SRS MapCHECK®
SUN NUCLEAR corporation
MODEL #1179
PATENT PENDING
REF
PDI
STATUS

Joint affine and deformable three-dimensional networks for brain MRI registration

Zhenyu Zhu, Yiqin Cao, Chenchen Qin and Yi Rao
National-Regional Key Technology Engineering Laboratory for Medical Ultrasound, Guangdong Key Laboratory for Biomedical Measurements and Ultrasound Imaging, School of Biomedical Engineering, Health Science Center, Shenzhen University, Shenzhen, China

Di Lin
The College of Intelligence and Computing, Tianjin University, Tianjin, China

Qi Dou
Computer Science and Engineering, The Chinese University of Hong Kong, Hong Kong, China

Dong Ni and Yi Wang[a)]
National-Regional Key Technology Engineering Laboratory for Medical Ultrasound, Guangdong Key Laboratory for Biomedical Measurements and Ultrasound Imaging, School of Biomedical Engineering, Health Science Center, Shenzhen University, Shenzhen, China

(Received 30 June 2020; revised 11 December 2020; accepted for publication 11 December 2020; published xx xxxx xxxx)

Purpose: Volumetric medical image registration has important clinical significance. Traditional registration methods may be time-consuming when processing large volumetric data due to their iterative optimizations. In contrast, existing deep learning-based networks can obtain the registration quickly. However, most of them require independent rigid alignment before deformable registration; these two steps are often performed separately and cannot be end-to-end.
Methods: We propose an end-to-end joint affine and deformable network for three-dimensional (3D) medical image registration. The proposed network combines two deformation methods; the first one is for obtaining affine alignment and the second one is a deformable subnetwork for achieving the nonrigid registration. The parameters of the two subnetworks are shared. The global and local similarity measures are used as loss functions for the two subnetworks, respectively. Moreover, an anatomical similarity loss is devised to weakly supervise the training of the whole registration network. Finally, the trained network can perform deformable registration in one forward pass.
Results: The efficacy of our network was extensively evaluated on three public brain MRI datasets including Mindboggle101, LPBA40, and IXI. Experimental results demonstrate our network consistently outperformed several state-of-the-art methods with respect to the metrics of Dice index (DSC), Hausdorff distance (HD), and average symmetric surface distance (ASSD).
Conclusions: The proposed network provides accurate and robust volumetric registration without any pre-alignment requirement, which facilitates the end-to-end deformable registration. © *2020 American Association of Physicists in Medicine* [https://doi.org/10.1002/mp.14674]

Key words: affine alignment, brain MRI, convolutional neural networks, deformable image registration, end-to-end registration

1. INTRODUCTION

Registration plays an important role in the field of medical image computing to establish the pixel-wise correspondences between different images.[1] By doing so, mono-/multi-modality information can be fused into the same coordinate system, which provides more convenient and reliable guidance for doctors to make diagnosis, treatment plan, and follow-ups. Many algorithms[1–3] have been proposed over the past few decades. However, registration is still a challenging task. Traditional registration methods may be computationally expensive and time-consuming due to their iterative optimizations. Furthermore, most nonrigid registration methods require an independent rigid alignment before conducting the deformable registration[4]; these two steps are often performed separately thus cannot be jointly optimized. Therefore, efficient and accurate medical image registration is still our research objective.

As illustrated in Fig. 1, the aim of registration is to match all corresponding anatomical points in two images to the same coordinate system through plausible spatial transformation. The output of registration should insure all anatomical structures, or at least all of the diagnostic/surgical sites on both images are matched together. Let F and M denote a fixed and a moving image, respectively. The goal of registration is to predict the optimal deformation W that optimizes the energy function: $\mathscr{S}(F, M \circ W) + \mathscr{R}(W)$, where $\mathscr{S}$ defines similarity criterion and $\mathscr{R}$ regularizes the deformation to match any specific properties in the solution. To this end, several nonlinear deformation algorithms[5] have been proposed,

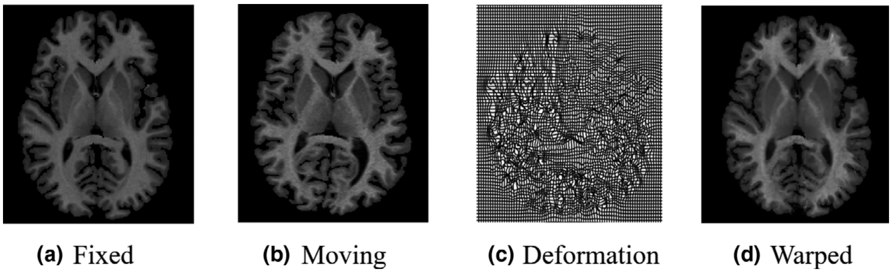

FIG. 1. Illustration of image registration. Given a fixed image (a) and a moving image (b), a deformation field (c) is predicted to warp the moving image so that warped image (d) and fixed image (a) are registered.

such as large diffeomorphic distance metric mapping (LDDMM),[6] standard symmetric normalization (SyN),[7] and Demons.[8] Most registration algorithms require both a measure of the similarity between moving and fixed images (such as sum of squared differences (SSD),[9] cross-correlation (CC),[10] normalized cross-correlation (NCC),[11] mutual information (MI),[12] and normalized mutual information (NMI),[13] etc.), and an optimization strategy to maximize the similarity between images. Therefore, these algorithms require a lot of iterations thus may take a relatively long time to deal with large data.

In recent years, deep learning models have been widely studied for the registration task.[14,15] Fan et al.[16] applied a supervised deep learning approach for image registration by using obtained ground-truth deformation fields as the supervision information of the network. Uzunova et al.[17] generated highly expressed models from very few training samples. Specifically, Uzunova et al. synthesized a large amount of realistic ground-truth data using model-based approach for the training of registration network. The problem with aforementioned supervised learning methods is that the quality of the registration relies heavily on the ground-truth data; however, unlike detection or segmentation tasks, it is always difficult to acquire registration ground-truth. In contrast, the common practice of unsupervised registration network is to generate the desired deformation field through similarity evaluation between images, and then obtain a warped image through spatial transformer networks (STN).[18] Rohé et al.[19] employed the U-net architecture[20] to predict the deformation field of the three-dimensional (3D) cardiac MR images and used the SSD as the similarity loss function. Li et al.[21] developed a learning-based method to predict deformation parameters using a fully convolutional network, but this is a two-dimensional (2D) network that tends to ignore the overall 3D information. Balakrishnan et al.[22] proposed an unsupervised 3D network with cross-correlation as its loss function. However, the prerequisite of this network was another rigid alignment. Cao et al.[23] proposed a convolutional neural network (CNN)-based regression model to directly learn the complex mapping from the input image pair but the input of this network was patch-wise data instead of the whole image. Duan et al.[24] proposed a multiscale framework to obtain the deformation field and added a discriminator to determine whether the registration was well enough. Hu et al.[25] used semi-supervised label similarity to train their network to perform MR-transrectal ultrasound registration. The weakly supervised network could provide more reliable registration but still requires a small amount of manual annotations.

In this study, we propose an end-to-end joint affine and deformable network for brain magnetic resonance image (MRI) registration. The proposed network combines affine alignment and deformable registration. Given two images to be registered, the affine alignment subnetwork is used to predict the affine transformation, and the deformable subnetwork is employed to conduct nonrigid registration. Two subnetworks are cascaded and they share network parameters to maximize network performance while reducing parameters. The whole network is trained using a weakly supervised manner by calculating the global and local similarities of image pairs, and also an anatomical similarity measure. Finally, the trained network can perform deformable registration in one forward pass. Experimental results show that our network can realize volume registration effectively without any pre-alignment requirement, which facilitates the end-to-end deformable registration.

Some of the preliminary results were previously published in an EMBC 2020 paper.[26] This article has considerable difference compared with the conference paper,[26] which consists of: (a) A new anatomical similarity loss. We devise this new loss to evaluate the structure similarity and to weakly supervise the training of the registration network. By simultaneously calculating the intensity and structure based similarities, the network performs more accurate registration. (b) Comparison of time efficiency. We show the comparison of inference and training time between our method and the state-of-the-art methods. (c) New experiments. We compare one more cutting-edge registration network.[27] We conduct ablation study to validate the contribution of the devised affine component and anatomical similarity loss. We show more detailed comparison results.

The rest of this article is described as follows. Section 2 introduces the specific details of the proposed registration network. Section 3 shows the registration performance of our network and several compared methods. Sections 4 and 5 present the discussion and conclusion of this study, respectively.

2. MATERIALS AND METHODS

The proposed registration network is illustrated in Fig. 2. We denote the input volumetric image pair of the registration network as a fixed image (F) and a moving image (M). The proposed network connects two subnetworks in a cascade manner to realize the deformable registration of F and M. It first learns the affine deformation, which stands for rotation, translation, scaling and shearing transformations. According to the affine alignment, the deformable subnetwork generates the 3D nonrigid deformation filed, which is then used by a spatial transformer to warp the moving image for the final registration.

2.A. Affine alignment subnetwork

The affine alignment subnetwork aims to align M with F by only considering the affine deformation. As shown in the green dash box of Fig. 2, let F and M be the input of this subnetwork, we model the affine alignment function $f_\Theta(F,M)=u$ using a CNN model, where u is an affine matrix containing 12 degrees of freedom. The CNN consists of four convolutional blocks and three downsampling blocks, a pooling layer and a linear layer. In our implementation, each convolutional block consists of a convolution layer with kernel size of 3 × 3 × 3 and stride of 1, a Leaky Relu activation layer,[28] and a batch normalization layer[29]; the number of filters in each convolutional layer are (16, 32, 32, 32). Each downsampling block consists of a convolution layer with kernel size of 3 × 3 × 3 and stride of 2. The subnetwork then generates an affine matrix through a global average pooling and a linear layer and finally outputs an affine-aligned image (A) volume after re-sampling. The affine alignment subnetwork is same with the downsampling part of the subsequent deformable subnetwork, which we design by means of parameter sharing. The detailed structure of the downsampling part is shown in the dotted black box in Fig. 3. The purpose is not only to reduce the parameters of the network but also to make the characteristics extracted by the two subnetworks consistently and robustly. In our implementation, we employ the cross-correlation of two images as the loss function of this affine alignment subnetwork. Taking into account the overall characteristics of the affine alignment, we adopt a global cross-correlation[10] similarity:

$$\mathcal{L}_{aff} = -\frac{\left(\sum_{p_i}(F(p_i)-\bar{F}(p_i))(A(p_i)-\bar{A}(p_i))\right)^2}{\left(\sum_{p_i}(F(p_i)-\bar{F}(p_i))\right)^2\left(\sum_{p_i}(A(p_i)-\bar{A}(p_i))\right)^2}, \quad (1)$$

where $A = u(M)$, F and A are fixed image and affine aligned image, respectively; $\bar{F}$ and $\bar{A}$ denote images with local mean intensities subtracted; and $p_i \in \Omega$ denotes each voxel in images, where Ω is the whole image domain.

