Recurrent Tissue-Aware Network for Deformable Registration of Infant Brain MR Images

Dongming Wei, Sahar Ahmad, Yuyu Guo, Liyun Chen, Yunzhi Huang, Lei Ma, Zhengwang Wu, Gang Li, *Senior Member, IEEE*, Li Wang, Weili Lin, Pew-Thian Yap, *Senior Member, IEEE*, Dinggang Shen, *Fellow, IEEE*, and Qian Wang, *Member, IEEE*

***Abstract*— Deformable registration is fundamental to longitudinal and population-based image analyses. However, it is challenging to precisely align longitudinal infant brain MR images of the same subject, as well as cross-sectional infant brain MR images of different subjects, due to fast brain development during infancy. In this paper, we propose a recurrently usable deep neural network for the registration of infant brain MR images. There are three main highlights of our proposed method. (i) We use brain tissue segmentation maps for registration, instead of intensity images, to tackle the issue of rapid contrast changes of brain tissues during the first year of life. (ii) A single registration network is trained in a one-shot manner, and then recurrently applied in inference for multiple times, such that the complex deformation field can be recovered incrementally. (iii) We also propose both the adaptive smoothing layer and the tissue-aware anti-folding constraint into the registration network to ensure the physiological plausibility of estimated deformations without degrading the registration accuracy. Experimental results, in comparison to the state-of-the-art registration methods, indicate that our proposed method achieves the highest registration accuracy while still preserving the smoothness of the deformation field. The implementation of our proposed registration network is available online https://github.com/Barnonewdm/ACTA-Reg-Net.**

***Index Terms*— Deformable registration, infant brain MR image, recurrent network, tissue-aware regularization.**

Manuscript received November 3, 2021; revised December 11, 2021; accepted December 15, 2021. Date of publication December 21, 2021; date of current version May 2, 2022. The work of Gang Li was supported in part by the United States National Institutes of Health (NIH) under Grant MH116225 and Grant MH117943. The work of Li Wang was supported in part by the United States National Institutes of Health (NIH) under Grant MH117943. The work of Pew-Thian Yap was supported in part by the United States National Institutes of Health (NIH) under Grant AG053867. The work of Qian Wang was supported in part by the Science and Technology Commission of Shanghai Municipality (STCSM) under Grant 19QC1400600. *(Corresponding authors: Dinggang Shen; Qian Wang.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Institutional Review Board (IRB) of the School of Medicine, University of North Carolina, Chapel Hill, NC, USA.

Dongming Wei is with the School of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai 200030, China, and also with the Department of Radiology and Biomedical Research Imaging Center (BRIC), University of North Carolina, Chapel Hill, NC 27599 USA (e-mail: dongming.wei@sjtu.edu.cn).

Sahar Ahmad, Lei Ma, Zhengwang Wu, Gang Li, Li Wang, Weili Lin, and Pew-Thian Yap are with the Department of Radiology and Biomedical Research Imaging Center (BRIC), University of North Carolina, Chapel Hill, NC 27599 USA (e-mail: sahar.ahmad967@gmail.com; malei.sky@gmail.com; wuzhengwang1984@gmail.com; gang_li@med.unc.edu; li_wang@med.unc.edu; weili_lin@med.unc.edu; ptyap@med.unc.edu).

Yuyu Guo and Liyun Chen are with the School of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai 200030, China (e-mail: yu.guo@sjtu.edu.cn; chen_li_yun@sjtu.edu.cn).

Yunzhi Huang is with the College of Biomedical Engineering, Sichuan University, Chengdu 610065, China, and also with the Department of Radiology and Biomedical Research Imaging Center (BRIC), University of North Carolina, Chapel Hill, NC 27599 USA (e-mail: huang_yunzhi@scu.edu.cn).

Dinggang Shen is with the School of Biomedical Engineering, ShanghaiTech University, Shanghai 201210, China, and also with Shanghai United Imaging Intelligence Company Ltd., Shanghai 201807, China (e-mail: dinggang.shen@gmail.com).

Qian Wang is with the School of Biomedical Engineering, ShanghaiTech University, Shanghai 201210, China (e-mail: wangqian2@shanghaitech.edu.cn).

This article has supplementary downloadable material available at https://doi.org/10.1109/TMI.2021.3137280, provided by the authors.

Digital Object Identifier 10.1109/TMI.2021.3137280

I. Introduction

INFANT brain image registration [1], [2] is crucial for longitudinal and population-based analysis. It can help study early brain development during infancy, often with dynamic volumetric and morphometric changes. For instance, the overall brain volume doubles to about 65% of the adult brain volume in the first year of life [3]. During this time span, the gray matter (GM) grows by 108% — 149%, while the white matter (WM) grows by around 11%. Cortical morphology also alters rapidly, with significant increases in cortical thickness and surface area. Thus, there is a long-standing interest in encoding the longitudinal trajectory of early brain development and characterizing brain maturation at population level. However, while deformable image registration serves as an important tool to quantify these changes, the rapid brain development during infancy makes the task challenging.

The traditional image registration methods (*e.g.*, SyN [4], diffeomorphic Demons [5], NiftyReg [6], etc.) usually maximize the similarity between the moving image and the fixed image, while constraining the deformation field to be physically smooth. Most existing brain image registration methods [4]–[6] are developed for general applications, which are not fully capable of handling the infant brain images. A major reason points to the lack of special consideration of dynamic appearance and structural changes of infant brain MR images. Particularly during the isointense phase (*i.e.*, $\sim$ 6 – 8 months in age) [7], the GM and WM tissues exhibit similar intensities, resulting in poor intensity contrast

and thus difficulty in aligning tissue boundaries during the registration process.

According to our knowledge, there are only a few registration methods that have been developed specifically for infant brain images. For example, Wu *et al.* [8] used the longitudinal trajectory and key points to complete the registration task. Hu *et al.* [9] applied random forest to predict the initial deformation field and also performed appearance transformation to facilitate the conventional registration methods (*e.g.*, SyN, diffeomorphic Demons, HAMMER [10]). However, both these methods involve many steps and hyper-parameters, and thus hard to be applied to infant MR images due to variable intensity contrasts.

More recently, deep learning has drawn much attention in image registration, as the well-trained network can efficiently generate the deformation field during the inference stage and achieve comparable accuracy with respect to the traditional optimization-based methods [11]. However, like the traditional methods, the deep-learning-based methods also come with a trade-off between registration accuracy and deformation field smoothness. For example, Balakrishnan *et al.* [12] proposed to train the network with the loss function comprising of both the image similarity and the L2-norm of the deformation field. The weight of each term in the loss function should be tuned in advance. The smoothness of the deformation field is typically guaranteed by sacrificing the similarity (*i.e.*, tissue overlap) between the moving and the fixed images. Dalca *et al.* [13] mitigated this issue by using the scaling and squaring integration to output the velocity field. The number of the steps of scaling and squaring integration varies in the inference stage, even though it is fixed in the training stage. However, the image similarity is also sacrificed with more scaling and squaring steps. Hu *et al.* [14] used a discriminator network to measure deformation field smoothness, instead of directly constraining the smoothness of deformation field in a non-adjustable and global manner. However, this method involves training an extra network with manual supervision, thus limiting its applicability in practice.

To further improve the registration accuracy, several cascaded deep learning methods are proposed. Following the coarse-to-fine strategy, Vos *et al.* [15] cascaded several networks to uncover the multi-resolution B-splines. Shen *et al.* [16] applied three cascaded networks after affine transformation to improve the inter- and intra-subject registration accuracy for osteoarthritis MR study. Zhao *et al.* [17] successfully improved registration performance over several datasets by cascading 10 networks. The above methods are superior in registration accuracy. However, they consume large CPU/GPU memory and training time, which severely degrades their usage in clinical applications.

For tackling the above issues, in this paper we propose a dedicated recurrent tissue-aware network (RTA-Net) for infant brain MR image registration. We specifically train a single deformable registration network (TA-Net) with tissue-aware smoothness regularization. Although the TA-Net is trained in one-shot manner, we recurrently use it to incrementally update the estimated deformation fields in the inference stage. Although each incremental deformation field has small magnitude, the final deformation field can be obtained by composing all the intermediate deformation fields. In contrast to those cascading multiple registration networks in the literature, our RTA-Net is convenient to use since only a single network needs to be trained for recurrent inference.

