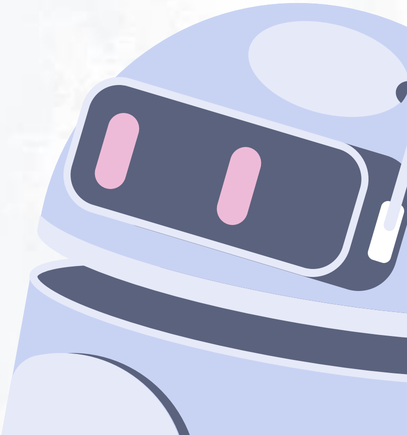
Conditional Temporal Attention Networks for Neonatal Cortical Surface Reconstruction

Qiang Ma, Liu Li, Vanessa Kyriakopoulou, Joseph Hajnal,
Emma C. Robinson, Bernhard Kainz, and Daniel Rueckert

MICCAI 2023



Neuro-iX

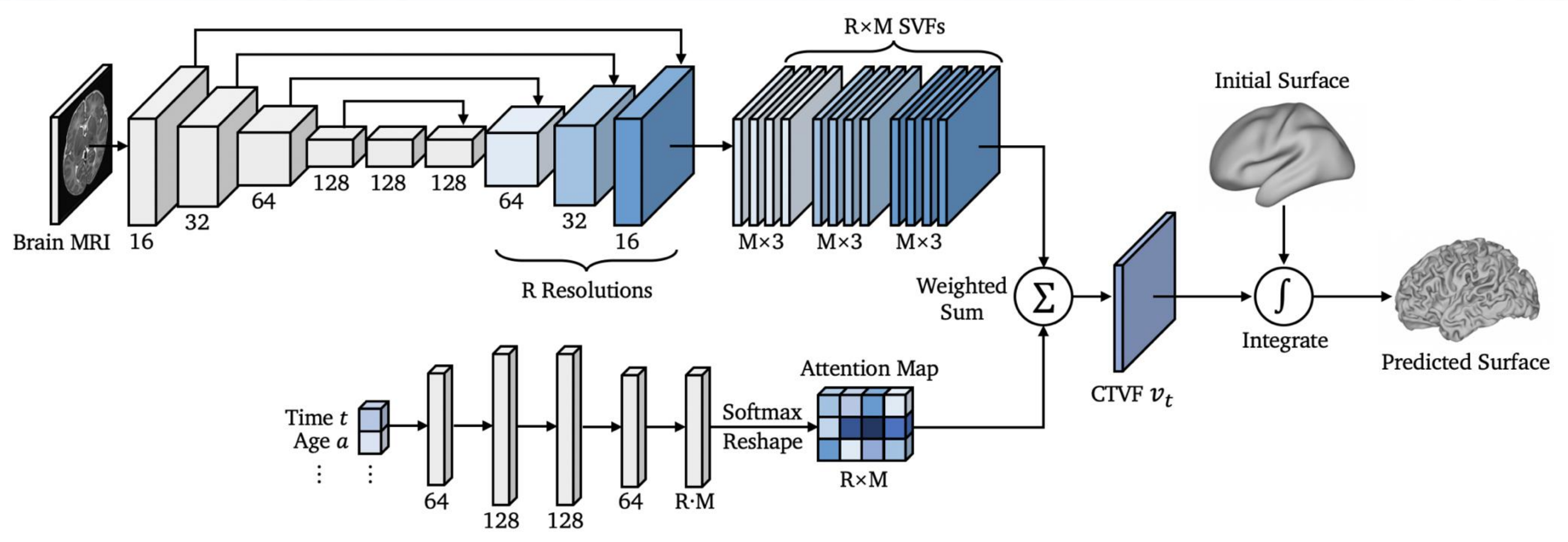
Why Cortical Surface Reconstruction Matters !!

