SurFlow: A Flow-Based Approach for Rapid and Accurate Cortical Surface Reconstruction from Infant Brain MRI

Xiaoyang Chen[1,2], Junjie Zhao[3], Siyuan Liu[4], Sahar Ahmad[1,2],
and Pew-Thian Yap[1,2(✉)]

[1] Department of Radiology, University of North Carolina, Chapel Hill, NC, USA
ptyap@med.unc.edu
[2] Biomedical Research Imaging Center, University of North Carolina,
Chapel Hill, NC, USA
[3] Department of Computer Science, University of North Carolina,
Chapel Hill, NC, USA
[4] College of Marine Engineering, Dalian Maritime University, Dalian, China

Abstract. The infant brain undergoes rapid changes in volume, shape, and structural organization during the first postnatal year. Accurate cortical surface reconstruction (CSR) is essential for understanding rapid changes in cortical morphometry during early brain development. However, existing CSR methods, designed for adult brain MRI, fall short in reconstructing cortical surfaces from infant MRI, owing to the poor tissue contrasts, partial volume effects, and rapid changes in cortical folding patterns. Here, we introduce an infant-centric CSR method in light of these challenges. Our method, *SurFlow*, utilizes three seamlessly connected deformation blocks to sequentially deform an initial template mesh to target cortical surfaces. Remarkably, our method can rapidly reconstruct a high-resolution cortical surface mesh with 360k vertices in approximately one second. Performance evaluation based on an MRI dataset of infants 0 to 12 months of age indicates that SurFlow significantly reduces geometric errors and substantially improves mesh regularity compared with state-of-the-art deep learning approaches.

Keywords: Infant Brain MRI · Cortical Surface Reconstruction · Topology Preservation

1 Introduction

The human brain undergoes dramatic changes in size, shape, and tissue architecture during the first postnatal year, driven by cellular processes [1] that lead to cortical folding and the formation of convoluted gyri and sulci on the cerebral

---

surface. Understanding cortical development in these early postnatal years is crucial for comprehending later-stage functional development. Cortical surface reconstruction (CSR) is a step necessary for functional and anatomical visualization, brain morphology, and quantification of cortical growth. However, reconstructing the cortical surface from infant MRI poses significant challenges due to the low signal-to-noise ratio, pronounced motion artifacts, complex folding patterns, and rapid brain expansion [2].

CSR methods developed so far include FreeSurfer [5], DeepCSR [4], Voxel2Mesh [13], PialNN [8], and CorticalFlow [7]. However, these methods are optimized primarily for adult brain MRI, typically with clear cortical gyral and sulcal patterns and good tissue contrasts. Their performance deteriorates significantly when applied to infant MRI due to challenges like narrow sulcal spaces, low tissue contrast, and partial volume effects. To address these issues, specific methods like the dHCP pipeline [9] and Infant FreeSurfer [14] have been developed. However, the dHCP pipeline is only effective for neonatal data and Infant FreeSurfer often fails to produce accurate cortical surfaces, especially for the first postnatal year when the brain undergoes dynamic changes in tissue contrasts and morphology. Additionally, the computational inefficiency of Infant FreeSurfer, which takes approximately 10 h to reconstruct cortical surfaces for a single subject, limits its applicability to large-scale studies.

In this paper, we propose *SurfFlow*, a geometric deep learning model designed to reconstruct accurate cortical surfaces from infant brain MRI. Our model comprises three cascaded deformation blocks, each responsible for predicting a flow field and constructing a diffeomorphic mapping for each vertex through solving a flow ordinary differential equation (ODE). The flow fields deform template meshes progressively to finally produce accurate genus-zero cortical surfaces. Our work offers a threefold contribution. First, we propose an efficient dual-modal flow-based CSR method, enabling the creation of high-resolution and high-quality mesh representations for complex cortical surfaces. Second, we propose a novel loss function that effectively regularizes the lengths of mesh edges, leading to substantial improvements in mesh quality. Third, our method represents the first attempt in tackling the challenging task of directly reconstructing cortical surfaces from infant brain MRI. Our method outperforms state-of-the-art methods by a significant margin, judging based on multiple surface evaluation metrics.

