



Edge-Aware Pyramidal Deformable Network for Unsupervised Registration of Brain MR Images

Yiqin Cao¹, Zhenyu Zhu¹, Yi Rao¹, Chenchen Qin², Di Lin³, Qi Dou⁴, Dong Ni¹ and Yi Wang^{1*}

¹ National-Regional Key Technology Engineering Laboratory for Medical Ultrasound, Guangdong Provincial Key Laboratory of Biomedical Measurements and Ultrasound Imaging, School of Biomedical Engineering, Health Science Center, Shenzhen University, Shenzhen, China, ² Tencent AI Lab, Shenzhen, China, ³ The College of Intelligence and Computing, Tianjin University, Tianjin, China, ⁴ Computer Science and Engineering, The Chinese University of Hong Kong, Hong Kong, China

OPEN ACCESS

Edited by:

Yuanpeng Zhang,
Nantong University, China

Reviewed by:

Bing Li,
Hong Kong Polytechnic University,
Hong Kong
Marco Fusella,
Veneto Institute of Oncology (IRCCS),
Italy

*Correspondence:

Yi Wang
onewang@szu.edu.cn

Specialty section:

This article was submitted to
Brain Imaging Methods,
a section of the journal
Frontiers in Neuroscience

Received: 22 October 2020

Accepted: 28 December 2020

Published: 21 January 2021

Citation:

Cao Y, Zhu Z, Rao Y, Qin C, Lin D, Dou Q, Ni D and Wang Y (2021) Edge-Aware Pyramidal Deformable Network for Unsupervised Registration of Brain MR Images. *Front. Neurosci.* 14:620235. doi: 10.3389/fnins.2020.620235

Deformable image registration is of essential importance for clinical diagnosis, treatment planning, and surgical navigation. However, most existing registration solutions require separate rigid alignment before deformable registration, and may not well handle the large deformation circumstances. We propose a novel edge-aware pyramidal deformable network (referred as EPReg) for unsupervised volumetric registration. Specifically, we propose to fully exploit the useful complementary information from the multi-level feature pyramids to predict multi-scale displacement fields. Such coarse-to-fine estimation facilitates the progressive refinement of the predicted registration field, which enables our network to handle large deformations between volumetric data. In addition, we integrate edge information with the original images as dual-inputs, which enhances the texture structures of image content, to impel the proposed network pay extra attention to the edge-aware information for structure alignment. The efficacy of our EPReg was extensively evaluated on three public brain MRI datasets including Mindboggle101, LPBA40, and IXI30. Experiments demonstrate our EPReg consistently outperformed several cutting-edge methods with respect to the metrics of Dice index (DSC), Hausdorff distance (HD), and average symmetric surface distance (ASSD). The proposed EPReg is a general solution for the problem of deformable volumetric registration.

Keywords: deformable image registration, convolutional neural networks, brain MR image, affine registration, 3D registration

1. INTRODUCTION

Deformable image registration is to perform spatial transformation between a pair of images and establish a non-linear point-wise correspondence to achieve spatial consistency (Sotiras et al., 2013). By doing so, mono-/multi-modality information can be fused into the same coordinate system. It plays a very important role in various medical imaging studies to provide complementary diagnostic information and investigate changes of anatomical structures. Although many algorithms have been proposed over the past few decades (Sotiras et al., 2013; Shen et al., 2017; Haskins et al., 2020), registration is still a challenging task. Traditional registration methods may be computationally expensive and time-consuming due to their iterative optimization during deformation estimation procedure. Moreover, most existing deformable registration solutions

require separate rigid alignment before non-rigid registration, and may not well handle the large deformation circumstances. Therefore, efficient and accurate deformable registration scheme is still greatly expected to compensate for complicated non-rigid deformations.

As illustrated in **Figure 1**, the goal of registration is to match all corresponding anatomical structures in two images to the same spatial system through plausible deformable transformation. To calculate the desired deformable transformation, several non-linear deformation algorithms (Klein et al., 2009) have been proposed, such as large diffeomorphic distance metric mapping (LDDMM) (Beg et al., 2005; Auzias et al., 2011), standard symmetric normalization (Avants et al., 2008), and Demons (Vercauteren et al., 2009). These methods treat deformable registration as a procedure of iterative optimization for maximizing the similarity [such as mean square error (MSE) (Wolberg and Zokai, 2000), normalized mutual information (NMI) (Knops et al., 2006), and normalized cross-correlation (NCC) (Rao et al., 2014), etc.] between fixed and warped moving images. However, the iterative optimization strategy may take a relatively long time to deal with complicated volumetric deformations.

To address aforementioned issue, deep neural networks have been widely investigated for the registration task in recent years (Haskins et al., 2020). The registration networks are beneficial to aggregate abundant features from paired images to effectively predict the deformation field. Eppenhof et al. (2018) employed synthetic random transformations to train a registration framework based on a convolutional neural network (CNN). Fan et al. (2019) also applied a supervised CNN for image registration by using obtained ground-truth deformation fields as the supervision information. Uzunova et al. (2017) synthesized a large amount of realistic ground-truth data using model-based strategy to train a registration network. Yang et al. (2017) proposed a patch-wise deformation prediction model, which is a deep encoder-decoder network devised to estimate the momentum-parameterization of LDDMM model. The major limitation of supervised registration networks (Cao et al., 2017; Sokooti et al., 2017; Uzunova et al., 2017; Yang et al., 2017; Eppenhof et al., 2018; Fan et al., 2019) is the prerequisite of the ground-truth registration fields, which would highly affect the network performance. However, unlike segmentation or detection tasks, it is always difficult to obtain registration ground-truth.

In contrast, some studies have focused on unsupervised deep learning algorithms which achieved great success in various registration tasks (Sheikhjafari et al., 2018; Kuang and Schmah, 2019). The mechanism of unsupervised registration networks is to build model to obtain the deformation fields based on maximizing the similarity between two images, thus is without the need of ground-truth deformations. Li and Fan (2018) proposed to predict deformation parameters using a fully convolutional network, but this is a 2D approach that tends to ignore the volumetric information. Rohé et al. (2017) utilized U-net (Ronneberger et al., 2015) to estimate the deformation field of 3D cardiac MR images and employed the sum of squared differences (SSD) as the similarity loss. Balakrishnan

et al. (2018) proposed an end-to-end network with cross-correlation as its loss function and spatial transformer networks (STN) (Jaderberg et al., 2015) as warping module. However, the prerequisite of this network was another rigid alignment. In addition to unsupervised learning, weakly supervised registration methods usually pay extra attention on the correspondences between structural information of two images, such as the extracted corresponding anatomical landmarks in prostate MR and ultrasound images (Hu et al., 2018). The weakly supervised network with structural similarity could provide more reliable registration but still requires a small amount of manual annotations.

One major challenge facing existing registration neural networks is the effective solution for large deformation compensation. To tackle this issue, Hu et al. (2019) proposed a registration network based on (Balakrishnan et al., 2018), which warps the multi-resolution feature maps to obtain the deformation field. However, in such a way, the low-resolution deformation field cannot be accurately acted on subsequent high-resolution features, thus may degrade the registration accuracy. At the same time, many other cascade/recursive networks (de Vos et al., 2019; Zhao et al., 2019a,b) have been proposed. The general idea of these networks is to progressively estimate the complicated transformation relationship between moving and fixed images, which is similar to the iterative optimization idea of traditional algorithms. For example, deep learning image registration (DLIR) (de Vos et al., 2019) combined affine and non-linear networks to calculate both affine alignment and non-linear registration. Volume tweening network (VTN) (Zhao et al., 2019b) cascaded several registration sub-networks, which deforms the moving images by multiple times according to the deformation estimation. The recursive cascade network (Zhao et al., 2019a) expanded the number of cascaded networks, and only calculated the similarity of the last cascade for training. In general, the cascade/recursive networks simplify the challenge of large deformation based on progressive deformation estimation. But the performance of these networks would be affected by the training strategies and the cumulative errors caused by the cascaded propagation.

In this study, we devise a novel edge-aware pyramidal deformable network (EPReg) for unsupervised volumetric registration. The proposed EPReg is a dual-stream pyramid framework, which utilizes original images and corresponding edge-aware maps to compose dual inputs, and generates multi-scale paired feature maps for recursively transforming the information between images into more representative features to predict more accurate deformation field. Finally, the trained EPReg can perform deformable registration in one forward pass. Extensive experiments on three 3D brain magnetic resonance imaging (MRI) datasets demonstrate that our proposed network achieves satisfactory registration performance.

The main contributions of our work are 2-fold.

