



Conditional Temporal Attention Networks for Neonatal Cortical Surface Reconstruction

Qiang Ma^{1(✉)}, Liu Li¹, Vanessa Kyriakopoulou², Joseph V. Hajnal²,
Emma C. Robinson², Bernhard Kainz^{1,2,3}, and Daniel Rueckert^{1,4}

¹ BioMedIA, Department of Computing, Imperial College London, London, UK
q.ma20@imperial.ac.uk

² King's College London, London, UK

³ FAU Erlangen-Nürnberg, Erlangen, Germany

⁴ Klinikum rechts der Isar, Technical University of Munich, Munich, Germany

Abstract. Cortical surface reconstruction plays a fundamental role in modeling the rapid brain development during the perinatal period. In this work, we propose Conditional Temporal Attention Network (CoTAN), a fast end-to-end framework for diffeomorphic neonatal cortical surface reconstruction. CoTAN predicts multi-resolution stationary velocity fields (SVF) from neonatal brain magnetic resonance images (MRI). Instead of integrating multiple SVFs, CoTAN introduces attention mechanisms to learn a conditional time-varying velocity field (CTVF) by computing the weighted sum of all SVFs at each integration step. The importance of each SVF, which is estimated by learned attention maps, is conditioned on the age of the neonates and varies with the time step of integration. The proposed CTVF defines a diffeomorphic surface deformation, which reduces mesh self-intersection errors effectively. It only requires 0.21 s to deform an initial template mesh to cortical white matter and pial surfaces for each brain hemisphere. CoTAN is validated on the Developing Human Connectome Project (dHCP) dataset with 877 3D brain MR images acquired from preterm and term born neonates. Compared to state-of-the-art baselines, CoTAN achieves superior performance with only 0.12 ± 0.03 mm geometric error and $0.07 \pm 0.03\%$ self-intersecting faces. The visualization of our attention maps illustrates that CoTAN indeed learns coarse-to-fine surface deformations automatically without intermediate supervision.

1 Introduction

Cortical surface reconstruction aims to extract 3D meshes of inner (white matter) and outer (pial) surfaces of the cerebral cortex from brain magnetic resonance images (MRI). These surfaces provide both 3D visualization and estimation of morphological features for the cortex [7, 11, 12]. In addition to accurately representing the highly folded cortex, the cortical surfaces of each hemisphere are

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/978-3-031-43901-8_30.

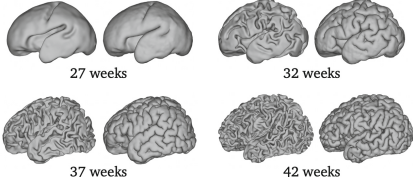


Fig. 1. Neonatal cortical surfaces at different post-menstrual ages.

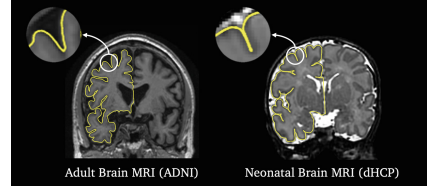


Fig. 2. Left: adult brain MRI from the ADNI dataset [20]. Right: neonatal brain MRI from the dHCP dataset [9].

required to be closed manifolds and topologically homeomorphic to a sphere [7]. Traditional neuroimage analysis pipelines [10, 13, 24, 29] such as FreeSurfer [10] comprise a series of processing steps to extract cortical surfaces from brain MRI. These pipelines provide an invaluable service to the research community. However, the current implementations have limited accuracy and require several hours to process a single MRI scan.

With the recent advances of geometric deep learning [32, 33], a growing number of fast learning-based approaches have been proposed to learn either implicit surface representation [6, 15] or explicit mesh deformation [4, 17, 21–23, 28] for cortical surface reconstruction. These approaches enhance the accuracy and reduce the processing time to a few seconds for a single subject. Recent studies focus on generating manifold cortical meshes [21, 22, 28] and preventing mesh self-intersections by learning diffeomorphic deformations, which have been widely adopted in medical image registration [1–3]. The basic idea is to learn stationary velocity fields (SVF) to deform an initial mesh template to target surfaces. Since a single SVF has limited representation capacity, several approaches [16, 21, 28, 30] have proposed to train multiple neural networks to predict a sequence of SVFs for coarse-to-fine surface deformation. This improves the geometric accuracy but increases the computational burden for both training and inference.