2.B. Deformable subnetwork

The fixed and moving images are roughly aligned by affine subnetwork. In order to improve the registration accuracy, the deformable subnetwork further provides the nonrigid registration. The orange dash box in Fig. 2 presents an overview of the deformable subnetwork. The input of this subnetwork is F and affine aligned image A. We model another function $g_\Theta(F,A)=\phi$, where ϕ is a nonrigid deformation field. The backbone architecture of this deformable subnetwork is based on U-Net,[20] which consists of an encoder-decoder with skip connections. This subnetwork generates the nonrigid deformation field associated with the whole volume. As shown in Fig. 3, this subnetwork contains three downsampling layers and three upsampling layers with intervening residual modules. In

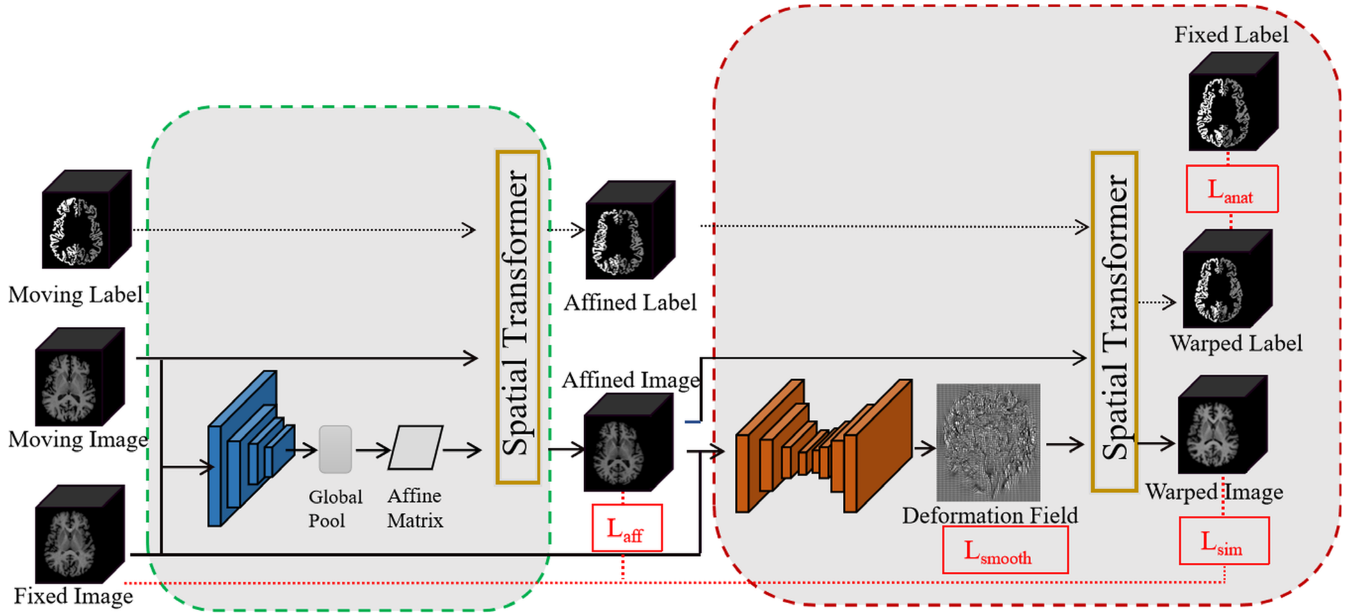

Fig. 2. The proposed joint affine and deformable 3D network for nonrigid image registration. The dashed lines indicate data flow only required in training phase.

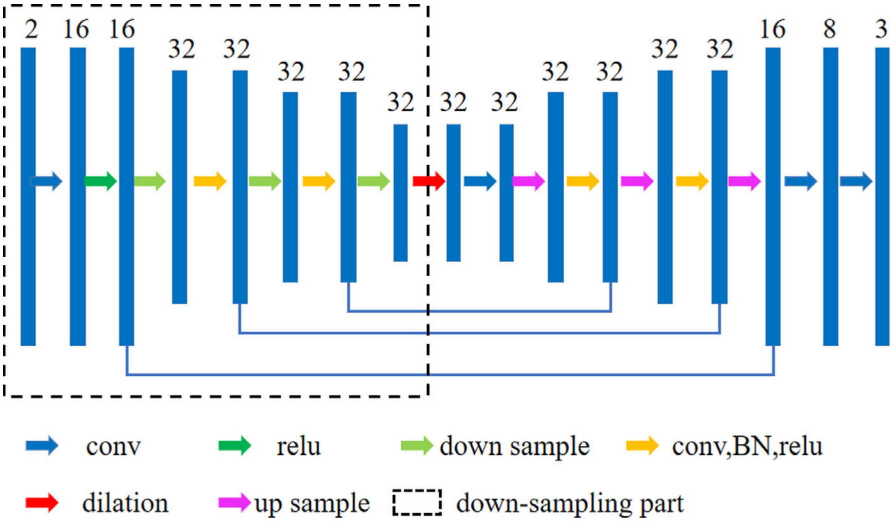

FIG. 3. The detailed design of the deformable subnetwork. The number of channels is shown above the rectangle and the initial channel number is 2.

TABLE I. The DSC (%), Hausdorff distance (HD), and average symmetric surface distance (ASSD) results (mean ± SD) of different affine alignment methods

	Methods	Frontal	Parital	Occipital	Temporal	Cingulate
DSC (%)	Before registration	16.3 ± 8.1	20.9 ± 6.3	16.6 ± 7.6	21.6 ± 6.6	15.6 ± 9.9
	ANTS	33.3 ± 4.3	29.9 ± 3.6	26.4 ± 4.8	33.4 ± 4.5	36.6 ± 6.6
	Freesurfer	32.9 ± 2.8	29.6 ± 3.0	27.9 ± 3.5	34.3 ± 3.3	34.5 ± 4.9
	Our affine alignment	**33.5** ± **4.0**	**31.4** ± **3.9**	**31.0** ± **3.1**	**36.0** ± **3.1**	**37.4** ± **7.5**
HD	Before registration	18.3 ± 4.9	18.5 ± 5.3	19.5 ± 4.5	10.6 ± 3.5	8.24 ± 0.9
	ANTS	14.6 ± 2.5	14.4 ± 3.0	15.9 ± 4.5	8.33 ± 2.9	9.12 ± 2.2
	Freesurfer	14.5 ± 4.1	**14.0** ± **4.2**	19.1 ± 5.9	10.0 ± 2.4	8.22 ± 1.2
	Our affine alignment	**14.3** ± **4.1**	14.6 ± 3.2	**14.5** ± **4.0**	**8.30** ± **2.3**	**8.20** ± **2.0**
ASSD	Before registration	4.28 ± 2.66	2.82 ± 1.09	3.99 ± 2.3	2.84 ± 1.03	1.99 ± 0.49
	ANTS	1.81 ± 0.27	2.28 ± 0.31	2.34 ± 0.99	1.82 ± 0.61	2.15 ± 0.83
	Freesurfer	**2.14** ± **0.32**	2.14 ± 0.36	2.13 ± 0.42	**1.71** ± **0.46**	1.57 ± 0.31
	Our affine alignment	2.46 ± 0.91	**1.95** ± **0.50**	**1.64** ± **0.32**	1.73 ± 0.61	**1.45** ± **0.60**

The best results are shown in bold.

the encoding path, the four convolutional blocks and three downsampling blocks are set alternatively to gradually reduce the spatial dimension, so that the network can learn abundant features from different scales. We add a dilated convolutional block at the last layer of the encoder to expand the receptive field. The dilated convolutional block consists of four convolutional subblocks, each with the dilation rate of 1, 2, 4, and 8, respectively; each subblock is composed of a convolution layer with kernel size of 3 × 3 × 3 and stride of 1, a Relu activation layer,[28] and a group normalization layer.[29] We warp A by using a spatial transformer networks (STN), and then evaluate the similarity between warped $\phi(A)$ and F. Considering that the deformable registration focuses on local similarities, we employ the patch-based cross-correlation as loss function:

$$\mathscr{L}_{sim} = -\sum_{p\varepsilon\Omega} \frac{\left(\sum_{p_i}(F(p_i)-\bar{F}(p))(W(p_i)-\bar{W}(p))\right)^2}{\left(\sum_{p_i}F(p_i)-\bar{F}(p)\right)^2\left(\sum_{p_i}(W(p_i)-\bar{W}(p))\right)^2}, \tag{2}$$

where $W = \phi(A)$, F and W are fixed image and registered image, respectively; $\bar{F}$ and $\bar{W}$ denote images with local mean intensities subtracted. Voxel p_i is the local neighborhood in n^3 ($n = 9$ in our implementation) volumetric patch at the center of voxel p.