To summarize, there are three major highlights of our work:

1) Our RTA-Net uses tissue segmentation, instead of intensities, for registration, to better handle the drastic appearance changes and poor intensity contrast in infant brain images.
2) The single registration network can recurrently infer the (incremental) deformation fields. Thus, it alleviates the training burden, in contrast to the existing methods that need to train multiple registration networks.
3) We propose an adaptive smoothing layer and a tissue-aware anti-folding constraint in the registration network, thus ensuring more smoothness of the generated deformation field than the state-of-the-art deep-learning-based registration methods, without sacrificing registration accuracy.

II. Related Work

A. Conventional Non-Deep Learning Based Registration Methods

Commonly used methods for brain magnetic resonance (MR) image registration include SyN [4], diffeomorphic Demons [5], NiftyReg [6], *etc*. Although comprehensive comparisons of these methods are reported in [18], [19], it is still difficult to assert the best algorithm for a general application especially when dealing with infant brain datasets. In addition, the iterative optimization method is typically used in the conventional registration methods, which consumes much time. By contrast, deep learning based methods are widely used to speed up the registration, resulting in comparable performance at the same time.

B. Deep One-Shot Registration Network

Deep neural network (DNN) for image registration is often trained to map an input pair of fixed/moving images to a deformation field. Many methods train a single network to complete registration in the one-shot manner. Generally, the one-shot network can be trained in three different ways:

1) Supervised Training — The similarity loss is defined over the deformation field, which requires ground-truth deformation field in the training stage. Several works either use the conventional methods to generate the deformation fields [20] or generate the fixed/moving image pairs using the pre-defined deformation fields [21].
2) Unsupervised Training — The loss function consisting of image similarity and regularization of the deformation field [12], [13] is incorporated to predict the deformation or the velocity. Furthermore, the discriminator network is invoked to measure the image similarity [22] and the smoothness of the predicted deformation field [14].
3) Weakly Supervised Training — The manual annotation is considered in the loss function [13], [23] to improve the alignment precision over the target regions. The accurate

annotation is required only in training, but not needed in inference.

Although many deep registration networks claim diffeomorphism theoretically, it is non-trivial to well preserve anatomical topology while deforming real MR images. As discussed in [13], the ratio of deformation folding for the adult brain MR images can decrease with more discretized steps in integrating velocity fields for the final deformation. The optimal number of the integration steps may vary for different datasets though. In SyN [4], when the integration steps go beyond four, many negative Jacobian determinants (associating with folding) are reported. To this end, we will use adaptive smoothing to regularize the deformation field based on the uncertainty of its magnitude. Moreover, we will also propose the tissue-aware Jacobian determinant regularizer to suppress folding-induced artifacts.

C. Cascaded Registration Networks

The cascaded deep registration networks improve registration accuracy, and several methods have been proposed based on the cascading approach. For instance, Vos *et al.* [15], Shen *et al.* [16], and Zhao *et al.* [17] cascaded different registration networks and trained them either separately (while keeping the weights of preceding networks fixed) or simultaneously to gradually improve the registration accuracy. A major limitation of these methods, however, is the extremely high CPU/GPU memory consumption, as each cascaded network has its own parameters to train. On the contrary, in this work, we propose RTA-Net by training only a single network. In the inference stage, the network is invoked recurrently. The scheme of one-shot training and recurrent inference can significantly reduce the amount of network parameters, making our network easier to train and deploy.

D. Infant Brain MR Image Segmentation

The segmentation of infant brain MR image into GM, WM, and cerebrospinal fluid (CSF) has progressed significantly in recent years, which enables us to establish the registration based on tissue segmentation maps. In the isointense phase, the segmentation is quite challenging. Early methods are mostly based on multi-atlas fusion [24]–[26]. Currently, efforts are dedicated to develop learning-based segmentation methods. For instance, Zhang *et al.* [27] proposed to use DNN with multi-modal MR images as input to improve the segmentation results. Wang *et al.* [7] further considered appearance features and estimated tissue probability maps to iteratively refine the tissue segmentation. As reported in the iSeg-2017 [28] and iSeg-2019 [29] challenges, current methods are able to gain around 93.6% Dice similarity coefficient (DSC) over GM, 93.9% over WM, and 95.9% over CSF.

III. Method

The proposed RTA-Net, as illustrated in Fig. 1 (a), aims to register the moving segmentation image I_m to a fixed segmentation image I_f by progressively refining the deformation field (c.f., Section III-A). The RTA-Net consists of a single tissue-aware TA-Net and a spatial transformer. The TA-Net (c.f., Section III-B) outputs a smooth (incremental) deformation field, given the input pair of the fixed image I_f and the moving image I_m. The spatial transformer is then used to resample the moving image (*i.e.*, the tissue segmentation map) using the tentatively estimated deformation field. After n-loop callbacks to TA-Net and spatial transformer, the RTA-Net outputs the final deformation field, as well as the deformed moving image.

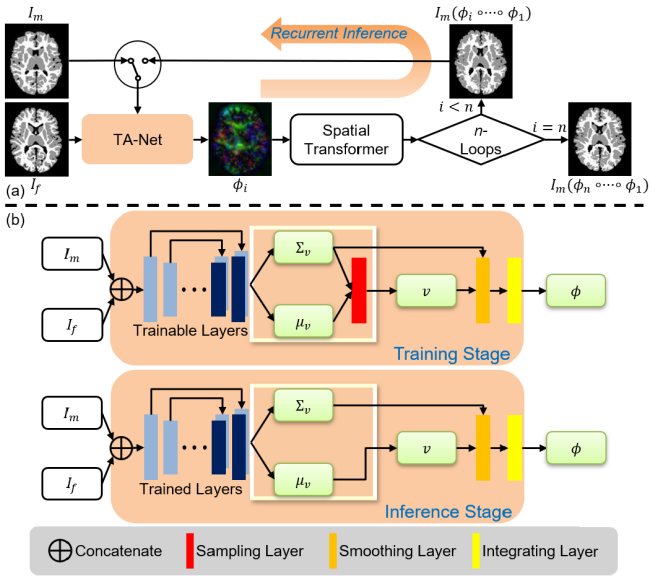


Fig. 1. (a) Overview of RTA-Net. The proposed RTA-Net consists of basic TA-Net (in orange box) and spatial transformer. With n-loops for recurrent inference, the final deformation field is composed from all incremental deformations ($\phi_i, i = 1, \cdots, n$). (b) Detailed architectures of TA-Net in the training and inference stages, respectively. The moving and fixed images $\{I_m, I_f\}$ are input into training layers to derive μ_v, Σ_v. Velocity v is sampled from $\mathcal{N}(\mu_v, \Sigma_v)$ in the training stage, and equals to μ_v directly in the inference stage. Deformation field ϕ can be obtained by the integration of v.

A. One-Shot Training and Recurrent Inference

In the training stage, only a single TA-Net is modeled, which makes our method very different from current registration techniques [15]–[17]. We particularly propose to use tissue-adaptive smoothing and tissue-aware Jacobian determinant regularization to help TA-Net output smooth and accurate deformation field. More details about TA-Net are shown in Section III-B.

In the inference stage, the TA-Net and spatial transformer are applied recurrently. At every iteration, a tentative incremental deformation field is output from TA-Net. Then, the moving image is updated by the spatial transformer according to the latest deformation field. The recurrent inference in general consists of the following three steps:

(i) The moving and fixed images $\{I_m, I_f\}$ are input to the pre-trained TA-Net to obtain the deformation field ϕ_1 and the deformed moving image $I_m(\phi_1)$;

(ii) The new pair $\{I_m(\phi_1), I_f\}$ is fed into the same TA-Net to get the new incremental deformation field ϕ_2. The moving image is then warped as $I_m(\phi_2 \circ \phi_1)$;

(iii) The TA-Net and spatial transformer are invoked for n times to obtain the final warped moving image $I_m(\phi)$ with the composed deformation field $\phi = \phi_n \circ \ldots \circ \phi_2 \circ \phi_1$.

Note that, to avoid error accumulation of repeated resampling, we compose the incremental deformation fields before warping the moving image at each iteration.