Cortical surface reconstruction plays an essential role in modeling and quantifying the brain development in fetal and neonatal neuroimaging studies such as the Developing Human Connectome Project (dHCP) [9, 24]. However, most learning-based approaches so far rely on adult MRI as training data [13, 20, 25]. Compared to adult data, neonatal brain MR images have lower resolution and contrast due to a smaller region of interest and the use of, *e.g.*, fast imaging sequences with sparse acquisition to minimize head motion artifacts of unsedated infants [5, 19]. Besides, the rapid growth and continuously increasing complexity of the cortex during the perinatal period lead to considerable variation in shape and scale between neonatal cortical surfaces at different post-menstrual ages (PMA) (see Fig. 1). Moreover, since the neonatal head is still small, the cortical sulci of term-born neonates are much narrower than adults as shown in Fig. 2. Hence, the neonatal pial surfaces are prone to be affected by partial volume effects and more likely to produce surface self-intersections.

Contribution. In this work, we present Conditional Temporal Attention Network (CoTAN). CoTAN adopts attention mechanism [18, 27] to learn a conditional time-varying velocity field (CTVF) for neonatal cortical surface reconstruction. Given an input brain MR image, CoTAN first predicts multiple SVFs at different resolutions. Rather than integrating all SVFs as [16, 21, 28], CoTAN learns conditional temporal attention maps to attend to specific SVFs for different time steps of integration and PMA of neonates. The CTVF is represented by the weighted sum of learned SVFs, and thus a single CoTAN model is sufficient to model the large deformation and variation of neonatal cortical surfaces. The evaluation on the dHCP neonatal dataset [9] verifies that CoTAN performs better in geometric accuracy, mesh quality and computational efficiency than state-of-the-art methods. The visualization of attention maps indicates that CoTAN learns coarse-to-fine deformations automatically without intermediate constraints. Our source code is released publicly at <https://github.com/m-qiang/CoTAN>.

2 Method

Diffeomorphic Surface Deformation. We define the diffeomorphic surface deformation $\phi_t : \mathbb{R}^3 \times \mathbb{R} \rightarrow \mathbb{R}^3$ as a flow ordinary differential equation (ODE) following previous work [1–3, 16, 21, 22, 28]:

$$\frac{\partial}{\partial t} \phi_t = v_t(\phi_t), \quad \phi_0 = Id, \quad t \in [0, T], \quad (1)$$

where v_t is a time-varying velocity field (TVF) and Id is the identity mapping. Given an initial surface $S_0 \subset \mathbb{R}^3$ with points $x_0 \in S_0$, we define $x_t := \phi_t(x_0)$ as the trajectories of the points on the deformable surface $S_t = \phi_t(S_0)$ for $t \in [0, T]$. Then the flow equation (1) can be rewritten as $\frac{d}{dt} x_t = v_t(x_t)$ with initial value x_0 . By the existence and uniqueness theorem for ODE solutions [16], if $v_t(x)$ is Lipschitz continuous with respect to x , the trajectories x_t will not intersect with each other, so that the surface self-intersections can be prevented effectively. By integrating the ODE (1), we obtain a diffeomorphism ϕ_T that deforms S_0 to a manifold surface S_T , on which the points are $x_T = \phi_T(x_0) = x_0 + \int_0^T v_t(x_t) dt$.

Conditional Temporal Attention Network (CoTAN). An overview of the CoTAN architecture is shown in Fig. 3. CoTAN first predicts multiple SVFs given a 3D brain MRI volume. A 3D U-Net [26] is used to extract feature maps with R resolution levels, each of which scales the input size by the factor of 2^{r-R} for $r = 1, \dots, R$. Then we upsample the multi-scale feature maps and learn M volumetric SVFs for each resolution. Let \mathbf{V} denote all $R \times M$ discrete SVFs. The continuous multi-resolution SVFs $\mathbf{v} : \mathbb{R}^3 \rightarrow \mathbb{R}^{R \times M \times 3}$ can be obtained by $\mathbf{v}(x) = \text{Lerp}(x, \mathbf{V})$, where $\text{Lerp}(\cdot)$ is the trilinear interpolation function. Each element $\mathbf{v}^{r,m} : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ is an SVF for $r = 1, \dots, R$ and $m = 1, \dots, M$. Note that $\mathbf{v}(x)$ is Lipschitz continuous since $\text{Lerp}(\cdot)$ is continuous and piece-wise linear.