1. We propose to fully exploit the useful complementary information from the multi-level feature pyramids to predict multi-scale deformation fields. Such coarse-to-fine estimation facilitates the progressive refinement of the predicted

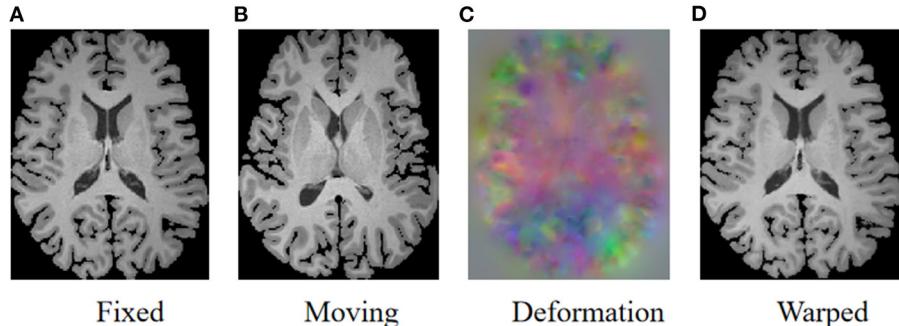


FIGURE 1 | Illustration of image registration. Given a fixed image (**A**) and a moving image (**B**), a deformation field (**C**) is calculated to warp the moving image so that the warped image (**D**) is registered to the fixed image.

registration field, which enables our network to handle large deformations between volumetric images.

2. We integrate edge information with the original images as dual-inputs, which enhances the texture structures of image content, to impel the proposed network pay extra attention to the edge-aware information for structure alignment.

The remainder of this paper is organized as follows. Section 2 presents the details of the edge-aware pyramidal deformable network. Section 3 shows the experimental results of the proposed EPReg for the application of brain MRI registration. Section 4 elaborates the discussion of the proposed network, and the conclusion of this study is given in section 5.

2. EDGE-AWARE PYRAMIDAL DEFORMABLE NETWORK

The proposed registration network is illustrated in **Figure 2** (A) with its affine alignment block (B) and deformable registration block (C). We denote the input volumetric image pair as a fixed volume (I_f) and a moving volume (I_m). Their edge maps are denoted as E_f and E_m , respectively. The EPReg network leverages deformable pyramid to progressively transform the information between I_f - I_m and E_f - E_m into more representative features to predict more accurate deformation field ($\phi_4 \sim \phi_1$). The trained EPReg can attain deformable registration in one forward pass.

The following subsections first give a brief introduction on the deformable registration and then present the details of our scheme and elaborate the novel edge-aware pyramidal deformable architecture.

2.1. Preliminaries

Volumetric registration is to establish the voxel-wise correspondences between different volumes (i.e., fixed volume $I_f \in \mathbb{R}^3$ and moving volume $I_m \in \mathbb{R}^3$). The goal is to predict the optimal deformation field ϕ , so that the warped moving volume $I_m \circ \phi \in \mathbb{R}^3$ can be matched with I_f . The optimization problem can be defined as:

$$\phi = \arg \min_{\phi} \mathcal{L}_{sim}(I_f, I_m \circ \phi) + \lambda \mathcal{L}_{smooth}(\phi), \quad (1)$$

where $I_m \circ \phi$ denotes I_m warped by ϕ . \mathcal{L}_{sim} defines similarity criterion and \mathcal{L}_{smooth} regularizes the deformation ϕ to match any specific properties in the solution, and λ is a regularization parameter. There are several conventional formulations for \mathcal{L}_{sim} and \mathcal{L}_{smooth} , respectively. Common similarity measures include MSE, NMI, NCC, and structural similarity index (SSIM) (Wang et al., 2004). \mathcal{L}_{smooth} is often formulated as a regularizer on the spatial gradients of the displacement field.

2.2. Pyramidal Deformable Network

The proposed EPReg is build on the dual-stream pyramid architecture as shown in **Figure 2A**. The dual-stream encoder part is with shared parameters. As shown in **Figure 3**, the encoder part consists of four down-sampling convolutional blocks. Each convolutional block contains a 3D strided convolution with stride of 2, to reduce the spatial dimension in half. For the second and third convolutional blocks, residual connections (He et al., 2016) are employed. Specifically, two residual blocks are successively employed, each of which consists of two convolutional layers with a residual connection. For the last convolutional block, a 3D atrous spatial pyramid pooling (ASPP) (Wang et al., 2019) module is used to resample features at different scales for the capture of more representative multi-scale information. Batch normalization and rectified linear unit (ReLU) operations are applied in each convolutional block.

The convolutional blocks capture hierarchical paired features (i.e., $F_1 \sim F_4$ and corresponding $M_1 \sim M_4$) of the input volumetric pair, which are then used to progressively predict multi-scale deformation field ($\phi_4 \sim \phi_1$). Such coarse-to-fine estimation based on paired feature pyramids enhances the capability for handing large-scale deformation estimation.

We first perform rigid alignment on the feature maps F_4 and M_4 with high-level semantic information. Specifically, we devise an affine block to achieve global alignment. As shown in **Figure 2B**, the affine block consists of an affine convolutional layer (a residual block and a $1 \times 1 \times 1$ convolution operation) and a global average pooling (GAP) layer. It takes paired feature maps F_4 - M_4 as input, and outputs the affine deformation field ϕ_4 , which contains 12 degrees of freedom:

$$\phi_4 = f_a(F_4, M_4; \theta_a), \quad (2)$$

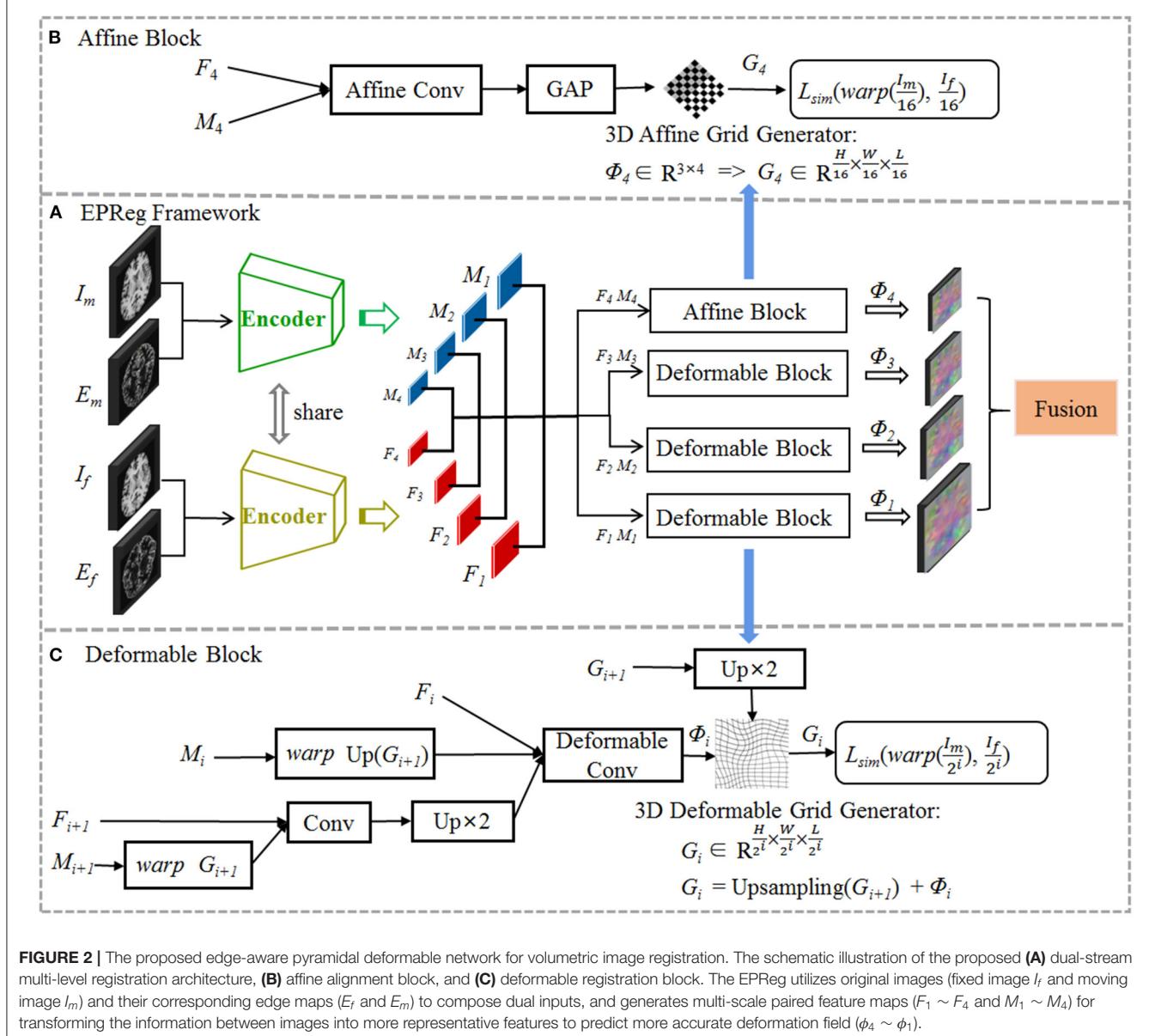


FIGURE 2 | The proposed edge-aware pyramidal deformable network for volumetric image registration. The schematic illustration of the proposed **(A)** dual-stream multi-level registration architecture, **(B)** affine alignment block, and **(C)** deformable registration block. The EPReg utilizes original images (fixed image I_f and moving image I_m) and their corresponding edge maps (E_f and E_m) to compose dual inputs, and generates multi-scale paired feature maps ($F_1 \sim F_4$ and $M_1 \sim M_4$) for transforming the information between images into more representative features to predict more accurate deformation field ($\phi_4 \sim \phi_1$).

where θ_a represents the parameter learned by affine block f_a . According to the estimated ϕ_4 , the 3D affine grid G_4 can be generated and then warp the moving volume to rigidly align with the fixed volume.

