IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received November 16, 2019, accepted November 28, 2019, date of publication December 3, 2019,
date of current version February 11, 2020.

Digital Object Identifier 10.1109/ACCESS.2019.2957233

Brain Deformable Registration Using Global and Local Label-Driven Deep Regression Learning in the First Year of Life

SHUNBO HU, LINTAO ZHANG, GUOQIANG LI, MINGTAO LIU, DEQIAN FU, AND WENYIN ZHANG

School of Information Science and Engineering, Linyi University, Linyi 27600, China
Linda Institute, Shandong Provincial Key Laboratory of Network Based Intelligent Computing, Linyi University, Linyi 276000, China

Corresponding author: Shunbo Hu (hushunbo@lyu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China (NSFC) under Grant 61771230 and Grant 61773244, in part by the Shandong Provincial Natural Science Foundation under Grant ZR2016FM40 and Grant ZR2019PF005, and in part by the Shandong Key Research and Development Program Project 2019GGX101006, 2019GNC106027.

ABSTRACT Accurate medical image registration is highly important for the quantitative analysis of infant brain dynamic development in the first year of life. However, the deformable registration of infant brain magnetic resonance (MR) images is highly challenging for the following two reasons: First, there are very large *anatomical and appearance variations* in these longitudinal images; Second, there is a *one-to-many correspondence* in appearance between global anatomical tissues and the small local tissues therein. In this paper, we use a CNN (convolution neural network)-based global-and-local-label-driven deformable registration scheme. Two to-be-registered image patches are input into the UNet-style regression network. Then, a dense displacement field (DDF) between them is obtained by optimizing the total loss function between two corresponding label patches. Global and local label patches are used only during training. During inference, two new MR images are divided into many patch pairs and fed into the trained network. By averaging the deformation of the patches at the same location, the final 3D DDF between the two whole images is obtained. The highlight is that the global (white matter (WM), gray matter (GM), and cerebrospinal fluid (CSF)) and local tissues can be registered simultaneously without any prior ground-truth deformation. Especially for the local hippocampal tissues, the Dice ratios are substantially improved after registration via our method. Experimental results are presented on the intrasubject and intersubject registration of infant brain MR images between different time points, and the intersubject registration of brain T1-weighted MR images on the OASIS-1 dataset, according to which the proposed method realizes higher accuracy on both global and local tissues compared with state-of-the-art registration methods.

INDEX TERMS Infant brain MR images, deformable registration, label-driven learning.

I. INTRODUCTION

When we study early brain development, diagnose developmental diseases, or establish population atlases, infant brain deformable registration is a highly important step. As a non-invasive method for obtaining high-contrast soft-tissue information, magnetic resonance imaging (MRI) is widely used in neuroimaging research on the infant brain. Unfortunately, due to the rapid brain development and the maturation of the brain myelin, there are dynamic and nonlinear *anatomical and appearance variations* in infant brain MR images [1], [2].

The associate editor coordinating the review of this manuscript and approving it for publication was Kezhi Li.

Moreover, local tissues are typically part of one global tissue, and they have similar appearances in MRI images. As shown in Fig. 1, the hippocampi are mainly located in the GM region, and they are strongly statistically correlated with other parts of the GM. Hence, the global GM tissue alignment does not correspond to the local hippocampal tissue alignment, which is a *one-to-many correspondence* problem. In short, deformable image registration of infant brain MR images is difficult due to two issues: large *anatomical and appearance variations* and *one-to-many correspondences*.

To resolve these two issues, we proposed a double-learning deformable registration scheme for infant brain MR images, and this scheme uses random regression forest to

learn an *appearance-displacement* model and an *appearance-appearance* model [3], [4]. This method yields more accurate registration results than three groups of methods: intensity-based methods [5]–[7], feature-based methods [8] and a sparse-representation-based method [9]. However, the double-learning methods [3], [4] require supervised deformation information. For example, the predefined voxel-level ground-truth deformations for learning are obtained via traditional registration methods, which may introduce new registration errors in the process of obtaining the ground-truth deformations. To avoid over reliance on the deformation supervision, Fan et al. proposed dual-supervised (ground-truth deformation and image density dissimilarity) BIRNet [10].

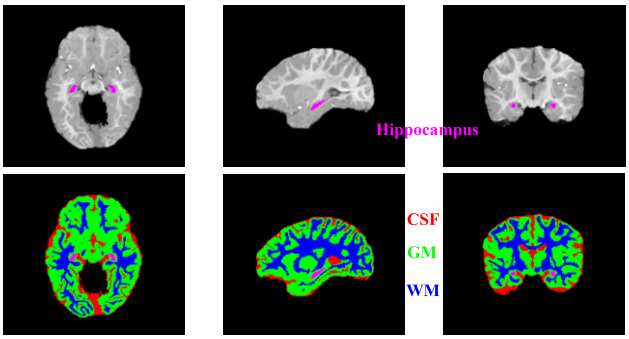


FIGURE 1. Slices of infant brain MR images and four labels: CSF, GM, WM and hippocampus.

In recent years, increasingly many deep-learning-based registration methods have been proposed [11]–[21] and reviewed [22]–[26]. Some of these methods can overcome the shortcomings of supervised transformation estimation. They can maintain the voxel-to-voxel spatial correspondence during deformable registration without any ground-truth deformation supervision. Most are based on CNN, including unsupervised [11]–[17] and weakly supervised [20], [21] deep learning methods. Cao et al. used the contextual cues to train the voxel correspondence at the patch center and applied a thin-plate spline to obtain the DDF [11]. In contrast to Cao's patch-based method, Balakrishnan et al. proposed VoxelMorph, which directly obtains the whole DDF between two registering images by optimizing the total loss function, and this function sums over the cross-correlation (CC) similarity and deformation regularization [12], [13]. They also extended their method to probabilistic diffeomorphic registration [27], [28] and brain MRI segmentation [29]. According to Hu YP et al., these image-similarity-driven unsupervised learning methods would inherit the key shortcomings of classical intensity-based image registration algorithms. For example, the intensity similarities, namely, SSD (sum of squared intensity differences) and SAD (sum of absolute differences), are only suitable for mono-modality image registration with small intensity variations and are not suitable for multimodality or mono-modality image registration with large *intensity and appearance variations*. Even though CC and MI (mutual information) are suitable for multimodality image registration, they are time-consuming and sensitive to image noise and intensity inhomogeneity (bias field). Hence, Hu YP et al. proposed a label-driven weakly-supervised registration framework that does not consider intensity variations or inhomogeneity. They selected the Dice similarity of the anatomical labels as the loss function for optimizing the DDF between two unlabeled real images during the training stage, and they used two new unlabeled real images to acquire the DDF during inference [20], [21]. The method of Hu YP et al. is regarded as a segmentation-based registration method[49]. In addition to the approaches that are discussed above, without requiring any intensity similarity, generative adversarial networks (GANs) were used to assess the quality of image alignment [16], [22], [30]–[32]. The segmentation task and the registration task can be improved via competition with each other. Sometimes one task is focused, e.g., registration-based or atlas-based segmentation, and segmentation-based registration [20], [21], [49]. Sometimes two tasks are combined, e.g., deep-learning-based networks are used to simultaneously optimize two tasks [50], to hybridlike optimize them [49], and to semi-supervised learn them[51]. Since these networks can output segmentation labels, their methods are suitable for registration with parts of labels or without labels.

Most of these methods can accurately align large global tissues; however, they aren't used directly for small local tissue registration where there are *one-to-many correspondences* between the MR appearances of local and global tissues. Hence, we have applied a global-and-local-label-driven deformable registration scheme for brain MR images. The global labels are labels of the global tissues, such as WM, GM and CSF in brain, and the local labels are labels that are located in small regions, e.g., subcortical structures such as the caudate, putamen, pallidum, hippocampus, amygdala, thalamus and accumbens area. The major contributions are summarized as follows:

1) Both global tissues and local tissues are accurately aligned simultaneously. During training, not only the global labels, such as WM, GM and CSF, but also the local labels are utilized. This utilization overcomes the *one-to-many* issue and establishes the voxel correspondences of global tissues and local tissues. Furthermore, the deformation folding in the neighborhoods of global regions will be alleviated due to the alignment between the local tissues.

2) The local tissue alignment is subject-independent and time-point-independent. The registration (dis)similarity between the local tissues is incorporated as a part of the total loss function, and the DDF between the local tissues is also trained during the training stage. Therefore, regardless of the intrasubject registration or intersubject registration of infant brain MR images, small local tissue regions will be accurately aligned for any age gap. This approach is also valid for adult brain MR image registration.

3) The deformable registration that is based on global-and-local-label-driven deep regression learning can help address

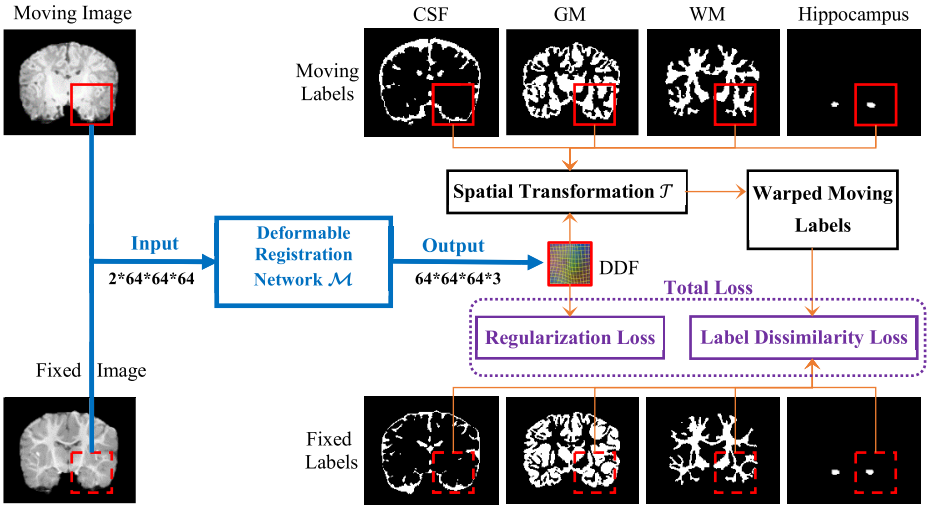


FIGURE 2. Framework of the proposed deformable image registration scheme, which is based on global-and-local-label-driven deep regression learning.

the huge appearance and morphological variations in infant brain MR images without any ground-truth deformation.

In addition, we use DDF regularization, multi strides and surface discarding to increase the accuracy and the smoothness of the DDF, and we use data augmentations and shuffle orders to increase the generalization performance.

