.

[52] L. Wang, Z. Wu, L. Chen, Y. Sun, W. Lin, and G. Li. ibeat v2.0: a multisite-applicable, deep learning-based pipeline for infant cerebral cortical surface reconstruction. *Nature protocols*, 18(5):1488–1509, 2023.

[53] N. Wang, Y. Zhang, Z. Li, Y. Fu, H. Yu, W. Liu, X. Xue, and Y.-G. Jiang. Pixel2Mesh: 3D mesh model generation via image guided deformation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(10):3600–3613, 2020.

[54] H. Zheng, H. Li, and Y. Fan. Coupled reconstruction of cortical surfaces by diffeomorphic mesh deformation. *Advances in Neural Information Processing Systems*, 37, 2023.

A Model Details

A.1 Cortical Ribbon Segmentation Network Architecture

Fig. 4 (Left) shows the detailed network architecture of our cortical ribbon segmentation network, which is a 5-level hierarchical encoder-decoder with skip connections. The network processes a 3D brain MRI to produce a cortical ribbon segmentation map. The white matter (WM) segmentation includes the interior of the WM surface, encompassing cortical WM, deep gray matter, ventricles, hippocampus, and other tissues within the surface. Similarly, the gray matter (GM) segmentation includes the interior of the pial surface. The output map has five classes: left hemisphere WM and GM, right hemisphere WM and GM, and background. In the encoder, each level uses a $3 \times 3 \times 3$ convolutional layer with a stride of 2 to downsample the features. In the decoder, features are upsampled by $2\times$ at each scale, concatenated with the corresponding features from the encoder via skip connections, and then fused using a $3 \times 3 \times 3$ convolutional layer with a stride of 1. For feature extraction at the input, a $3 \times 3 \times 3$ convolutional layer with a stride of 1 is used. Before the final prediction, three consecutive convolutional layers are applied. Each convolutional layer is followed by a leaky ReLU activation function, except for the last one, which uses a Softmax function before computing the cross-entropy loss with the ground truth.

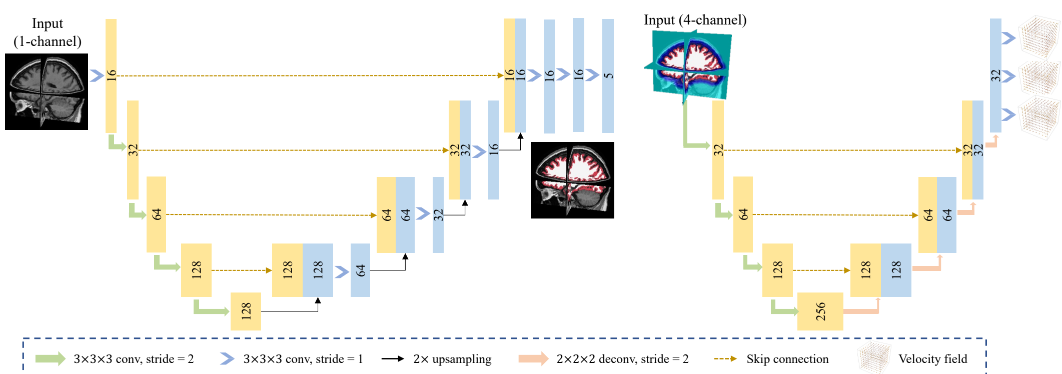


Figure 4: Left: 3D U-Net architecture for ribbon segmentation. The output, i.e., the cortical ribbon map, is overlaid on the input image for illustration. Right: 3D U-Net architecture for cortical surface reconstruction. The learned velocity fields are used to calculate deformations.

A.2 Cortical Surface Reconstruction Network Architecture and Training details

As shown in Fig.4 (Right), our cortical surface reconstruction (CSR) network operates at five scales. To conserve memory, we downsample the input image using a $3 \times 3 \times 3$ convolution with a stride of 2 and skip complex feature fusion via skip connections in the decoding path at this scale. To improve the accuracy of the velocity fields (VFs), we use $2 \times 2 \times 2$ deconvolutions with a stride of 2 in the decoding path instead of $2\times$ trilinear upsampling. At the output stage, we employ three parallel $3 \times 3 \times 3$ convolutional layers to generate VFs for the white matter (WM), midthickness, and pial surfaces, respectively. ReLU activation functions are used after each convolutional layer, except for the three parallel layers, where Softsign functions are applied. The VFs are then utilized to compute diffeomorphic deformations.

B Experimental Settings

B.1 Dataset Preprocessing

We preprocessed all the MRIs of the ADNI-1 [24] and OASIS-1 [35] datasets with the same protocols as following: Based on the standard processing protocol in FreeSurfer V7.2.0 [17], the original images were conformed and normalized (saved as `orig.mgz`), affinely registered to the MNI152 template [8] using the NiftyReg toolbox [37]. The respective ribbon segmentation maps, SDFs, and pseudo-ground-truth surfaces were also transformed using the computed transformation. Similarly, we utilize iBEAT V2.0 [52] to process the BCP [23] dataset and merge the brain tissue segmentation results as the ribbon segmentation maps.