CoTAN adopts a channel-wise attention mechanism [18, 27] to focus on specific SVFs since it is time-consuming to integrate all $R \times M$ SVFs [16, 21, 28].

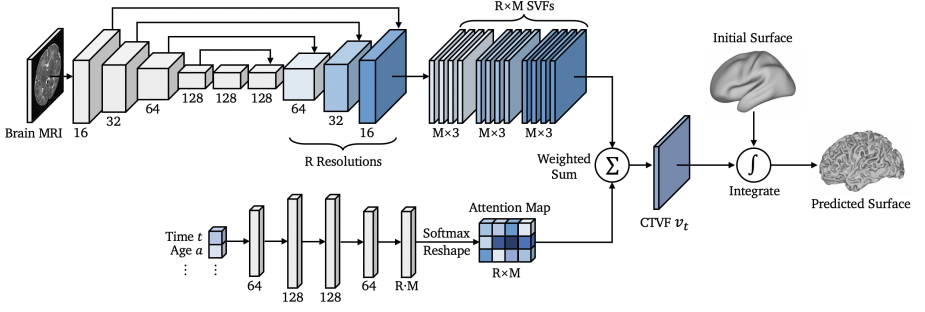


Fig. 3. The architecture of the proposed CoTAN framework. Given an input 3D brain MRI, CoTAN uses a U-Net to predict M SVFs for each resolution level R . An attention map is learned to focus on specific SVFs varying with the input integration time t and conditioned on the age a of the neonatal subjects. Further conditioning could be achieved with minimal effort, *e.g.*, biological sex, diagnoses, etc. For each time step t , the CTVF v_t is represented by the weighted sum of all $R \times M$ SVFs. By integrating the CTVF, CoTAN deforms an input initial surface to the predicted cortical surface.

The attention is conditioned on both integration time $t \in [0, T]$ and information about the subject. To model the high variation between infant brains, we consider the post-menstrual ages (PMA) $a \in \mathbb{R}$ of the neonates at scan time as the conditioning variable in this work. Note that we do not use a self-attention module [8, 31] to learn key and query pairs. Instead, we learn a probability attention map to measure the importance of each SVF. More precisely, as shown in Fig. 3, we use a fully connected network (FCN) to encode the input time t and PMA a into a $(R \cdot M) \times 1$ feature vector. After reshaping and softmax activation, the FCN learns conditional temporal attention maps $\mathbf{p}(t, a) \in \mathbb{R}^{R \times M}$ which satisfy $\sum_{r=1}^R \sum_{m=1}^M \mathbf{p}^{r,m}(t, a) = 1$ for any t and a . Then a *conditional time-varying velocity field* (CTVF) is predicted by computing the weighted sum of all SVFs:

$$v_t(x; a) = \sum_{r=1}^R \sum_{m=1}^M \mathbf{p}^{r,m}(t, a) \cdot \mathbf{v}^{r,m}(x). \quad (2)$$

The CTVF is adaptive to the integration time and the age of subjects, which can handle the large deformation and variation for neonatal cortical surfaces. Such an attention mechanism encourages CoTAN to learn coarse-to-fine surface deformation by attending to SVFs at different resolutions.

To deform the initial surface S_0 to the target surface, we integrate the flow ODE (1) with the CTVF through the forward Euler method. For $k = 0, \dots, K-1$, the surface points are updated by $x_{k+1} = x_k + h v_k(x_k; a)$, where K is the total integration steps and $h = T/K$ is the step size with $T = 1$. For each step k , we only need to recompute the attention maps $\mathbf{p}(hk, a)$ and update the CTVF $v_k(x_k; a)$ accordingly by Eq. (2). CoTAN only integrates a single CTVF which saves considerable runtime compared to integrating multiple SVFs directly as [16, 21, 28].

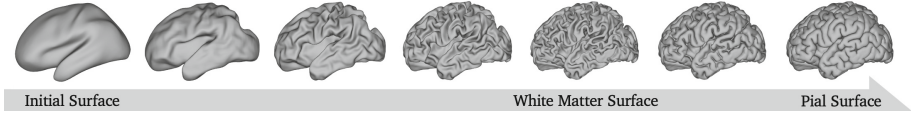


Fig. 4. Diffeomorphic deformation from an initial template mesh to cortical surfaces.