This paper extends a preliminary version of this work that was presented at the MLMI 2019 conference [33]. We build on that work by providing additional comparisons, results, analysis, and discussion. We compare five registration methods and conduct experiments not only on T1-weighted MR images but also on T2-weighted MR images and adult brain MR images.

II. METHODS

In this work (as illustrated in Fig. 2), we use a deep regression architecture to model the infant brain deformable registration network $\mathcal{M}$: $\left(I_m^A, I_m^B\right) \Rightarrow \phi_m$ for the m-th patch pair, where $m = 1, 2, \ldots, M$ and M is the total number of patches. The 3D patch pair $\left(I_m^A, I_m^B\right)$ is extracted from the moving image A and the fixed image B at the same position with a fixed stride. Images A and B have been affinely aligned in preprocessing. ϕ_m is the DDF that indicates the voxel dense correspondence, and this DDF is the output of the registration network $\mathcal{M}$ by minimizing the total loss function between all the label pairs, namely, the warped moving label patch l_{mn}^W and the fixed label patch l_{mn}^B, $(n = 1, 2, \ldots, N)$ within the m-th patch, where N is the total number of labels. The total loss function includes the label dissimilarity loss and the deformation regularization term. The label dissimilarity is measured by calculating the multiscale modality-independent Dice dissimilarity between the same label patch pairs. We use Gaussian kernels of various scales to filter the labels to obtain multiscale filtered label images. Then, we calculate the Dice dissimilarity between these filtered label images and sum the Dice dissimilarities to obtain the multiscale Dice dissimilarity. We use the bending energy as a deformation regulation term to preserve the deformation smoothness.

According to the total loss function, the network $\mathcal{M}$ is trained under the weak supervision of global and local labels. Firstly, each $64 \times 64 \times 64$ patch is sampled from a moving image and a fixed image that are centered at the same position. Then, they are concentrated together and inputted into a UNet-style and CNN-based network $\mathcal{M}$. The network output is a 3D deformation field in 3 coordinate directions with patches of the same size as the inputs. Then, the output transformation is applied to warp the same position global-label patches of CSF, WM and GM, and the local-label patch of hippocampus. Furthermore, the multiscale label dissimilarity loss is calculated between these warped moving label patches and the corresponding fixed label patches. At the same time, the regularization loss is calculated according to the output dense deformation. The two losses that are specified above are summed to calculate the total loss. The network parameters are optimized by minimizing the total loss. Last, the pipeline is repeated for all sampled patches of all training images with their global and local labels until the network performance is acceptable.

Once the registration network has been trained, we apply it during inference. By feeding a pair of patches from one new testing MR image and the fixed image into the network, the DDF between them is obtained without using any label information. As illustrated in Fig. 2, only the blue paths are needed, and the brown parts of the labels and the violet parts of the loss functions are not used during inference. Then the deformation registration network will output the 3D DDF for that patch pair. Considering the influence of the patch boundary, we discard the surfaces of the 3D DDF patch and only use the middle patch part. By averaging the middle patch outputs that cover each grid position, we will obtain the whole 3D DDF for the whole images. The details are presented in the following sections.

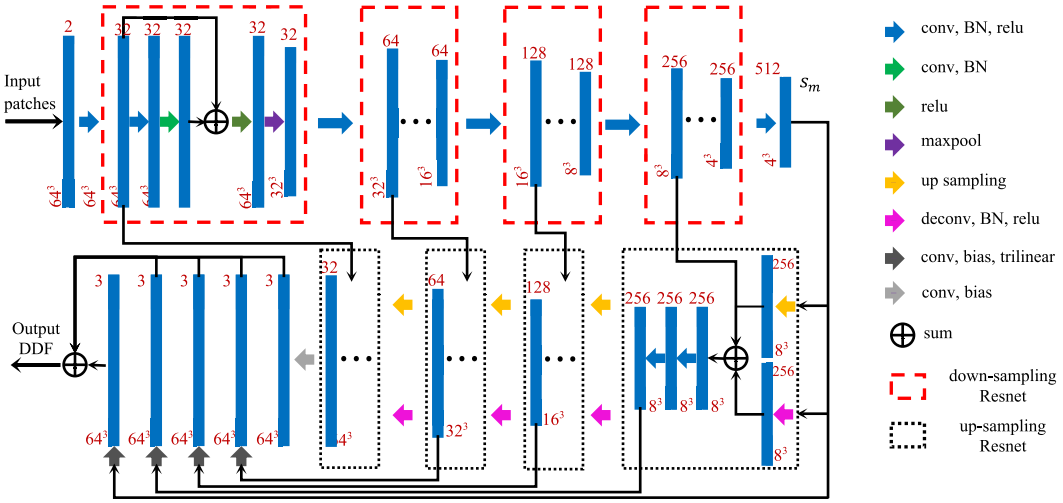


FIGURE 3. Detailed architecture of the patch-based deformable registration network.

A. LOSS FUNCTION USING GLOBAL AND LOCAL LABELS

To train $\mathcal{M}$, the total loss function of the m-th pair of label patches is defined as:

$$E = \frac{1}{|\{i\}|}\sum_i E_G\left(l_{mi}^W, l_{mi}^B\right) + \frac{1}{N_1}\sum_j \alpha_j E_L\left(l_{mj}^W, l_{mj}^B\right) + \beta E_R\left(\phi_m\right), \quad (1)$$

where i is the sequence number of the global label, j is the local label sequence number after $\{i\}$, and $|\{i\}| + |\{j\}| = N$. E_G measures the global dissimilarity between the i-th warped moving label patch l_{mi}^W and the corresponding fixed label patch l_{mi}^B. E_L represents the local dissimilarity between two local label patches, namely, l_{mj}^W and l_{mj}^B. $l_m^W = \mathcal{T}(\phi_m, l_m^A)$ is the warped label patch that is obtained by transforming the moving label patch l_m^A using DDF ϕ_m. $\alpha_j = 1$ if the voxel number of local label j in patch m is larger than a predefined threshold; otherwise, $\alpha_j = 0$. $N_1 = \max(1, \sum_j \alpha_j)$. β is a hyperparameter for regularization, which controls the smoothness of the DDF. We empirically select $\beta = 0.5$ if $\sum_j \alpha_j = 0$, as there are only global label loss; otherwise, $\beta = 1$ since two losses occur, namely, global and local label losses, in order to maintain a constant ratio between the loss term and the regularization term. Therefore, we utilize all possible global labels and the local labels that are mainly located in the selected patches.

The multiscale dissimilarity loss function E_G is defined as follows:

$$E_G\left(l_{mi}^W, l_{mi}^B\right) = 1 - \frac{1}{N_2}\sum_\sigma \delta_{Dice}\left(g_\sigma\left(l_{mi}^W\right), g_\sigma\left(l_{mi}^B\right)\right), \quad (2)$$

where N_2 is the number of scales, which is set to 6 in this work, with a scale set of σ {0,1,2,4,8,16}. σ is the isotropic standard deviation of a 3D Gaussian filter kernel. If $\sigma = 0$, no low-pass Gaussian filtering is applied. As σ is increased, the filtered label images become smoother. We select the largest value, namely, 16, to ensure that the main part of the Gaussian kernel covers the patch. The Gaussian filter kernel can be expressed as follows:

$$g_\sigma = \frac{1}{\left(\sqrt{2\pi}\sigma\right)^3} e^{-\frac{|\boldsymbol{x}|^2}{2\sigma^2}}, \quad (3)$$

where $|\boldsymbol{x}|$ represents the distance to the kernel center. E_L is similar to (2), except that it is defined between two local labels. $g_\sigma\left(l_{mi}^W\right)$ and $g_\sigma\left(l_{mi}^B\right)$ represent the filtered labels for l_{mi}^W and l_{mi}^B by Gaussian kernel g_σ, which can be obtained via 3D convolution. The values in $g_\sigma\left(l_{mi}^W\right)$ and $g_\sigma\left(l_{mi}^B\right)$ are between 0 and 1.

δ_{Dice} measures the Dice similarity between two labels, which is defined as follows:

$$\delta_{Dice}\left(g_\sigma\left(l_{mi}^W\right), g_\sigma\left(l_{mi}^B\right)\right) = \frac{2\left\|g_\sigma\left(l_{mi}^W\right) \circ g_\sigma\left(l_{mi}^B\right)\right\|_1}{\left\|g_\sigma\left(l_{mi}^W\right)\right\|_1 + \left\|g_\sigma\left(l_{mi}^B\right)\right\|_1}, \quad (4)$$

where $\|\ \ \|_1$ denotes the ℓ^1 norm, and '$\circ$' denotes Hadamard multiplication.

Additionally, the DDF should be smooth to preserve the topological correspondence. A regularization term of the bending energy is used in the following discrete form:

$$E_R(\phi) = \frac{1}{|V|}\sum\left[\left(\frac{\partial^2\phi}{\partial x^2}\right)^2 + \left(\frac{\partial^2\phi}{\partial y^2}\right)^2 + \left(\frac{\partial^2\phi}{\partial z^2}\right)^2 + 2\left(\frac{\partial^2\phi}{\partial xy}\right)^2 + 2\left(\frac{\partial^2\phi}{\partial yz}\right)^2 + 2\left(\frac{\partial^2\phi}{\partial xz}\right)^2\right], \quad (5)$$

where $|V|$ represents the number of voxels in a patch.

B. DEFORMABLE REGISTRATION NETWORK

The architecture of the deformable registration network for brain MRI registration is illustrated in detail in Fig. 3. The inputs are a pair of patches that are extracted from infant brain MR images at two time points. The size of each input patch is

$64\times64\times64$. The output DDF patch is of size $64\times64\times64\times3$, and is located at the same location as the input patch.

The network consists of four upsampling Resnet blocks and four downsampling Resnet blocks. Four summation skip layers are used to shortcut the entire network and to alleviate the gradient vanishing issue. The output DDF is calculated via summation over five resolution levels. Four of the levels are obtained via trilinear interpolation after convolving feature data with a bias term from the three end layers of the upsampling blocks and the s_m layer. The last part of the DDF is obtained by convolving from the end layer of the last Resnet block without using trilinear interpolation.