2 Methods

2.1 Overview

As depicted in Fig. 1a, SurfFlow consists of three deformation blocks, each consisting of a 3D Unet [3] and a diffeomorphic mesh deformation (DMD) module. The final mesh, either the pial surface or the white surface, is achieved by composing the three diffeomorphic deformations generated by the DMD modules. Each Unet takes a T1w-T2w image pair as input and, except the first network, also receives flow fields predicted by previous deformation blocks to predict a new
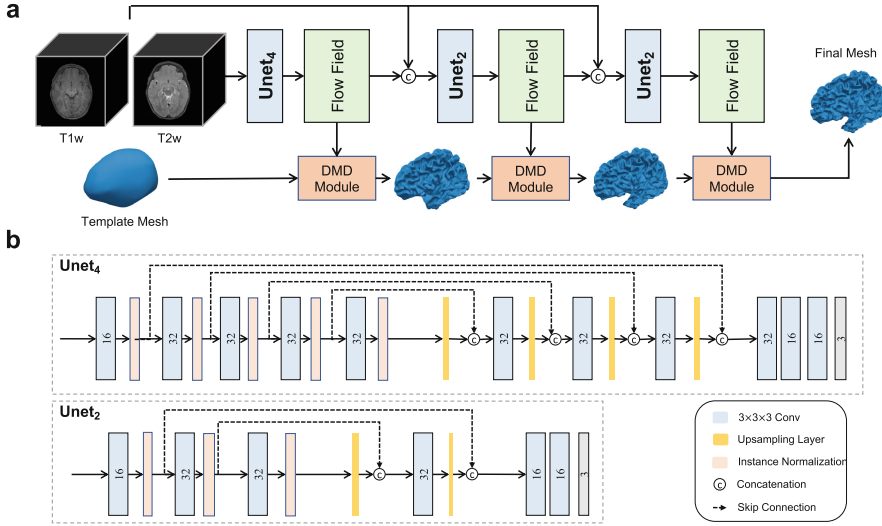
Fig. 1. (a) SurfFlow network structure. The network has three deformation blocks, each consisting of a 3D Unet and a diffeomorphic mesh deformation (DMD) module. (b) Unet architectures.

flow field. Figure 1b provides a detailed illustration of the design of the two different 3D Unets used. Each DMD module employs the flow field predicted within the same block and computes a diffeomorphic mapping $\phi_\theta(t; \mathbf{x})$ for each vertex $\mathbf{x}$ in the mesh generated by the previous deformation block through integration, assuming a stationary flow field. Formally, the dynamics of mesh deformation are formulated as a stationary flow ODE specified by a 3D Unet:

$$\frac{\mathrm{d}\phi_\theta(t; \mathbf{x})}{\mathrm{d}t} = \mathbf{V}(\phi_\theta(t; \mathbf{x})), \quad \phi_\theta(0; \mathbf{x}) = \mathbf{x}_0, \tag{1}$$

where θ denotes the parameters of the 3D Unet, $\mathbf{x}_0$ is the initial mesh before deformation in each stage. SurfFlow is trained stage-wise: a deformation block is trained with all previous deformation blocks frozen. Following CorticalFlow++ [10], the Runge-Kutta method is used for numerical integration.

2.2 Dual-Modal Input

Infant MRI exhibits three distinct phases during the first year of life. In the *infantile phase* ($\leq$5 months), gray matter (GM) shows higher signal intensity than white matter (WM) in T1w images. The *isointense phase* (5–8 months) corresponds to an increase in intensity of WM owing to myelination associated with brain maturation. This significantly lowers the contrast between GM and WM in T1w images and similarly T2w images. In the *early adult-like phase* ($\geq$ 8 months), the GM intensity is lower than WM in T1w images, similar to the

tissue contrast in adult MRI. We propose to use both T1w and T2w images for complementary information needed for accurate surface reconstruction.