Neonatal Cortical Surface Reconstruction. We train two CoTAN models on the dHCP neonatal dataset [9] to 1) deform an initial surface into a white matter surface and to 2) expand the white matter surface into a pial surface as shown in Fig. 4. We use the same initial surface (leftmost in Fig. 4) for all subjects, which is created by iteratively applying Laplacian smoothing to a Conte-69 surface atlas [14]. We generate pseudo ground truth (GT) surfaces by the dHCP structural neonatal pipeline [24], which has been fully validated through quality control performed by clinical experts.

For white matter surface reconstruction, we consider loss functions that have been widely used in previous work [4, 32, 33]: the Chamfer distance loss \mathcal{L}_{cd} computes the distance between two point clouds, the mesh Laplacian loss \mathcal{L}_{lap} regularizes the smoothness of the mesh, and the normal consistency loss \mathcal{L}_{nc} constrains the cosine similarity between the normals of two adjacent faces. The final loss is weighted by $\mathcal{L} = \mathcal{L}_{cd} + \lambda_{lap}\mathcal{L}_{lap} + \lambda_{nc}\mathcal{L}_{nc}$. We train CoTAN in two steps for white matter surface extraction. First, we pre-train the model using relatively large weights λ_{lap} and λ_{nc} for regularization. The Chamfer distance is computed between the vertices of the predicted and pseudo-GT surfaces. These ensure that the initial surface can be deformed robustly during training. Then, we fine-tune CoTAN using small weights to increase geometric accuracy. The distances are computed between 150k uniformly sampled points on the surfaces.

For pial surface reconstruction, we follow [22, 23] and use the pseudo-GT white matter surfaces as the input for training. Then the MSE loss $\mathcal{L} = \sum_i \|\hat{x}_i - x_i^*\|^2$ can be computed between the vertices of predicted and pseudo-GT pial meshes. No point matching is required since the pseudo-GT white matter and pial surfaces have the same mesh connectivity. Therefore, the MSE loss provides stronger supervision than the Chamfer distance, while the latter is prone to mismatching the points in narrow sulci, resulting in mesh self-intersections.

3 Experiments

Implementation Details. We evaluate CoTAN on the third release of dHCP neonatal dataset [9] (<https://biomedica.github.io/dHCP-release-notes/>), which includes 877 T2-weighted (T2w) brain MRI scanned from newborn infants at PMA between 27 to 45 weeks. The MRI images are affinely aligned to the MNI152 space and clipped to the size of $112 \times 224 \times 160$ for each brain hemisphere. The dataset is split into 60/10/30% for training/validation/testing.

For the CoTAN model, we set the resolution $R=3$ and the number of SVFs $M=4$ for each resolution. For integration, we set the total number of steps to

Table 1. Comparative results of neonatal cortical surface reconstruction on the dHCP dataset. The geometric accuracy (ASSD and HD90) and mesh quality (the ratio of SIFs) are reported for white matter and pial surfaces. Smaller values mean better results. *CoTAN (ours) shows significant improvement ($p < 0.05$) compared to baselines.

Method	White Matter Surface			Pial Surface		
	ASSD (mm)	HD90 (mm)	SIF (%)	ASSD (mm)	HD90 (mm)	SIF (%)
CoTAN	0.107 \pm 0.026	0.217 \pm 0.076	0.001 \pm 0.004	0.121 \pm 0.029	0.259 \pm 0.075	0.071 \pm 0.034
CortexODE	0.109 \pm 0.052	0.231 \pm 0.326	0.001 \pm 0.002	0.134 \pm 0.052*	0.306 \pm 0.358*	0.221 \pm 0.114*
CFPP	0.118 \pm 0.028*	0.241 \pm 0.085*	0.075 \pm 0.057*	0.124 \pm 0.031*	0.273 \pm 0.086*	2.457 \pm 1.003*
CorticalFlow	0.122 \pm 0.029*	0.247 \pm 0.080*	0.048 \pm 0.032*	0.157 \pm 0.031*	0.331 \pm 0.089*	9.798 \pm 2.902*
Vox2Cortex	0.115 \pm 0.035*	0.233 \pm 0.110*	0.253 \pm 0.169*	0.130 \pm 0.039*	0.291 \pm 0.141*	14.366 \pm 2.262*
DeepCSR	0.129 \pm 0.047*	0.276 \pm 0.211*	—	0.299 \pm 0.070*	1.214 \pm 0.337*	—

Table 2. Comparative results of runtime, GPU memory cost, and the number of model parameters for both training and testing.