B. TA-Net for Tissue-Aware Deformable Registration

The TA-Net takes as the input a pair of moving and fixed images, and outputs the deformation field. We show the architecture of TA-Net in Fig. 1 (b), consisting of trainable layers, sampling layer, smoothing layer, and integrating layer. A single TA-Net is trained for all inference loops.

The trainable layers are similar to those in 3D U-Net, which particularly output the velocity field v to further generate deformation field ϕ following $\phi = Id + \int_0^1 v\, dt$. Here Id indicates the identity transformation, and $\int_0^1 v\, dt$ derives the displacement vector field from the velocity.

As the velocity field v is first output by the trainable layers, it is numerically integrated in finite discretized steps (*i.e.*, seven steps implemented in the integrating layer) to obtain the deformation field ϕ. The integration here is done by the scaling and squaring operations [13], [30] in the integrating layer as illustrated by Fig. 1 (b). We have found that, while the magnitude of v is typically smaller than ϕ numerically, the trainable layers can predict v more accurately than predicting ϕ directly.

The loss function for the entire TA-Net consists of image similarity between the moving image warped by ϕ and the fixed image, and the regularization terms for both velocity field and deformation field:

$$\begin{aligned} \mathcal{L} &= -\mathrm{Sim}(I_m(\phi), I_f) + \mathrm{Reg}(v) + \mathrm{Reg}(\phi), \quad \phi \\ &= Id + \int_0^1 v\, dt, \end{aligned} \tag{1}$$

$\mathrm{Sim}(\cdot,\cdot)$ is implemented here as the localized normalized cross-correlation between the two images. $\mathrm{Reg}(v)$ and $\mathrm{Reg}(\phi)$ will be elaborated in the following.

1) Variational Regularization of the Velocity Field: We model the velocity field as a multivariate Gaussian field $\mathcal{N}(\mu_v, \Sigma_v)$ [13], where each voxel has the mean velocity estimate and the variance. In the training stage, the trainable layers estimate μ_v and Σ_v for every voxel, and then instantiate the velocity field v by drawing from $\mathcal{N}(\mu_v, \Sigma_v)$ via the sampling layer. The instance of the velocity field is further smoothed and then integrated to derive ϕ, which allows the computation of the loss function in Eq. 1. In the inference stage, however, we can assign the inferred value of μ_v as the instance of v by neglecting Σ_v. Note that the uncertainty of the velocity field encoded in Σ_v will assist adaptive smoothing as described in Section III-B.2.

For obtaining smooth velocity field v, we define $\mathrm{Reg}(v)$ as

$$\mathrm{Reg}(v) = ||\nabla \mu_v||^2 + ||\Sigma_v||^2. \tag{2}$$

However, such conventional regularization is often insufficient in avoiding folding artifacts in infant MR image registration. Therefore, we further propose the uncertainty based adaptive smoothing and the tissue-aware Jacobian determinant regularization.

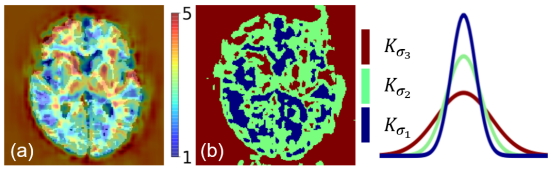


Fig. 2. (a) Uncertainty of the displacement magnitude is overlaid with the moving image. The value in color bar represents the uncertainty value. (b) The voxels are classified into three groups (blue, green, red). Each voxel will be smoothed by its group's smoothing kernel K_{σ_i}, $i = 1,2,3$.

2) Uncertainty Based Adaptive Smoothing: The deformation should exhibit local smoothness, such that the displacement vectors in a small neighborhood are spatially consistent. Conventional registration methods often use a Gaussian filter with fixed kernel size for this purpose. The filter is defined as $K_\sigma = e^{-\frac{1}{2}||\frac{p-p'}{\sigma}||^2}$, where p and p' are voxel positions, and p' is the neighbour of p. The Gaussian kernel then applies to the displacement field. While bigger σ will better suppress folding artifacts, it may result in over-smoothing of the displacement field and thus reduce registration accuracy. Adaptive smoothing [31] has been invested to avoid homogeneous and isotropic regularization to avoid over smoothing.

To tackle this issue, we implement adaptive Gaussian kernels in the smoothing layer of our network. We first calculate the uncertainty, $\Sigma_{|d|}$, of the displacement magnitude $|d|$. Then, different Gaussian filters are used according to $\Sigma_{|d|}$. Principally, the region with higher $\Sigma_{|d|}$ is smoothed by the filter with a bigger σ. On the contrary, if $\Sigma_{|d|}$ is small, then a small σ is applied to help preserve the accurately estimated velocity.

a) Computation of $\boldsymbol{\Sigma_{|d|}}$: The uncertainty of displacement magnitude $\Sigma_{|d|}$ per voxel can be computed by:

$$\Sigma_{|d|} = \int_0^1 3\sqrt{3} \cdot \Sigma_v^{\max}\, dt, \tag{3}$$

where $\Sigma_v^{\max} = \max(\Sigma_{v,x}, \Sigma_{v,y}, \Sigma_{v,z})$ indicates the maximum velocity uncertainty component, and the integration is performed using scaling and squaring operations in the integrating layer. The detailed derivation of Eq. 3 is provided in *Supplementary Materials*.

b) Adaptive smoothing based on $\boldsymbol{\Sigma_{|d|}}$: Voxels are then classified into three equally-sized groups based on $\Sigma_{|d|}$. For each group, a corresponding kernel is used. Particularly, we define the three following kernels from small to large: $\sigma_1 = \frac{\max(\Sigma_{|d|}) + 2\cdot\min(\Sigma_{|d|})}{3}$, $\sigma_2 = \frac{2\cdot\max(\Sigma_{|d|}) + \min(\Sigma_{|d|})}{3}$, $\sigma_3 = \max(\Sigma_{|d|})$. We show an example of the uncertainty map overlaid on the moving image in Fig. 2. If a voxel p belongs to the blue group in the figure, $v(p)$ is smoothed by K_{σ_1} with small kernel size.

3) Tissue-Aware Jacobian Determinant Regularizer: In addition to the regularization and smoothing operation over the velocity field v, we further propose the Jacobian determinant regularizer to obtain physiologically plausible deformation field $J(\phi)$. The proposed tissue-aware regularizer adaptively

constrains the three tissue types of GM, WM, and CSF, which is defined as:

$$\mathrm{Reg}(\phi) = \begin{cases} \exp(|\min(J(\phi)) - 1|) - 1; & \text{if GM or WM} \\ \exp(|\mathrm{mean}(J(\phi)) - 1|) - 1; & \text{otherwise.} \end{cases} \tag{4}$$

For GM and WM, the minimum of the Jacobian determinant should be positive to avoid folding. For the background and CSF, the Jacobian determinant should be close to one, as the displacement over these regions should be around zero. Note that we use the exponential operation to penalize the regions with very large or negative Jacobian determinants.

C. Implementation

We have implemented the TA-Net by Keras and Tensorflow, and trained on a single 12GB NVIDIA Titan X GPU. Adam optimizer is used with a learning rate of 1e-4. Only a pair of segmentation maps is input in a batch. The TA-Net is trained by 1.5e3 epochs (100 iterations in each epoch).

IV. Experiments and Results

A. Dataset

With increasing interest in studying the early brain development, many longitudinal infant brain MR images have been collected in the past years. Our dataset consists of longitudinal T1w and T2w images (acquired at 2 weeks, 3, 6 and 12 months after birth) of 47 healthy infant subjects, who are enrolled as part of the Multi-visit Advanced Pediatric Brain Imaging (MAP) study.

The imaging parameters for T1w MR images were: TR = 1900 ms, TE = 4.38 ms, flip angle = 7°, 144 sagittal slices, and 1 mm isotropic voxel resolution. The imaging parameters for T2w MR images were TR = 7380 ms, TE = 119 ms, flip angle = 150°, 64 sagittal slices, and $1.25 \times 1.25 \times 1.95\,\mathrm{mm}^3$ voxels resolution.

The number of 3D images for each subject varies in acquisition due to missed imaging session. Particularly, the training dataset comprises 56 scans of 27 subjects, and 57 scans of the remaining 20 subjects are used for inference and validation. For efficiently using the available data, we have simulated deformation fields to warp the available tissue segmentation maps for data augmentation. The data augmentation discussion can be checked from *Supplementary Materials*.