Conventional intensity-based similarity measures only evaluate intensity-based features, but do not take structure similarity into consideration. Considering the aim of registration is to match all corresponding anatomical structures from two images, in this study, an anatomical similarity loss is devised to weakly supervise the training of the whole registration network. The anatomical similarity loss is defined by the Dice index (DSC) between the segmentation masks of the fixed image and the warped moving image, and can be calculated as follows:

$$\mathscr{L}_{anat} = 1 - \frac{2|S_F \cap \phi(u(S_M))|}{|S_F| + |\phi(u(S_M))|}, \tag{3}$$

where S_F and S_M are the segmentation masks of the fixed and moving images, respectively; $\phi(u(S_M))$ is the registered moving image. Note that the segmentation masks are only used in

TABLE II. The DSC (%), Hausdorff distance (HD), and average symmetric surface distance (ASSD) results (mean ± SD) on MindBoggle101 dataset

	Methods	Frontal	Parital	Occipital	Temporal	Cingulate
DSC (%)	Before	16.3 ± 8.1	20.9 ± 6.3	16.6 ± 7.6	21.6 ± 6.6	15.6 ± 9.9
	SyN[33]	54.4 ± 4.5	46.8 ± 6.8	49.8 ± 5.0	48.1 ± 6.2	49.2 ± 9.0
	VM[22]	53.4 ± 8.1	52.7 ± 6.2	51.0 ± 7.3	43.3 ± 7.6	48.3 ± 9.2
	FAIM[34]	57.2 ± 6.8	55.1 ± 7.1	53.7 ± 6.5	46.9 ± 6.9	50.8 ± 9.5
	Multi-FC[24]	58.3 ± 7.1	50.4 ± 7.3	55.4 ± 6.9	47.7 ± 7.9	54.3 ± 8.9
	SYMNet[27]	63.3 ± 5.2	55.2 ± 7.4	59.3 ± 6.3	49.0 ± 6.3	59.7 ± 8.5
	Ours *w.o.* affine	57.2 ± 6.2	54.2 ± 7.1	60.3 ± 6.1	46.0 ± 7.2	45.7 ± 8.2
	Ours *w.o.* $\mathscr{L}_{anat}$	60.5 ± 6.0	54.2 ± 6.3	58.9 ± 6.6	50.0 ± 8.0	58.6 ± 8.9
	Ours	**63.7 ± 7.3**	**56.8 ± 6.6**	**62.6 ± 6.7**	**53.3 ± 7.8**	**60.2 ± 7.7**
HD	Before	18.3 ± 4.9	18.5 ± 5.3	19.5 ± 4.5	10.6 ± 3.5	8.2 ± 0.9
	SyN[33]	**11.2 ± 2.2**	**3.0 ± 1.9**	14.8 ± 3.8	**6.8 ± 2.3**	7.9 ± 2.0
	VM[22]	13.2 ± 3.6	13.0 ± 1.8	13.5 ± 4.6	8.6 ± 3.0	7.9 ± 1.9
	FAIM[34]	12.9 ± 2.7	13.3 ± 1.7	13.3 ± 4.4	8.7 ± 2.3	7.8 ± 2.1
	Multi-FC[24]	12.6 ± 2.9	13.2 ± 2.0	13.7 ± 4.4	8.0 ± 2.4	8.0 ± 2.2
	SYMNet[27]	11.5 ± 4.2	13.4 ± 3.4	14.3 ± 2.1	8.1 ± 3.4	8.1 ± 2.4
	Ours *w.o.* affine	15.5 ± 5.8	14.4 ± 5.7	17.3 ± 5.1	9.8 ± 3.4	8.2 ± 1.4
	Ours *w.o.* $\mathscr{L}_{anat}$	14.4 ± 3.5	13.7 ± 2.6	13.8 ± 3.8	8.7 ± 2.9	8.7 ± 2.2
	Ours	13.8 ± 3.1	13.2 ± 2.5	**12.9 ± 3.6**	7.9 ± 2.3	**7.8 ± 2.0**
ASSD	Before	4.28 ± 2.66	2.82 ± 1.09	3.99 ± 2.30	2.84 ± 1.03	1.99 ± 0.49
	SyN[33]	1.32 ± 0.53	1.45 ± 0.32	1.48 ± 0.35	1.11 ± 0.27	1.37 ± 0.66
	VM[22]	1.51 ± 0.33	1.32 ± 0.33	1.36 ± 0.32	1.32 ± 0.72	1.43 ± 0.43
	FAIM[34]	1.42 ± 0.32	1.29 ± 0.28	1.44 ± 0.66	1.12 ± 0.46	1.20 ± 0.58
	Multi-FC[24]	1.34 ± 0.28	1.29 ± 0.23	1.39 ± 0.53	1.06 ± 0.53	1.20 ± 0.63
	SYMNet[27]	1.90 ± 0.74	1.17 ± 0.31	1.69 ± 0.41	1.02 ± 0.38	1.11 ± 0.33
	Ours *w.o.* affine	1.94 ± 0.78	1.40 ± 0.36	2.69 ± 0.51	1.66 ± 0.34	1.14 ± 0.23
	Ours *w.o.* $\mathscr{L}_{anat}$	1.96 ± 1.39	1.37 ± 0.51	1.34 ± 0.76	1.39 ± 0.79	1.33 ± 0.61
	Ours	**1.43 ± 0.71**	**1.16 ± 0.42**	**1.11 ± 0.37**	**0.96 ± 0.37**	**1.00 ± 0.49**

The best results are shown in bold.

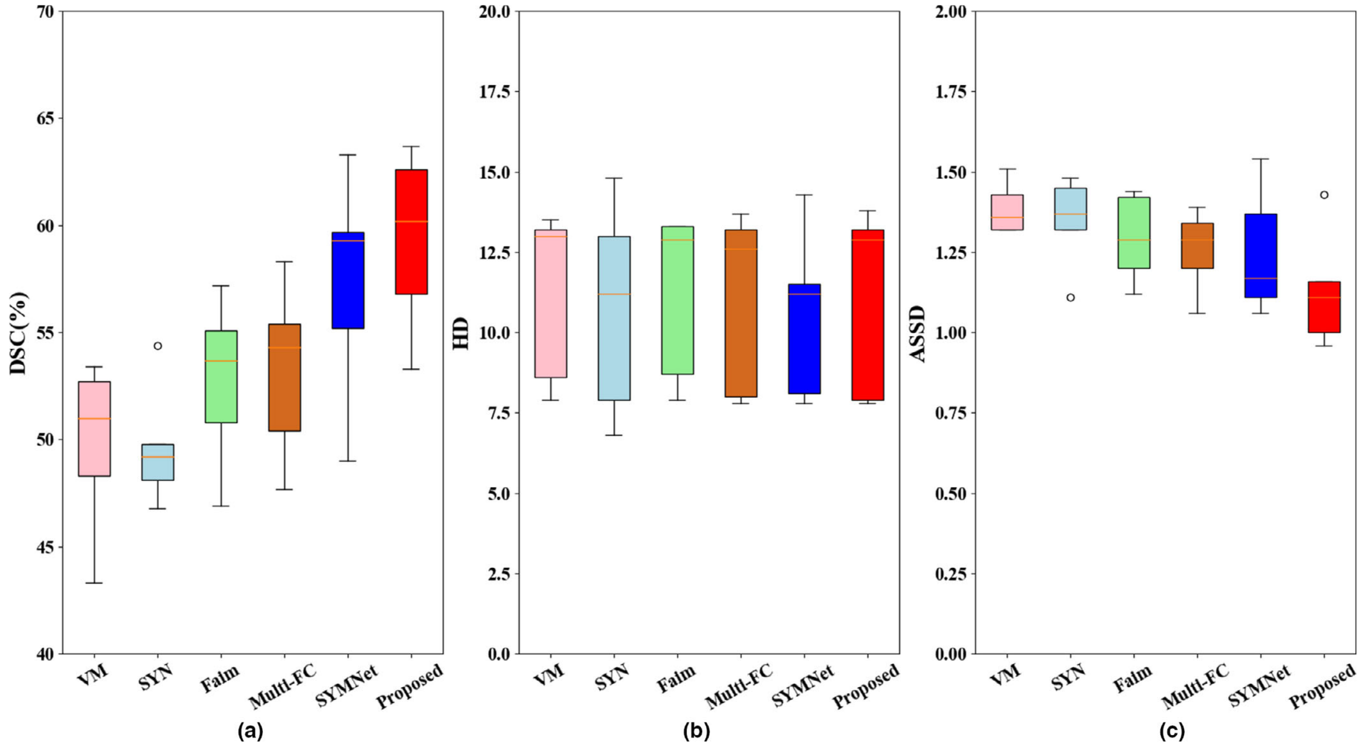

FIG. 4. The (a) DSC (%), (b) Hausdorff distance (HD), and (c) average symmetric surface distance (ASSD) results from SyN,[33] VoxelMorph,[22] FAIM,[34] Multi-FC,[24] SYMNet,[27] and our proposed network on MindBoggle101 dataset.

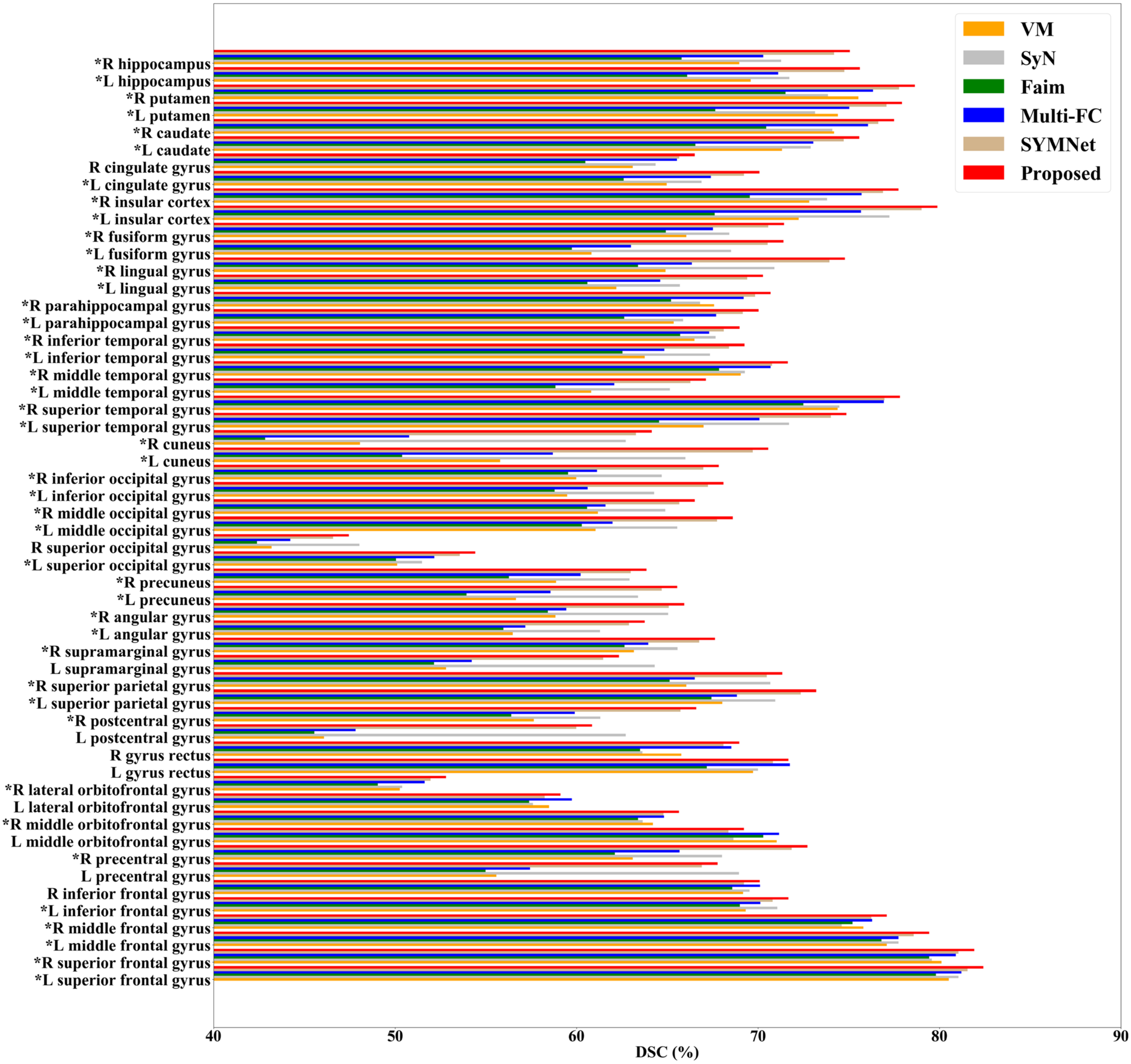

FIG. 5. Comparisons of the DSC (%) results by SyN,[33] VoxelMorph,[22] FAIM,[34] Multi-FC,[24] SYMNet,[27] and our proposed network, respectively. The results were evaluated in terms of DSC values across the 54 ROIs in LPBA40 dataset, "*" indicates that the proposed network outperformed the state-of-the-art methods.

the network training phase; segmentation is not necessary in the inference phase.