2.3   Loss Function

The loss function used for training *SurfFlow* is a weighted sum of Chamfer distance (CD; L_{cd}) and a piecewise edge length loss (PELL; L_e), balanced by parameter λ:

$$L_{total} = L_{cd} + \lambda L_e. \tag{2}$$

Chamfer Distance Loss. The Chamfer distance is commonly used as the loss function for surface reconstruction. It measures the distance from a vertex in one mesh P to the closest vertex in another mesh Q bidirectionally:

$$L_{cd} = \frac{1}{|P|} \sum_{p \in P} \min_{q \in Q} \|p - q\|_2^2 + \frac{1}{|Q|} \sum_{q \in Q} \min_{p \in P} \|p - q\|_2^2, \tag{3}$$

where p and q are vertices in P and Q, respectively.

Piecewise Edge Length Loss. We observed that vertices tend to be clustered in surfaces generated by CorticalFlow++ (Fig. 3). One possible cause is that the edge length loss used in CorticalFlow++ pushes edges to zero length. To address this problem, we propose a piecewise edge length loss that encourages edge lengths to lie within a suitable range. The proposed loss is formulated as

$$L_e = \frac{1}{|P|} \sum_{p \in P} \frac{1}{|\mathcal{N}(p)|} \sum_{k \in \mathcal{N}(p)} \begin{cases} \epsilon - \|p - k\|_2^2 & \text{if } \|p - k\|_2^2 < \epsilon, \\ 0 & \text{if } \epsilon \leqslant \|p - k\|_2^2 \leqslant \gamma \epsilon, \\ \|p - k\|_2^2 - \gamma \epsilon & \text{otherwise,} \end{cases} \tag{4}$$

where P denotes the predicted mesh, p is a vertex in P, $\mathcal{N}(p)$ consists of the neighbors of p, and ϵ and γ are two tunable hyper-parameters that control the range position and width, respectively.

2.4   Deformation Computation in DMD Modules

Following CorticalFlow++, we use the fourth-order Runge-Kutta method to solve the stationary flow ODE over time interval $[0, 1]$ for both accuracy and stability. The solution $\mathbf{x}_1$ is obtained iteratively with

$$\mathbf{x}_{t_{n+1}} = \mathbf{x}_{t_n} + \frac{h}{6} \Big(\mathbf{k}_1 + 2\mathbf{k}_2 + 2\mathbf{k}_3 + \mathbf{k}_4 \Big), \tag{5}$$

where $\mathbf{k}_1 = \mathbf{V}(\mathbf{x}_{t_n})$, $\mathbf{k}_2 = \mathbf{V}(\mathbf{x}_{t_n} + \mathbf{k}_1 \frac{h}{2})$, $\mathbf{k}_3 = \mathbf{V}(\mathbf{x}_{t_n} + \mathbf{k}_2 \frac{h}{2})$, and $\mathbf{k}_4 = \mathbf{V}(\mathbf{x}_{t_n} + \mathbf{k}_3 h)$. $\mathbf{V}(\mathbf{x})$ represents the trilinear interpolation of the flow field at position $\mathbf{x}$. h is the step size and is set to $1/30$. t_n is the time at n-th step, and $t_{n+1} = t_n + h$.

2.5   Implementation Details

Our network was trained stage-wise. We froze the parameters of one deformation block once trained and then start the training of the next block. Each deformation block was trained for 27k iterations. Adam optimizer was used with an initial learning rate of 0.0001. The parameter λ of the loss function was set to 3.0 to balance the two loss terms. We set ϵ for PELL based on the average edge length determined from the training set. We set the γ to 4.0 and ϵ to 5×10^{-5}. Instance normalization (IN) [11] layers were added between convolutional and activation layers for faster convergence and improved performance.

3   Results

3.1   Data

The dataset includes aligned T1w and T2w image pairs from 121 subjects, 2 weeks to 12 months of age, from the Baby Connectome Project (BCP) [6]. Among them, 90 cases were used for training, 12 for validation, and 19 for testing. Ground truth cortical surfaces were generated with iBEAT v2.0 [12].

To obtain a smooth starting template, an average convex hull computed from the training dataset was re-meshed and triangularized. The Catmull-Clark subdivision algorithm was then applied to generate enough faces and vertices. Decimation was used to control the number of vertices. These steps were carried out in Blender[1].