C. CALCULATION OF THE TOTAL DDF DURING INFERENCE

During inference, the total DDF is calculated by averaging the output DDF ϕ_m from the network $\mathcal{M}$. The total DDF at location $\boldsymbol{u}$, which is denoted as $\phi_{total}(\boldsymbol{u})$, is obtained as follows:

$$\phi_{total}(\boldsymbol{u}) = \frac{1}{N_3}\sum\nolimits_{\{m|\boldsymbol{u}\in\boldsymbol{P}_m\}}\phi_m(\boldsymbol{u}), \tag{6}$$

where $N_3 = |\{m|\ \boldsymbol{u}\in\boldsymbol{P}_m\}|$ is the number of patches that cover location $\boldsymbol{u}$. N_3 is related to the stride value n_s and the position $\boldsymbol{u}$.

$$N_3(\boldsymbol{u}) = \prod\nolimits_{i=d,w,h} min\left(\left\lfloor 1+\frac{u_i}{n_s}\right\rfloor, \left\lfloor 1+\frac{n_i-1-u_i}{n_s}\right\rfloor, \left\lfloor\frac{n_{pi}}{n_s}\right\rfloor\right), \tag{7}$$

where $n_d\times n_h\times n_w$ is the image size, $n_{pd}\times n_{ph}\times n_{pw}$ is the patch size, and $\boldsymbol{u}=(u_d, u_w, u_h)$ with $0\le u_i\le n_i-1$.

We extract overlapping patches via a sliding window with a stride of 32 during training and with a stride of 4 during inference. We choose the stride of 32 during training since it is half of the patch size of 64 and the chosen patches cover the whole images with enough overlap. During inference smaller stride can produce smoother DDF, so we choose a stride of 4 for efficiency. At the same time, to further increase the smoothness of the DDF and to ensure that it contains sufficient neighborhood information, surface-discarding is adopted. For example, the patch size of the DDF is $64\times64\times64\times3$ during training, while during inference, we only utilize the middle part of the patch with a size of $58\times58\times58\times3$ by discarding 3-voxel-width surfaces on each face of the patch.

The DDFs and their Jacobian determinants are presented in Fig. 4 and are obtained by Label-reg [20], [21] via the global label (LRG) method and via our method, which combines global and local label registration. The intrasubject registration deformations of MR images from 2 weeks old to 12 months old are presented. The coronal slices of two DDFs and their detailed grids are presented in the top row, and the axial slices of the two DDFs and their Jacobian determinant images are presented in the bottom row with pseudo color. In the top row, the DDF on the left uses only the global WM, GM and CSF labels and no local labels, while the DDF on the right utilizes the global labels and the local hippocampal label simultaneously. In the bottom row, the first two left images are the DDF and its Jacobian determinant via the LRG method, and the last two images are sliced from the results of our method. The background value is 0 in the DDF image, and 1 in the Jacobian determinant image. From Fig. 4, we conclude that there are three advantages of combining global and local label information in the optimization of the DDF. First, higher anti-folding performance is realized and is illustrated in the two regions that are indicated by the green arrows in the top row. Hence, the correct local tissue alignment may provide smooth regularization to the global neighborhood deformations. Second, as shown by the yellow arrows, the local transformation alignment is more accurate. This accuracy is significant when we analyze the development of local tissues, lesions or tumors, and in the registration of small regions of interest (ROIs). Third, the DDF in global regions via our method is smoother than via the LRG method. As shown in the bottom row, there are fewer large deformations in the global label regions, and the values of Jacobian determinant in many regions are closer to 1 on the right images.

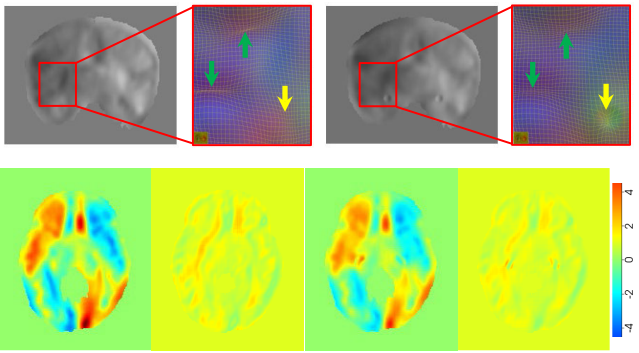


FIGURE 4. DDFs and their Jacobian determinants, which were calculated without (LRG method, in the left two columns) or with (our method, in the right two columns) the local hippocampal labels.

III. EXPERIMENTAL RESULTS

The first experimental dataset included infant brain MR images of 24 subjects, where each subject had T1-weighted and T2-weighted MR images at 2 weeks old and at 3, 6, 9 and 12 months old. T1-weighted MR images were obtained using a head-only 3T MR scanner with a resolution of $1\times1\times1$ mm^3, and T2-weighted images were acquired with a resolution of $1.25\times1.25\times1.95$ mm^3. In preprocessing, each T2-weighted MR image was linearly aligned to its corresponding T1-weighted MR images for each subject at the same time point with FLIRT [34]–[36] and isotopically upsampled to $1\times1\times1$ mm^3 resolution. Then, from the affinely registered images, the skull is removed to obtain brain images, bias correction is used to remove intensity inhomogeneity by a N4 bias field correction method in ANTs [52], and histogram equalization is used to enhance the images. Last, the corresponding segmentation labels, namely, WM, GM and hippocampus, were obtained using the iBEAT

TABLE 1. Means and standard deviations of the Dice ratios of three labels for intrasubject registration of T1-weighted MR images using five registration methods. The GM and WM labels are global labels, and the hippocampal label is a local label.

Labels	Methods	2 week to 12-month	3-Month to 12-Month	6-Month to 12-Month	9-month to 12-month
GM	Demons	0.611 ± 0.017	0.668 ± 0.030	0.823 ± 0.013	**0.879** ± 0.010
	MI	0.677 ± 0.027	0.643 ± 0.038	0.825 ± 0.010	0.873 ± 0.010
	SyN	0.633 ± 0.016	0.681 ± 0.035	0.829 ± 0.012	**0.879** ± 0.009
	LRG	0.837 ± 0.012	**0.828** ± 0.021	**0.871** ± 0.006	0.876 ± 0.007
	our	**0.849** ± 0.009	0.796 ± 0.017	0.867 ± 0.005	0.876 ± 0.005
WM	Demons	0.703 ± 0.008	0.695 ± 0.039	0.746 ± 0.038	0.834 ± 0.043
	MI	0.741 ± 0.024	0.663 ± 0.040	0.781 ± 0.041	0.862 ± 0.047
	SyN	0.537 ± 0.025	0.601 ± 0.040	0.823 ± 0.018	0.868 ± 0.011
	LRG	0.834 ± 0.014	**0.804** ± 0.070	0.855 ± 0.006	**0.869** ± 0.006
	our	**0.835** ± 0.010	**0.804** ± 0.049	**0.859** ± 0.006	0.864 ± 0.006
hippocampus	Demons	0.447 ± 0.090	0.589 ± 0.064	0.664 ± 0.077	0.708 ± 0.035
	MI	0.124 ± 0.086	0.228 ± 0.094	0.640 ± 0.090	0.722 ± 0.048
	SyN	0.348 ± 0.140	0.601 ± 0.096	0.713 ± 0.077	0.740 ± 0.046
	LRG	0.467 ± 0.173	0.637 ± 0.066	0.631 ± 0.112	0.735 ± 0.046
	our	**0.850** ± 0.015	**0.851** ± 0.012	**0.842** ± 0.010	**0.870** ± 0.019

toolbox [37] and via experts' manual refinement. Next, all the images were cropped to the same size of 160 × 192×128. Finally, data augmentations were leveraged by applying a random affine transformation to each image and the corresponding label prior to training, which were implemented in the image argumentation layer using open-source code in NiftyNet [38], [39].

The second experimental dataset is public OASIS-1 (Open Access Series of Imaging Studies) [47], which consists of a cross-sectional collection of 416 subjects of ages 18 to 96. For each subject, the dataset contains T1-weighted MRI images and their segmentation label images. The labels include three global tissues: WM, GM and CSF. Since the MR and label images are affinely aligned to the atlas, we continue to segment small local subcortical labels, i.e., hippocampus and caudate, via the run_first_all command with the corresponding structure options in FSL packages. All images are resampled to the same dimensions of 176 × 208×176 with 1-mm isotropic voxels.

For intrasubject registration, we used 22 subjects for training, 1 subject for validation and 1 subject for inference; for intersubject registration, we used 21 subjects for training, 1 subject for validation, 1 subject for inference and 1 subject for the fixed images. The network was trained on a 12GB Nvidia TitanX GPU with 10000 iterations. In each iteration, a pair of randomly augmented MR images were generated and further sampled to 60 patches, which can be calculated as follows:

$$N_4 = \left\lfloor \frac{\left(n_d - \left(n_{pd} - n_s\right)\right)\left(n_h - \left(n_{hd} - n_s\right)\right)\left(n_w - \left(n_{wd} - n_s\right)\right)}{n_s^3} \right\rfloor, \tag{8}$$

where $n_d \times n_h \times n_w$ is the image size, $n_{pd} \times n_{ph} \times n_{pw}$ is the patch size, and n_s is the stride. For cropped experimental images, the number of patches is $N_4 = ((160\text{-}(64\text{-}32))/32)((192\text{-}(64\text{-}32))/32)((128\text{-}(64\text{-}32))/32) = 60$ with a patch size of 64×64×64 and a stride of 32. Hence, during training, the total number of patch pairs of MR images was 10000×60 = 600000. In each cycle, the orders of subjects and patches were shuffled. It took approximately 72 hours using an Adam optimizer starting at a learning rate 10^{-5} with a minibatch size of 4 during training, and it took only ∼11 seconds to calculate the total DDF during inference with the same stride.

A. REGISTRATION RESULTS OF INFANT BRAIN MRI

The Dice ratios of WM, GM and hippocampus are calculated to evaluate the registration performance in Tables 1-4. We list the Dice ratio results of intra/intersubject registration from 2-week, 3-, 6- and 9-month-old to 12-month-old subjects. These ratio results are obtained via Demons [40]–[42], via Mutual Information (MI) [5], [6], [43], via symmetric normalization (SyN) in ANTs [48] with the CC loss function (a python version seen in DIPY package), via Label-reg [20], [21] with global labels (LRG) and via our method, respectively. The Demons and MI methods are implemented by SimpleITK [42]–[44].

The Dice ratios between the i-th warped moving label l_i^W and the corresponding fixed label l_i^B are calculated as the following without any Gaussian filtering:

$$\delta_{Dice}\left(l_i^W, l_i^B\right) = \frac{2\left|l_i^W \cap l_i^B\right|}{\left|l_i^W\right| + \left|l_i^B\right|}. \tag{9}$$

The target registration error (TRE) is also used to measure the image registration accuracy, which is defined as the arithmetic-mean-centroid distance between the warped

TABLE 2. Means and standard deviations of the Dice ratios of three labels for intersubject registration of T1-w MR images using five registration methods.