Method	Runtime		GPU (GB)		Model
	Train	Test	Train	Test	#Param
CoTAN (Ours)	57.9h	0.21s	8.71	4.05	2.47 M
CortexODE	51.6h	1.88 s	9.24	4.61	1.99 M
CFPP	131.5h	0.57 s	9.86	3.56	1.03 M
CorticalFlow	105.7h	0.51 s	9.12	3.56	1.03 M
Vox2Cortex	63.7h	1.48 s	6.96	6.26	6.54 M
DeepCSR	15.1 h	10.69 s	5.26	2.33	4.65 M

Table 3. The results of ablation experiments for CoTAN on white matter surface reconstruction.

Method	ASSD (mm)	HD90 (mm)
CoTAN (Ours)	0.107 \pm 0.026	0.217 \pm 0.076
Pre-train	0.116 \pm 0.029	0.238 \pm 0.095
$R=1$	0.112 \pm 0.027	0.232 \pm 0.081
$M=1$	0.111 \pm 0.030	0.231 \pm 0.097
SVF ($a=t=0$)	0.138 \pm 0.037	0.291 \pm 0.109
CVF ($t=0$)	0.135 \pm 0.036	0.285 \pm 0.111
TVF ($a=0$)	0.108 \pm 0.028	0.222 \pm 0.090

$K=50$ with step size $h=0.02$. We re-mesh the initial mesh to 140k vertices, of which the coordinates are normalized to $[-1, 1]$. We use the Adam optimizer for training. For the white matter surface, we first pre-train CoTAN for 100 epochs using a learning rate of $\gamma=10^{-4}$ and weights $\lambda_{lap}=0.5$, $\lambda_{nc}=5 \times 10^{-4}$ for the loss function. Then we fine-tune for 100 epochs using smaller weights $\lambda_{lap}=0.1$ and $\lambda_{nc}=10^{-4}$ with $\gamma=2 \times 10^{-5}$. For the pial surface, we set the maximum channel size of CoTAN as 32 to avoid overfitting and train for 200 epochs with $\gamma=10^{-4}$. We only consider left brain hemisphere in the experiments. All experiments are conducted on a Nvidia RTX3080 GPU with 12GB memory.

Comparative Results. We compare the performance of CoTAN with existing learning-based cortical surface extraction approaches including CortexODE [22], CorticalFlow++ (CFPP) [28], CorticalFlow [21], Vox2Cortex [4] and DeepCSR [6]. We employ the fast topology correction [22] for DeepCSR. CoTAN can guarantee spherical topology and the Euler number is 2 for all predicted surfaces.

—*Geometric Accuracy:* We measure the geometric accuracy by commonly used metrics [4, 6, 22]: average symmetric surface distance (ASSD) and 90th percentile of Hausdorff distance (HD90). The distances are computed between uniformly sampled 100k points on the predicted and pseudo-GT surfaces. For fair com-

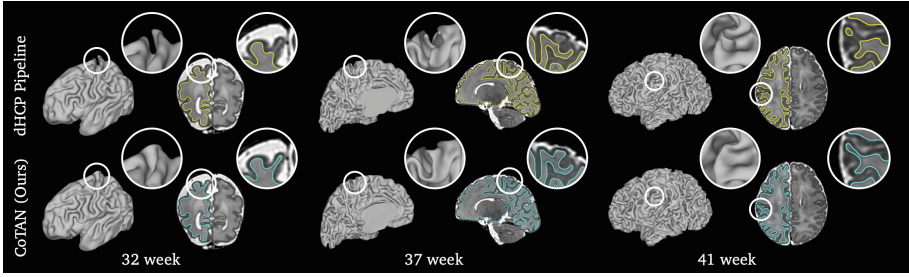


Fig. 5. Visualization of neonatal cortical surfaces generated by the dHCP structural pipeline [24] and CoTAN for different ages. CoTAN shows better anatomical accuracy as highlighted in both surface meshes and corresponding brain MRI images.

parison, the predicted cortical meshes have around 140k vertices for all baseline approaches. Note that CoTAN and [6, 21, 28] can generalize on high-resolution meshes with up to 600k vertices (see Appendix). We conduct a paired t-test to examine the statistical significance. As reported in Table 1, CoTAN achieves significantly superior geometric accuracy (p -value <0.05) compared to all state-of-the-art baselines, except for the white matter surfaces of CortexODE.