3.2   Evaluation Metrics

For performance evaluation and comparison between different methods, we utilized the following metrics: Chamfer distance (CD), average symmetric surface distance (ASSD), 90% Hausdorff distance (HD), and normal consistency (NC). Their definitions are as follows:

$$\text{CD} = \frac{1}{2}\Big(\sum_{p \in P}\min_{q \in Q}\frac{\|p-q\|^2}{|P|} + \sum_{q \in Q}\min_{p \in P}\frac{\|p-q\|^2}{|Q|}\Big),$$

$$\text{ASSD} = \frac{1}{2}\Big(\sum_{p \in P}\min_{q \in Q}\frac{\|p-q\|}{|P|} + \sum_{q \in Q}\min_{p \in P}\frac{\|p-q\|}{|Q|}\Big),$$

$$\text{HD} = \max\Big(\max_{p \in P}\min_{q \in Q}\|p-q\|, \max_{q \in Q}\min_{p \in P}\|p-q\|\}\Big),$$

$$\text{NC} = \frac{1}{2}\Big(\frac{\sum_{p \in P}(\mathbf{n_p} \cdot \mathbf{n_{pq}})}{|P|} + \frac{\sum_{q \in Q}(\mathbf{n_q} \cdot \mathbf{n_{qp}})}{|Q|}\Big),$$

where P and Q are respectively the predicted and ground truth (GT) meshes, $\mathbf{n_p}$ and $\mathbf{n_q}$ are the normals at p and q, $\mathbf{n_{pq}}$ is the normal of the vertex in Q that is closest to p, and $\mathbf{n_{qp}}$ is the normal of the vertex in P that is closest to q.

---

[1] https://www.blender.org.

Table 1. Comparison of different CSR methods in reconstructing the white and pial surfaces of the left (L) and right (R) hemispheres. ↓ and ↑ indicate better performance with lower and higher metric values, respectively.

	CD ↓		ASSD ↓		HD ↓		NC ↑	
Pial Surface								
	L	R	L	R	L	R	L	R
DeepCSR	4.69	3.85	1.26	1.13	5.10	4.58	0.75	0.76
	(±2.10)	(±2.00)	(±0.25)	(±0.27)	(±1.13)	(±1.37)	(±0.02)	(±0.02)
CorticalFlow++	1.02	0.67	0.74	0.63	2.15	1.55	0.81	0.84
	(±0.17)	(±0.10)	(±0.03)	(±0.04)	(±0.25)	(±0.12)	(±0.01)	(±0.01)
SurfFlow	0.39	0.36	0.48	0.46	0.98	0.80	0.90	0.91
	(±0.14)	(±0.11)	(±0.03)	(±0.05)	(±0.08)	(±0.10)	(±0.01)	(±0.01)
White Surface								
	CD ↓		ASSD ↓		HD ↓		NC ↑	
	L	R	L	R	L	R	L	R
DeepCSR	0.74	0.50	0.58	0.60	1.28	1.27	0.90	0.90
	(±0.50)	(±0.90)	(±0.11)	(±0.16)	(±0.58)	(±0.68)	(±0.02)	(±0.02)
CorticalFlow++	1.13	0.91	0.86	0.78	1.96	1.66	0.72	0.73
	(±0.43)	(±0.12)	(±0.04)	(±0.04)	(±0.11)	(±0.13)	(±0.13)	(±0.04)
SurfFlow	0.37	0.41	0.47	0.49	0.86	0.90	0.93	0.92
	(±0.12)	(±0.15)	(±0.05)	(±0.06)	(±0.10)	(±0.12)	(±0.01)	(±0.01)