B.2 Baselines

We compared our SegCSR with representatives from the two categories of existing DL-based CSR methods and evaluated their performance for both WM and pial surface reconstruction. DeepCSR [13] and 3D U-Net [44] represent implicit surface reconstruction methods, while others fall into the category of explicit methods. Note that we modify the 3D U-Net method to first generate SDFs based on the ribbon segmentation results, then perform topology correction, and finally utilize the Marching Cubes algorithm to extract the mesh. Since it does not require pGT surfaces from FreeSurfer for training supervision, it can be treated as a weakly supervised learning-based baseline. CorticalFlow++[47] utilizes smoothed convex hulls as the initialization template, trains a chain of deformation fields, and employs a fourth-order Runge-Kutta (RK4) solver to compute the integration for the initial value problem. CortexODE[31] uses WM segmentation for surface initialization and Neural ODE for deformation computation. Vox2cortex [8] deforms averaged surface templates with a GNN-based network to reconstruct multiple surfaces. CoCSR [54] integrates multiple cortical surface reconstructions into a single network. A summary of the state-of-the-art CSR methods is provided in Table 4.

Table 4: Summary of baseline methods in terms of surface representation, supervision in training, and loss functions.

Method	Representation	Supervision	Primary Loss function
3D U-Net [44]	Implicit	Ribbon segmentation	Cross Entropy
DeepCSR [13]		SDFs	L1 Loss
CorticalFlow++ [47]	Explicit	Mesh	Bi-directional Chamfer Loss
cortexODE [31]		Mesh	Bi-directional Chamfer Loss
Vox2Cortex [8]		Mesh	Bi-directional Chamfer Loss
CoCSR [54]		Mesh	Bi-directional Chamfer Loss
SegCSR (Ours)	Explicit	Ribbon segmentation	Weak Supervision

C More Experimental Results

C.1 Quantitative comparison of our methods with Related Works

Due to space limit, we only showcase the quantitative results on left hemisphere in the main paper. Quantitative comparison results on the right hemisphere are summarized as a supplement to Table 1 in the main paper.

Table 5: Quantitative analysis of cortical surface reconstruction on geometric accuracy and self-intersections. The Chamfer distance (CD), average symmetric surface distance (ASSD), Hausdorff distance (HD), and the ratio of the self-intersecting faces (SIF) were measured for WM and pial surfaces on three datasets. The mean value and standard deviation are reported. Lower scores indicate better results for all metrics. "S" denotes the use of pGT surfaces from conventional pipelines, while "W" represents weak supervision by pGT ribbon segmentations. In each supervision setting, the best results are in bold, and the second best results are underlined.