—*Mesh Quality*: We evaluate the mesh quality by the ratio of self-intersecting faces (SIF) as shown in Table 1. Note that due to the narrower sulcal gaps of neonatal cortex compared to adult [20, 25] (see Fig. 2), all baseline methods produce more SIFs than reported in their original papers. Except that DeepCSR produces no SIFs since it uses the Marching Cubes algorithm, CoTAN achieves the best mesh quality with only 0.001% SIFs in white matter surfaces and 0.071% SIFs in pial surfaces. These remaining SIFs are likely introduced by the discretization of triangular mesh representation and ODE integration.

CoTAN produces fewer SIFs for three reasons. Firstly, CoTAN employs diffeomorphic deformation, while the non-diffeomorphic Vox2Cortex creates 14% SIFs. We further set the integration steps $K=5$ for CoTAN so that the deformation is no longer diffeomorphic. The SIFs of pial surfaces are increased to $2.99 \pm 1.19\%$. Secondly, CoTAN reconstructs the pial surface by expanding the input white matter surface. It is difficult to avoid collisions during the deformation from a smooth template into deep sulci, *e.g.*, in Vox2Cortex and CorticalFlow. Lastly, CoTAN uses MSE loss rather than Chamfer distance for pial surface extraction, which alleviates the mismatch between points in the narrow sulci.

—*Computational Efficiency*: We report the runtime and GPU memory cost for both training and testing, as well as the number of learnable parameters for CoTAN and all baseline approaches in Table 2. The runtime includes both model inference and post-processing. CoTAN only requires 0.21 s to extract cortical surfaces for each hemisphere, which is $2\times$ faster than the best baseline. CoTAN can be trained end-to-end and reduces training time by 46% compared to CorticalFlow and CFPP, which have to train three U-Nets consecutively to parameterize three SVFs for a single surface. Although CoTAN uses relatively

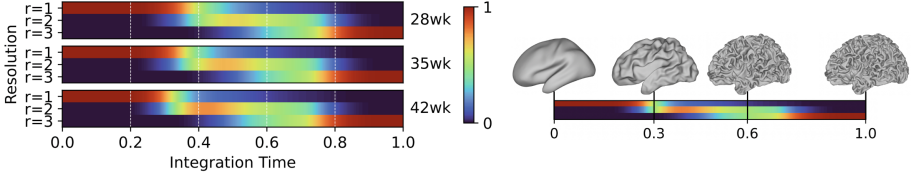


Fig. 6. Visualization of attention maps predicted by CoTAN for white matter surface reconstruction. Left: conditional temporal attention maps $\mathbf{p}^r(t, a)$ at different time and ages. Right: coarse-to-fine white matter surface deformations learned by CoTAN.

more parameters to learn surface deformations, it is still memory efficient which only costs 8.7GB and 4GB GPU memory for training and testing.

Ablation Study. We conduct ablation studies on CoTAN and evaluate the geometric accuracy on the white matter surface reconstruction. First, without fine-tuning, the geometric errors increase by 8% as reported in Table 3. Then we consider CoTAN with single resolution ($R=1$) or only predict a single SVF ($M=1$) for each resolution. The geometric distances increase in both cases.

Next, we examine the effectiveness of the CTVF $v_t(x; a)$ by fixing the input time $t=0$ or age $a=0$. We train CoTAN models to predict the TVF ($a=0$), CVF ($t=0$) and SVF ($a=t=0$), which are degraded from the CTVF. Table 3 shows that the SVF increases geometric errors by 30% due to its limited representation capacity. The CVF increases accuracy slightly by learning conditional deformations adaptive to the age of neonates. The TVF exploits temporal information to model a wider range of deformations and effectively improves the performance.

Lastly, we train a U-Net to predict $R \times M$ SVFs and integrate them directly without attention. Since the gradients are backpropagated through all SVFs, it requires >140 h training time which is $2.4 \times$ slower than our attention-based CoTAN. The model is also sensitive to small updates, which can affect all SVFs. This results in exploding gradients which we have observed in the training, whereas CoTAN can be trained robustly by integrating a single CTVF.