Based on paired feature pyramids, we then progressively carry out the non-rigid registration via the devised deformable block (see Figure 2C). The deformable block contains a deformable convolutional layer, which consists of two residual blocks and a $1 \times 1 \times 1$ convolution operation. To estimate ϕ_i , the input of the deformable block includes three components, i.e., F_i , warped M_i using previously estimated ϕ_{i+1} , and the fused previous feature maps (F_{i+1} and warped M_{i+1} using ϕ_{i+1}):

$$\phi_i = f_{d_i}(F_i, M_i, F_{i+1}, M_{i+1}, \phi_{i+1}; \theta_{d_i}), \quad (3)$$

where θ_{d_i} represents the parameter learned by the i -th deformable block f_{d_i} , and $i = 1, 2, 3$. In such a way, based on previous deformation estimation (i.e., ϕ_{i+1}) and paired feature pyramids, each deformable block further estimates extra deformation ϕ_i , which can integrate with ϕ_{i+1} to attain more accurate non-rigid registration. For the i -th deformable block, the 3D deformable grid G_i is the combination of G_{i+1} (with $2 \times$ upsampling) and ϕ_i (see Figure 2C). Figure 4 illustrates one example of the progressive deformation estimation and registration. It can be observed that the progressive deformation estimation can gradually refine the whole deformation field. The low-resolution deformation field ϕ_3 contains coarse and global deformation information, while the high-resolution deformation field ϕ_1 captures more detailed local displacements. Thus, the whole deformation field can attain accurate registration.

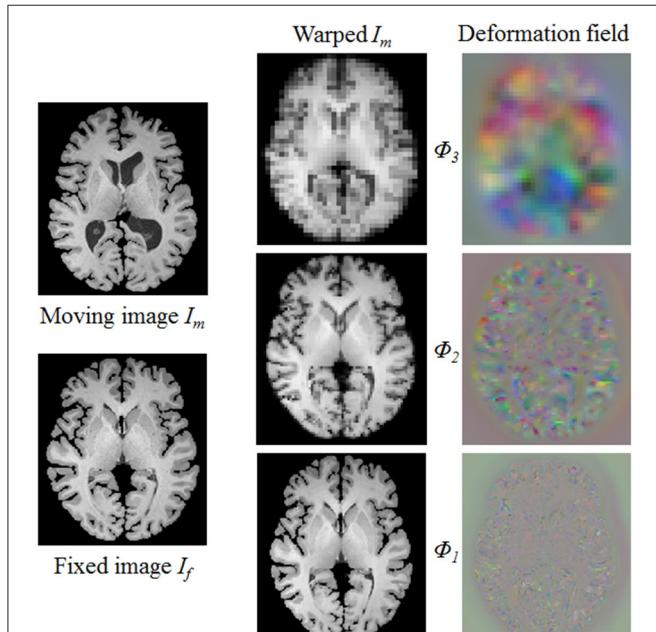
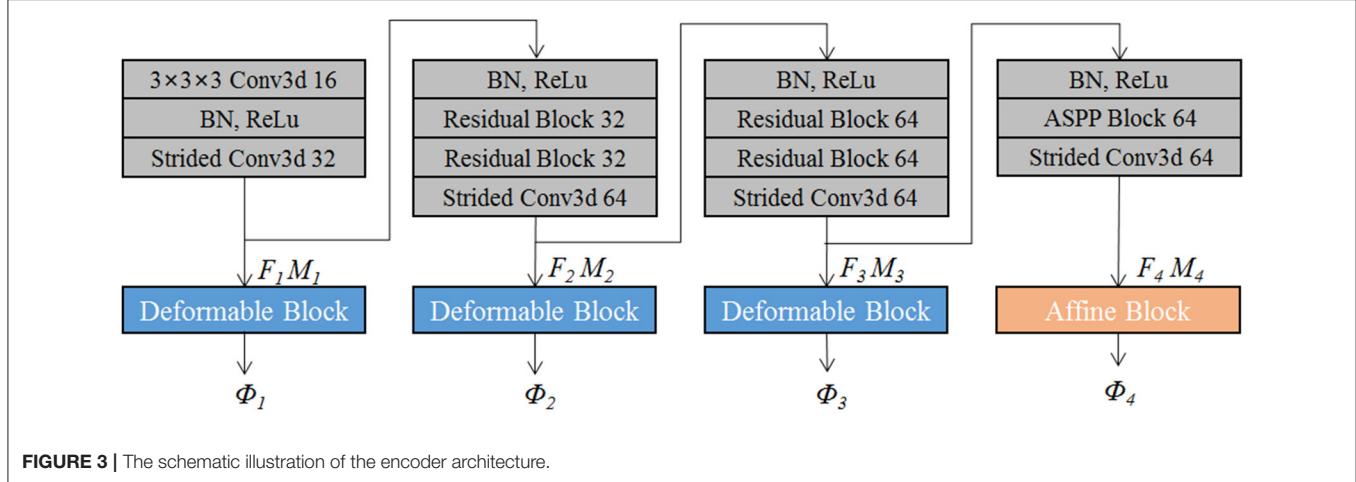


FIGURE 4 | One example to illustrate the multi-scale deformable registration fields and the procedure of progressive registration. It can be observed that the progressive deformation estimation can gradually generate more explicit local displacements, thus provide more and more accurate registration even for the large deformation case.

In summary, we utilize multi-level feature pyramids generated from paired volumes to estimate the multi-scale deformation fields. Our network generates the multi-scale deformation fields in a coarse-to-fine manner, which aggregates both high-level context information and low-level details. High-level context information is applied to the coarse and rigid alignment, while low-level details are devoted to the non-rigid registration.

2.3. Edge-Aware Dual Inputs

We integrate edge information with the original image as dual-inputs for each encoding stream (see **Figure 5**), which enhances the texture structures of image content, to impel the proposed network pay extra attention to the edge-aware information for structure alignment.

Considering the effectiveness and easy implementation of the Sobel edge detector, 3D Sobel operator is designed to extract the edges of original volume. The 3D Sobel operator contains three filtering kernels as S_x , S_y , and S_z . Each kernel is a $3 \times 3 \times 3$ tensor, and is responsible for the calculation of image gradient along x -/ y -/ z -axis. The kernel S_z is shown as an example:

$$S_z(:,:, -1) = \begin{bmatrix} +1 & +2 & +1 \\ +2 & +4 & +2 \\ +1 & +2 & +1 \end{bmatrix}, S_z(:,:, 0) = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \\ S_z(:,:, +1) = \begin{bmatrix} -1 & -2 & -1 \\ -2 & -4 & -2 \\ -1 & -2 & -1 \end{bmatrix}. \quad (4)$$

Kernels S_x and S_y are with the same kernel weights as S_z , but along different directions. S_x , S_y , and S_z are applied to convolve a volumetric image I , and further generate its corresponding edge map E as follows:

$$E = \sqrt{(S_x * I)^2 + (S_y * I)^2 + (S_z * I)^2}, \quad (5)$$

where $*$ denotes the convolution operation. The generated edge map is beneficial to impel the network leverage edge-aware information for structure alignment.

2.4. Training Loss

We adopt the patch-based cross-correlation (Rao et al., 2014) as the similarity function:

$$\mathcal{L}_{sim}(I_f, I_m) = \\ - \sum_{p \in \Omega} \frac{\left(\sum_{p_n} (I_f(p_n) - \bar{I}_f(p))(I_m(p_n) - \bar{I}_m(p)) \right)^2}{\sum_{p_n} (I_f(p_n) - \bar{I}_f(p)) \sum_{p_n} (I_m(p_n) - \bar{I}_m(p))}, \quad (6)$$

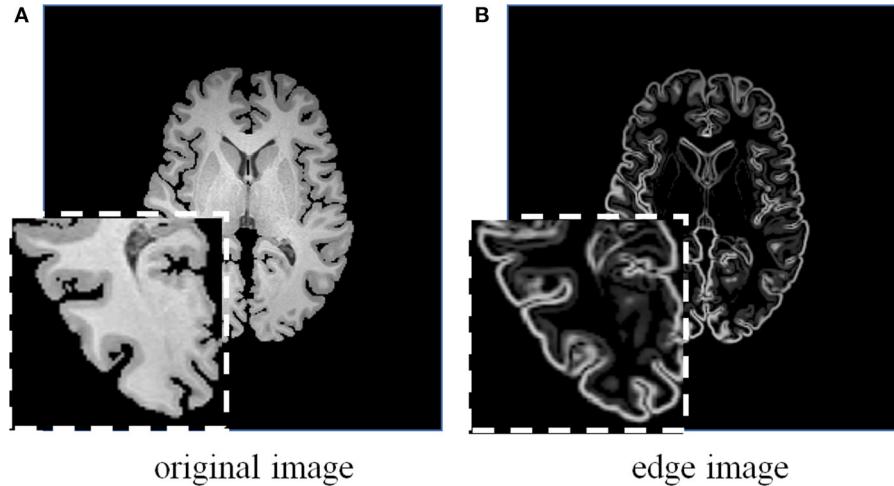


FIGURE 5 | A brain MRI image (**A**) and its corresponding edge map (**B**) generated using 3D Sobel operator. The edge map can more explicitly represent the structural information.

where \bar{I}_f and \bar{I}_m denote volumes with local mean intensities subtracted. $p \in \Omega$ denotes each voxel in volumetric image, where Ω is the whole image domain. Voxel p_n is the local neighborhood in v^3 ($v = 9$ in our implementation) volumetric patch at the center of voxel p .