Data	Sup.	Method	R-Pial Surface				R-WM Surface			
			CD (mm)	ASSD (mm)	HD (mm)	SIF (%)	CD (mm)	ASSD (mm)	HD (mm)	SIF (%)
ADNI	S	CorticalFlow++ [47]	0.550±0.038	0.413±0.034	0.891±0.071	0.101±0.069	0.548±0.035	0.403±0.032	0.883±0.068	0.071±0.042
		cortexODE [31]	0.482±0.019	0.220±0.022	0.461±0.060	0.033±0.017	0.470±0.020	0.207±0.019	0.444±0.018	0.023±0.016
		Vox2Cortex [8]	0.593±0.032	0.382±0.029	0.755±0.061	0.071±0.045	0.588±0.029	0.363±0.024	0.741±0.057	0.059±0.035
		CoCSR [54]	**0.326**±0.023	**0.126**±0.012	**0.271**±0.024	**0.015**±0.013	**0.320**±0.020	**0.124**±0.012	**0.265**±0.022	**0.006**±0.003
	W	DeepCSR [13]	0.948±0.080	0.597±0.068	1.154±0.207	\	0.942±0.077	0.589±0.065	1.140±0.195	\
		3D U-Net [44]	0.601±0.048	0.342±0.037	0.784±0.166	\	0.476±0.014	0.268±0.016	0.563±0.031	\
		SegCSR (Ours)	**0.582**±0.021	**0.328**±0.022	**0.751**±0.050	0.009±0.009	**0.470**±0.015	**0.261**±0.021	**0.548**±0.038	0.011±0.010
OASIS	S	CorticalFlow++ [47]	0.540±0.037	0.405±0.032	0.834±0.060	0.095±0.052	0.536±0.035	0.402±0.031	0.830±0.058	0.088±0.049
		cortexODE [31]	0.497±0.023	0.225±0.024	0.473±0.065	0.038±0.027	0.481±0.021	0.214±0.021	0.450±0.022	0.025±0.019
		Vox2Cortex [8]	0.598±0.033	0.386±0.031	0.761±0.064	0.072±0.040	0.592±0.031	0.379±0.028	0.752±0.061	0.061±0.037
		CoCSR [54]	**0.411**±0.034	**0.144**±0.017	**0.284**±0.022	**0.018**±0.015	**0.353**±0.026	**0.130**±0.021	**0.272**±0.024	**0.009**±0.004
	W	DeepCSR [13]	0.989±0.086	0.619±0.071	1.336±0.215	\	0.980±0.082	0.601±0.069	1.175±0.202	\
		3D U-Net [44]	0.613±0.070	0.333±0.050	0.777±0.268	\	0.456±0.014	0.249±0.020	0.493±0.033	\
		SegCSR (Ours)	**0.584**±0.018	**0.323**±0.019	**0.728**±0.041	0.012±0.011	**0.452**±0.012	**0.224**±0.016	**0.465**±0.030	0.012±0.010
BCP	S	CorticalFlow++ [47]	0.926±0.271	0.729±0.035	1.940±0.174	1.113±0.374	0.892±0.240	0.721±0.033	1.877±0.148	0.531±0.105
		cortexODE [31]	0.758±0.081	0.394±0.032	0.820±0.102	0.121±0.060	0.676±0.069	0.346±0.029	0.814±0.098	0.098±0.033
		CoCSR [54]	**0.575**±0.038	**0.214**±0.022	**0.464**±0.059	**0.060**±0.037	**0.542**±0.038	**0.198**±0.020	**0.446**±0.049	**0.056**±0.030
	W	DeepCSR [13]	2.672±1.131	1.222±0.214	3.101±1.209	\	1.437±0.519	0.426±0.049	0.927±0.116	\
		3D U-Net [44]	1.174±0.312	0.790±0.058	2.136±1.020	\	0.687±0.118	0.376±0.039	0.788±0.063	\
		SegCSR (Ours)	**0.926**±0.070	**0.497**±0.060	**1.287**±0.142	0.058±0.056	**0.875**±0.067	**0.476**±0.050	**1.203**±0.130	0.054±0.055

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: We clearly summarize the contributions in Section 1 Introduction.

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: We discuss the limitations of the work in Section 5 Conclusions.

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors

should reflect on how these assumptions might be violated in practice and what the implications would be.

- The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
- The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
- The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
- If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
- While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory Assumptions and Proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [NA]

   Justification: This is not a theoretical paper.

   Guidelines:

   - The answer NA means that the paper does not include theoretical results.
   - All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
   - All assumptions should be clearly stated or referenced in the statement of any theorems.
   - The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
   - Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
   - Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental Result Reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: We clearly illustrate the methodology in Section 3, the experimental setups in Section 4.1, and more implementation details in the Supplementary Materials.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may

be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.

- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [No]

   Justification: (1) We used and cited public datasets and gave pre-processing steps in Section 4.1 and more details in the Supplementary Materials. (2) We provided code links for public baselines in the Supplementary Materials. (3) Code of our proposed method will be made public upon acceptance of the paper.

   Guidelines:

   - The answer NA means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (https://nips.cc/public/guides/CodeSubmissionPolicy) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (https://nips.cc/public/guides/CodeSubmissionPolicy) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental Setting/Details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

Answer: [Yes]

Justification: We specify all the training and test details in Section 4.1 and the Supplementary Materials.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment Statistical Significance**

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [Yes]

Justification: We reported mean and standard deviation of each experimental setting and computed the statistical significance.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments Compute Resources**

Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

Answer: [Yes]

Justification: We specify the computation resources for training and testing in Section 4.1 and more details in the Supplementary Materials.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
- The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
- The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code Of Ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics https://neurips.cc/public/EthicsGuidelines?

   Answer: [Yes]

   Justification: We review and conform with the NeurIPS Code of Ethics.

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader Impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

    Justification: We discuss the societal impacts of the work in Section 5 Conclusions.

    Guidelines:

    - The answer NA means that there is no societal impact of the work performed.
    - If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
    - The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
    - The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
    - If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

    Answer: [NA]

    Justification: Our paper poses no such risks.

    Guidelines:

    - The answer NA means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.

- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: We properly cite relevant baseline methods and datasets.

    Guidelines:

    - The answer NA means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
    - For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
    - If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New Assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [Yes]

    Justification: We communicate the details of the dataset/code/model in our main paper and Supplementary Materials. We will make code and model public upon the acceptance of the paper.

    Guidelines:

    - The answer NA means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and Research with Human Subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [NA]

    Justification: Our paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional Review Board (IRB) Approvals or Equivalent for Research with Human Subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [NA]

    Justification: Our paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.