Attention Maps. We explore the attention maps $\mathbf{p}^{r,m}(t, a)$ learned by CoTAN. We define $\mathbf{p}^r = \sum_{m=1}^M \mathbf{p}^{r,m}$ to reflect the importance of the SVFs at each resolution level $r=1, \dots, R$, where $R=3$ and larger r means higher resolution. Figure 6 visualizes the attention maps $\mathbf{p}^r(t, a)$ for white matter surface reconstruction for integration time $t \in [0, 1]$ and age $a \in \{28, 35, 42\}$. It shows that the attention maps focus on low resolution ($r=1$) at the beginning of the integration, and then attends to high resolution ($r=3$) when $t \rightarrow 1$. Furthermore, Fig. 6 (Right) shows that an initial white matter surface deforms into a coarse shape for $t \leq 0.3$ and learns fine details for $t \geq 0.6$. This matches the attention maps and demonstrates that CoTAN learns coarse-to-fine deformations automatically without any supervision on the intermediate deformations. Additionally, Fig. 6 shows that CoTAN pays more attention to low resolution for younger subjects (28 week) whose brains have not fully developed yet. More deformations at higher resolutions ($r \geq 2$) are required for older neonates (≥ 35 week) with highly folded cortex.

Discussion. One limitation of our experiments is that we only train and evaluate CoTAN based on the pseudo-GT generated by the dHCP structural pipeline [24]. Previous approaches [4, 6, 22] have been validated by the test-retest experiments. However, this is infeasible for neonates, whose brain develops rapidly even within a short period. To verify the superior anatomical accuracy of CoTAN, we provide qualitative comparison between the pseudo-GT and CoTAN as visualized in Fig. 5. It shows that CoTAN can effectively mitigate corruptions introduced by the dHCP pipeline for neonatal subjects at different ages. In addition, the dHCP pipeline requires 4 h to process a single subject [24], while CoTAN extracts cortical surfaces in only 0.21 s for each brain hemisphere.

4 Conclusion

In this work, we propose CoTAN for diffeomorphic neonatal cortical surface reconstruction. CoTAN employs an attention mechanism to combine multiple SVFs to a CTVF, which outperforms existing baselines in geometric accuracy and mesh quality. CoTAN can also be extended and applied to extract adult cortical surfaces conditioned on the age, gender or pathological information of the subjects. Our future work will integrate CoTAN into a learning-based pipeline for universal cortical surface analysis across all age groups.

Acknowledgements. This work was supported by the President’s PhD Scholarship at Imperial College London. Support was also received from the ERC project MIA-NORMAL 101083647 and ERC project Deep4MI 884622. Data were provided by the developing Human Connectome Project, KCL-Imperial-Oxford Consortium funded by the ERC under the European Union Seventh Framework Programme (FP/2007-2013) / ERC Grant Agreement no. [319456]. We are grateful to the families who generously supported this trial.

References

1. Ashburner, J.: A fast diffeomorphic image registration algorithm. *Neuroimage* **38**(1), 95–113 (2007)
2. Balakrishnan, G., Zhao, A., Sabuncu, M.R., Gutttag, J., Dalca, A.V.: VoxelMorph: a learning framework for deformable medical image registration. *IEEE Trans. Med. Imaging* **38**(8), 1788–1800 (2019)
3. Beg, M.F., Miller, M.I., Trounev, A., Younes, L.: Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *Int. J. Comput. Vision* **61**, 139–157 (2005)
4. Bongratz, F., Rickmann, A.M., Pölsterl, S., Wachinger, C.: 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*, pp. 20773–20783 (2022)
5. Cordero-Grande, L., Hughes, E.J., Hutter, J., Price, A.N., Hajnal, J.V.: Three-dimensional motion corrected sensitivity encoding reconstruction for multi-shot multi-slice MRI: application to neonatal brain imaging. *Magn. Reson. Med.* **79**(3), 1365–1376 (2018)