Labels	Methods	2 week to 12-month	3-Month to 12-Month	6-Month to 12-Month	9-month to 12-month
GM	Demons	0.600 ± 0.016	0.639 ± 0.038	0.746 ± 0.033	0.793 ± 0.032
	MI	0.632 ± 0.027	0.643 ± 0.038	0.716 ± 0.041	0.743 ± 0.046
	SyN	0.619 ± 0.018	0.649 ± 0.038	0.747 ± 0.032	0.784 ± 0.034
	LRG	**0.770** ± 0.023	**0.759** ± 0.020	**0.810** ± 0.022	**0.814** ± 0.027
	our	0.759 ± 0.027	0.726 ± 0.025	0.772 ± 0.030	0.788 ± 0.032
WM	Demons	**0.727** ± 0.012	**0.714** ± 0.035	0.695 ± 0.017	0.689 ± 0.042
	MI	**0.727** ± 0.015	0.663 ± 0.040	0.687 ± 0.037	0.688 ± 0.056
	SyN	0.492 ± 0.019	0.530 ± 0.040	0.690 ± 0.046	0.718 ± 0.047
	LRG	0.725 ± 0.035	0.687 ± 0.041	**0.752** ± 0.034	**0.760** ± 0.037
	our	0.716 ± 0.038	0.663 ± 0.045	0.732 ± 0.036	0.740 ± 0.039
hippocampus	Demons	0.465 ± 0.104	0.552 ± 0.058	0.597 ± 0.071	0.612 ± 0.059
	MI	0.141 ± 0.095	0.228 ± 0.094	0.522 ± 0.077	0.602 ± 0.080
	SyN	0.390 ± 0.135	0.573 ± 0.099	0.639 ± 0.062	0.635 ± 0.069
	LRG	0.402 ± 0.175	0.522 ± 0.080	0.541 ± 0.092	0.652 ± 0.055
	our	**0.843** ± 0.020	**0.850** ± 0.010	**0.856** ± 0.011	**0.851** ± 0.007

TABLE 3. Means and standard deviations of the Dice ratios of three labels for intrasubject registration of T2-w MR images using five registration methods.

Labels	Methods	2 week to 12-month	3-Month to 12-Month	6-Month to 12-Month	9-month to 12-month
GM	Demons	0.616 ± 0.018	0.635 ± 0.027	0.717 ± 0.045	0.772 ± 0.068
	MI	0.678± 0.028	0.656 ± 0.033	0.748 ± 0.042	0.781 ± 0.078
	SyN	0.667± 0.022	0.668 ± 0.027	0.750 ± 0.038	0.791 ± 0.060
	LRG	0.803 ± 0.009	**0.810** ± 0.014	**0.870** ± 0.006	**0.873** ± 0.007
	our	**0.838** ± 0.010	0.788 ± 0.022	0.850 ± 0.008	0.870 ± 0.009
WM	Demons	0.719 ± 0.015	0.677 ± 0.034	0.690 ± 0.044	0.739 ± 0.095
	MI	0.731 ± 0.035	0.666 ± 0.066	0.722 ± 0.062	0.759 ± 0.130
	SyN	0.633± 0.040	0.623 ± 0.076	0.720 ± 0.048	0.745 ± 0.073
	LRG	0.821 ± 0.007	**0.801**± 0.065	**0.858** ± 0.006	**0.861** ± 0.008
	our	**0.823** ± 0.008	0.798 ± 0.050	0.842 ± 0.007	**0.861** ± 0.008
hippocampus	Demons	0.428 ± 0.116	0.526 ± 0.083	0.594 ± 0.082	0.628 ± 0.146
	MI	0.382 ± 0.152	0.463 ± 0.124	0.623 ± 0.077	0.624 ± 0.175
	SyN	0.539± 0.145	0.561 ± 0.072	0.629 ± 0.099	0.683 ± 0.060
	LRG	0.433 ± 0.167	0.640 ± 0.053	0.635 ± 0.127	0.734 ± 0.050
	our	**0.852** ± 0.018	**0.841** ± 0.020	**0.865** ± 0.014	**0.848** ± 0.011

moving label and the corresponding fixed label.

$$TRE_{amc}\left(l_i^W, l_i^B\right) = \left\| \frac{\sum_{\boldsymbol{u}_w \in l_i^W} \boldsymbol{u}_w}{\left|l_i^W\right|} - \frac{\sum_{\boldsymbol{u}_B \in l_i^B} \boldsymbol{u}_B}{\left|l_i^B\right|} \right\|_2, \quad (10)$$

where $\boldsymbol{u}_w$ and $\boldsymbol{u}_B$ are voxel coordinates in labels l_i^W and l_i^B, respectively, and $||\quad||_2$ denotes the ℓ^2-norm. The TRE results are listed in Tables 5-8.

In Tables 1-4, for the global GM and WM labels, our registration scheme has similar Dice ratios to the LRG method and has higher Dice ratios than the Demons, MI and SyN methods in most cases. For the local hippocampal label, our registration method yields the best results among the five registration methods and realizes Dice ratios of more than 83.8% in all cases. For example, in comparison with the **best results** of the other four methods in terms of the mean Dice ratios of hippocampus: 1) for the intrasubject registration of T1-weighted MR images in Table 1, our method improves the ratios by approximately 82%, 34%, 18% and 18% for images of 2-week-old and 3-, 6- and 9-month-old subjects, respectively, to 12-month-old subjects; 2) for the intersubject registration of T1-weighted MR images in Table 2, our method enhances the results by approximately 81%, 48%, 34% and 31%, respectively; 3) for the intrasubject registration of T2-weighted MR images in Table 3, our method improves the results by approximately 58%, 31%, 36% and 16%, respectively; 4) for the intersubject registration of T2-weighted MR images in Table 4, our method enhances the results by approximately 93%, 63%, 37% and 35%, respectively; 5) Comparing the global WM, GM and local hippocampus region registration results for intersubject registration for any time gap in Tables 2 and 4, the Dice

TABLE 4. Means and standard deviations of the Dice ratios of three labels for intersubject registration of T2-w MR images using five registration methods.

Labels	Methods	2 week to 12-month	3-Month to 12-Month	6-Month to 12-Month	9-month to 12-month
GM	Demons	0.583 ± 0.021	0.594 ± 0.037	0.675 ± 0.028	0.718 ± 0.039
	MI	0.617 ± 0.030	0.601 ± 0.040	0.684 ± 0.030	0.710 ± 0.039
	SyN	0.622 ± 0.023	0.618 ± 0.036	0.691 ± 0.029	0.723 ± 0.037
	LRG	**0.767** ± 0.019	**0.764** ± 0.021	**0.793** ± 0.023	**0.795** ± 0.027
	our	0.754 ± 0.024	0.738 ± 0.022	0.780 ± 0.027	0.787 ± 0.033
WM	Demons	**0.720** ± 0.015	0.685 ± 0.032	0.695 ± 0.017	0.669 ± 0.031
	MI	0.714 ± 0.030	0.649 ± 0.047	0.704 ± 0.026	0.708 ± 0.034
	SyN	0.547 ± 0.022	0.534 ± 0.062	0.605 ± 0.046	0.616 ± 0.050
	LRG	**0.720** ± 0.032	0.684 ± 0.047	**0.730** ± 0.035	**0.734** ± 0.036
	our	0.705 ± 0.035	**0.686** ± 0.040	0.719 ± 0.039	0.714 ± 0.049
hippocampus	Demons	0.438 ± 0.109	0.515 ± 0.088	0.589 ± 0.054	0.584 ± 0.080
	MI	0.289 ± 0.171	0.320 ± 0.132	0.564 ± 0.094	0.586 ± 0.093
	SyN	0.444 ± 0.169	0.504 ± 0.116	0.621 ± 0.062	0.637 ± 0.044
	LRG	0.407 ± 0.145	0.511 ± 0.072	0.594 ± 0.070	0.620 ± 0.066
	our	**0.858** ± 0.013	**0.838** ± 0.009	**0.848** ± 0.013	**0.862** ± 0.007

TABLE 5. Means and standard deviations of the TREs (mm) of three labels for intrasubject registration of T1-w MR images using five registration methods.

Labels	Methods	2 week to 12-month	3-Month to 12-Month	6-Month to 12-Month	9-month to 12-month
GM	Demons	1.596 ± 0.451	1.684± 1.182	0.529 ± 0.325	0.387 ± 0.169
	MI	1.670 ± 0.440	1.688 ± 1.089	0.579 ± 0.202	0.388 ± 0.185
	SyN	1.365 ± 0.360	1.384 ± 1.042	0.617 ± 0.189	0.421 ± 0.229
	LRG	0.764 ± 0.239	**0.960** ± 0.298	0.525 ± 0.177	0.414 ± 0.152
	our	**0.678** ± 0.228	1.234 ± 0.418	**0.366** ± 0.123	**0.385** ± 0.134
WM	Demons	1.050 ± 0.255	1.278 ± 0.367	1.261 ± 0.262	0.860 ± 0.310
	MI	**0.549** ± 0.181	0.777 ± 0.237	0.379 ± 0.163	**0.257** ± 0.074
	SyN	1.255 ± 0.777	1.819 ± 1.535	0.673 ± 0.265	0.376 ± 0.273
	LRG	0.615 ± 0.372	0.578 ± 0.262	0.428 ± 0.207	0.338 ± 0.088
	our	0.616 ± 0.303	**0.568** ± 0.369	**0.245** ± 0.095	0.310 ± 0.151
hippocampus	Demons	4.435 ± 2.496	2.885 ± 1.516	2.577 ± 1.625	2.309 ± 1.187
	MI	3.390 ± 2.392	1.968 ± 1.073	2.240 ± 1.612	2.131 ± 0.747
	SyN	4.403 ± 2.567	2.268 ± 1.139	1.920 ± 1.164	1.961 ± 0.743
	LRG	3.988 ± 1.375	2.135 ± 0.960	3.100 ± 1.844	1.898 ± 0.921
	our	**0.632** ± 0.288	**0.616** ± 0.326	**0.705** ± 0.352	**0.537** ± 0.252

ratios of the hippocampus that are obtained using our method are the largest (best) among all 15 values in each column. For intrasubject registration from 2-week-old to 12-month-old subjects and form 3-month-old to 12-month-old subjects in Tables 1 and 3, the Dice ratios of the hippocampus that are obtained using our method are also the largest among all 15 values; and for the remaining time gap (last two columns), our hippocampus registration results are better than most of those of the other four methods.