To avoid obtaining an unpractical or discontinuous deformation field, we also add a diffusion regularizer \mathcal{L}_{smooth} to impose smooth constraint on the spatial gradients of the overall deformation field ϕ :

$$\mathcal{L}_{smooth}(\phi) = \sum_{p \in \Omega} \|\nabla \phi(p)\|^2, \quad (7)$$

where $\phi = \sum_{i=1}^4 upsample2^i(\phi_i)$ is the aggregation of multi-scale deformation fields.

As the deformation is estimated progressively, we consider similarity loss for each scale of the registration pyramid. Therefore, the total loss is defined as:

$$\mathcal{L} = \sum_{i=1}^4 \mathcal{L}_{sim}(down2^i(\bar{I}_f), down2^i(\bar{I}_m) \circ G_i) + \lambda \mathcal{L}_{smooth}(\phi), \quad (8)$$

where $down2^i$ denotes a down-sample operation with a factor of 2^i . G_i is the 3D deformation grid generated to warp the moving volume.

3. EXPERIMENTS AND RESULTS

3.1. Materials

The study protocol was reviewed and approved by the Institution's Ethical Review Board. Experiments were carried on three public brain MRI datasets with manually labeled region of interests (ROIs), including Mindboggle101 (Klein and Tourville, 2012), LPBA40 (Shattuck et al., 2008), and IXI30 (Serag et al., 2012).

- Mindboggle101 (101 T1-weighted MRI volumes): 62 volumes were involved to conduct experiments as described in Kuang and Schmah (2019). Specifically, 42 volumes (i.e., 1,722 pairs) from subsets of NKI-RS-22 and NKI-TRT-20 were used for training, and 20 volumes (i.e., 380 pairs) from OASIS-TRT-20 were involved for testing.
- LPBA40 (40 T1-weighted MRI volumes): 30 volumes were randomly selected for training and the remaining 10 volumes were used for testing.
- IXI30 (30 T1-weighted MRI volumes): all 30 volumes were used for testing. In order to investigate the generalization ability of the network, we employed the model trained on LPBA40 to register images from IXI30 dataset.

All volumes were pre-processed by histogram and intensity normalization, and skull-stripping using FreeSurfer (Fischl, 2012).

3.2. Implementation Details

In our experiments, each input volumetric image was resized into the dimension of $192 \times 192 \times 192$. The network was trained on a GPU of NVIDIA Tesla V100. The value of the regularization parameter λ was set empirically as 1000. For the whole registration network, the number of epochs was set to 300. The network was implemented using Pytorch and Adam optimization (Kingma and Ba, 2014), and the learning rate was initially set to 2e-4, with 0.5 weight decay after every 10 epoch.

3.3. Evaluation Metrics

To quantitatively evaluate the registration performance, Dice index (DSC) (Dice, 1945), Hausdorff distance (HD) (Huttenlocher et al., 1993), and average symmetric surface distance (ASSD) (Taha and Hanbury, 2015) were

calculated. The DSC is defined as:

$$DSC = \frac{2|R_{I_f} \cap R_{I_m}|}{|R_{I_f}| + |R_{I_m}|}, \quad (9)$$

where R_{I_f} and R_{I_m} are the segmented ROIs of I_f and I_m , respectively. The HD measures the longest distance over the shortest distances between the segmented ROIs of I_f and I_m . The ASSD is defined as:

$$ASSD = \frac{1}{|B_{I_f}| + |B_{I_m}|} \left(\sum_{x \in B_{I_f}} d(x, B_{I_m}) + \sum_{y \in B_{I_m}} d(y, B_{I_f}) \right) \quad (10)$$

where B_{I_f} and B_{I_m} are the segmented surfaces of I_f and I_m , respectively. The operator $d(.,)$ is the shortest Euclidean distance operator.

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

3.4. Registration Accuracy

We compared our EPReg network with four cutting-edge brain MRI registration schemes: SyN (Avants et al., 2008), VoxelMorph (Balakrishnan et al., 2018), FAIM (Kuang and Schmah, 2019), and MSNet (Duan et al., 2019). 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. In addition, we also compared the network without edge-aware input, which is denoted as PReg.

Figure 6 shows the visual comparisons from different registration methods on Mindboggle101 dataset. Our network can generate more accurate registered images, and the internal structures can be preserved consistently by using our network. **Table 1** further reports the numerical results on five regions of the images from dataset Mindboggle101. It can be observed that our EPReg consistently achieved best registration performance with respect to DSC and HD metrics. Regarding the ASSD evaluation, our network obtained the best ASSD values on occipital and temporal regions; and the second best ASSD on the frontal, parietal and cingulate regions. It is worth noting that both our EPReg and PReg networks outperformed other cutting-edge methods by a large margin in terms of DSC values, which demonstrates the proposed deformable pyramid contributed to the improvement of registrations.

For the LPBA40 dataset, the visualization and quantitative results (including seven regions) are shown in **Figure 7** and **Table 2**, respectively. It can be observed that our network achieved overall satisfactory registration performance on LPBA40 dataset. We further calculated DSC, HD, ASSD values of 54 corresponding sub-regions from warped and fixed volumes. The comparison results are illustrated in **Figure 8**. For the 54 sub-regions, the proposed EPReg achieved the best DSC, HD, and ASSD values on 41, 32, 38 sub-regions, respectively.

The numerical results on IXI30 are illustrated in **Figure 9**. IXI30 dataset has 95 subregions but 30 of them are extremely small regions. Thus we calculated DSC, HD, and ASSD values

of the remaining 65 subregions. It can be observed that the proposed EPReg achieved the best DSC, HD, and ASSD values on 49, 35, 49 sub-regions, respectively, which shows that our network has satisfactory generalization ability. **Figure 10** visualizes registered images from different methods. Our network again attained overall satisfactory registration performance on this dataset.

3.5. Statistical Analyses

To investigate the statistical significance of the proposed network over other compared registration methods, a student test was conducted. Specifically, the two-sample, two-tailed *t*-test was employed to pairwisely compare the registration performance between our method and the other five methods on three different datasets (see **Table 3**). It can be observed that the null hypotheses for the five comparing pairs on the metric of DSC were not accepted at the 0.05 level. As a result, our method can be regarded as significantly better than the other five compared methods on DSC metric. In addition, the null hypotheses for the pairs of SyN-EPReg, VM-EPReg, and FAIM-EPReg on all three metrics were not accepted at the 0.05 level, which demonstrates our method was significantly better than these three methods on all three metrics. The p-values of PReg-EPReg on metrics of HD and ASSD from dataset Mindboggle101 were beyond the 0.05 level, which indicates that our method and PReg achieved similar performance with regard to the HD and ASSD evaluation on dataset Mindboggle101. The pair of MSNet-EPReg held the similar results on metrics of HD and ASSD from dataset IXI. In general, the statistical analyses prove that our method had an overall better registration performance than other compared cutting-edge methods.

3.6. Comparison of Time Efficiency

We further compared the time efficiency. **Table 4** lists the inference time for registering a pair of images using different methods. It can be observed that for the affine alignment, the time spent by our affine component (less than 0.3 s) was much less than that of the traditional affine alignment method (i.e., ANTS, Avants et al., 2008) using iterative optimization. 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, Avants et al., 2008) before deformable registration.

4. DISCUSSION

Deformable registration is an important task of medical image computing and has various clinical applications. It is to search for the inhomogeneous point-wise displacements to match homologous locations from the moving domain to the fixed domain. Due to the large search space for the complicated non-rigid deformation, most existing registration schemes require separate rigid alignment before deformable registration to reduce the search space, or iteratively optimize the estimated deformation field. Even so, they may not well handle the large deformation cases. We have attempted to tackle this issue by devising a dual-stream deformable pyramid architecture.