In addition, to avoid obtaining an unpractical or discontinuous deformation field, we also add a diffusion regularizer $\mathscr{L}_{smooth}$ to impose smooth constraint of the nonrigid deformation field ϕ:

$$\mathscr{L}_{smooth} = \sum_{p\in\Omega} \| \bigtriangledown \phi(p) \|^2. \tag{4}$$

Therefore, the total loss is:

$$\mathscr{L}(F,M) = \mathscr{L}_{aff} + \mathscr{L}_{sim} + \mathscr{L}_{anat} + \lambda \mathscr{L}_{smooth}, \tag{5}$$

where λ is a regularization parameter.

2.C. Implementation details

In our experiments, each input volumetric image is resized to size 192 × 192 × 192. The network is trained on a GPU of NVIDIA Tesla V100. The value of the regularization parameter λ is 1000[†]. For the whole registration network, the number of epochs is set to 300. The network is implemented using Pytorch and Adam optimization,[30] and the learning rate is initially set to 1e-4, with 0.5 weight decay after every 10 epoch.

[†] In our implementation, we conducted hyperparameter tuning. By considering both the registration performance and computational efficiency, the regularization parameter λ is set as 1000.

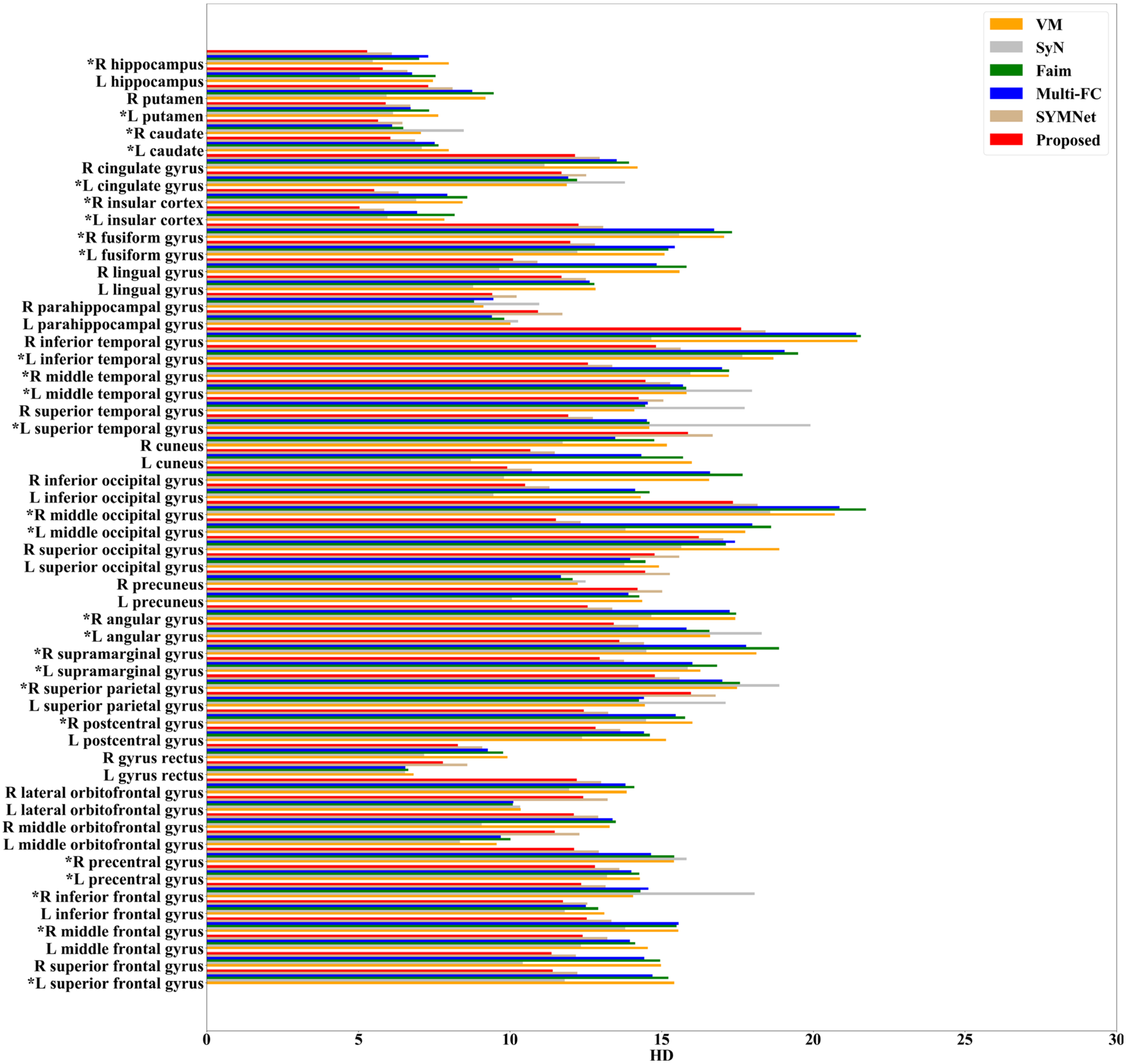

FIG. 6. Comparisons of the Hausdorff distance (HD) results by different methods on LPBA40 dataset. "*" indicates that our network outperformed the state-of-the-art methods.

3. EXPERIMENTS AND RESULTS

3.A. Materials

The Institution's Ethical Review Board approved all experimental procedures involving human subjects. Experiments were carried on three brain MRI datasets, including Mindboggle101,[31] LPBA40,[32] and IXI.[33]

1. Mindboggle101 (101 brain MRI images, each with 62 manually labeled region of interests (ROIs)): 62 images (42 for training and 20 for testing) were involved to conduct experiments as described in.[34]
2. LPBA40 (40 brain MRI images, each with 54 manually labeled ROIs): 30 images were randomly selected for training and the remaining 10 images were used as the testing set.
3. IXI (30 brain MRI images, each with 95 manually labeled ROIs): all 30 images were used for testing. In order to investigate the generalization ability of the network, we employed the model trained on other dataset (here we used Mindboggle101 dataset) to evaluate images from IXI dataset.

All images were pre-processed by histogram and intensity normalization.

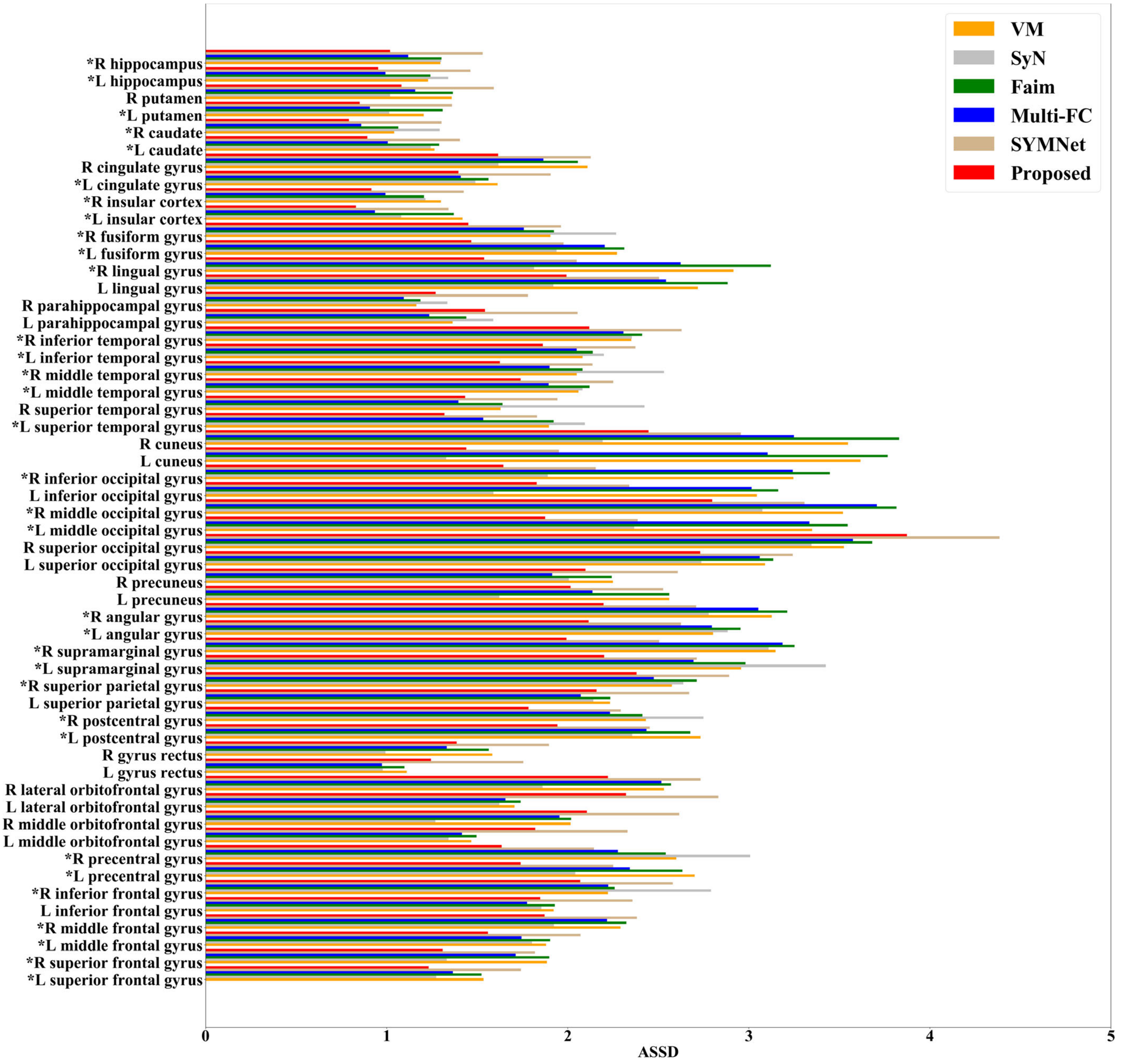

FIG. 7. Comparisons of the average symmetric surface distance (ASSD) results by different methods on LPBA40 dataset. "*" indicates that our network outperformed the state-of-the-art methods.