In Tables 5-8, we list the TREs results of intra/intersubject registration from 2-week-old and 3-, 6- and 9-month-old to 12-month-old subjects that were obtained via the five methods. Under most cases for the global labels, the TREs that are obtained by our method are smaller than 1mm; are lower than those by the Demons, MI and SyN methods; and are similar to those by the LRG method. Under all cases for the local hippocampus region, our method realizes the smallest TREs in comparison with the other four methods. For example, in comparison with the **best** hippocampus TRE results of the other four methods: 1) in Table 5, for the intra-subject registration of T1-weighted MR images, the TREs are reduced to approximately 19%, 31%, 37% and 28% of the corresponding smallest mTRE (mean of TREs) of the other four methods from images of 2-week-old and 3-, 6- and 9-month-old subjects, respectively, to 12-month-old subjects; 2) for the intersubject registration of T1-weighted MR images in Table 6, our method reduces mTRE to approximately 18%, 55%, 31% and 16%, respectively; 3) for the intrasubject registration of T2-weighted MR images in Table 7, our method reduces mTRE to approximately 32%, 41%, 48% and 24%, respectively; and 4) for the intersubject registration of T2-weighted MR images in Table 8, our method

TABLE 6. Means and standard deviations of the TREs (mm) of three labels for intersubject registration of T1-w MR images using five registration methods.

Labels	Methods	2 week to 12-month	3-Month to 12-Month	6-Month to 12-Month	9-month to 12-month
GM	Demons	1.591± 0.460	1.879 ± 1.277	1.135 ± 0.309	0.976 ± 0.304
	MI	1.778 ± 0.504	1.915 ± 1.196	0.954 ± 0.323	0.762 ± 0.371
	SyN	1.513 ± 0.386	1.682 ± 1.189	1.091 ± 0.316	0.855 ± 0.373
	LRG	**0.943** ± 0.332	**1.553** ± 0.449	0.842 ± 0.299	**0.541** ± 0.234
	our	1.099 ± 0.377	2.490 ± 0.568	**0.626** ± 0.282	0.765 ± 0.284
WM	Demons	1.260 ± 0.275	1.259 ± 0.371	1.048 ± 0.381	0.916 ± 0.309
	MI	0.602 ± 0.220	**0.757** ± 0.442	0.775 ± 0.251	0.659 ± 0.271
	SyN	1.523 ± 0.472	2.178 ± 1.417	0.826 ± 0.331	1.037 ± 0.339
	LRG	**0.568** ± 0.210	0.917 ± 0.339	0.820 ± 0.191	0.568 ± 0.205
	our	0.715 ± 0.253	1.253 ± 0.216	**0.592** ± 0.184	**0.454** ± 0.189
hippocampus	Demons	3.714 ± 3.136	3.061 ± 1.432	2.105 ± 1.282	2.394 ± 1.411
	MI	3.796 ± 2.703	3.486 ± 1.781	2.184 ± 1.125	2.490 ± 0.768
	SyN	4.649 ± 3.902	2.177 ± 1.107	2.162 ± 1.049	2.803 ± 1.566
	LRG	4.299 ± 2.612	2.236 ± 0.901	3.448 ± 1.263	3.649 ± 2.471
	our	**0.673** ± 0.360	**1.199** ± 0.405	**0.645** ± 0.351	**0.375** ± 0.197

TABLE 7. Means and standard deviations of the TREs (mm) of three labels for intrasubject registration of T2-w MR images using five registration methods.

Labels	Methods	2 week to 12-month	3-Month to 12-Month	6-Month to 12-Month	9-month to 12-month
GM	Demons	1.952 ± 0.852	2.496 ± 0.839	1.362 ± 0.540	1.007 ± 1.003
	MI	2.398± 0.921	3.514 ± 0.857	0.879 ± 0.222	0.929 ± 1.206
	SyN	1.462± 0.666	2.176 ± 0.711	0.944 ± 0.276	0.893 ± 0.870
	LRG	**0.541** ± 0.291	**1.100** ± 0.335	0.573 ± 0.144	**0.357** ± 0.126
	our	0.621 ± 0.309	1.165 ± 0.341	**0.465** ± 0.242	0.427 ± 0.181
WM	Demons	2.011 ± 0.593	1.986 ± 0.808	1.640 ± 0.814	1.805 ± 1.255
	MI	1.891 ± 0.377	2.367± 1.108	1.064 ± 0.271	1.278 ± 1.770
	SyN	2.231± 0.650	2.915 ± 1.046	1.007 ± 0.321	1.446 ± 1.641
	LRG	**0.627** ± 0.279	0.445 ± 0.293	**0.378** ± 0.130	**0.304** ± 0.139
	our	0.630 ± 0.480	**0.360** ± 0.219	0.381 ± 0.158	0.415 ± 0.186
hippocampus	Demons	3.461± 2.458	3.125 ± 1.437	2.407 ± 1.133	2.508 ± 0.575
	MI	3.472 ± 1.723	1.987 ± 0.932	2.399 ± 1.452	3.100 ± 1.409
	SyN	3.598± 1.705	2.379 ± 1.148	2.674 ± 1.470	2.135 ± 0.695
	LRG	3.892 ± 1.300	2.020 ± 0.836	2.832 ± 1.772	2.042 ± 0.714
	our	**1.091** ± 0.771	**0.820** ± 0.414	**1.159** ± 0.366	**0.492** ± 0.219

reduces mTRE to approximately 23%, 19%, 48% and 20%, respectively.

We draw the following conclusions: 1) The registration of the local tissue, namely, the hippocampus, is subject-independent, modality-independent and time-point-independent. Our method realizes satisfactory alignment performance for both intrasubject and intersubject registration of T1- or T2-weighted MR images for every time-point pair, namely, it realizes the largest Dice ratios and the smallest TREs; 2) For global tissues (WM and GM), our scheme yields satisfactory registration results, which are similar to those of LRG and superior to those of Demons and MI in most cases; 3) With a smaller longitudinal age gap, the infant brain MR images are more similar; hence, the registration become easier and higher Dice ratios can be realized. However, among the five registration methods, namely, the learning-based methods, LRG and our method, these tendencies are not readily observed since the Dice ratio values for the larger time gap from the low time point to 12 months old are closer to those for the smaller time gap from the high time point to 12 months old. Furthermore, for the local hippocampus region, this variation tendency is not observed for our registration method, which realizes larger Dice ratios under any time interval; 4) Among the five methods, only our method can yield convincing registration results for local tissue. For example, under the same registration condition, the Dice ratios and TREs of the hippocampus from our method are the best among the five methods; 5) Comparing the Dice ratios or TREs between global and local regions, the registration accuracy of local hippocampus region that is realized

TABLE 8. Means and standard deviations of the TREs (mm) for three labels for intersubject registration of T2-w MR images using five registration methods.

Labels	Methods	2 week to 12-month	3-Month to 12-Month	6-Month to 12-Month	9-month to 12-month
GM	Demons	2.907 ± 0.798	3.405 ± 0.892	1.758 ± 0.476	1.446 ± 0.731
	MI	3.051 ± 1.078	4.534 ± 1.186	1.498 ± 0.500	1.345 ± 1.045
	SyN	2.723 ± 0.885	3.303 ± 1.070	1.274 ± 0.343	1.099 ± 0.724
	LRG	**0.880** ± 0.379	**1.387** ± 0.518	**0.575** ± 0.238	1.111 ± 0.296
	our	1.182 ± 0.359	1.536 ± 0.346	0.771 ± 0.244	**1.101** ± 0.350
WM	Demons	1.803 ± 0.487	1.678 ± 0.647	1.197 ± 0.488	1.512 ± 1.582
	MI	2.306 ± 0.756	2.311 ± 1.369	2.074 ± 0.633	1.782 ± 1.242
	SyN	2.740 ± 0.864	4.102 ± 1.108	1.760 ± 0.693	1.849 ± 0.977
	LRG	**0.555** ± 0.236	1.137 ± 0.347	0.807 ± 0.269	**0.443** ± 0.195
	our	0.740 ± 0.263	**1.130** ± 0.247	**0.694** ± 0.232	0.631 ± 0.236
hippocampus	Demons	4.123 ± 3.623	2.968 ± 1.285	3.513 ± 1.086	2.638 ± 0.916
	MI	4.021 ± 2.487	2.823 ± 1.018	2.385 ± 1.150	2.444 ± 0.942
	SyN	4.991 ± 2.981	3.194 ± 1.194	3.252 ± 1.340	2.303 ± 0.917
	LRG	4.130 ± 1.989	2.340 ± 1.347	2.983 ± 1.744	3.066 ± 2.206
	our	**0.941** ± 0.383	**0.434** ± 0.272	**1.143** ± 0.372	**0.450** ± 0.204