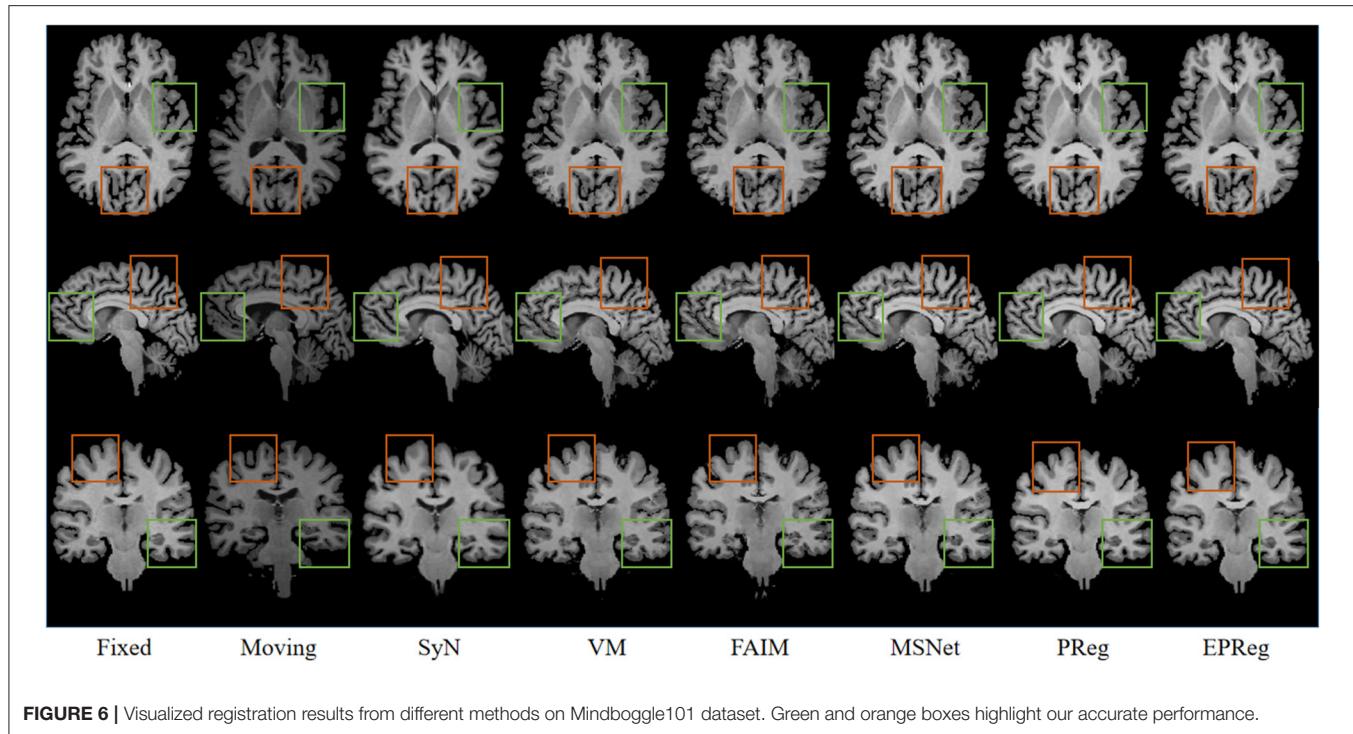


FIGURE 6 | Visualized registration results from different methods on Mindboggle101 dataset. Green and orange boxes highlight our accurate performance.

TABLE 1 | The DSC (%), Hausdorff distance (HD), and average symmetric surface distance (ASSD) results (mean \pm SD) from SyN (Avants et al., 2008), VM (Balakrishnan et al., 2018), FAIM (Kuang and Schmah, 2019), MSNet (Duan et al., 2019), and our network on MindBoggle101 dataset.

	Methods	Frontal	Parietal	Occipital	Temporal	Cingulate
DSC (%)	SyN	54.4 \pm 4.5	46.8 \pm 6.8	49.8 \pm 5.0	48.1 \pm 6.2	49.2 \pm 9.0
	VM	53.4 \pm 8.1	52.7 \pm 6.2	51.0 \pm 7.3	43.3 \pm 7.6	48.3 \pm 9.2
	FAIM	57.2 \pm 6.8	55.1 \pm 7.1	53.7 \pm 6.5	46.9 \pm 6.9	50.8 \pm 9.5
	MSNet	58.3 \pm 7.1	50.4 \pm 7.3	55.4 \pm 6.9	47.7 \pm 7.9	54.3 \pm 8.9
	PReg	65.6 \pm 9.3	60.5 \pm 8.1	59.9 \pm 7.5	67.5 \pm 7.0	62.4 \pm 11.2
	EPReg	67.1 \pm 9.0	61.5 \pm 7.5	60.6 \pm 7.0	67.9 \pm 7.2	64.9 \pm 9.5
HD	SyN	13.2 \pm 2.2	14.0 \pm 1.9	14.8 \pm 3.8	6.8 \pm 2.3	7.9 \pm 2.0
	VM	13.2 \pm 3.6	13.6 \pm 1.8	13.5 \pm 4.6	8.6 \pm 3.0	7.9 \pm 1.9
	FAIM	12.9 \pm 2.7	13.3 \pm 1.7	13.3 \pm 4.4	8.7 \pm 2.3	7.8 \pm 2.1
	MSNet	12.6 \pm 2.9	13.2 \pm 2.0	13.7 \pm 4.4	8.0 \pm 2.4	8.0 \pm 2.2
	PReg	12.7 \pm 3.2	13.5 \pm 2.8	13.1 \pm 3.7	7.4 \pm 2.5	8.4 \pm 2.1
	EPReg	12.3 \pm 3.1	13.2 \pm 2.5	12.9 \pm 3.6	7.9 \pm 2.3	7.8 \pm 2.0
ASSD	SyN	1.32 \pm 0.53	1.45 \pm 0.32	1.48 \pm 0.35	1.11 \pm 0.27	1.37 \pm 0.66
	VM	1.51 \pm 0.33	1.32 \pm 0.33	1.36 \pm 0.32	1.32 \pm 0.72	1.43 \pm 0.43
	FAIM	1.42 \pm 0.32	1.29 \pm 0.28	1.44 \pm 0.66	1.12 \pm 0.46	1.20 \pm 0.58
	MSNet	1.34 \pm 0.28	1.29 \pm 0.23	1.39 \pm 0.53	1.06 \pm 0.53	1.20 \pm 0.63
	PReg	0.42 \pm 0.13	0.73 \pm 0.22	0.96 \pm 0.32	0.62 \pm 0.21	1.00 \pm 0.37
	EPReg	0.43 \pm 0.14	0.75 \pm 0.26	0.93 \pm 0.30	0.61 \pm 0.21	1.03 \pm 0.39

Best results are highlighted in bold.

The proposed deformable pyramid leverages multi-scale paired features to progressively estimate residual deformation field with a reduced search space rather than a large one, which facilitates the estimation of large/complicated deformation field. In addition, the affine alignment is also integrated within the deformable pyramid in a seamless manner, thus enables

the trained network performing non-rigid registration in one forward pass.

Most unsupervised registration networks optimize deformation fields based on the maximization of the intensity-based image similarity. Considering the boundary/shape information is commonly used to constrain the registration,

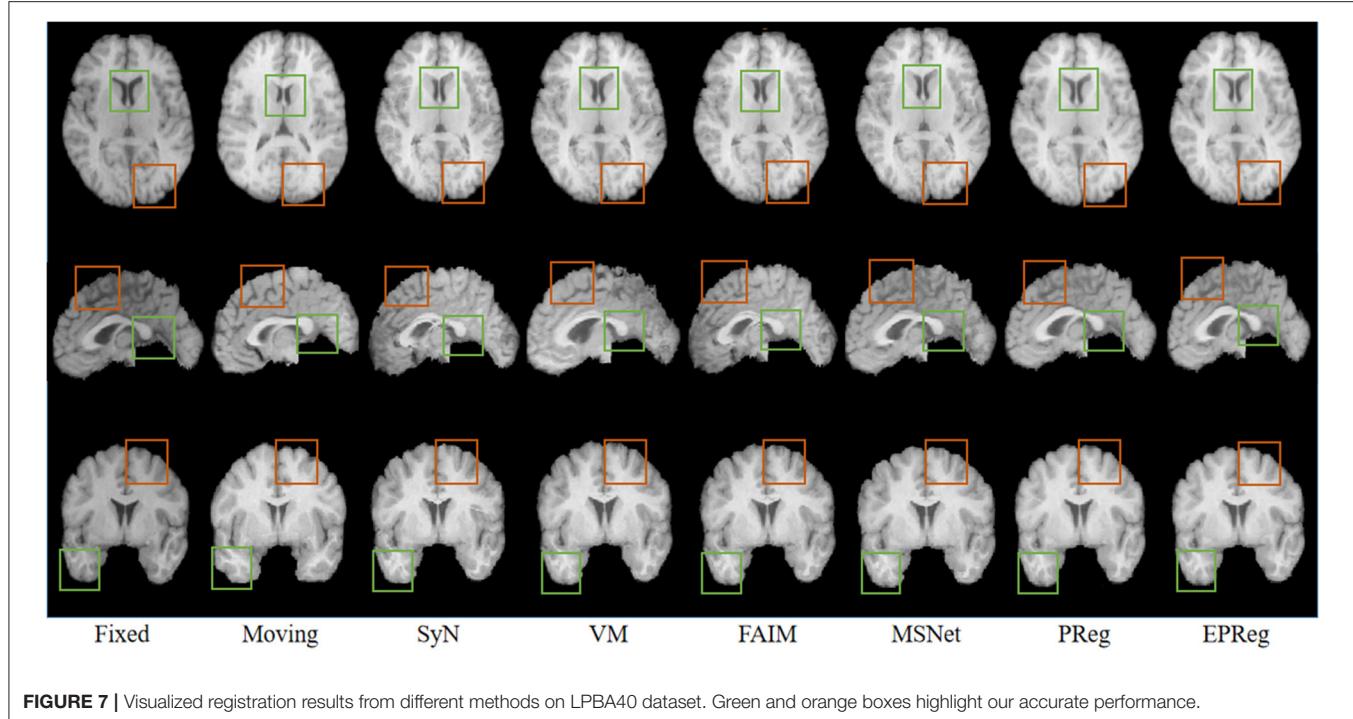


FIGURE 7 | Visualized registration results from different methods on LPBA40 dataset. Green and orange boxes highlight our accurate performance.