3.B. Evaluation metrics

To evaluate the performance of the registration, DSC,[35] Hausdorff distance (HD),[36] and average symmetric surface distance (ASSD)[37] were employed as the quantitative metrics. The DSC is defined as:

$$DSC = \frac{2|S_F \cap S_M|}{|S_F| + |S_M|}, \tag{6}$$

where S_F and S_M are the segmented ROIs of the fixed and moving images, respectively. The HD measures the longest distance over the shortest distances between the segmented ROIs of the fixed and moving images. The ASSD can be calculated as:

$$ASSD = \frac{1}{|B_F| + |B_M|}\left(\sum_{x \epsilon B_F} d(x, B_M) + \sum_{y \epsilon B_M} d(y, B_F)\right), \tag{7}$$

where B_F and B_M are the segmented surfaces of the fixed and moving images, respectively. The operator $d(,)$ is the shortest Euclidean distance operator.

All evaluations were calculated in 3D. A better registration shall have larger DSC, and smaller HD and ASSD.

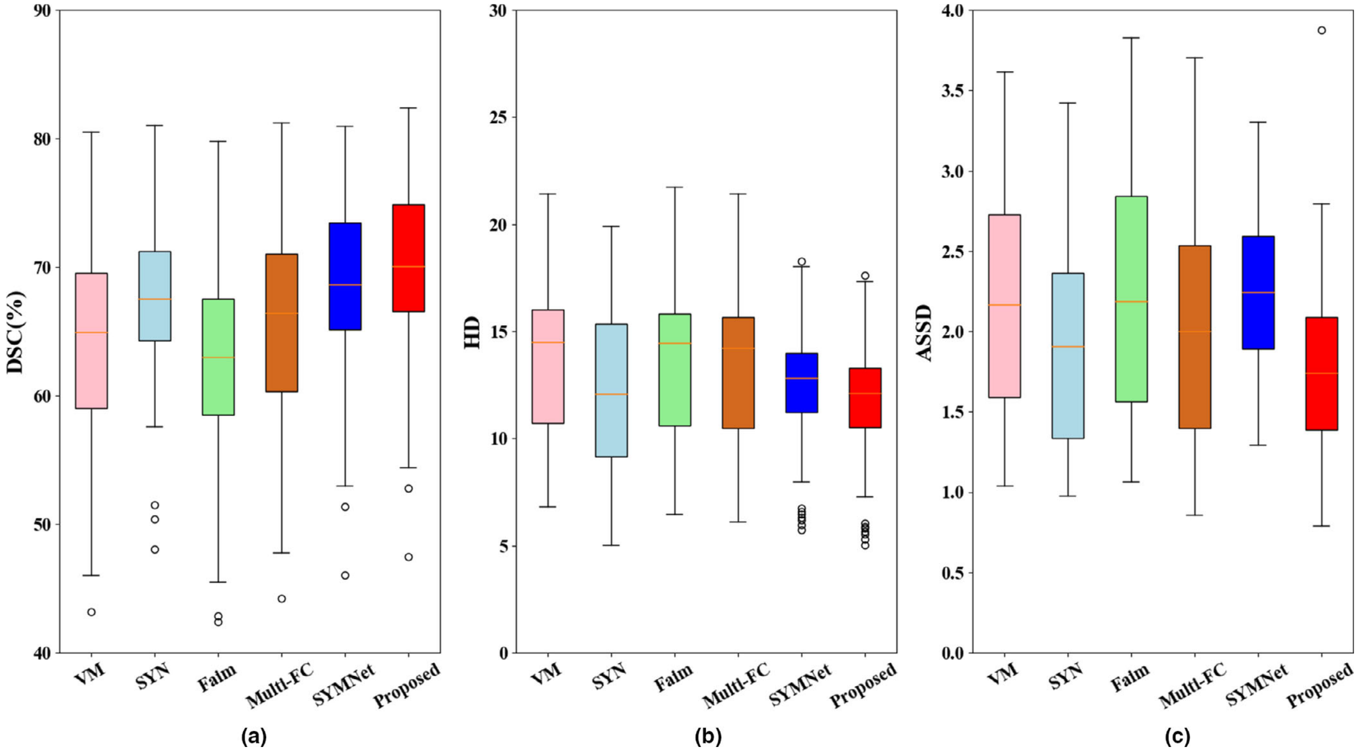

FIG. 8. The (a) DSC (%), (b) Hausdorff distance (HD), and (c) average symmetric surface distance (ASSD) results from SyN,[33] VoxelMorph,[22] FAIM,[34] Multi-FC,[24] SYMNet,[27] and our proposed network on LPBA40 dataset.

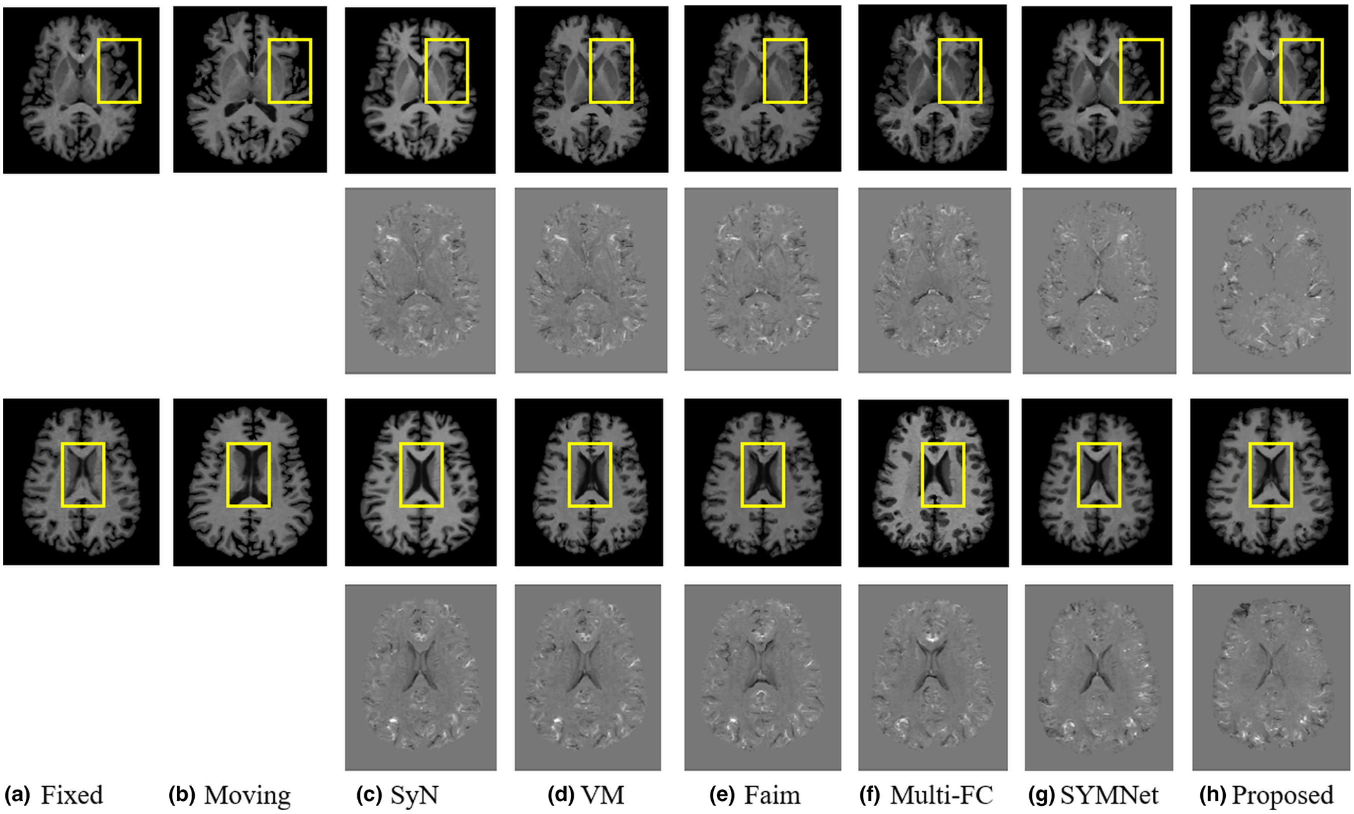

FIG. 9. Some example registration results (rows 1 and 3) and their difference maps (rows 2 and 4) from different registration methods on Mindboggle101 dataset.