B. Pre-Processing

The dataset is pre-processed by the UNC infant pre-processing pipeline [32] for the subsequent image registration. The steps involve: (i) N3 intensity inhomogeneity correction [33]; (ii) learning-based skull stripping [34]; (iii) removing cerebellum and brain stem via registration [10]; (iv) rigid alignment of all the longitudinal images of the same subject, and rigid alignment of T2w image with its corresponding T1w image; (v) tissue segmentation (CSF, GM and WM) using a learning-based multi-source integration framework [7]; (vi) rigid alignment of all the tissue segmentation maps with the fixed tissue segmentation map using FLIRT [35]; (vii) resampling of all the tissue segmentation maps and intensity images were then resampled to have a size of $256 \times 256 \times 256$ with $1 \times 1 \times 1\,\mathrm{mm}^3$ voxels resolution; (viii) masking and filling of sub-cortical structures, and left/right hemisphere separation of brain [36]; (ix) construction of topologically correct and geometrically accurate white matter (WM/GM interface) and pial (GM/CSF interface) cortical surfaces. (x) delineation of 34 cortical ROIs (see *Supplementary Materials*) by first parcellating the constructed cortical surfaces, and then performing the surface-to-volume mapping of the labels. The mapping was carried out by assigning the label from the closest vertex to each voxel.

C. Evaluation Metrics

We employ the following three measures to assess the performance of our registration method, namely Dice similarity coefficient (DSC), max average Hausdorff distance (MAHD) [37], and ratio of the folding points (RFP). In general, higher DSC, smaller MAHD, and smaller RFP represent better performance of the registration method. Detailed definitions of the measures can be found in *Supplementary Materials*.

D. Comparative Analysis

We have evaluated our proposed method to perform inter-subject, intra-subject, and spatiotemporal registration. The proposed method is compared with state-of-the-art registration methods including ANTs [4], diffeomorphic Demons [5], NiftyReg [38] and VoxelMorph [13]. In terms of the cascaded networks, we are unfortunately unable to reproduce their methods over the infant dataset in this paper. For Vos *et al.* [15], there is no public code. For Shen *et al.* [16] and Zhao *et al.* [17], we are unable to train all the networks due to large GPU memory requirements. Note that, before performing deformable registration, all the images are aligned using FLIRT for affine registration. We trained VoxelMorph using its default parameters. The parameter details of the traditional optimization-based registration methods are given in *Supplementary Materials*. All the methods are used to register tissue segmentation maps instead of intensity images for fair comparison.

1) Inter-Subject Registration: We first evaluate inter-subject registration. To train the network for inter-subject registration, the moving and fixed time-points are randomly selected from different subjects in the training dataset. In this way, we have a total of 1,161 inter-subject pairs for training. For inference, the 12-month-old scan of a randomly selected inference subject is chosen as the fixed image. All other subjects at 2-week-old, 3-month-old, and 6-month-old in the inference dataset are registered to the selected fixed scan, resulting in a total of 54 inter-subject pairs for evaluation.

We report DSC over GM and WM, MAHD for inner/outer cortical surfaces to evaluate the registration accuracy, and RFP of the deformation fields to evaluate the smoothness of deformation field. As shown in Table I, our RTA-Net obtains significant improvement ($p < 0.05$ in paired t-tests) over all the

TABLE I

INTER-SUBJECT REGISTRATION RESULTS OF DIFFERENT METHODS, INCLUDING MEAN±STD DSC (%) OVER GM AND WM, MEAN±STD MAHD (MM) FOR INNER AND OUTER CORTICAL SURFACES, AND MEAN RFP (%) FOR THE DEFORMATION FIELD

		2-week-old to 12-month-old		3-month-old to 12-month-old		6-month-old to 12-month-old		RFP (%)
		GM	WM	GM	WM	GM	WM	
DSC (%)	FLIRT	59.94±1.64	64.28±0.91	61.15±1.32	53.61±1.16	61.85±1.47	56.78±1.40	0
	ANTs	79.26±0.84	75.45±1.01	79.75±0.67	75.68±1.14	80.57±0.61	77.73±0.73	0.0453
	Demons	73.60±0.87	68.99±0.73	74.03±0.68	69.01±0.84	74.70±0.67	71.52±0.71	0.0006
	NiftyReg	72.65±1.56	68.29±1.60	73.61±1.23	68.70±0.91	74.56±0.70	72.94±3.98	**<1e-4**
	VoxelMorph	77.23±6.60	72.74±6.70	79.89±4.75	75.78±4.65	80.01±5.53	76.34±6.22	0.6915
	RTA-Net	**85.01±0.47**	**82.78±0.55**	**85.16±0.36**	**82.95±0.21**	**85.17±0.34**	**83.29±0.17**	0.0027
		Inner	Outer	Inner	Outer	Inner	Outer	
MAHD (mm)	FLIRT	1.49±0.04	3.65±0.17	1.47±0.09	3.19±0.17	1.42±0.08	2.56±0.17	
	ANTs	0.82±0.03	3.30±0.18	0.82±0.05	2.88±0.18	0.76±0.04	2.21±0.13	
	Demons	0.91±0.03	3.31±0.20	0.89±0.02	2.94±0.20	0.86±0.03	2.33±0.13	
	NiftyReg	1.06±0.03	3.62±0.22	1.01±0.03	3.14±0.20	0.97±0.05	2.38±0.15	
	VoxelMorph	0.83±0.01	2.97±0.11	0.82±0.02	2.76±0.15	0.79±0.01	2.28±0.12	
	RTA-Net	**0.65±0.03**	**2.58±0.10**	**0.63±0.01**	**2.41±0.13**	**0.60±0.01**	**1.98±0.11**	

comparing methods for DSC and MAHD at all three settings of time-points (*i.e.*, registering from 2-week-old to 12-month-old fixed image, from 3-month-old to 12-month-old, and from 6-month-old to 12-month-old). The results clearly indicate higher registration accuracy of our method over other methods under comparison.

In terms of RFP, RTA-Net also delivers more smooth deformation fields than ANTs and VoxelMorph. Although Demons and NiftyReg obtain smaller RFP, they actually sacrifice DSC and MAHD —- we attribute this phenomenon as that the two registration methods are inclined to under-estimate or over-smooth the deformation fields. The registration results and the deformation fields can be visually inspected in Fig. 3, confirming that RTA-Net obtains accurate alignment with smooth deformation fields.

In addition to the segmented tissues, we also evaluate DSC over 34 ROIs using all the inference moving/fixed pairs (as shown in Fig. 4). The overall performance over small ROIs is improved by the proposed RTA-Net, *i.e.*, 54.43 ± 13.06 by ANTs, 54.13±11.69 by diffeomorphic Demons, 51.37±11.69 by NiftyReg, 47.36 ± 13.32 by VoxelMorph, and 55.38 ± 12.96 by RTA-Net. The proposed RTA-Net shows significant improvement over 19 ROIs (p <0.05, paired t-tests) compared with each of the four alternative methods, as indicated by "*" in Fig. 4.

2) Intra-Subject Registration: Intra-subject registration is key to measure temporal changes of infant cortical structures. To further evaluate RTA-Net for intra-subject registration, we have validated on all intra-subject pairs over the inference dataset. The network for intra-subject registration is refined from the previous inter-subject registration network. Specifically, in addition to the aforementioned 1,161 inter-subject training pairs, we have used 64 intra-subject pairs to refine the network, as the moving and fixed images in an intra-subject pair should come from the same subject. Then, for each inference subject, we choose the 12-month-old time-point as the fixed, and register all other time-points of the subject to the fixed one-by-one. Thus, there are 57 intra-subject pairs for inference evaluation in total.

The quantitative results for the intra-subject registration are presented in Table II. RTA-Net has achieved significantly higher DSC (p <0.05, paired t-tests) than VoxelMorph for both GM and WM. Regarding the non-deep-learning methods, RTA-Net obtains comparable or improved DSC, except for WM in registering from 3-month-old to 12-month-old and from 6-month-old to 12-month-old. The MAHD results are generally consistent with the DSC results. Particularly, RTA-Net has significantly improved over the deep-learning-based VoxelMorph (p <0.05, paired t-test). It has also shown better alignment for outer surfaces compared to the non-deep-learning methods (*i.e.*, ANTs and NiftyReg, p <0.05, paired t-tests). However, for the inner cortical surface, RTA-Net yields larger MAHD, compared to the conventional methods.