TABLE 2 | The DSC (%), Hausdorff distance (HD), and average symmetric surface distance (ASSD) results (mean \pm SD) from SyN (Avants et al., 2008), VM (Balakrishnan et al., 2018), FAIM (Kuang and Schmah, 2019), MSNet (Duan et al., 2019), and our network on LPBA40 dataset.

	Methods	Frontal	Parietal	Occipital	Temporal	Cingulate	Putamen	Hippocampus
DSC (%)	SyN	68.1 \pm 4.0	57.1 \pm 5.8	50.7 \pm 6.8	58.2 \pm 4.9	58.9 \pm 6.2	64.2 \pm 7.3	65.2 \pm 7.0
	VM	71.2 \pm 3.1	57.4 \pm 7.0	58.8 \pm 6.0	61.7 \pm 4.4	54.9 \pm 6.2	66.1 \pm 9.4	60.9 \pm 9.0
	FAIM	73.1 \pm 3.1	59.4 \pm 7.3	60.6 \pm 6.1	65.1 \pm 4.2	61.2 \pm 5.8	70.8 \pm 10.0	68.0 \pm 6.4
	MSNet	73.4 \pm 3.3	60.0 \pm 7.0	61.0 \pm 6.1	65.9 \pm 4.0	62.1 \pm 5.5	72.7 \pm 8.5	68.8 \pm 5.7
	PReg	76.8 \pm 2.7	66.6 \pm 6.2	66.7 \pm 4.8	71.6 \pm 3.4	65.9 \pm 6.6	78.6 \pm 2.6	75.2 \pm 2.7
	EPReg	77.0 \pm 2.6	68.3 \pm 4.5	68.7 \pm 3.7	72.7 \pm 2.7	66.7 \pm 5.4	79.1 \pm 2.5	75.3 \pm 2.8
HD	SyN	13.2 \pm 1.8	15.8 \pm 2.5	16.2 \pm 2.9	16.9 \pm 3.0	13.0 \pm 3.1	10.9 \pm 9.9	7.1 \pm 1.4
	VM	13.6 \pm 1.9	19.2 \pm 5.0	15.1 \pm 2.9	16.8 \pm 3.4	13.6 \pm 2.8	8.9 \pm 2.7	8.5 \pm 2.0
	FAIM	13.5 \pm 2.0	19.1 \pm 4.8	15.1 \pm 3.0	16.5 \pm 3.4	13.2 \pm 2.6	8.4 \pm 2.9	7.9 \pm 1.9
	MSNet	13.5 \pm 1.9	19.0 \pm 4.8	15.0 \pm 2.9	16.4 \pm 3.4	13.1 \pm 2.7	8.2 \pm 3.1	7.6 \pm 1.8
	PReg	12.3 \pm 1.5	18.1 \pm 4.8	14.6 \pm 2.9	15.8 \pm 3.4	13.2 \pm 2.7	7.6 \pm 3.1	6.4 \pm 1.8
	EPReg	12.3 \pm 1.4	17.5 \pm 4.0	14.1 \pm 2.3	15.0 \pm 2.7	12.5 \pm 2.1	7.4 \pm 3.2	6.2 \pm 1.3
ASSD	SyN	2.44 \pm 0.44	2.71 \pm 0.40	3.18 \pm 0.58	2.55 \pm 0.40	2.22 \pm 0.31	1.70 \pm 0.63	1.56 \pm 0.33
	VM	1.41 \pm 0.27	2.11 \pm 0.41	1.87 \pm 0.30	1.49 \pm 0.19	2.02 \pm 0.31	1.28 \pm 0.32	1.41 \pm 0.32
	FAIM	1.35 \pm 0.27	2.06 \pm 0.41	1.86 \pm 0.32	1.38 \pm 0.18	1.87 \pm 0.29	1.12 \pm 0.32	1.22 \pm 0.23
	MSNet	1.31 \pm 0.26	2.04 \pm 0.41	1.86 \pm 0.31	1.35 \pm 0.18	1.82 \pm 0.27	1.05 \pm 0.28	1.18 \pm 0.21
	PReg	1.11 \pm 0.23	1.72 \pm 0.37	1.60 \pm 0.21	1.16 \pm 0.15	1.68 \pm 0.31	0.85 \pm 0.17	0.98 \pm 0.11
	EPReg	1.08 \pm 0.19	1.62 \pm 0.26	1.54 \pm 0.18	1.11 \pm 0.11	1.65 \pm 0.25	0.85 \pm 0.16	0.96 \pm 0.11

weakly supervised registration networks pay extra attention on leveraging the correspondences between structural information. However, obtaining such corresponding structures requires manual annotations. Instead of utilizing the structural information to form an explicit loss function, we simply apply the Sobel edge map as an extra input of the network. The purpose is to enhance the texture

structures of image content, thus to pay attention for the edge-aware alignment. The comparison results listed in Tables 1, 2 (evaluations on some relatively larger regions) show that the network with edge-aware input (i.e., EPReg) achieved overall better registration performance than PReg. The comparisons on lots of small sub-regions shown in Figures 8, 9 demonstrate that EPReg attained more

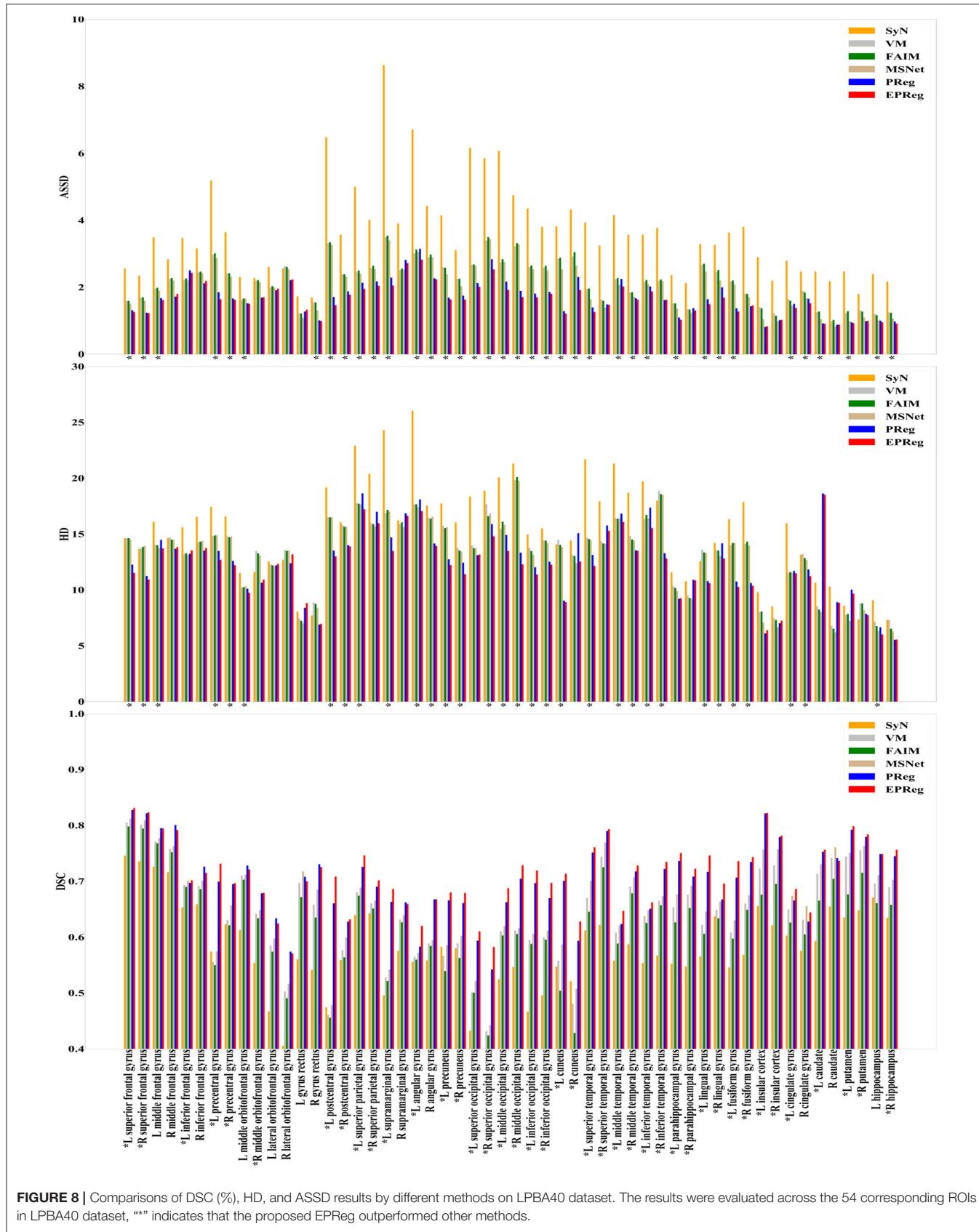


FIGURE 8 | Comparisons of DSC (%) and HD, ASSD results by different methods on LPBA40 dataset. The results were evaluated across the 54 corresponding ROIs in LPBA40 dataset, “*” indicates that the proposed EPReg outperformed other methods.