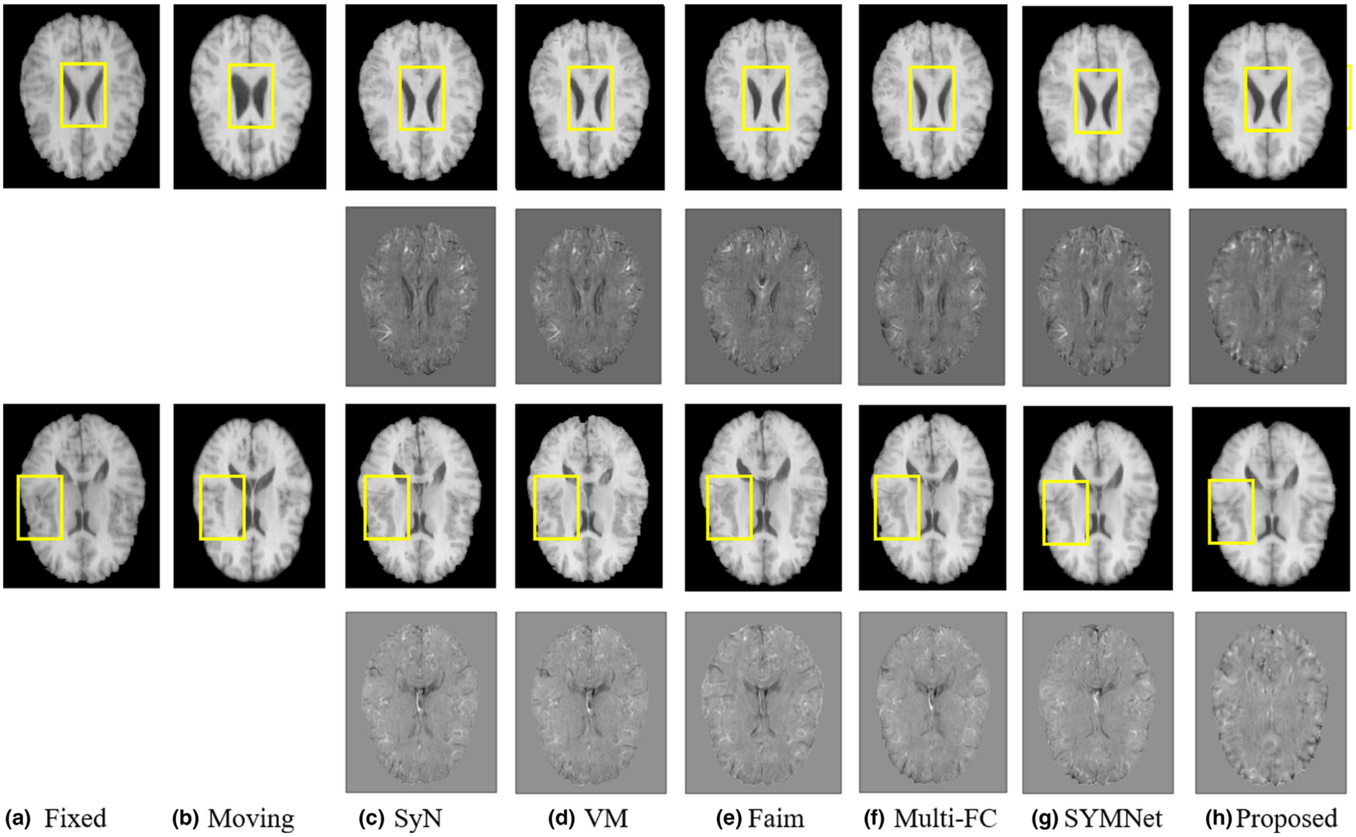

FIG. 10. Some example registration results (rows 1 and 3) and their difference maps (rows 2 and 4) from different registration methods on LPBA40 dataset.

3.C. Evaluation on affine alignment

We first investigated the intermediate affine alignment results obtained by our network, and compared them with advanced normalization tools (ANTS)[33] and Freesufer Toolkit.[2] The average DSC, HD, and ASSD results of five regions obtained on Mindboggle101 are shown in Table I. From Table I, it can be observed that our subnetwork consistently provided more accurate alignment than traditional rigid registration methods did, which could be a better initialization for the subsequent deformable registration.

3.D. Comparisons with state-of-the-art methods

We then compared our network with four state-of-the-art brain MRI registration methods: SyN,[7] VoxelMorph,[22] FAIM,[34] Multi-FC, [24] and SYMNet.[27] For a fair comparison, we obtained their results either by directly taking the results from their papers or by generating the results from the public codes provided by the authors using the recommended parameter setting.

The registration results on five regions of the brain images from dataset MindBoggle101 are reported in Table II. It can be observed that our network consistently achieved best registration performance with respect to DSC and ASSD metrics. Regarding the HD evaluation, our network obtained the best HD values on occipital and cingulate regions; and the second best HD on the partial and temporal regions. In Table II, it is worth noting that our full network consistently outperformed the independent deformable subnetwork (i.e., Ours *w.o.* affine) with respect to all evaluation metrics. This comparison demonstrates the integrated affine subnetwork contributed to the improvement of registration accuracy. Note that Table II also reports the results from the proposed network without the using of anatomical similarity loss (i.e., Ours *w.o.* $\mathscr{L}_{anat}$). It is demonstrated that the using of anatomical similarity loss contributed to the improvement of registrations. Figure 4 plots the average DSC, HD, and ASSD results from compared state-of-the-art methods and our network on MindBoggle101 dataset. Our network achieved overall satisfactory registration performance.

For the LPBA40 dataset, we composed 10 images into 90 image pairs for testing, and calculated DSC values of 54 corresponding subregions from warped and fixed images. The comparison results with respect to DSC, HD, and ASSD metrics are shown in Figs. 5–7. For the 54 subregions, the proposed network achieved the best DSC, HD,and ASSD values on 44, 26, and 33 subregions, respectively. Figure 8 plots the average DSC, HD and ASSD results from different methods on LPBA40 dataset. It can be observed that our network achieved overall satisfactory registration performance on LPBA40 dataset.

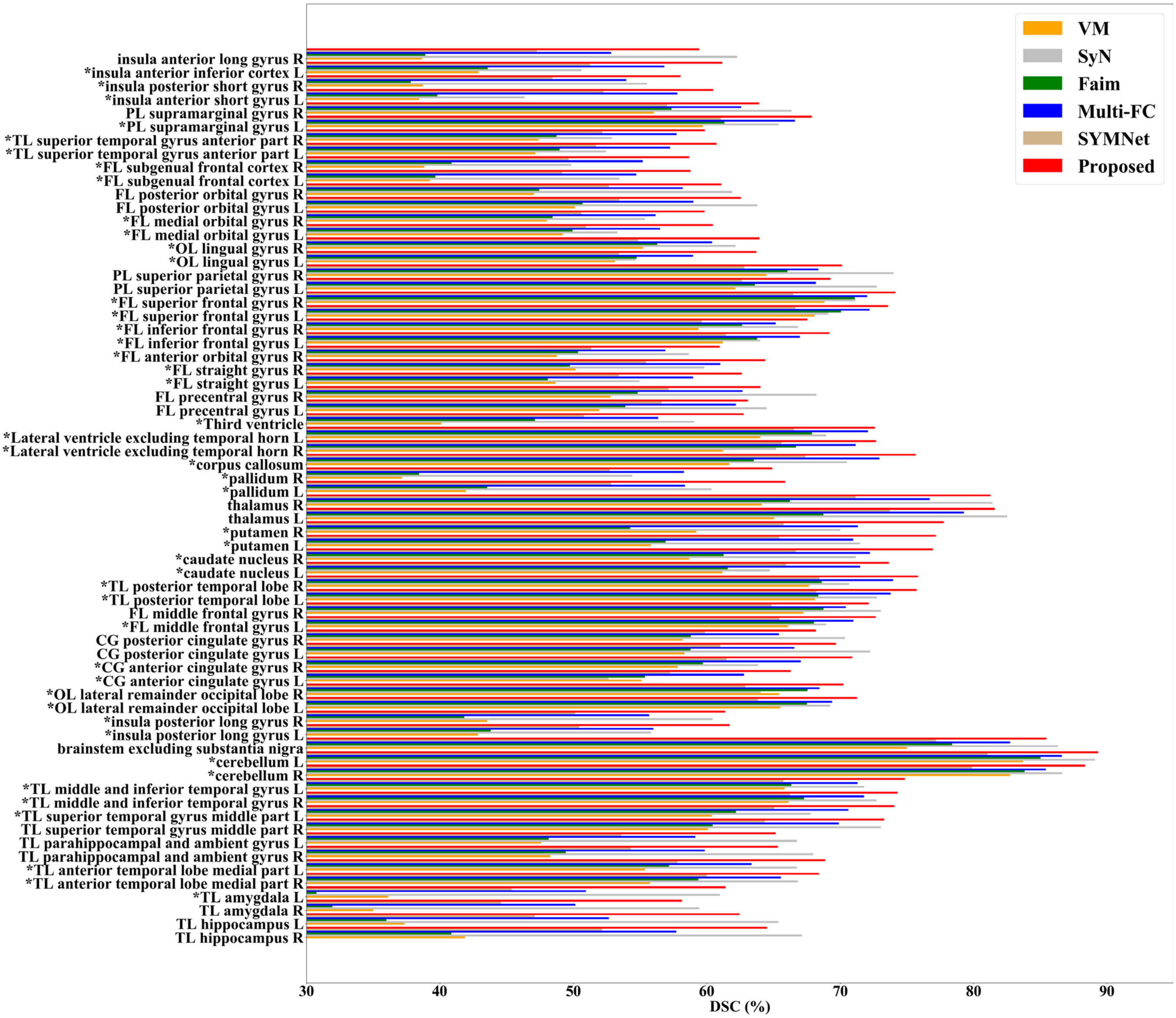

FIG. 11. Comparisons of the DSC (%) results by different methods. The results were evaluated in terms of DSC values across the 66 ROIs in IXI dataset, "*" indicates that the proposed network outperformed the state-of-the-art methods.

Figures 9 and 10 show the visual comparisons from different registration methods for two datasets, respectively. Our network can generate more accurate registered images, and the internal structures can be preserved consistently by using our network.

3.E. Investigation of the generalization ability on IXI dataset

To evaluate the model generalization ability, we directly employed the trained model using Mindboggle101 to register images from the IXI dataset. Note that SyN is not a learning-based method thus we directly conducted it. IXI dataset has 95 subregions but 29 of them are extremely small regions. Thus we measured DSC, HD, and ASSD values of the remaining 66 subregions. The comparison results with respect to different metrics are shown in Figs. 11–13. For the 66 subregions, the proposed network achieved the best DSC, HD, and ASSD values on 46, 32, and 49 subregions, respectively, which shows that our network has satisfactory generalization ability. Figure 14 plots the average DSC, HD, and ASSD results from different methods on IXI dataset. Our network again attained overall satisfactory registration performance on this dataset.