Meanwhile, RTA-Net obtains very few folding points (RFP<1e-4%), which is comparable with the conventional methods. Compared to VoxelMorph with quite implausible deformation field (RFP>0.2%), RTA-Net is reliable in intra-subject registration with remarkable RFP.

Based on the DSC, MAHD and RFP results, it can be concluded that RTA-Net performs better in registering both GM and outer cortical surface than the comparing methods, and can also obtain comparable results over WM and inner surface with the conventional methods. Note that there is always a significant improvement of RTA-Net than the state-of-the-art deep-learning-based VoxelMorph.

3) Spatiotemporal Registration: We further combine the aforementioned inter- and intra-subject registration, and apply to the setting of spatiotemporal registration of all subjects in the inference set. For each inference subject, the spatiotemporal registration consists of two steps: (1) all time-points of a subject are first registered with the 12-month-old time-point of the subject, by using the intra-subject RTA-Net; (2) all subjects estimate the deformation fields toward the fixed subject, by using their 12-month-old time-points and the inter-subject registration network. Here the fixed subject is the same as described in Section IV-D.1, and the two networks are the same as described in Section IV-D.1 and Section IV-D.2, respectively. Note, for avoiding the interpolation error, deformation field composition is conducted, such that each time-point of a subject will be warped for a single time to reach the fixed image space. The above spatiotemporal registration scheme is widely used in the literature [39] for handling population of sequences.

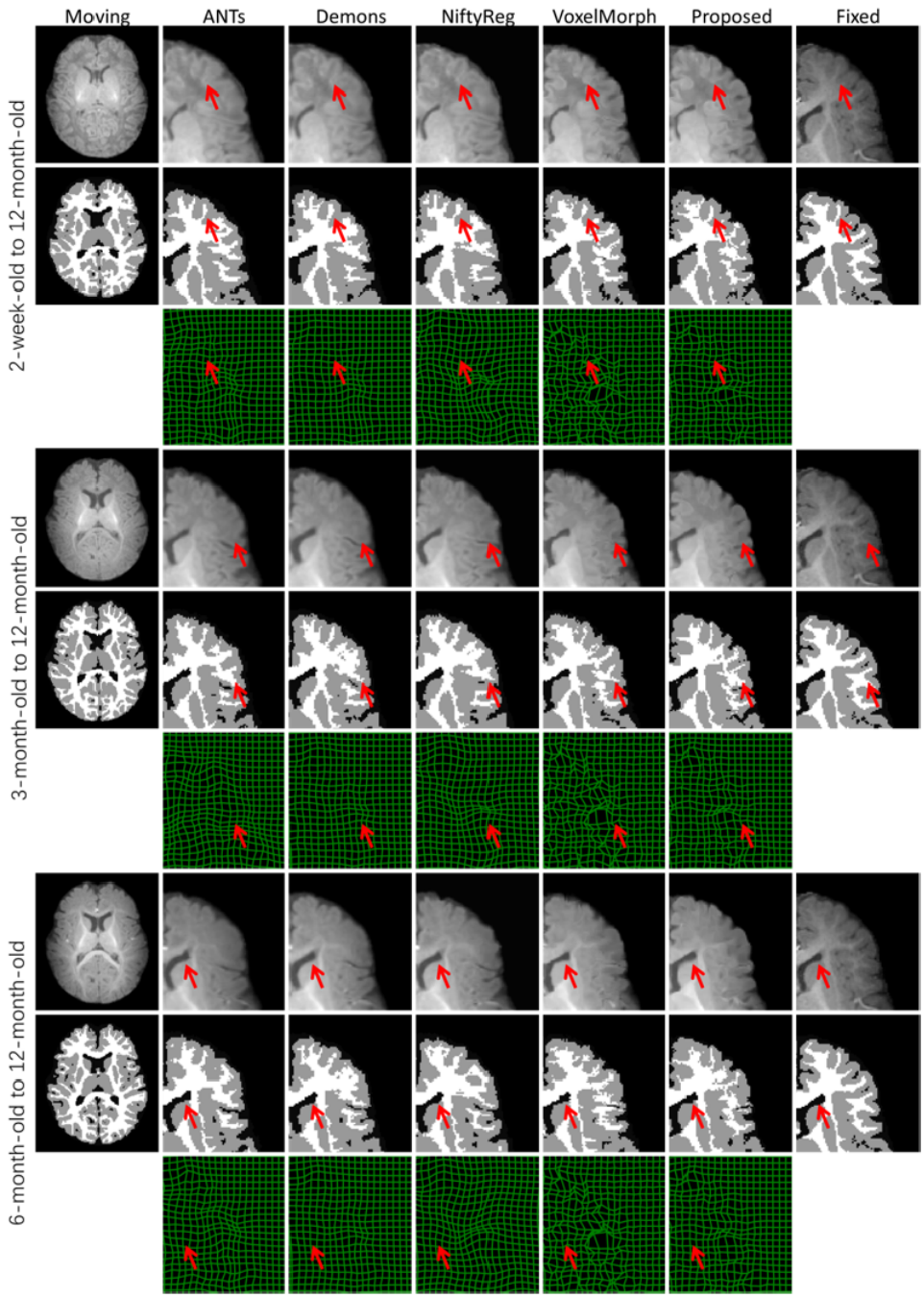


Fig. 3. Inter-subject registration results obtained with the comparing methods. For each time-point setting, the three rows illustrate the T1w images, the tissue segmentation maps, and the deformation fields. Examples of improved regions are indicated by red arrows.

We also compare with ANTs, Demons, NiftyReg, and VoxelMorph. Each of the comparing registration methods is applied in both of intra- and inter-subject registration steps following the spatiotemporal registration scheme above and using the same configurations as in Section IV-D.1 and Section IV-D.2, respectively.

The DSC comparison over GM and WM of different methods is reported in Table III. It can be seen that RTA-Net still obtains better DSC than all other methods. In terms of RFP, RTA-Net performs better than ANTs, NiftyReg and VoxelMorph. Although Demons obtains better RFP than RTA-Net, it actually sacrifices the DSC results by a large margin (4.7%-6.7% decrease in DSC compared to RTA-Net), which is similar to the case in the previous inter-subject validation. The results indicate that our proposed RTA-Net obtains superior registration accuracy without sacrificing the smoothness of deformation field in the spatiotemporal registration task.

When comparing the results in Table III with the inter-subject registration results in Table I, it can also be found that, for Demons and NiftyReg, the spatiotemporal registration can steadily improve the DSC results than direct inter-subject registration (*i.e.* each time-point of a subject is independently

TABLE II

INTRA-SUBJECT REGISTRATION RESULTS OF DIFFERENT METHODS, INCLUDING MEAN±STD DSC (%) OVER GM AND WM, MEAN±STD MAHD (MM) FOR INNER AND OUTER CORTICAL SURFACES, AND MEAN RFP (%) FOR THE DEFORMATION FIELD

		2-week-old to 12-month-old		3-month-old to 12-month-old		6-month-old to 12-month-old		RFP (%)
		GM	WM	GM	WM	GM	WM	
DSC (%)	FLIRT	62.17±1.70	57.19±0.59	65.78±2.85	59.43±1.25	68.94±0.41	66.60±0.41	0
	ANTs	87.09±0.09	85.25±1.69	88.37±0.35	**87.83±0.89**	89.55±0.86	88.61±0.46	**<1e-4**
	Demons	85.43±0.34	84.09±1.39	86.34±0.36	85.63±1.01	88.44±0.63	88.07±0.15	**<1e-4**
	NiftyReg	85.73±0.99	85.15±0.08	87.68±1.05	87.10±0.19	88.98±0.81	**88.87±0.44**	**<1e-4**
	VoxelMorph	84.81±1.14	80.68±0.27	85.88±1.72	82.40±0.80	87.18±0.79	84.62±0.37	0.2081
	RTA-Net	**88.13±0.23**	**85.65±1.35**	**89.18±1.04**	86.77±0.40	**89.93±0.31**	87.95±0.16	**<1e-4**
		Inner	Outer	Inner	Outer	Inner	Outer	
MAHD (mm)	FLIRT	1.42±0.01	2.26±0.01	1.29±0.07	1.88±0.01	1.09±0.09	1.42±0.16	
	ANTs	**0.49±0.05**	1.22±0.11	0.45±0.04	0.92±0.10	0.39±0.01	0.58±0.05	
	Demons	0.53±0.05	1.27±0.09	0.48±0.04	0.98±0.08	0.38±0.01	0.64±0.05	
	NiftyReg	0.52±0.01	1.31±0.09	**0.45±0.01**	0.98±0.08	**0.38±0.01**	0.64±0.05	
	VoxelMorph	0.83±0.03	1.03±0.01	0.57±0.01	0.88±0.01	0.51±0.02	0.59±0.06	
	RTA-Net	0.51±0.03	**0.80±0.01**	0.46±0.01	**0.67±0.03**	0.47±0.03	**0.51±0.06**	