3.3   Results

We compared SurfFlow with CorticalFlow++ and DeepCSR. To ensure a fair comparison, we modified both CorticalFlow++ and DeepCSR to use dual-modal inputs. Specifically, for DeepCSR, we utilized the finest possible configuration, generating a 512^3 3D grid to predict the signed distance field for surface reconstruction using the marching cube algorithm. As depicted in Table 1, our evaluation demonstrates that SurfFlow outperforms the other two methods in all metrics. SurfFlow stands out with an average Chamfer distance of less than 0.5 mm, demonstrating significantly smaller errors compared with CorticalFlow++ and DeepCSR, which result in 22% to 1100% larger errors. Similar improvements were observed for evaluations with the ASSD, HD, and NC metrics. Furthermore, during our evaluation, we noticed that DeepCSR yielded higher errors for pial surface reconstruction due to numerous mesh topological artifacts. The utilization of implicit representation in DeepCSR does not guarantee a genus zero manifold. Visual comparisons (Figs. 2 and 3) further confirm that SurfFlow excels in reconstructing cortical gyri and sulci compared with CorticalFlow++ and DeepCSR. Moreover, SurfFlow demonstrates superior robustness with more consist results, as indicated lower standard deviations.

SurfFlow utilizes PELL to ensure that mesh edge lengths are within the desired range. We observed that this optimization leads to smoother, more uniform, and accurate meshes. In contrast, CorticalFlow++ shows limited accuracy in certain mesh areas, primarily due to its inability to ensure mesh uniformity.

This is evident in the zoomed-in areas depicted in Fig. 4, where the deficiency in mesh faces hinders accurate predictions. The superiority of SurfFlow in terms of mesh smoothness is further supported by the NC results.

Table 2. CSR performance with respect to modalities (T1w and T2w), instance normalization (IN), and piecewise edge length loss (PELL) for white and pial surfaces of the left (L) and right (R) hemispheres. ↓ and ↑ indicate better performance with lower and higher metric values, respectively.