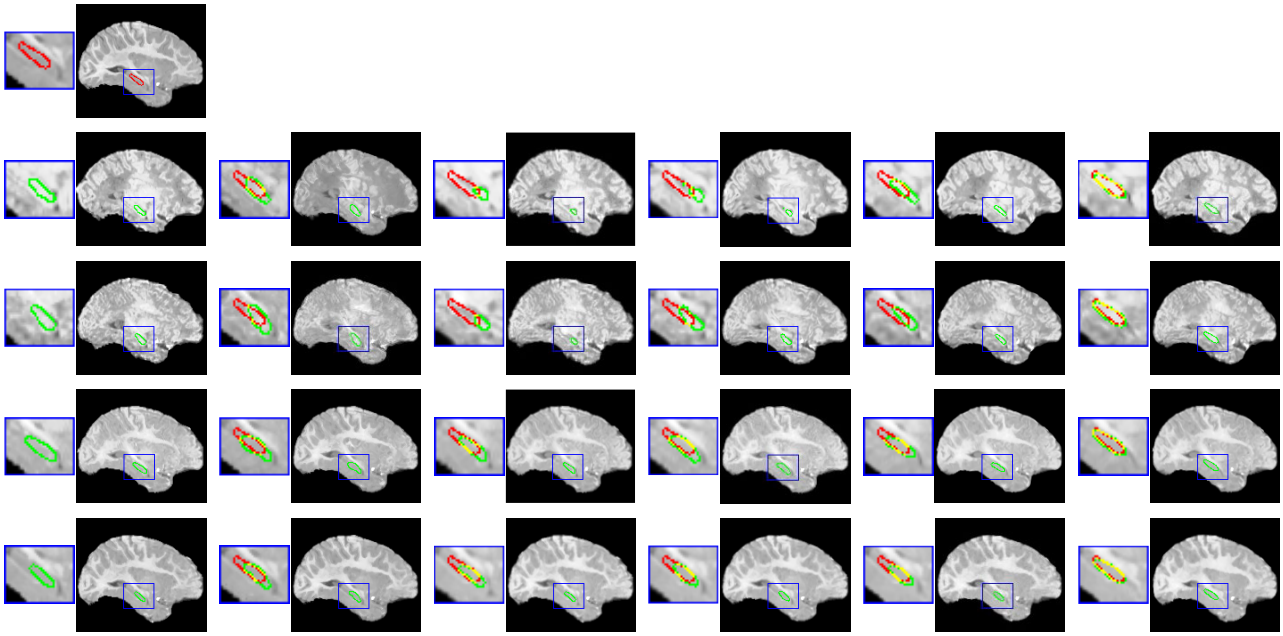


FIGURE 5. Visualization of the intrasubject registration of T1-weighted MR images. We present intrasubject registration results from 2-week-old, 3-, 6-, and 9-month-old to 12-month-old infant brain MR images from the second row to the fifth row, respectively. The top sagittal slice is from a 12-month fixed image, and the slices in remaining four rows from left to right are a moving image and warped images by Demons, MI, SyN, LRG and our method. In this figure, the red color denotes the fixed hippocampal contours, the green color denotes the moving or warped contours, and the yellow denotes the overlap contours between the fixed and the warped contours. Our proposed method can realize the largest hippocampus overlaps (shown in the last column).

using our method is higher than that for the global WM or GM region that is realized using any of the five methods in most cases. Therefore, our method is suitable not only for large or global region registration but also for small or local tissue registration.

Qualitative comparisons are presented in Fig. 5-8, which show the sagittal slices of MR images from 2-week-old and 3-, 6-, and 9-month-old to 12-month-old intra- or intersubject registration. In these figures, sagittal slice in top row comes from 12-month fixed images, and the sagittal slices from left to right in the four bottom rows are moving images and warped images by Demons, MI, SyN, LRG and our method, respectively. In these figures, the red color denotes the fixed hippocampal contours, the green color denotes the moving or warped contours, and the yellow represents the overlap contours between the fixed and the warped contours. Our

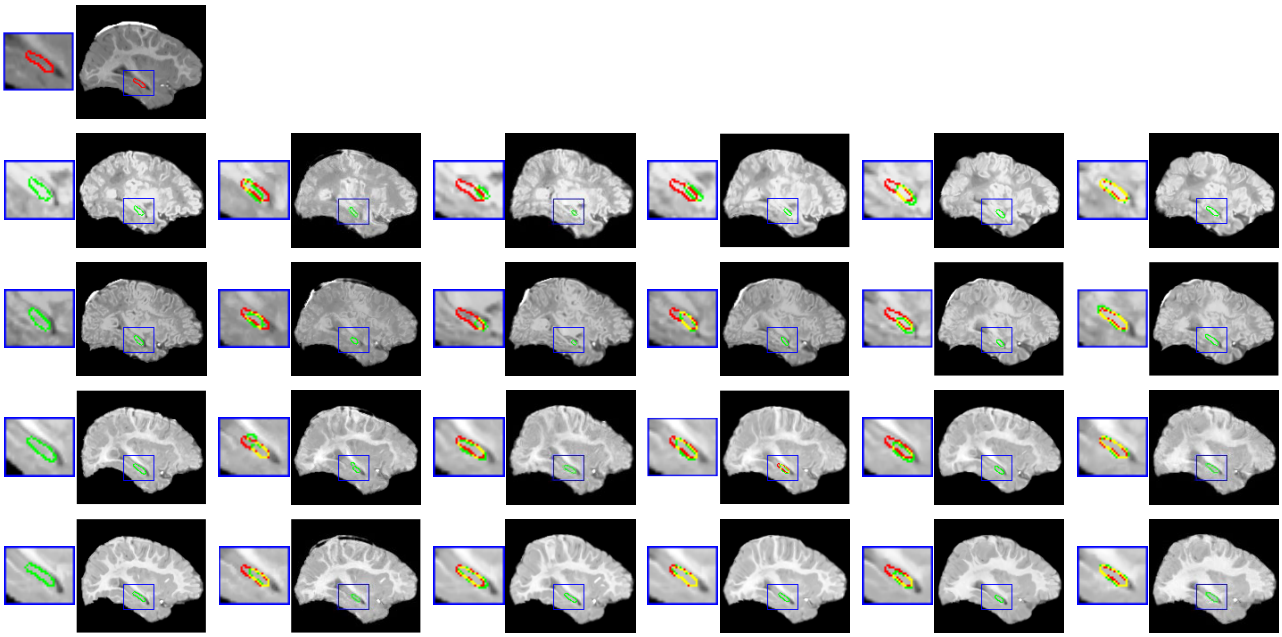


FIGURE 6. Visualization of the intersubject registration of T1-weighted MR images. The sagittal slices are in the same order as in Fig. 5. Our proposed method still realizes the largest hippocampus overlaps (shown in the last column).

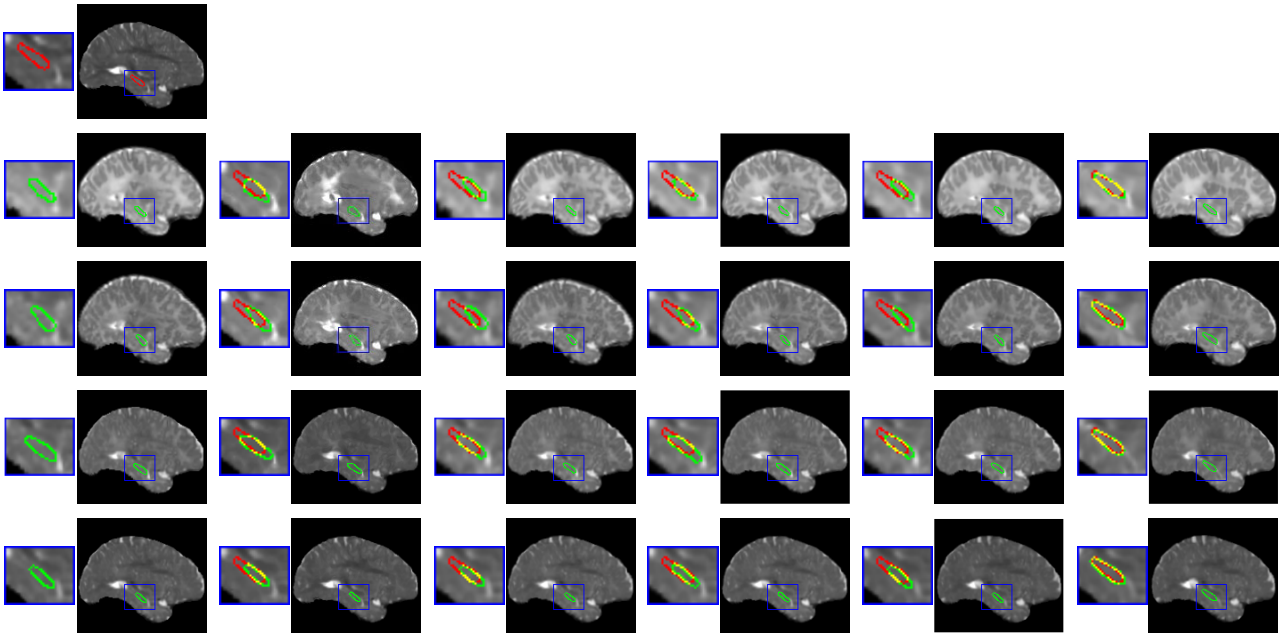


FIGURE 7. Visualization of the intrasubject registration of T2-weighted MR images. The arrangement of the sagittal slices is the same as in Fig. 5. Our proposed method still realizes the largest hippocampus overlaps (shown in the last column).

method obtains the largest yellow overlap contours among the five methods, which shows the best registration performance for the local tissue, namely, the hippocampus.

B. REGISTRATION RESULTS OF OASIS-1

In order to evaluate the generalization performance of our method, we apply it on the public OASIS-1 dataset, and we compare the performances of our registration method on two global tissues (WM and GM) and two local tissues (hippocampus and caudate) with those of the LRG method. In this intersubject registration experiment, 416 subjects are used, where one subject is randomly chosen for the fixed images, 30 for the testing images, and the remainder for the moving images. The other parameters are the same as in the first experiment.

The quantitative results are presented in Fig. 9. For the global regions, namely, WM and GM, the Dice ratios and TREs of our method are slightly better than those of the LRG

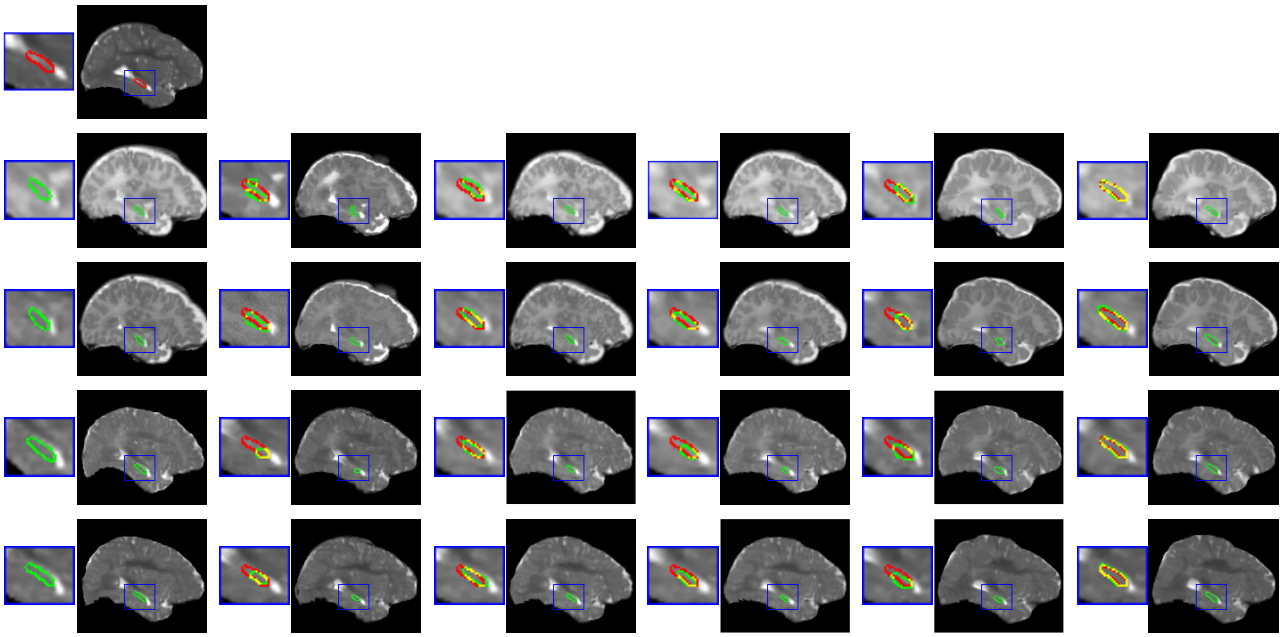


FIGURE 8. Visualization of the intersubject registration of T2-weighted MR images. The sagittal slices are in the same order as in Fig. 5. Our proposed method can realizes the largest hippocampus overlaps (shown in the last column).