TABLE III

SPATIOTEMPORAL REGISTRATION RESULTS OF DIFFERENT METHODS, INCLUDING MEAN±STD DSC (%) OVER GM AND WM, MEAN±STD MAHD (MM) FOR INNER AND OUTER CORTICAL SURFACES, AND MEAN RFP (%) FOR THE DEFORMATION FIELD

	2-week-old to 12-month-old		3-month-old to 12-month-old		6-month-old to 12-month-old		RFP (%)
	DSC (%)		DSC (%)		DSC (%)		
	GM	WM	GM	WM	GM	WM	
ANTs	79.95±0.36	75.85±1.22	80.40±0.33	76.49±1.16	80.13±0.12	76.65±0.72	0.0390↑
Demons	77.64±0.03↑	73.09±0.97↑	77.99±0.15↑	73.46±0.80↑	77.73±0.30↑	74.06±0.47↑	**0.0017**↓
NiftyReg	79.12±0.09↑	76.99±0.91↑	79.42±0.26↑	76.93±1.30↑	79.75±0.50↑	77.89±0.42↑	0.0232↓
VoxelMorph	81.11±0.01↑	77.13±0.44↑	81.39±0.09↑	77.71±0.32↑	80.84±0.28	77.74±0.15	0.4145↑
RTA-Net	**82.33±0.06**↓	**78.51±1.04**↓	**82.87±0.47**↓	**79.32±0.21**↓	**83.48±0.18**↓	**80.74±0.20**↓	0.0124↓

↑ indicates the result is much improved and ↓ indicates much decreased than inter-subject registration, as shown in Table I.

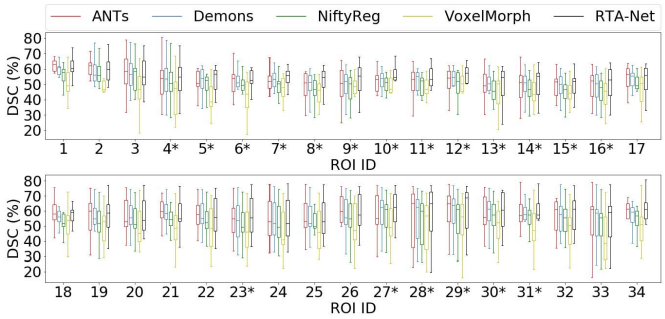


Fig. 4. DSC (%) over 34 cortical ROIs achieved by the comparing methods for inter-subject registration.

registered to the fixed image space). ANTs gets comparable spatiotemporal and inter-subject registration results. For VoxelMorph, the DSC results can be improved by spatiotemporal registration, except for registering 6-month to 12-month, which is slightly improved (<1.5% improvement) than inter-subject registration in Table I. For RTA-Net, the spatiotemporal results decreased compared to the direct inter-subject registration, though our method still performs much better than all other methods in the spatiotemporal setting.

The above comparison suggests that RTA-Net gains more in performance of inter-subject registration than of intra-subject registration. Note that, in the inter-subject registration scheme, each inference time-point is independently registered with the fixed destination. On the contrary, in the spatiotemporal scheme, each time-point will use the 12-month-old time-point in the same subject as the intermediate guidance, and further deforms to the fixed destination. This may be the reason behind the decreased performance by RTA-Net in spatiotemporal registration than in direct inter-subject registration.

Combining the registration accuracy of RTA-Net in Section IV-D, it can be found that RTA-Net obtains much improved performance than the comparing methods, when coping with more dynamic deformation (i.e., inter-subject registration, or intra-registration with large time span from 2-week-old to 12-month-old). However, RTA-Net's advantage shrinks, when coping with subtel deformation (3-month-old or 6-month-old to 12-month-old in intra-subject registration). This inidcates that the inter-subject registration stage in spatitemporal registration helps the performance improvement, while the proposed RTA-Net can better handle difficult registration cases.

Finally, based on the above experiment results, it can be concluded that (1) Demons and NiftyReg may under-estimate the deformation fields by direct inter-subject registration, which can be mitigated by the spatiotemporal registration; (2) ANTs obtains comparable accuracy for inter-subject registration and spatiotemporal registration, while the folding of the deformation fields can be severe; (3) RTA-Net still outperforms all of the comparing methods, even though the spatiotemporal registration yields a relatively worse outcome than the direct inter-subject registration.

E. Ablation Study

We further perform the ablation studies to verify the individual contributions of the four key components, *i.e.*, recurrent

TABLE IV
MEAN±STD DSC (%) AND MEAN RFP (%) OVER GM AND WM RESULTS OF RTA-NET IN THE ABLATION STUDIES

	Recurrent Inference	Reg(v)	Reg(ϕ)	Smoothing layer	2-week-old to 12-month-old DSC (%)		3-month-old to 12-month-old DSC (%)		6-month-old to 12-month-old DSC (%)		RFP (%)
					GM	WM	GM	WM	GM	WM	
Inter-	✗	✓	✓	✓	77.62±0.96	71.69±0.85	78.21±0.57	71.99±0.51	79.12±0.74	74.27±0.87	0.0073
	✓	✗	✓	✓	84.70±0.47	81.95±0.54	84.70±0.26	82.92±0.19	84.97±0.30	82.86±0.30	0.0158
	✓	✓	✗	✓	84.98±0.56	80.80±0.95	83.99±0.34	81.76±0.25	85.07±0.66	81.14±0.37	0.2953
	✓	✓	✓	✗	85.00±0.48	**82.86±0.54**	85.15±0.36	**83.02±0.23**	85.14±0.35	**83.32±0.18**	0.0275
	✓	✓	✓	✓	**85.01±0.47**	82.78±0.55	**85.16±0.36**	82.95±0.21	**85.17±0.34**	83.29±0.17	**0.0027**
Intra-	✗	✓	✓	✓	82.06±0.85	77.03±1.48	84.24±1.49	80.23±0.16	86.68±0.18	84.21±1.73	0.0013
	✓	✗	✓	✓	87.17±0.53	84.15±0.87	88.26±1.07	85.79±0.15	88.77±0.79	86.76±0.14	0.0233
	✓	✓	✗	✓	84.80±1.14	80.68±0.25	85.86±1.72	82.39±0.81	87.18±0.79	84.63±0.38	0.2301
	✓	✓	✓	✗	88.03±0.45	85.45±1.46	89.09±0.34	86.68±0.21	89.91±0.67	87.91±0.12	0.0136
	✓	✓	✓	✓	**88.13±0.23**	**85.65±1.35**	**89.18±1.04**	**86.77±0.40**	**89.93±0.31**	**87.95±0.16**	**<1e-4**

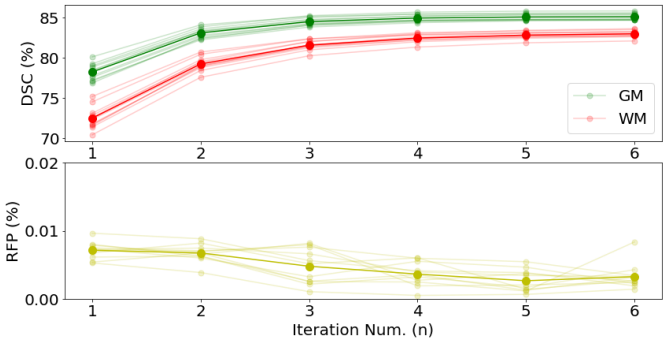


Fig. 5. The mean DSC of GM and WM over the inference dataset (top), and RFP of the deformation field (bottom) with different number of iterations in the recurrent framework of RTA-Net.

inference, Reg(v), Reg(ϕ), and smoothing layer, in RTA-Net. The different combinations of these components used in our ablation studies are presented in Table IV.