|			T1w ✓ / ✗ / ✓ / ✓ / ✓ / ✗ / ✓						
			T1w ✓	**✗**	**✓**	**✓**	**✓**	**✗**	**✓**
			T2w ✗	**✓**	**✓**	**✓**	**✗**	**✓**	**✓**
			IN ✗	**✗**	**✗**	**✓**	**✓**	**✓**	**✓**
			PELL ✗	**✗**	**✗**	**✓**	**✗**	**✓**	**✓**
CD ↓	Pial	L	0.95 (±0.15)	0.89 (±0.16)	1.02 (±0.17)	0.66 (±0.21)	0.90 (±0.42)	0.98 (±0.23)	0.39 (±0.14)
		R	1.05 (±0.18)	1.08 (±0.24)	0.67 (±0.10)	0.66 (±0.17)	0.78 (±0.26)	0.88 (±0.31)	0.36 (±0.11)
	White	L	1.19 (±0.17)	1.31 (±0.27)	1.13 (±0.43)	0.85 (±0.17)	0.82 (±0.27)	0.93 (±0.27)	0.37 (±0.12)
		R	1.28 (±0.14)	1.32 (±0.24)	0.91 (±0.12)	0.90 (±0.18)	0.88 (±0.29)	0.99 (±0.36)	0.41 (±0.15)
ASSD ↓	Pial	L	0.74 (±0.04)	0.72 (±0.07)	0.74 (±0.03)	0.58 (±0.03)	0.64 (±0.11)	0.71 (±0.08)	0.48 (±0.03)
		R	0.77 (±0.05)	0.78 (±0.07)	0.63 (±0.04)	0.60 (±0.05)	0.60 (±0.05)	0.68 (±0.10)	0.46 (±0.05)
	White	L	0.86 (±0.04)	0.94 (±0.09)	0.86 (±0.04)	0.70 (±0.04)	0.68 (±0.10)	0.74 (±0.11)	0.47 (±0.05)
		R	0.92 (±0.05)	0.94 (±0.08)	0.78 (±0.04)	0.73 (±0.04)	0.70 (±0.10)	0.76 (±0.13)	0.49 (±0.06)
HD ↓	Pial	L	1.86 (±0.20)	1.79 (±0.17)	2.15 (±0.24)	1.36 (±0.16)	1.69 (±0.60)	1.89 (±0.28)	0.98 (±0.08)
		R	2.04 (±0.21)	2.09 (±0.19)	1.55 (±0.12)	1.42 (±0.18)	1.41 (±0.31)	1.51 (±0.33)	0.80 (±0.10)
	White	L	1.88 (±0.15)	1.92 (±0.22)	1.96 (±0.11)	1.57 (±0.12)	1.50 (±0.34)	1.62 (±0.26)	0.86 (±0.10)
		R	1.98 (±0.13)	1.94 (±0.23)	1.66 (±0.13)	1.70 (±0.18)	1.55 (±0.30)	1.67 (±0.35)	0.90 (±0.12)
NC ↑	Pial	L	0.80 (±0.02)	0.80 (±0.02)	0.81 (±0.01)	0.86 (±0.01)	0.85 (±0.04)	0.83 (±0.02)	0.90 (±0.01)
		R	0.80 (±0.01)	0.80 (±0.02)	0.84 (±0.01)	0.85 (±0.01)	0.82 (±0.05)	0.82 (±0.03)	0.91 (±0.01)
	White	L	0.66 (±0.06)	0.62 (±0.02)	0.72 (±0.04)	0.83 (±0.02)	0.81 (±0.08)	0.80 (±0.04)	0.93 (±0.01)
		R	0.65 (±0.06)	0.63 (±0.02)	0.73 (±0.04)	0.80 (±0.02)	0.79 (±0.08)	0.80 (±0.04)	0.92 (±0.01)
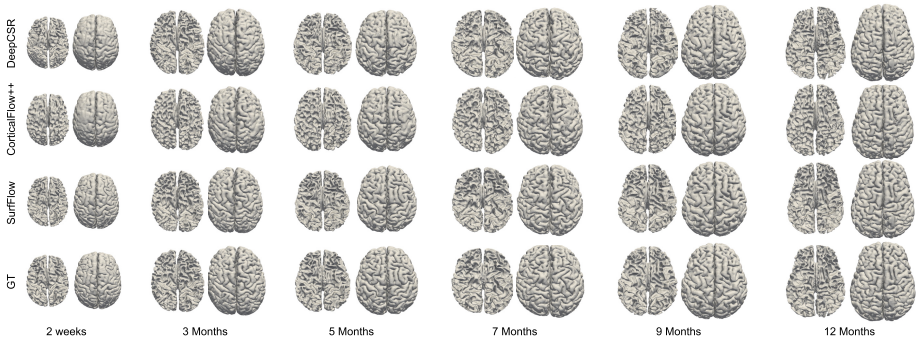
Fig. 2. Comparison of white and pial surfaces reconstructed via SurfFlow, CorticalFlow++, and DeepCSR for different time points.

3.4   Ablation Study

The results of an ablation study (Table 2) confirm that surface prediction performance improves (i) when both modalities are concurrently used; (ii) when
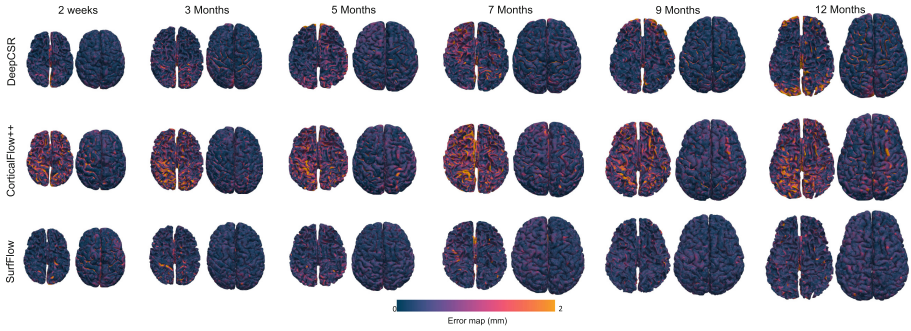
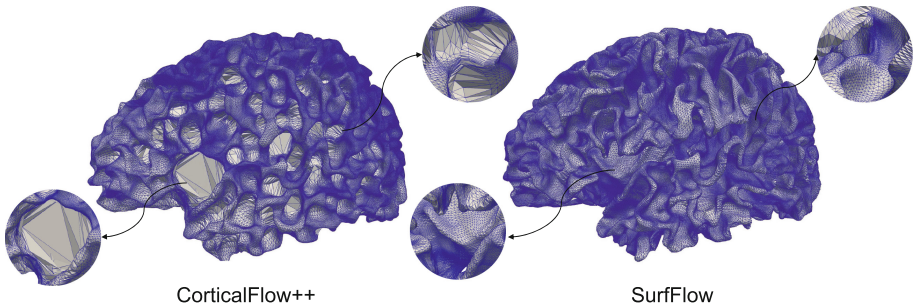**Fig. 3.** Error map comparison between SurfFlow, CorticalFlow++, and DeepCSR.