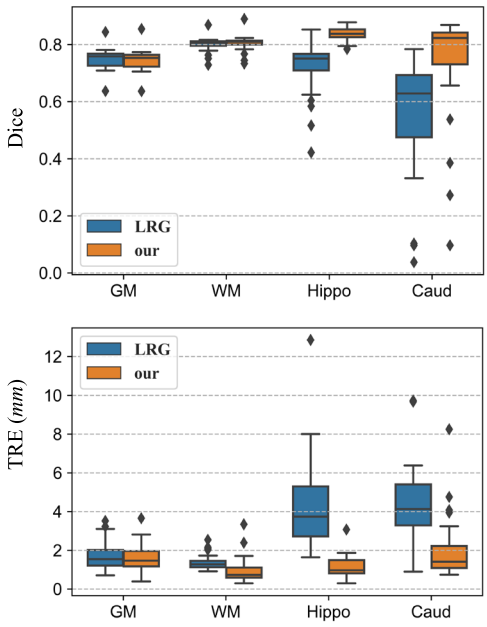


FIGURE 9. Quantitative results for four labels.

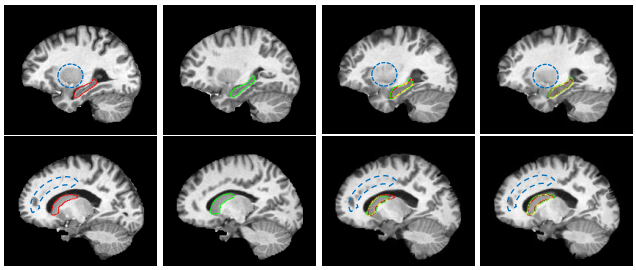


FIGURE 10. Visualization of the intersubject registration of T1-weighted MR images. The first row corresponds to the 60th sagittal slice, and the second row corresponds to the 103th sagittal slice. The slices of four columns, from left to right, are from a fixed image, a moving image, and the warped moving images that are obtained by the LRG method and our method. Our proposed method can realize the largest hippocampus overlaps (as shown in the top-right slice) and the largest caudate overlaps (as shown in the bottom-right slice).

method, except the Dice ratio on GM, which is slightly lower. For the local regions, namely, the hippocampus and caudate, the registration accuracy of our method is much better than that of the LRG method. For example, the Dice ratios of our method are higher than those of the LRG method, and the TREs of our method are lower than those of the LRG method.

Qualitative results of intersubject registration of T1-weighted MR images are presented in Fig. 10. Not only for the hippocampus tissue in the top row but also for the caudate region in the bottom row, our method realizes larger contour overlaps than the LRG method. In addition, according to the inner dot blue line region, slices of the warped image that is obtained using our method (4th column) are more similar to those of the fixed image (first column) than those of the warped image that was obtained via the LRG method (third column).

From the experiments of infant brain and OASIS-1 dataset, we can draw the following advantages using our method. Both LRG and our method provide highly satisfactory results for the registration of the global tissues, namely, WM and GM. However, in the local tissue, namely, the hippocampus or the caudate, the LRG method does not yield the same satisfactory registration results as in the global regions; hence, the registration of the global regions does not guarantee that the local region can also be aligned. As our method uses

both global labels and local labels, it provides satisfactory registration results in global and local tissue regions. Thus, our method can overcome the *one-to-many correspondence* problem between local regions and global regions. From the results of registration between large-time-gap images, e.g., from 2-week-old and 3-month-old to 12-month-old MR images, although there are substantial *anatomical and appearance variations* between those time intervals, our method still yields satisfactory registration results; hence, our method can overcome the problem of *anatomical and appearance variations*.

Our method also has several shortcomings: First, the computational burden is increased in the training stage since local labels are also used. Second, if local tissues are too small and too far away, then they cannot be found in the same position patches. Hence, it is highly difficult to align them. Third, the deformation field should be smoother around the local tissues that are being learned. Global deformation and local deformation should be balanced.

IV. CONCLUSION

In this work, we have proposed a novel patch-based deformable registration method for infant brain MR images, which combines global and local label information for training the DDF via a UNet-style [45] regression network without any ground-truth deformation. Then, during inference, the total DDF between two new longitudinal infant brain MR images is calculated by averaging the network outputs for denser pairwise patches. Our registration scheme can overcome the problems of large *anatomical and appearance variations* and *one-to-many correspondences* between global tissues and local tissues. Experimental results are reported for intrasubject and intersubject registration of T1- or T2-weighted MR images of infant brains at various time points, and they demonstrate promising registration performance not only for global tissues but also for local tissues. An additional experiment on the OASIS-1 dataset results in the same conclusions. The registration performance of our method on local tissues is satisfactory and is independent of the subjects, modalities, and time points.

ACKNOWLEDGMENT

We thank for the open source code of Label-reg published by Y. Hu et al. [46]. Data I were provided by Iseg Challenge and UNC. Data II were provided in part by OASIS: Longitudinal: Principal Investigators: D. Marcus, R, Buckner, J. Csernansky, J. Morris; P50 AG05681, P01 AG03991, P01 AG026276, R01 AG021910, P20 MH071616, U24 RR021382.

REFERENCES

[1] 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.

[2] L. Wang, Y. Gao, F. Shi, G. Li, J. H. Gilmore, W. Lin, and D. Shen, "LINKS: Learning-based multi-source integration framework for segmentation of infant brain images," *NeuroImage*, vol. 108, pp. 160–172, Mar. 2015.

[3] S. Hu, L. Wei, Y. Gao, Y. Guo, G. Wu, and D. Shen, "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.

[4] L. Wei, X. Cao, Z. Wang, Y. Gao, S. Hu, L. Wang, G. Wu, and D. Shen, "Learning-based deformable registration for infant MRI by integrating random forest with auto-context model," *Med. Phys.*, vol. 44, no. 12, pp. 6289–6303, Dec. 2017.

[5] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, "Multimodality image registration by maximization of mutual information," *IEEE Trans. Med. Imag.*, vol. 16, no. 2, pp. 187–198, Apr. 1997.

[6] W. M. Wells, III, P. Viola, H. Atsumi, S. Nakajima, and R. Kikinis, "Multi-modal volume registration by maximization of mutual information," *Med. Image Anal.*, vol. 1, no. 1, pp. 35–51, Jan. 1996.

[7] A. Roche, G. Malandain, X. Pennec, and N. Ayache, "The correlation ratio as a new similarity measure for multimodal image registration," in *Medical Image Computing and Computer-Assisted Intervention-MICCAI*, vol. 1496. Berlin, Germany: Springer, 1998, pp. 1115–1124.

[8] 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.

[9] Y. Wu, G. Wu, L. Wang, B. C. Munsell, Q. Wang, W. Lin, Q. Feng, W. Chen, and D. Shen, "Hierarchical and symmetric infant image registration by robust longitudinal-example-guided correspondence detection," *Med. Phys.*, vol. 42, no. 7, pp. 4174–4189, Jul. 2015.

[10] 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.

[11] X. Cao, J. Yang, J. Zhang, D. Nie, M. Kim, Q. Wang, and D. Shen, "Deformable image registration based on similarity-steered CNN regression," in *Medical Image Computing and Computer Assisted Intervention-MICCAI*, vol. 10433. Berlin, Germany: Springer, 2017, pp. 300–308.

[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] G. Balakrishnan, A. Zhao, M. R. Sabuncu, J. Guttag, and A. V. Dalca, "An unsupervised learning model for deformable medical image registration," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Apr. 2018, pp. 9252–9260.

[14] B. D. de Vos, F. F. Berendsen, M. A. Viergever, M. Staring, and I. Išgum, "End-to-end unsupervised deformable image registration with a convolutional neural network," Apr. 2017, *arXiv:1704.06065*. [Online]. Available: https://arxiv.org/abs/1704.06065

[15] J. Krebs, H. Delingette, B. Mailhé, N. Ayache, and T. Mansi, "Learning a probabilistic model for diffeomorphic registration," *IEEE Trans. Med. Imag.*, vol. 38, no. 9, pp. 2165–2176, Sep. 2019, doi: 10.1109/TMI.2019.2897112.

[16] C. Qin, B. Shi, R. Liao, T. Mansi, D. Rueckert, and A. Kamen, "Unsupervised deformable registration for multi-modal images via disentangled representations," in *Proc. Int. Conf. Inf. Process. Med. Imag. (IPMI)*, vol. 11492. Berlin, Germany: Springer, 2019, pp. 249–261.

[17] J. Zhang, "Inverse-consistent deep networks for unsupervised deformable image registration," Sep. 2018, *arXiv:1809.03443*. [Online]. Available: https://arxiv.org/abs/1809.03443

[18] S. Guan, C. Meng, Y. Xie, Q. Wang, K. Sun, and T. Wang, "Deformable cardiovascular image registration via multi-channel convolutional neural network," *IEEE Access*, vol. 7, pp. 17524–17534, 2019.

[19] Z. Yang, T. Dan, and Y. Yang, "Multi-temporal remote sensing image registration using deep convolutional features," *IEEE Access*, vol. 6, pp. 38544–38555, 2018.

[20] Y. Hu, M. Modat, E. Gibson, L. Wenqi, N. Ghavami, E. Bonmati, W. Guotai, S. Bandula, C. M. Moore, M. Emberton, S. Ourselin, J. A. Noble, D. C. Barratt, and T. Vercauteren, "Weakly-supervised convolutional neural networks for multimodal image registration," *Med. Image Anal.*, vol. 49, pp. 1–13, Oct. 2018.