We compare DSC over GM and WM, and the RFP of deformation fields for inter- and intra-subject registration. For each combination, we use the same inter- and intra-subject training dataset in Section IV-D. Then, we train the networks using the same configuration, except for the component involved in comparison. It can be seen from Table IV that, without the recurrent inference, DSC is degraded by a large margin in both inter- and intra-subject registration. Without Reg(v), RTA-Net obtains a decreased DSC and worse RFP (*i.e.,* 5 times higher). Without Reg(ϕ), RTA-Net obtains a decreased DSC and much worse RFP, especially for the intra-subject registration. Without the smoothing layer, RTA-Net achieves comparable DSC in both inter- and intra-subject registration, yet with a cost of much worse RFP (*i.e.*, 10 times higher).

In this way, it can be concluded that the recurrent inference boosts the registration alignment accuracy (*i.e.*, DSC). Reg(ϕ) mainly helps generate topology-preserved deformation field. Reg(v) and the smoothing layer further adds smoothness constraint to the deformation field, without sacrificing registration accuracy. The above results validate the effectiveness of the proposed four components in RTA-Net.

V. DISCUSSION

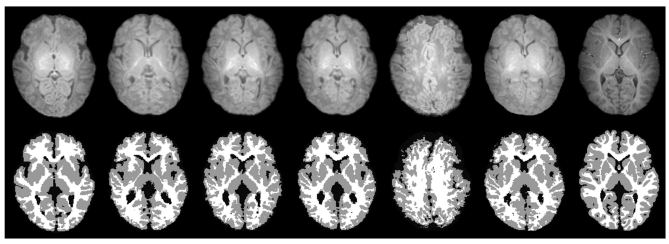
A. Recurrent Inference Analysis

The steps of recurrent inference directly relates to the overall performance of the RTA-Net. We compare DSC over GM and WM, and RFP of the inference dataset, by changing the number of iterations for inter-subject registration inference. The results are illustrated in Fig. 5, where the mean DSC of GM and WM improves sharply from the 1st to the 2nd iteration, and then becomes stable after the 5th iteration. On the other hand, RFP is slightly decreasing alongside all the iterations. This indicates that the recurrent inference *not only* incrementally refines the estimated deformation field for accurate registration, *but also* well preserves the smoothness of the deformation field. Concerning the computation time, we have decided to use five iterations for recurrent inference in our implementation.

In contrast to the experiment in [17], which amplifies the folding points with recurrent inference, our method has achieved robust improvement in both the registration accuracy and well-constrained deformation field smoothness. The reason points to the carefully designed network used for the recurrent inference. Our network involves the tissue-aware Jacobian regularization and adaptive smoothing layer, which could help suppress the points of folded deformation without sacrificing the registration accuracy (*i.e.*, DSC over WM and WM). In addition to the registration performance improvement, our proposed RTA-Net is lightweight, which can be trained and used for inference without the need for huge GPU memory.

B. Registration Over Intensity Images

Someone may have concern about using the intensity images directly for infant brain MR image registration. Here, we have validated the traditional registration methods like Demons and ANTs, and deep-learning-based VoxelMorph to register intensity images directly. For ANTs, we have used two different similarity metrics, *i.e.*, cross-correlation (CC) and mutual information (MI). For VoxelMorph, we have used the intensity images of the same training dataset in Section IV-A to train the VoxelMorph model with the reported network parameters in [13]. We have also reproduced the weakly supervised version of VoxelMorph (VoxelMorph-w), which uses not only the intensity images but also the GM and WM segmentation maps at the training stage. Generally, the performance of the direction registration of the intensity images is quite low (around 60% DSC over GM or WM). We have shown an example from the testing dataset in Fig. 6. Specifically, VoxelMorph fails to output a reasonable warped image. With the help of the segmentation maps in the training, VoxelMorph-w can predict



GM DSC (%)	56.38	60.32	59.92	41.99	61.10
WM DSC (%)	47.76	52.94	57.09	33.76	53.07

Fig. 6. The inter-subject registration results and the DSC values for GM and WM by different registration methods (*i.e.*, Demons, ANTs CC, ATNs MI, VoxelMorph, and weakly supervised VoxelMorph) based on intensity images directly. The DSC results of GM and WM are listed in the bottom table.

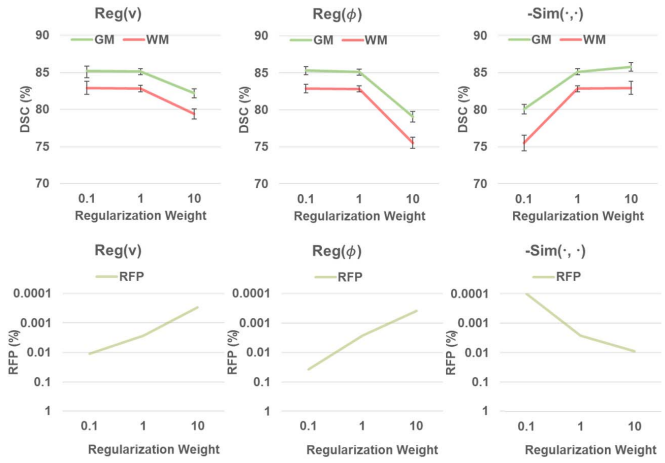


Fig. 7. The results for the inter-subject registration pair in the testing dataset with various weights, including DSC for GM and WM, and RFP of deformation fields.

reasonable output, yet the DSCs over GM and WM are still low.

C. Parameter Analysis

For clarifying the contribution of each term in Eq. 1, we have used different weights to train the network. For each comparison experiment, we set two weights to be equal to 1, and vary the left one between 0.1 and 10. Fig. 7 shows the DSC and RFP results over the inter-subject testing dataset. It can be found that the DSC would gradually decrease when increasing the weights for Reg(υ) and Reg(ϕ) or decreasing the weight for -Sim(·,·). By contrast, RFP would be improved with bigger weights for Reg(υ) and Reg(ϕ) or smaller weight for -Sim(·,·).

D. Limitations and Future Trends

One of main novelties is the recurrent inference strategy, which is potential to improve accuracy of the deep learning registration methods when coping with the large deformation image registration. We found that it is typically hard for deep learning methods to finish the registration by a single inference step. With the help of curated network structure, the registration performance can be much improved by recurrent inference. In our experiments, we focus on the infant brain MR image and evaluated the proposed RTA-Net can improve the registration performance over the segmentation images. In our experiments, we focus on the infant brain MR image and evaluate the proposed RTA-Net that can improve the registration performance based on the segmentation images. Our proposed method has high potential to better register other datasets. For example, as the adult brain MR images have relatively consistent intensity distribution for GM and WM, as well as clear contrast along the tissue boundaries, our proposed RTA-Net should be well generalized to register the original intensity images directly.

In terms of the similarity metric, except for the cross-correlation, mean square error (MSE), mutual information, soft Dice ratio and other widely used metrics can also be used for registering segmentation maps. In our experiment, our proposed RTA-Net has improved the registration performance compared to the baseline methods (*i.e.*, ANTs, Demons, NiftyReg, and VoxelMorph). Yet, we do not analyze the influence of various metrics for the final performance. While those compared methods use different metrics in optimizing the deformation fields as their recommended configurations, we find no significant difference between MSE and cross-correlation, both of which are considered in our experiment.

VI. Conclusion

In this paper, we have presented a deformation registration network (RTA-Net) for infant brain MR images. RTA-Net is recurrently used in inference stage to improve the registration performance. We have also incorporated adaptive smoothing layer and imposed the tissue-aware regularization to obtain smooth deformation field without sacrificing registration accuracy. We have extensively evaluated our method against the state-of-the-art registration methods using inter- and intra-subject, spatiotemporal registration. The experimental results have validated that our proposed method can achieve good registration accuracy, with higher overlap accuracy for the GM/WM tissues and majority cortical ROIs, and also minimized the folding in the deformation field.