- Extracts 3D surfaces (white matter & pial) from brain MRI
- Crucial for visualizing and analyzing brain development
- Surfaces must be accurate, smooth, and topologically correct
- Traditional tools like FreeSurfer are slow (hours per MRI)
- New learning-based methods are fast (seconds), but have challenges:
  - Mesh self-intersections
  - Poor handling of neonatal brain variation
- Neonatal MRIs are lower quality & more variable due to age
- Need a fast, accurate, neonatal-specific method

Overview of CoTAN

- **CoTAN** = Conditional Temporal Attention Network
- Designed for fast, accurate neonatal brain surface reconstruction
- Starts from a smooth template and deforms it using MRI input
- Predicts **multiple velocity fields (SVFs)** at different resolutions
- Learns **when and which SVFs to use** through an attention mechanism
- Attention is conditioned on:
  - **Time** in the deformation process
  - **Age** of the infant **(PMA)**
- Produces clean, detailed surfaces in a **single pass**
- Trained and validated on the **dHCP neonatal dataset**
- Learns **coarse-to-fine deformations automatically**

CoTAN Architecture



CoTAN takes a 3D brain MRI as input and uses a U-Net to predict multiple motion fields (SVFs) at different resolutions. An attention network selects which fields to focus on, based on time and the baby’s age. These are combined into a smart velocity field (CTVF), which deforms an initial surface step-by-step into the final brain surface

Theoretical Foundations of CoTAN

Diffeomorphic Deformation via ODE: $\frac{\partial \phi_t}{\partial t} = v_t(\phi_t), \ \phi_0 = \mathrm{Id}, \ t \in [0, T]$

Multi-Resolution SVFs

SVFs are interpolated (using trilinear interpolation or *Lerp*) to obtain a continuous function v(x)

Conditional Temporal Attention:

An FCN encodes the current **time (t)** and the neonate's **age (a)** to produce attention maps $\mathrm{P}_{(t,\alpha)}^{r,m}$

These maps satisfy $\sum_{r=1}^{R} \sum_{m=1}^{M} \mathrm{P}_{(t,\alpha)}^{r,m} = 1$ and determine the weight for each SVF

Conditional Time-Varying Velocity Field (CTVF)

The final velocity is a weighted sum of SVFs: $v_t(x;\alpha) = \sum_{r=1}^{R} \sum_{m=1}^{M} \mathrm{P}_{(t,\alpha)}^{r,m} \cdot v^{r,m}(x)$

This CTVF is then integrated via forward Euler: $x_{\kappa+1} = x_\kappa + h v_\kappa(x_\kappa;\alpha)$

Neonatal Cortical Surface Reconstruction Overview

- **Two-Stage Approach:**
  - **Stage 1:** Deform initial surface into a **white matter** surface
  - **Stage 2:** Expand white matter into a **pial** surface
- Use a **common initial mesh** (generated via Laplacian smoothing of a Conte-69 atlas)
- **Pseudo Ground Truth (GT)** obtained from the validated dHCP structural pipeline

White Matter Surface Reconstruction

Loss Functions:

Chamfer Distance Loss ($\mathcal{L}_{cd}$): Measures point cloud similarity

Mesh Laplacian Loss ($\mathcal{L}_{lap}$): Enforces smoothness.

Normal Consistency Loss ($\mathcal{L}_{nc}$): Aligns face normals

$$\mathcal{L} = \mathcal{L}_{cd} + \lambda_{lap}\mathcal{L}_{lap} + \lambda_{nc}\mathcal{L}_{nc}$$

Training Strategy:

Pre-Training: High λ values for smooth, robust deformation

Fine-Tuning: Lower λ values; loss computed on 150k sampled points for enhanced accuracy

Pial Surface Reconstruction

Input: Pseudo GT white matter surface is used as the starting point

Loss Function:

Mean Squared Error (MSE): Measures vertex-wise differences between predicted and pseudo GT pial surfaces

Why MSE?

Provides stronger supervision by taking advantage of identical mesh connectivity

Avoids mismatches common with Chamfer loss in narrow sulci

Outcome:

Better refined pial surfaces with accurate, detailed deformations

Dataset & Model Setup

Dataset:

dHCP neonatal dataset

877 T2-weighted MRI scans

PMA: 27 to 45 weeks

Clipped to 112×224×160 per hemisphere

Split: 60% training / 10% validation / 30% testing

Model Configuration:

Resolution $R = 3$ (3 scales)

$M = 4$ SVFs per scale → 12 SVFs total

Integration: $K = 50$ steps, $h = 0.02$ per step

Geometric Accuracy & Mesh Quality

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

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.