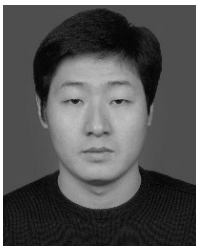
[21] Y. Hu, M. Modat, E. Gibson, N. Ghavami, E. Bonmati, C. M. Moore, M. Emberton, J. A. Noble, D. C. Barratt, and T. Vercauteren, "Label-driven weakly-supervised learning for multimodal deformable image registration," in *Proc. IEEE 15th Int. Symp. Biomed. Imag. (ISBI)*. Washington, DC, USA: IEEE, Apr. 2018, pp. 1070–1074.

[22] G. Haskins, U. Kruger, and P. Yan, "Deep learning in medical image registration: A survey," Mar. 2019, *arXiv:1903.02026*. [Online]. Available: https://arxiv.org/abs/1903.02026

[23] M. A. Viergever, J. B. A. Maintz, S. Klein, K. Murphy, M. Staring, and J. P. W. Pluim, "A survey of medical image registration–under review," *Med. Image Anal.*, vol. 33, pp. 140–144, Oct. 2016.
[24] D. Shen, G. Wu, and H. Suk, "Deep learning in medical image analysis," *Annu. Rev. Biomed. Eng.*, vol. 19, pp. 221–248, Jun. 2017.
[25] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. W. M. van der Laak, B. van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, Dec. 2017.
[26] J. Ker, L. Wang, J. Rao, and T. Lim, "Deep learning applications in medical image analysis," *IEEE Access*, vol. 6, pp. 9375–9389, 2017.
[27] A. V. Dalca, G. Balakrishnan, J. Guttag, and M. R. Sabuncu, "Unsupervised learning for fast probabilistic diffeomorphic registration," in *Medical Image Computing and Computer Assisted Intervention-MICCAI*, vol. 11070. Berlin, Germany: Springer, 2018, pp. 729–738.
[28] A. V. Dalca, G. Balakrishnan, J. Guttag, and M. R. Sabuncu, "Unsupervised learning of probabilistic diffeomorphic registration for images and surfaces," Mar. 2019, *arXiv:1903.03545*. [Online]. Available: https://arxiv.org/abs/1903.03545
[29] A. V. Dalca, E. Yu, P. Golland, B. Fischl, M. R. Sabuncu, and J. E. Iglesias, "Unsupervised deep learning for Bayesian brain MRI segmentation," Apr. 2019, *arXiv:1904.11319*. [Online]. Available: https://arxiv.org/abs/1904.11319
[30] J. Fan, X. Cao, Z. Xue, P.-T. Yap, and D. Shen, "Adversarial similarity network for evaluating image alignment in deep learning based registration," in *Medical Image Computing and Computer Assisted Intervention-MICCAI*, vol. 11070. Berlin, Germany: Springer, 2018, pp. 739–746.
[31] X. Yi, E. Walia, and P. Babyn, "Generative adversarial network in medical imaging: A review," Sep. 2018, *arXiv:1809.07294*. [Online]. Available: https://arxiv.org/abs/1809.07294
[32] C. Tanner, F. Ozdemir, R. Profanter, V. Vishnevsky, E. Konukoglu, and O. Goksel, "Generative adversarial networks for mr-ct deformable image registration," Jul. 2018, *arXiv:1807.07349*. [Online]. Available: https://arxiv.org/abs/1807.07349
[33] S. Hu, L. Zhang, G. Li, M. Liu, D. Fu, and W. Zhang, "Infant brain deformable registration using global and local label-driven deep regression learning," in *Proc. Int. Workshop Mach. Learn. Med. Imag. (MICCAI)*, vol. 11861. Berlin, Germany: Springer, 2019, pp. 106–114.
[34] 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.
[35] M. Jenkinson, C. F. Beckmann, T. E. Behrens, M. W. Woolrich, and S. M. Smith, "FSL," *NeuroImage*, vol. 62, no. 2, pp. 782–790, 2012.
[36] M. W. Woolrich, S. Jbabdi, B. Patenaude, M. Chappell, S. Makni, T. Behrens, C. Beckmann, M. Jenkinson, and S. M. Smith, "Bayesian analysis of neuroimaging data in FSL," *NeuroImage*, vol. 45, pp. S173–S186, Mar. 2009.
[37] Y. Dai, F. Shi, L. Wang, G. Wu, and D. Shen, "iBEAT: A toolbox for infant brain magnetic resonance image processing," *Neuroinformatics*, vol. 11, no. 2, pp. 211–225, Apr. 2013.
[38] E. Gibson, W. Li, C. Sudre, L. Fidon, D. I. Shakir, G. Wang, Z. Eaton-Rosen, R. Gray, T. Doel, Y. Hu, T. Whyntie, P. Nachev, M. Modat, D. C. Barratt, S. Ourselin, M. J. Cardoso, and T. Vercauteren, "NiftyNet: A deep-learning platform for medical imaging," *Comput. Methods Programs Biomed.*, vol. 158, pp. 113–122, May 2018.
[39] *NiftyNet*. Accessed: Nov. 9, 2019. [Online]. Available: https://github.com/NifTK/NiftyNet
[40] T. Vercauteren, X. Pennec, A. Perchant, and N. Ayache, "Diffeomorphic demons: Efficient non-parametric image registration," *Neuroimage*, vol. 45, pp. S61–S72, Mar. 2009.
[41] *Demons*. Accessed: Nov. 6, 2019. [Online]. Available: https://itk.org/SimpleITKDoxygen/html/DemonsRegistration1_2DemonsRegistrtion1_8py-example.html
[42] J.-P. Thirion, "Image matching as a diffusion process: An analogy with Maxwell's demons," *Med. Image Anal.*, vol. 2, no. 3, pp. 243–260, 1998.
[43] *Mutual Information*. Accessed: Nov. 6, 2019. [Online]. Available: https://itk.org/SimpleITKDoxygen/html/ImageRegistrationMethodBSpline2_2ImageRegistrationMethodBSpline2_8py-example.html
[44] B. C. Lowekamp, D. T. Chen, L. I. Ibanez, and D. Blezek, "The design of simpleiTK," *Frontiers Neuroinform.*, vol. 7, p. 45, Dec. 2013.
[45] O. Ronneberger, P. Fischer, T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention-MICCAI*, vol. 9351. Berlin, Germany: Springer, 2015, pp. 234–241.
[46] *Label-Reg*. Accessed: Nov. 6, 2019. [Online]. Available: https://github.com/YipengHu/label-reg
[47] D. S. Marcus, A. F. Fotenos, J. G. Csernansky, J. C. Morris, and R. L. Buckner, "Open access series of imaging studies: Longitudinal MRI data in nondemented and demented older adults," *J. Cogn. Neurosci.*, vol. 22, no. 12, pp. 2677–2684, 2010.
[48] 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, 2008.
[49] B. Li, W. J. Niessen, S. Klein, M. de Groot, M. A. Ikram, M. W. Vernooij, and E. E. Bron, "A hybrid deep learning framework for integrated segmentation and registration: Evaluation on longitudinal white matter tract changes," in *Medical Image Computing and Computer Assisted Intervention-MICCAI*, vol. 11766. Berlin, Germany: Springer, 2019, pp. 645–653.
[50] T. Estienne, M. Vakalopoulou, S. Christodoulidis, E. Battistela, M. Lerousseau, A. Carre, G. Klausner, R. Sun, C. Robert, S. Mougiakakou, N. Paragios, and E. Deutsch, "U-ReSNet: Ultimate coupling of registration and segmentation with deep nets," in *Medical Image Computing and Computer Assisted Intervention-MICCAI*, vol. 11766. Berlin, Germany: Springer, 2019, pp. 310–319.
[51] Z. Xu and M. Niethammer, "DeepAtlas: Joint semi-supervised learning of image registration and segmentation," in *Medical Image Computing and Computer Assisted Intervention-MICCAI*, vol. 11765. Berlin, Germany: Springer, 2019, pp. 420–429.
[52] N. J. Tustison, B. B. Avants, P. A. Cook, Y. Zheng, A. Egan, P. A. Yushkevich, and J. C. Gee, "N4ITK: Improved N3 bias correction," *IEEE Trans. Med. Imag.*, vol. 29, no. 6, pp. 1310–1320, Jun. 2010.

SHUNBO HU received the M.S. degree from Shandong University, Jinan, in 2003, and the Ph.D. degree from Shandong University, Jinan, in 2008. He is currently a Professor with the Department of Information Science and Engineering, Linyi University, China. He has received two grants from the National Natural Science Foundation of China. He has published more than 50 articles. His current research interests include medical image registration, image segmentation, brain science, machine learning, and deep learning.

LINTAO ZHANG received the Ph.D. degree from the Institute of Electrics, Chinese Academy of Sciences, Beijing, in 2011. From 2011 to 2016, he was an Engineer at ZTE Corporation and has researched on real-time video coding, video image processing, and algorithmic optimization. He is currently a Lecturer with the Department of Information Science and Engineering, Linyi University, China. His current research interests include medical image registration, image segmentation, machine learning, and deep learning.



GUOQIANG LI received the M.S. degree from Sun Yat-sen University, Guangzhou, in 2007, and the Ph.D. degree from the Beijing Institute of Technology, Beijing, in 2014. He is currently an Associate Professor with the Department of Information Science and Engineering, Linyi University, China. He has received one grant from the Natural Science Foundation of Shandong Province, China. He has published more than ten articles. His current research interests include medical image registration, machine learning, and deep learning.

MINGTAO LIU received the M.S. degree from the Ocean University of China, Qingdao, in 2009, and the Ph.D. degree from the Ocean University of China, Qingdao, in 2016. He is currently an Instructor with the Department of Information Science and Engineering, Linyi University, China. His current research interests include medical image registration and segmentation, machine learning, and deep learning.

DEQIAN FU received the M.S. degree from the Shandong University of Science and Technology (SDUST), China, in 2005, and the Ph.D. degree from The University of Suwon (USW), South Korea, in 2014. He is currently an Associate Professor with the Department of Information Science and Engineering, Linyi University, China. He has published more than 20 academic articles. His current research interests include medical image processing, machine learning, deep learning, and big data.

WENYIN ZHANG received the M.S. degree from the Shandong University of Science and Technology, in 2002, and the Ph.D. degree from the Chengdu Institute of Computing Technology, Chinese Academy of Science, in 2005. He is currently a Professor with the School of Information Sciences and Technology, Linyi University, China. He has published more than 50 articles. His current research interests include image processing, pattern recognition, steganography, information security, and block chain.