Fig. 4. Comparison of mesh uniformity.

instance normalization is used, and (iii) when PELL is used as opposed to the edge length loss in CorticalFlow++.

4   Conclusion

We presented SurfFlow—a flow-based deep-learning network to accurately reconstruct cortical surfaces. SurfFlow predicts a flow field to deform a template surface toward a target surface. It produces smooth and uniform surface meshes with sub-millimeter accuracy and outperforms CorticalFlow++ and DeepCSR in terms of surface accuracy and regularity.

References

1. Ahmad, S., et al.: Multifaceted atlases of the human brain in its infancy. Nat. Meth. **20**, 55–64 (2023)
2. Barkovich, M.J., Li, Y., Desikan, R.S., Barkovich, A.J., Xu, D.: Challenges in pediatric neuroimaging. Neuroimage **185**, 793–801 (2019)

3. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3D U-Net: learning dense volumetric segmentation from sparse annotation. In: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W. (eds.) MICCAI 2016. LNCS, vol. 9901, pp. 424–432. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46723-8_49

4. Cruz, R.S., Lebrat, L., Bourgeat, P., Fookes, C., Fripp, J., Salvado, O.: DeepCSR: a 3D deep learning approach for cortical surface reconstruction. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 806–815 (2021)

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

6. Howell, B.R., et al.: The UNC/UMN baby connectome project (BCP): an overview of the study design and protocol development. Neuroimage **185**, 891–905 (2019)

7. Lebrat, L., et al.: CorticalFlow: a diffeomorphic mesh deformation module for cortical surface reconstruction. arXiv preprint arXiv:2206.02374 (2022)

8. Ma, Q., Robinson, E.C., Kainz, B., Rueckert, D., Alansary, A.: PialNN: a fast deep learning framework for cortical Pial surface reconstruction. In: Abdulkadir, A., et al. (eds.) MLCN 2021. LNCS, vol. 13001, pp. 73–81. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87586-2_8

9. Makropoulos, A., et al.: The developing human connectome project: a minimal processing pipeline for neonatal cortical surface reconstruction. Neuroimage **173**, 88–112 (2018)

10. Santa Cruz, R., et al.: CorticalFlow++: boosting cortical surface reconstruction accuracy, regularity, and interoperability. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds.) MICCAI 2022. LNCS, vol. 13435, pp. 496–505. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-16443-9_48

11. Ulyanov, D., Vedaldi, A., Lempitsky, V.: Instance normalization: the missing ingredient for fast stylization. arXiv preprint arXiv:1607.08022 (2016)

12. Wang, L., Wu, Z., Chen, L., Sun, Y., Lin, W., Li, G.: iBEAT v2.0: a multisite-applicable, deep learning-based pipeline for infant cerebral cortical surface reconstruction. Nat. Protoc. **18**, 1488–1509 (2023)

13. Wickramasinghe, U., Remelli, E., Knott, G., Fua, P.: Voxel2Mesh: 3D mesh model generation from volumetric data. In: Martel, A.L., et al. (eds.) MICCAI 2020. LNCS, vol. 12264, pp. 299–308. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59719-1_30

14. Zöllei, L., Iglesias, J.E., Ou, Y., Grant, P.E., Fischl, B.: Infant FreeSurfer: an automated segmentation and surface extraction pipeline for T1-weighted neuroimaging data of infants 0–2 years. Neuroimage **218**, 116946 (2020)