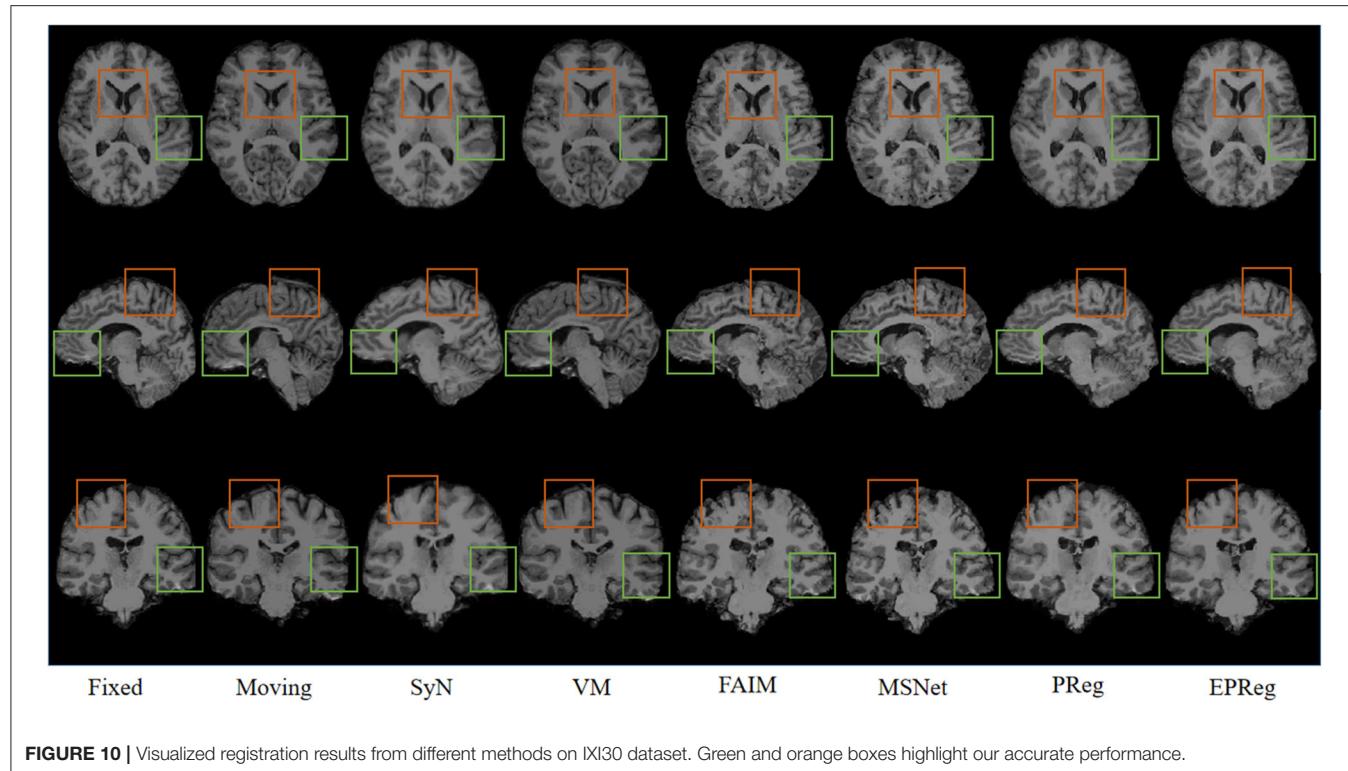


FIGURE 10 | Visualized registration results from different methods on IXI30 dataset. Green and orange boxes highlight our accurate performance.

TABLE 3 | *P*-values of student tests between different methods on different metrics.

Method	Mindboggle101			LPBA40			IXI		
	DSC	HD	ASSD	DSC	HD	ASSD	DSC	HD	ASSD
SyN-EPReg	8.35e-13	4.01e-02	9.02e-06	3.84e-05	2.13e-07	4.79e-03	1.30e-03	3.15e-02	7.60e-03
VM-EPReg	1.47e-14	2.45e-03	9.80e-09	6.93e-16	7.16e-16	7.88e-09	1.11e-04	2.16e-02	1.17e-04
FAIM-EPReg	9.70e-17	1.72e-03	5.32e-08	1.22e-14	5.73e-16	2.50e-08	6.09e-05	4.54e-02	6.72e-03
MSNet-EPReg	5.93e-12	2.08e-03	6.91e-08	3.56e-10	2.07e-16	1.58e-10	1.56e-03	0.91	0.71
PReg-EPReg	1.84e-06	0.63	0.65	2.00e-05	3.74e-05	2.40e-05	4.50e-03	6.89e-03	2.49e-02

TABLE 4 | Inference time (second) for registering a pair of images using different methods.

Methods	Mindboggle101	LPBA40	IXI
ANTs (affine alignment)	8.19 ± 0.30	7.73 ± 0.27	9.13 ± 0.29
SyN	39.24 ± 2.07	33.32 ± 1.46	40.10 ± 2.01
VM	0.66 ± 0.01	0.11 ± 0.01	0.73 ± 0.01
FAIM	1.17 ± 0.01	0.49 ± 0.01	1.26 ± 0.01
MSNet	1.34 ± 0.01	0.45 ± 0.01	1.41 ± 0.01
EPReg (affine only)	0.09 ± 0.01	0.09 ± 0.01	0.23 ± 0.01
EPReg	0.25 ± 0.01	0.23 ± 0.01	0.61 ± 0.01

accurate registration than PReg, due to the usage of the edge-aware information.

We evaluated the proposed edge-aware pyramidal deformable network on three different brain MR datasets and compared with four cutting-edge registration methods. The DSC, HD, and ASSD were employed to evaluate all methods. Specifically,

HD and ASSD were employed to provide evaluation on the differences in boundary shape between two volumes. The numerical results and statistical analyses show that our method had an overall better registration accuracy than other compared cutting-edge methods. Furthermore, our registration network provided very efficient inference procedure,

which achieved accurate volumetric registration within 1 s. From the experimental results, it can be demonstrated that the proposed progressive deformation estimation scheme contributed to the improvement of registration accuracy and efficiency.

The proposed registration method focuses on the large deformation estimation by progressively predicting the residual deformation. It is beneficial for the estimation of large-scale deformation. Further validations on other body parts (e.g., chest or abdominal CT images) will be our future work.

The problem of registration validation in clinical settings is still an open issue. Most of recent research articles focusing on developing new registration approaches employ DSC as a primary metric to evaluate the registration accuracy (Balakrishnan et al., 2018; Cao et al., 2018; Loi et al., 2018, 2020; Fan et al., 2019; Huang et al., 2021). For a fair comparison, we also applied DSC to evaluate the registration performance. However, the DSC may suffer of the limitation to be dependent on the volume of structures. Thus we further employed HD and ASSD to evaluate boundary differences between two volumetric regions. Although there is no guaranteed thresholds w.r.t. DSC, HD, ASSD for quality assessment of registration on brain MR images, the comparison results on a series of sub-regions (over 50 sub-regions) show the proposed network consistently outperformed other cutting-edge registration methods with respect to metrics of DSC, HD, and ASSD. In addition, the visual results in **Figures 6, 7, 10** illustrate the satisfactory registration performance obtained by our method. In our future work, the phantom study which could generate the ground truth deformation fields for straightforward registration validation would be conducted.

5. CONCLUSION

In this paper, we have presented an **edge-aware pyramidal deformable network for unsupervised volumetric registration**. The proposed network focuses on the large deformation estimation by progressively predicting the residual deformation. Our key idea is to fully exploit the useful complementary

multi-level information from paired features to predict multi-scale deformation fields. We achieve this by developing a deformable pyramid architecture, which can generate multi-scale paired feature maps for progressively transforming the paired information into more representative features to predict more accurate deformation field. In addition, we leverage extra edge information to impel the network pay attention to the edge-aware alignment. Extensive experiments on several 3D MRI datasets demonstrate that our edge-aware pyramidal deformable network achieves satisfactory registration performance. The coarse-to-fine progressive registration procedure is beneficial to compensate for the large-scale deformation, and can be regarded as a general solution for deformable volumetric registration.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

YC, ZZ, YR, CQ, DL, QD, DN, and YW response for study design. YC and CQ implemented the research. YC, ZZ, QD, and YW conceived the experiments. YC, ZZ, and YR conducted the experiments. YC, ZZ, YR, DL, and YW analyzed the results. YC and YW wrote the main manuscript text and prepared the figures. All authors reviewed the manuscript.

FUNDING

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).