References

[1] A. Sotiras, C. Davatzikos, and N. Paragios, "Deformable medical image registration: A survey," *IEEE Trans. Med. Imag.*, vol. 32, no. 7, pp. 1153–1190, Jul. 2013.
[2] H. Lester and S. R. Arridge, "A survey of hierarchical non-linear medical image registration," *Pattern Recognit.*, vol. 32, no. 1, pp. 129–149, 1998.
[3] R. C. Knickmeyer *et al.*, "A structural MRI study of human brain development from birth to 2 years," *J. Neurosci.*, vol. 28, no. 47, pp. 12176–12182, Nov. 2008.
[4] B. B. Avants, C. L. Epstein, M. Grossman, and J. C. Gee, "Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain," *Med. Image Anal.*, vol. 12, no. 1, pp. 26–41, Feb. 2008.
[5] T. Vercauteren, X. Pennec, A. Perchant, and N. Ayache, "Diffeomorphic demons: Efficient non-parametric image registration," *NeuroImage*, vol. 45, no. 1, pp. 62–71, 2009.
[6] 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," *J. Med. Imag.*, vol. 1, no. 2, p. 24003, 2014.

[7] L. Wang *et al.*, "LINKS: Learning-based multi-source integration framework for segmentation of infant brain images," *NeuroImage*, vol. 108, pp. 160–172, Mar. 2015.

[8] Y. Wu *et al.*, "Hierarchical and symmetric infant image registration by robust longitudinal-example-guided correspondence detection," *Med. Phys.*, vol. 42, no. 7, pp. 4174–4189, Jul. 2015.

[9] S. Hu *et al.*, "Learning-based deformable image registration for infant MR images in the first year of life," *Med. Phys.*, vol. 44, no. 1, pp. 158–170, Jan. 2017.

[10] D. Shen and C. Davatzikos, "HAMMER: Hierarchical attribute matching mechanism for elastic registration," *IEEE Trans. Med. Imag.*, vol. 21, no. 11, pp. 1421–1439, Nov. 2002.

[11] G. Haskins, U. Kruger, and P. Yan, "Deep learning in medical image registration: A survey," *Mach. Vis. Appl.*, vol. 31, nos. 1–2, pp. 1–18, Feb. 2020.

[12] G. Balakrishnan, A. Zhao, M. R. Sabuncu, J. Guttag, and A. V. Dalca, "VoxelMorph: A learning framework for deformable medical image registration," *IEEE Trans. Med. Imag.*, vol. 38, no. 8, pp. 1788–1800, Aug. 2019.

[13] A. V. Dalca, G. Balakrishnan, J. Guttag, and M. Sabuncu, "Unsupervised learning of probabilistic diffeomorphic registration for images and surfaces," *Med. Image. Anal.*, vol. 57, pp. 226–236, Oct. 2019.

[14] Y. Hu *et al.*, "Adversarial deformation regularization for training image registration neural networks," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervent.*, 2018, pp. 774–782.

[15] B. D. de Vos, F. F. Berendsen, M. A. Viergever, H. Sokooti, M. Staring, and I. Išgum, "A deep learning framework for unsupervised affine and deformable image registration," *Med. Image Anal.*, vol. 52, pp. 128–143, Feb. 2018.

[16] Z. Shen, X. Han, Z. Xu, and M. Niethammer, "Networks for joint affine and non-parametric image registration," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 4224–4233.

[17] S. Zhao, Y. Dong, E. Chang, and Y. Xu, "Recursive cascaded networks for unsupervised medical image registration," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 10599–10609.

[18] A. Klein *et al.*, "Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration," *NeuroImage*, vol. 46, no. 3, pp. 786–802, Jul. 2009.

[19] Y. Ou, H. Akbari, M. Bilello, X. Da, and C. Davatzikos, "Comparative evaluation of registration algorithms in different brain databases with varying difficulty: Results and insights," *IEEE Trans. Med. Imag.*, vol. 33, no. 10, pp. 2039–2065, Oct. 2014.

[20] J. Fan, X. Cao, P.-T. Yap, and D. Shen, "BIRNet: Brain image registration using dual-supervised fully convolutional networks," *Med. Image Anal.*, vol. 54, pp. 193–206, May 2019.

[21] K. A. J. Eppenhof and J. P. W. Pluim, "Pulmonary ct registration through supervised learning with convolutional neural networks," *IEEE Trans. Med. Imag.*, vol. 38, no. 5, pp. 1097–1105, Dec. 2019.

[22] J. Fan, X. Cao, Q. Wang, P.-T. Yap, and D. Shen, "Adversarial learning for mono or multi-modal registration," *Med. Image Anal.*, vol. 58, Dec. 2019, Art. no. 101545.

[23] Y. Hu *et al.*, "Weakly-supervised convolutional neural networks for multimodal image registration," *Med. Image Anal.*, vol. 49, pp. 1–13, Oct. 2018.

[24] S. K. Warfield, M. Kaus, F. A. Jolesz, and R. Kikinis, "Adaptive, template moderated, spatially varying statistical classification," *Med. Image Anal.*, vol. 4, no. 1, pp. 43–55, 2000.

[25] M. Prastawa, J. H. Gilmore, W. Lin, and G. Gerig, "Automatic segmentation of MR images of the developing newborn brain," *Med. Image Anal.*, vol. 9, no. 5, pp. 457–466, 2005.

[26] C. A. Cocosco, A. P. Zijdenbos, and A. C. Evans, "A fully automatic and robust brain MRI tissue classification method," *Med. Image Anal.*, vol. 7, no. 4, pp. 513–527, Dec. 2003.

[27] W. Zhang *et al.*, "Deep convolutional neural networks for multi-modality isointense infant brain image segmentation," *NeuroImage*, vol. 108, pp. 214–224, Mar. 2015.

[28] L. Wang *et al.*, "Benchmark on automatic six-month-old infant brain segmentation algorithms: The iSeg-2017 challenge," *IEEE Trans. Med. Imag.*, vol. 38, no. 9, pp. 2219–2230, Sep. 2019.

[29] Y. Sun *et al.*, "Multi-site infant brain segmentation algorithms: The iSeg-2019 challenge," 2020, *arXiv:2007.02096*.

[30] J. Ashburner, "A fast diffeomorphic image registration algorithm," *Neuroimage*, vol. 38, no. 1, pp. 95–113, 2007.

[31] N. D. Cahill, J. A. Noble, and D. J. Hawkes, "A demons algorithm for image registration with locally adaptive regularization," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Berlin, Germany: Springer, 2009, pp. 574–581.

[32] G. Li, L. Wang, F. Shi, J. H. Gilmore, W. Lin, and D. Shen, "Construction of 4D high-definition cortical surface atlases of infants: Methods and applications," *Med. Image Anal.*, vol. 25, no. 1, pp. 22–36, 2015.

[33] J. G. Sled, A. P. Zijdenbos, and A. C. Evans, "A nonparametric method for automatic correction of intensity nonuniformity in MRI data," *IEEE Trans. Med. Imag.*, vol. 17, no. 1, pp. 87–97, Feb. 1998.

[34] F. Shi, L. Wang, Y. Dai, J. H. Gilmore, W. Lin, and D. Shen, "LABEL: Pediatric brain extraction using learning-based meta-algorithm," *NeuroImage*, vol. 62, no. 3, pp. 1975–1986, 2012.

[35] M. Jenkinson and S. Smith, "A global optimisation method for robust affine registration of brain images," *Med. Image Anal.*, vol. 5, no. 2, pp. 143–156, Jun. 2001.

[36] G. Li *et al.*, "Mapping longitudinal development of local cortical gyrification in infants from birth to 2 years of age," *J. Neurosci.*, vol. 34, no. 12, pp. 4228–4238, 2014.

[37] M.-P. Dubuisson and A. K. Jain, "A modified Hausdorff distance for object matching," in *Proc. 12th IAPR Int. Conf. Pattern Recognit. (ICPR)*, vol. 1, Oct. 1994, pp. 566–568.

[38] D. Rueckert, L. I. Sonoda, C. Hayes, D. L. G. Hill, M. O. Leach, and D. J. Hawkes, "Nonrigid registration using free-form deformations: Application to breast MR images," *IEEE Trans. Med. Imag.*, vol. 18, no. 8, pp. 712–721, Aug. 1999.

[39] H. Jia, G. Wu, Q. Wang, and D. Shen, "ABSORB: Atlas building by self-organized registration and bundling," *NeuroImage*, vol. 51, no. 3, pp. 1057–1070, Jul. 2010.