Computational Efficiency

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

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

Runtime:

Only **0.21 seconds per hemisphere** (including inference and post-processing)

Approximately 2× faster than the best alternative methods

Training Efficiency:

End-to-end training reduces time by 46% compared to methods like CorticalFlow and CFPP, which require multiple networks

Memory Footprint:

GPU usage: 8.7GB for training and 4GB for testing

Overall:

Although CoTAN uses relatively more parameters, it is both fast and memory-efficient

Ablation Study

Fine-Tuning Matters:

Skipping fine-tuning increases geometric error by 8%

Multiscale & Multiple SVFs:

Using a single resolution (R = 1) or a single SVF (M = 1) degrades accuracy

CTVF Variants:

SVF (a = t = 0): 30% higher errors due to limited representation

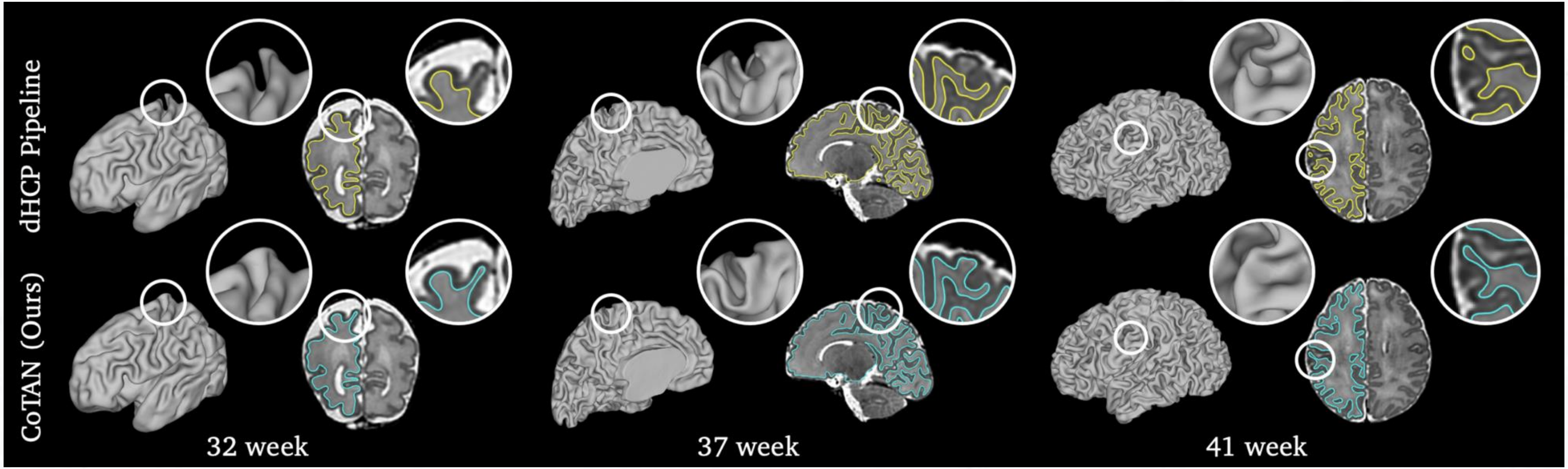
CVF (t = 0): Improves accuracy slightly by considering age only

TVF (a = 0): Leverages time to better model deformations

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

CoTAN on White Mattersurface reconstruction

Visual Comparisons



Visual comparisons show CoTAN mitigates errors from the dHCP pipeline

Conclusion

- Introduces a diffeomorphic, attention-based framework for neonatal cortical surface reconstruction
- Combines multiple motion fields (SVFs) into a single, time- and age-adaptive velocity field (CTVF)
- Dramatically reduced runtime—from hours to seconds per subject

Thanks!

Any questions?

Moradkhanikaveh94@gmail.com