3.F. Comparison of time efficiency

We further compared the time efficiency. Table III lists the inference time for registering a pair of images using different methods. It can be observed that for the affine alignment, the

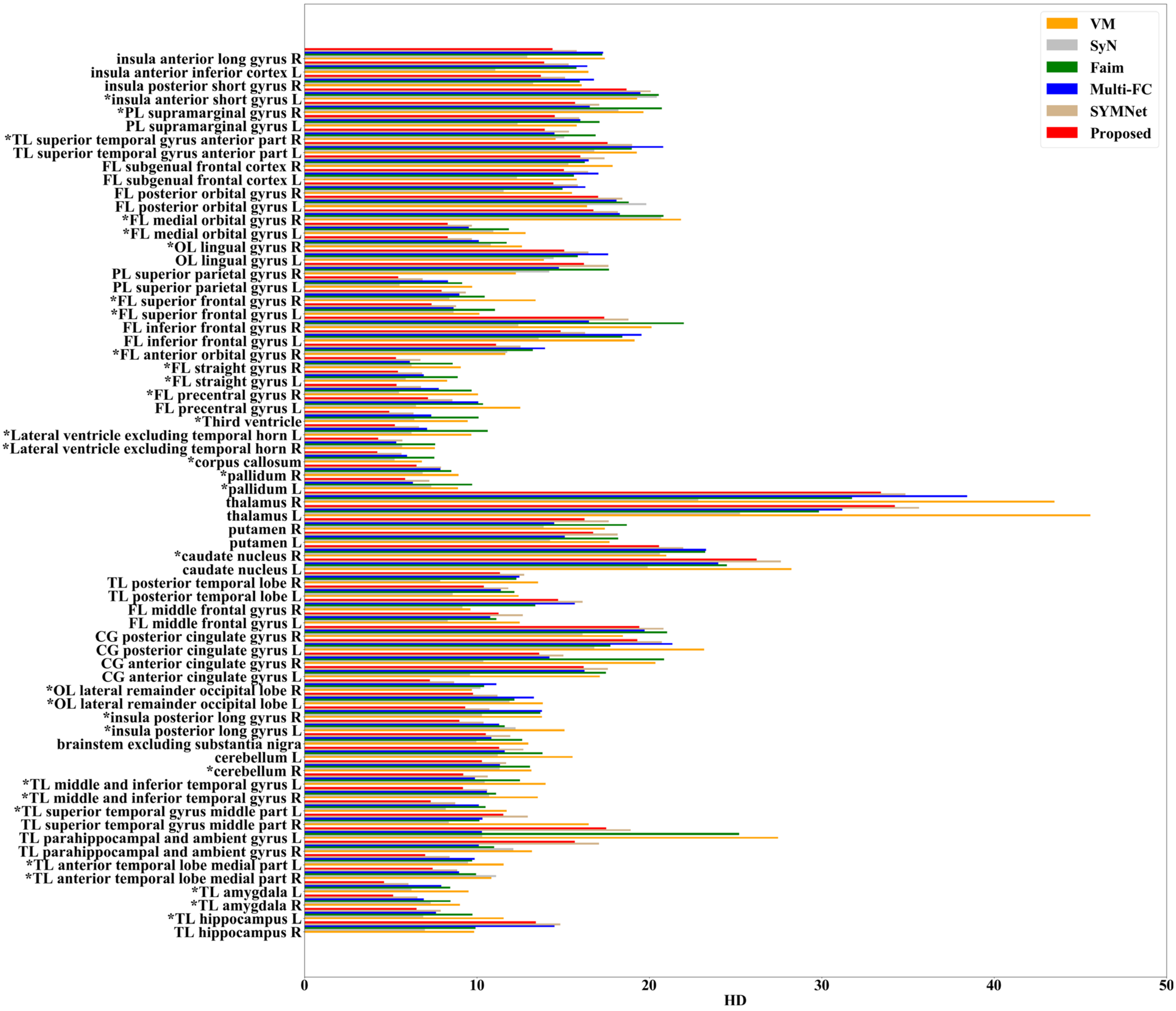

FIG. 12. Comparisons of the Hausdorff distance (HD) results by different methods on IXI dataset. "*" indicates that our network outperformed the state-of-the-art methods.

time spent by our affine subnetwork was much less than that of the traditional affine alignment method (i.e., ANTS[33]) using iterative optimization. For the deformable registration procedure, our deformable subnetwork was a litter bit slower than other deformable networks, which is mainly due to the deeper architecture of our network design. However, considering the overall registration time, our end-to-end registration network was much faster than other registration networks which require extra affine alignment (i.e., alignment using ANTS[33]) before deformable registration.

The training time for VoxelMorph,[22] FAIM,[34] Multi-FC,[24] SYMNet[27] and our network on datasets Mindboggle101 and LPBA40 were (134h, 49h), (254h, 134h), (299h, 144h), (187h, 91h), and (188h, 104h), respectively. The training time for our network was slower than that for VM,[22] but faster than that for FAIM[34] and Multi-FC.[24]

4. DISCUSSION

We have proposed a joint affine and deformable network to facilitate the workflow of 3D medical image registration. The nonrigid registration is to search for the point-wise displacement field to map homologous locations from the target domain to the source domain. Due to the large searching space and complicated deformation imposed, most conventional nonrigid registration methods require an independent rigid alignment before conducting the deformable registration. These two steps are often performed separately thus cannot be jointly optimized. We have attempted to tackle this issue by devise a deep neural network for realizing affine and deformable registrations simultaneously. Experimental results on Section 3.C demonstrate that our affine subnetwork can provide more accurate rigid

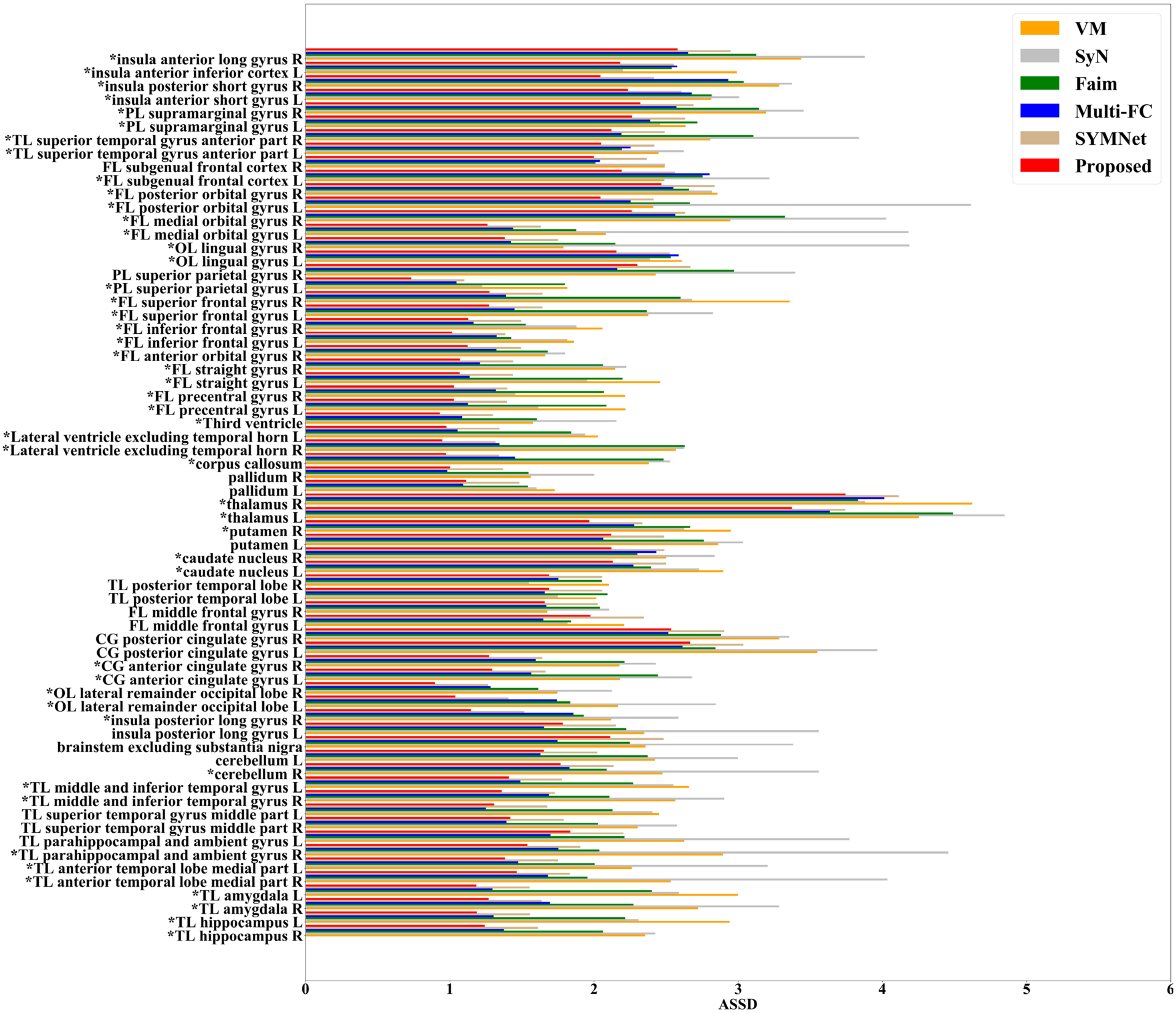

FIG. 13. Comparisons of the average symmetric surface distance (ASSD) results by different methods on IXI dataset. "*" indicates that our network outperformed the state-of-the-art methods.

alignment than traditional methods with respect to all evaluation metrics, thus could be a better initialization for the subsequent deformable registration. The more important impact is that the affine and deformable registration is a seamless integration, so as to enable the trained network performing nonrigid registration in one forward pass. In such a way, our proposed network not only provides satisfactory registration accuracy (see Table II), but also achieves end-to-end registration thus is more efficient than other deformable networks which require extra affine alignment for initialization (see Table III).

Conventional unsupervised registration networks optimize the desired deformation field by evaluating the similarity between images. Intensity-based similarity measures including NCC, MI, and MSE are often used as the training loss. Inspired by the boundary/surface-based registration, we have devised an anatomical similarity loss to weakly supervise the training of the whole registration network. The efficacy of the devised anatomical similarity loss can be demonstrated from Table II (Ours *w.o.* $\mathscr{L}_{anat}$ vs Ours). By simultaneously evaluating the intensity- and structure-based similarities, the network can provide more accurate and robust registration results.

Compared to unsupervised registration networks, the main limitation of our method is that network training needs extra segmentation masks. However, the segmentation masks are only needed in the network training phase. When the network is well trained, the registration inference can be conducted without the presence of the image segmentation. To more effectively using the segmentation masks, our future study will focus on the joint registration and segmentation tasks.