6. 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)
7. Dale, A.M., Fischl, B., Sereno, M.I.: Cortical surface-based analysis: I. segmentation and surface reconstruction. *Neuroimage* **9**(2), 179–194 (1999)
8. Dosovitskiy, A., et al.: An image is worth 16×16 words: transformers for image recognition at scale. In: International Conference on Learning Representations (2021)
9. Edwards, A.D., et al.: The developing human connectome project neonatal data release. *Front. Neurosci.* **16**, 886772 (2022)
10. Fischl, B.: FreeSurfer. *Neuroimage* **62**(2), 774–781 (2012)
11. Fischl, B., Dale, A.M.: Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proc. Natl. Acad. Sci.* **97**(20), 11050–11055 (2000)
12. Fischl, B., Sereno, M.I., Dale, A.M.: Cortical surface-based analysis: II: inflation, flattening, and a surface-based coordinate system. *Neuroimage* **9**(2), 195–207 (1999)
13. Glasser, M.F., Sotiropoulos, S.N., Wilson, J.A., et al.: The minimal preprocessing pipelines for the human connectome project. *Neuroimage* **80**, 105–124 (2013)
14. Glasser, M.F., Van Essen, D.C.: Mapping human cortical areas in vivo based on myelin content as revealed by T1- and T2-weighted MRI. *J. Neurosci.* **31**(32), 11597–11616 (2011)
15. Gopinath, K., Desrosiers, C., Lombaert, H.: SEGReCON: learning joint brain surface reconstruction and segmentation from images. In: de Bruijne, M., et al. (eds.) MICCAI 2021. LNCS, vol. 12907, pp. 650–659. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87234-2_61
16. Gupta, K., Chandraker, M.: Neural mesh flow: 3D manifold mesh generation via diffeomorphic flows. In: Proceedings of the 34th International Conference on Neural Information Processing Systems, pp. 1747–1758 (2020)
17. Hoopes, A., Iglesias, J.E., Fischl, B., Greve, D., Dalca, A.V.: TopoFit: rapid reconstruction of topologically-correct cortical surfaces. In: Medical Imaging with Deep Learning (2021)
18. Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7132–7141 (2018)
19. Hughes, E.J., Winchman, T., Padormo, F., Teixeira, R., et al.: A dedicated neonatal brain imaging system. *Magn. Reson. Med.* **78**(2), 794–804 (2017)
20. Jack, C.R., Jr., et al.: The Alzheimer’s disease neuroimaging initiative (ADNI): MRI methods. *J. Magn. Reson. Imaging* **27**(4), 685–691 (2008)
21. Lebrat, L., et al.: CorticalFlow: a diffeomorphic mesh transformer network for cortical surface reconstruction. *Adv. Neural. Inf. Process. Syst.* **34**, 29491–29505 (2021)
22. Ma, Q., Li, L., Robinson, E.C., Kainz, B., Rueckert, D., Alansary, A.: Cortex-ODE: learning cortical surface reconstruction by neural ODEs. *IEEE Trans. Med. Imaging* **42**, 430–443 (2022)
23. 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

24. Makropoulos, A., et al.: The developing human connectome project: a minimal processing pipeline for neonatal cortical surface reconstruction. *Neuroimage* **173**, 88–112 (2018)
25. Marcus, D.S., Wang, T.H., Parker, J., Csernansky, J.G., Morris, J.C., Buckner, R.L.: Open access series of imaging studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. *J. Cogn. Neurosci.* **19**(9), 1498–1507 (2007)
26. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) *MICCAI 2015*. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
27. Roy, A.G., Navab, N., Wachinger, C.: Concurrent spatial and channel ‘squeeze & excitation’ in fully convolutional networks. In: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds.) *MICCAI 2018*. LNCS, vol. 11070, pp. 421–429. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00928-1_48
28. 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
29. Shattuck, D.W., Leahy, R.M.: BrainSuite: an automated cortical surface identification tool. *Med. Image Anal.* **6**(2), 129–142 (2002)
30. Sun, S., Han, K., Kong, D., Tang, H., Yan, X., Xie, X.: Topology-preserving shape reconstruction and registration via neural diffeomorphic flow. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 20845–20855 (2022)
31. Vaswani, A., et al.: Attention is all you need. In: *Advances in Neural Information Processing Systems*, vol. 30 (2017)
32. Wang, N., Zhang, Y., Li, Z., Fu, Y., Liu, W., Jiang, Y.G.: Pixel2Mesh: generating 3D mesh models from single RGB images. In: *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 52–67 (2018)
33. 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