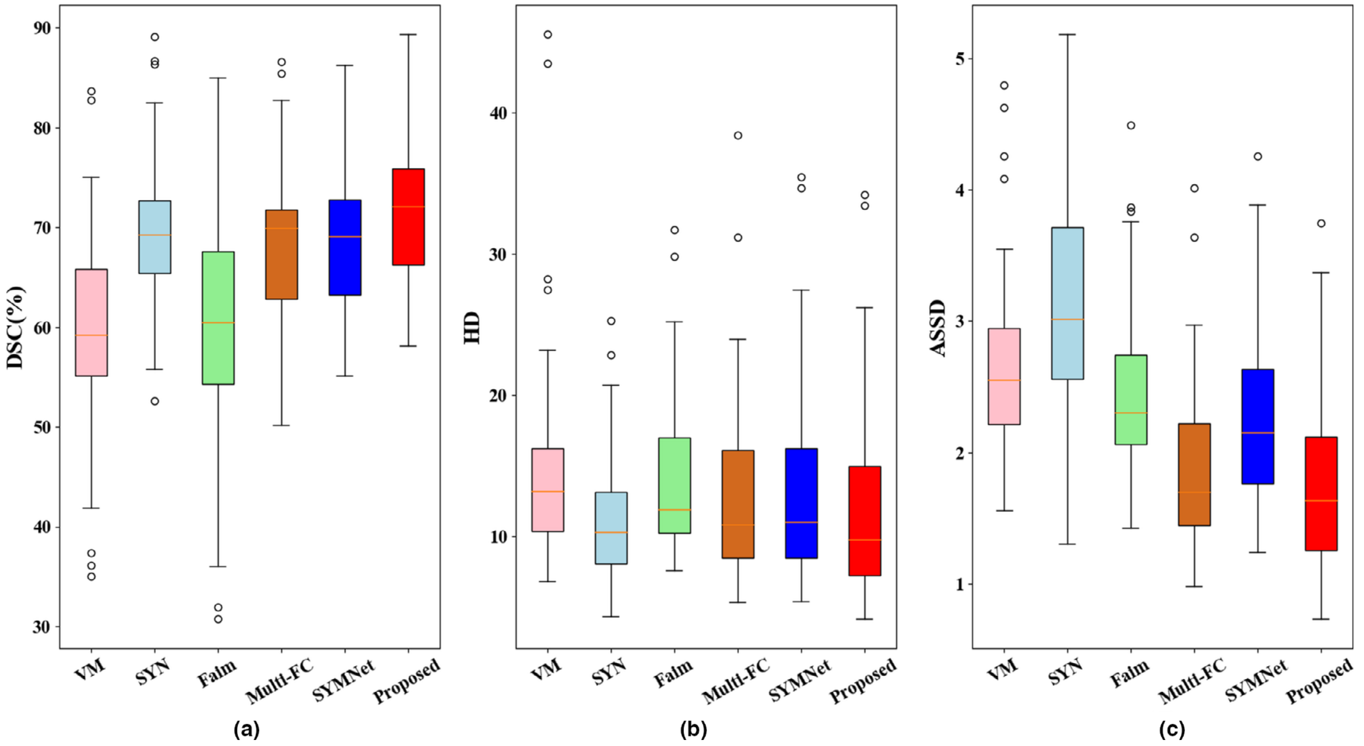

FIG. 14. The (a) DSC (%), (b) Hausdorff distance (HD), and (c) average symmetric surface distance (ASSD) results from SyN,[33] VoxelMorph,[22] FAIM,[34] Multi-FC,[24] SYMNet,[27] and our proposed network on IXI dataset.

TABLE III. Inference time (*second*) for registering a pair of images using different methods

Methods	Mindboggle101	LPBA40	IXI
ANTS[33] (affine alignment)	8.19 ± 0.30	7.73 ± 0.27	9.13 ± 0.29
SyN[33]	39.24 ± 2.07	33.32 ± 1.46	40.10 ± 2.01
VM[22]	0.66 ± 0.01	0.11 ± 0.01	0.73 ± 0.01
FAIM[34]	1.17 ± 0.01	0.49 ± 0.01	1.26 ± 0.01
Multi-FC[24]	1.34 ± 0.01	0.45 ± 0.01	1.41 ± 0.01
SYMNet[27]	1.12 ± 0.01	0.66 ± 0.01	1.24 ± 0.01
Ours (affine only)	0.27 ± 0.01	0.29 ± 0.01	0.49 ± 0.01
Ours	1.78 ± 0.01	1.18 ± 0.01	1.81 ± 0.01

5. CONCLUSION

In this paper, we have introduced a 3D end-to-end medical image registration strategy. Our network cascades the affine alignment and deformable registration subnetworks. These two subnetworks share network parameters to maximize registration performance while reducing parameters. The network was trained in a weakly supervised manner by calculating global and local image similarities, and the devised anatomical similarity. Finally, the trained network can perform deformable registration in one forward pass. Extensive experiments on various brain MRI datasets demonstrate that our network achieved volumetric registration effectively and robustly, and consistently outperformed state-of-the-art methods. The proposed network provides accurate and robust volumetric registration without any pre-alignment requirement, which facilitates the end-to-end deformable registration.

ACKNOWLEDGMENTS

This work was supported in part by the National Key R&D Program of China (No. 2019YFC0118300), in part by the National Natural Science Foundation of China under Grants 62071305 and 61701312, in part by the Guangdong Basic and Applied Basic Research Foundation (2019A1515010847), in part by the Medical Science and Technology Foundation of Guangdong Province (B2019046), in part by the Natural Science Foundation of SZU (No. 860-000002110129), and in part by the Shenzhen Peacock Plan (KQTD2016053112051497).

CONFLICT OF INTEREST

The authors have no conflicts to disclose.

[a] Author to whom correspondence should be addressed. Electronic mail: onewang@szu.edu.cn.

REFERENCES

1. Sotiras A, Davatzikos C, Paragios N. Deformable medical image registration: A survey. *IEEE Trans Med Imaging*. 2013;32:1153–1190.

2. Fischl B, FreeSurfer. *NeuroImage*. 2012;62:774–781.
3. Avants BB, Tustison NJ, Song G, Cook PA, Klein A, Gee JC. A reproducible evaluation of ANTs similarity metric performance in brain image registration. *NeuroImage*. 2011;54:2033–2044.
4. Ourselin S, Roche A, Prima S, Ayache N. Block matching: A general framework to improve robustness of rigid registration of medical images, in International Conference on Medical Image Computing And Computer-Assisted Intervention, Springer; 2000:557–566.
5. Klein A, Andersson J, Ardekani BA, et al. Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. *NeuroImage*. 2009;46:786–802.
6. Beg MF, Miller MI, Trouvé A, Younes L. Computing large deformation metric mappings via geodesic ows of diffeomorphisms. *Int J Comput Vision*. 2005;61:139–157.
7. Avants BB, Epstein CL, Grossman M, Gee JC. Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. *Med Image Anal*. 2008;12:26–41.
8. Vercauteren T, Pennec X, Perchant A, Ayache N. Diffeomorphic demons: Efficient non-parametric image registration. *NeuroImage*. 2009;45:S61–S72.
9. Wolberg G, Zokai S. Robust image registration using log-polar transform, in Proceedings 2000 International Conference on Image Processing (Cat No. 00CH37101), volume 1. IEEE; 2000:493–496.
10. Cideciyan AV. Registration of ocular fundus images: an algorithm using cross correlation of triple invariant image descriptors. *IEEE Eng Med Biol Mag*. 1995;14:52–58.
11. Rao YR, Prathapani N, Nagabhooshanam E. Application of normalized cross correlation to image registration, International Journal of Research. *Eng Tech*. 2014;3:12–16.
12. Viola P, Wells WM. III Alignment by maximization of mutual information. *Int J Comput Vision*. 1997;24:137–154.
13. Knops ZF, Maintz JA, Viergever MA, Pluim JP. Normalized mutual information based registration using k-means clustering and shading correction. *Med Image Anal*. 2006;10:432–439.
14. de Vos BD, Berendsen FF, Viergever MA, Sokooti H, Staring M, Išgum I. A deep learning framework for unsupervised affine and deformable image registration. *Med Image Anal*. 2019;52:128–143.
15. Zhao S, Dong Y, Chang EI, et al. Recursive cascaded networks for unsupervised medical image registration, in Proceedings of the IEEE International Conference on Computer Vision; 2019:10600–10610.
16. Fan J, Cao X, Yap P-T, Shen D. BIRNet: Brain image registration using dual-supervised fully convolutional networks. *Med Image Anal*. 2019;54:193–206.
17. Uzunova H, Wilms M, Handels H, Ehrhardt J. Training CNNs for image registration from few samples with model-based data augmentation, in International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer; 2017:223–231.
18. Jaderberg M, Simonyan K, Zisserman A. et al. Spatial transformer networks, in Advances in neural information processing systems, 2015:2017–2025.
19. Rohé M-M, Datar M, Heimann T, Sermesant M, Pennec X. SVF-Net: Learning deformable image registration using shape matching, in International Conference on Medical Image Computing and Computer-Assisted Intervention Springer; 2017:266–274.
20. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation, in International Conference on Medical image computing and computer-assisted intervention. Springer; 2015:234–241.
21. Li H, Fan Y. Non-rigid image registration using self-supervised fully convolutional networks without training data, in 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), IEEE, 2018:1075–1078.
22. Balakrishnan G, Zhao A, Sabuncu MR, Guttag J, Dalca AV. An unsupervised learning model for deformable medical image registration, in Proceedings of the IEEE conference on computer vision and pattern recognition; 2018:9252–9260.
23. Cao X, Yang J, Zhang J, et al. Deformable image registration based on similarity-steered CNN regression, in International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer; 2017:300–308.
24. Duan L, Yuan G, Gong L, et al. Adversarial learning for deformable registration of brain MR image using a multi-scale fully convolutional network. *Biomed Signal Process Control*. 2019;53:101562.
25. Hu Y, Modat M, Gibson E, et al. Label-driven weakly-supervised learning for multimodal deformable image registration, in 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), IEEE, 2018:1070–1074.
26. Zhu Z, Cao Y, Qin C, Rao Y, Ni D, Wang Y. Unsupervised 3D End-to-end Deformable Network for Brain MRI Registration, in 2020 42nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); 2020:1355–1359.
27. Mok TCW, Chung ACS. Fast Symmetric Diffeomorphic Image Registration with Convolutional Neural Networks, in 2020 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2020:4643–4652.
28. Maas AL, Hannun AY, Ng AY. Rectifier nonlinearities improve neural network acoustic models, in Proc. icml. 2013; 30:3.
29. Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift, arXiv preprint arXiv:1502.03167; 2015.
30. Kingma DP, Ba J, Adam: A method for stochastic optimization, arXiv preprint arXiv:1412.6980; 2014.
31. Klein A, Ghosh SS, Bao FS, et al. Mindboggling morphometry of human brains. *PLoS Comput Biol*. 2017;13:e1005350.
32. Shattuck DW, Mirza M, Adisetiyo V, et al. Construction of a 3D probabilistic atlas of human cortical structures. *NeuroImage*. 2008;39:1064–1080.
33. Avants BB, Tustison N, Song G. Advanced normalization tools (ANTS). *Insight J*. 2009;2:1–35.
34. Kuang D, Schmah T. Faima convnet method for unsupervised 3d medical image registration, in International Workshop on Machine Learning in Medical Imaging, Springer; 2019:646–654.
35. Dice LR. Measures of the amount of ecologic association between species. *Ecology*. 1945;26:297–302.
36. Huttenlocher DP, Klanderman GA, Rucklidge WJ. Comparing images using the Hausdorff distance. *IEEE Trans Pattern Anal Mach Int*. 1993;15:850–863.
37. Taha AA, Hanbury A. Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. *BMC Medical*. 2015;15:29.