REFERENCES

- Auzias, G., Colliot, O., Glaunes, J. A., Perrot, M., Mangin, J.-F., Trouve, A., et al. (2011). Diffeomorphic brain registration under exhaustive sulcal constraints. *IEEE Trans. Med. Imaging* 30, 1214–1227. doi: 10.1109/TMI.2011.2108665
- Avants, B. B., Epstein, C. L., Grossman, M., and Gee, J. C. (2008). Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. *Med. Image Anal.* 12, 26–41. doi: 10.1016/j.media.2007.06.004
- Balakrishnan, G., Zhao, A., Sabuncu, M. R., Guttag, J., and Dalca, A. V. (2018). “An unsupervised learning model for deformable medical image registration,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Salt Lake City, UT), 9252–9260.
- Beg, M. F., Miller, M. I., Trouvé, A., and Younes, L. (2005). Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *Int. J. Comput. Vision* 61, 139–157. doi: 10.1023/B:VISI.0000043755.93987.aa
- Cao, X., Yang, J., Zhang, J., Nie, D., Kim, M., Wang, Q., et al. (2017). “Deformable image registration based on similarity-steered cnn regression,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Quebec City, QC: Springer), 300–308.
- Cao, X., Yang, J., Zhang, J., Wang, Q., Yap, P. T., and Shen, D. (2018). Deformable image registration using a cue-aware deep regression network. *IEEE Trans. Biomed. Eng.* 65, 1900–1911. doi: 10.1109/TBME.2018.2822826
- de Vos, B. D., Berendsen, F. F., Viergever, M. A., Sokooti, H., Staring, M., and Isgum, I. (2019). A deep learning framework for unsupervised affine and deformable image registration. *Med. Image Anal.* 52, 128–143. doi: 10.1016/j.media.2018.11.010
- Dice, L. R. (1945). Measures of the amount of ecological association between species. *Ecology* 26, 297–302.
- Duan, L., Yuan, G., Gong, L., Fu, T., Yang, X., Chen, X., et al. (2019). Adversarial learning for deformable registration of brain mr image using

- a multi-scale fully convolutional network. *Biomed. Signal Process. Control* 53:101562. doi: 10.1016/j.bspc.2019.101562
- Epplenhoef, K. A., Lafarge, M. W., Moeskops, P., Veta, M., and Pluim, J. P. (2018). “Deformable image registration using convolutional neural networks,” in *Medical Imaging 2018: Image Processing*, Vol. 10574 (San Francisco, CA: International Society for Optics and Photonics).
- Fan, J., Cao, X., Yap, P.-T., and Shen, D. (2019). Birnet: brain image registration using dual-supervised fully convolutional networks. *Med. Image Anal.* 54, 193–206. doi: 10.1016/j.media.2019.03.006
- Fischl, B. (2012). Freesurfer. *Neuroimage* 62, 774–781. doi: 10.1016/j.neuroimage.2012.01.021
- Haskins, G., Kruger, U., and Yan, P. (2020). Deep learning in medical image registration: a survey. *Mach. Vis. Appl.* 31:8. doi: 10.1007/s00138-020-01060-x
- He, K., Zhang, X., Ren, S., and Sun, J. (2016). “Deep residual learning for image recognition,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (Las Vegas, NV: IEEE Computer Society), 770–778.
- Hu, X., Kang, M., Huang, W., Scott, M. R., Wiest, R., and Reyes, M. (2019). “Dual-stream pyramid registration network,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Shenzhen: Springer), 382–390.
- Hu, Y., Modat, M., Gibson, E., Li, W., Ghavami, N., Bonmati, E., et al. (2018). Weakly-supervised convolutional neural networks for multimodal image registration. *Med. Image Anal.* 49, 1–13. doi: 10.1016/j.media.2018.07.002
- Huang, Y., Ahmad, S., Fan, J., Shen, D., and Yap, P.-T. (2021). Difficulty-aware hierarchical convolutional neural networks for deformable registration of brain mr images. *Med. Image Anal.* 67:101817. doi: 10.1016/j.media.2020.101817
- Huttenlocher, D. P., Klanderman, G. A., and Ruckridge, W. J. (1993). Comparing images using the hausdorff distance. *IEEE Trans. Pattern Anal. Mach. Intell.* 15, 850–863. doi: 10.1109/34.232073
- Jaderberg, M., Simonyan, K., Zisserman, A., and Kavukcuoglu, K. (2015). “Spatial transformer networks,” in *Advances in Neural Information Processing Systems* (Montreal, QC), 2017–2025.
- Kingma, D. P., and Ba, J. (2014). Adam: a method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
- Klein, A., Andersson, J., Ardekani, B. A., Ashburner, J., Avants, B., Chiang, M.-C., et al. (2009). Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. *Neuroimage* 46, 786–802. doi: 10.1016/j.neuroimage.2008.12.037
- Klein, A., and Tourville, J. (2012). 101 labeled brain images and a consistent human cortical labeling protocol. *Front. Neurosci.* 6:171. doi: 10.3389/fnins.2012.00171
- Knops, Z. F., Maintz, J. A., Viergever, M. A., and Pluim, J. P. (2006). Normalized mutual information based registration using k-means clustering and shading correction. *Med. Image Anal.* 10, 432–439. doi: 10.1016/j.media.2005.03.009
- Kuang, D., and Schmah, T. (2019). “Faim—a convnet method for unsupervised 3d medical image registration,” in *International Workshop on Machine Learning in Medical Imaging* (Shenzhen: Springer), 646–654.
- Li, H., and Fan, Y. (2018). “Non-rigid image registration using self-supervised fully convolutional networks without training data,” in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)* (Washington, DC: IEEE), 1075–1078.
- Loi, G., Fusella, M., Lanzi, E., Cagni, E., Garibaldi, C., Iacoviello, G., et al. (2018). Performance of commercially available deformable image registration platforms for contour propagation using patient-based computational phantoms: a multi-institutional study. *Med. Phys.* 45, 748–757. doi: 10.1002/mp.12737
- Loi, G., Fusella, M., Vecchi, C., Menna, S., Rosica, F., Gino, E., et al. (2020). Computed tomography to cone beam computed tomography deformable image registration for contour propagation using head and neck, patient-based computational phantoms: a multicenter study. *Pract. Radiat. Oncol.* 10, 125–132. doi: 10.1016/j.prro.2019.11.011
- Rao, Y. R., Prathapani, N., and Nagabhooshanam, E. (2014). Application of normalized cross correlation to image registration. *Int. J. Res. Eng. Technol.* 3, 12–16. doi: 10.15623/ijret.2014.0317003
- Rohé, M.-M., Datar, M., Heimann, T., Sermesant, M., and Pennec, X. (2017). “Svf-net: Learning deformable image registration using shape matching,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Quebec City, QC: Springer), 266–274.
- Ronneberger, O., Fischer, P., and Brox, T. (2015). “U-net: convolutional networks for biomedical image segmentation,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Munich: Springer), 234–241.
- Serag, A., Aljabar, P., Ball, G., Counsell, S. J., Boardman, J. P., Rutherford, M. A., et al. (2012). Construction of a consistent high-definition spatio-temporal atlas of the developing brain using adaptive kernel regression. *Neuroimage* 59, 2255–2265. doi: 10.1016/j.neuroimage.2011.09.062
- Shattuck, D. W., Mirza, M., Adisetiyo, V., Hojatkashani, C., Salamon, G., Narr, K. L., et al. (2008). Construction of a 3d probabilistic atlas of human cortical structures. *Neuroimage* 39, 1064–1080. doi: 10.1016/j.neuroimage.2007.09.031
- Sheikhjafari, A., Noga, M., Punithakumar, K., and Ray, N. (2018). Unsupervised deformable image registration with fully connected generative neural network. in *1st Conference on Medical Imaging with Deep Learning (MIDL 2018)* (Amsterdam), 1–8.
- Shen, D., Wu, G., and Suk, H.-I. (2017). Deep learning in medical image analysis. *Annu. Rev. Biomed. Eng.* 19, 221–248. doi: 10.1146/annurev-bioeng-071516-044442
- Sokooti, H., De Vos, B., Berendsen, F., Lelieveldt, B. P., Isgum, I., and Staring, M. (2017). “Nonrigid image registration using multi-scale 3d convolutional neural networks,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Quebec City, QC: Springer), 232–239.
- Sotiras, A., Davatzikos, C., and Paragios, N. (2013). Deformable medical image registration: a survey. *IEEE Trans. Med. Imaging* 32, 1153–1190. doi: 10.1109/TMI.2013.2265603
- Taha, A. A., and Hanbury, A. (2015). Metrics for evaluating 3d medical image segmentation: analysis, selection, and tool. *BMC Med. Imaging* 15:29. doi: 10.1186/s12880-015-0068-x
- Uzunova, H., Wilms, M., Handels, H., and Ehrhardt, J. (2017). “Training cnns for image registration from few samples with model-based data augmentation,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Quebec City, QC: Springer), 223–231.
- Vercauteren, T., Pennec, X., Perchant, A., and Ayache, N. (2009). Diffeomorphic demons: efficient non-parametric image registration. *Neuroimage* 45, S61–S72. doi: 10.1016/j.neuroimage.2008.10.040
- Wang, Y., Dou, H., Hu, X., Zhu, L., Yang, X., Xu, M., et al. (2019). Deep attentive features for prostate segmentation in 3d transrectal ultrasound. *IEEE Trans. Med. Imaging* 38, 2768–2778. doi: 10.1109/TMI.2019.2913184
- Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.* 13, 600–612. doi: 10.1109/TIP.2003.819861
- Wolberg, G., and Zokai, S. (2000). “Robust image registration using log-polar transform,” in *Proceedings 2000 International Conference on Image Processing (Cat. No. 00CH37101)*, Vol. 1 (Vancouver, BC: IEEE), 493–496.
- Yang, X., Kwitt, R., Styner, M., and Niethammer, M. (2017). Quicksilver: fast predictive image registration—a deep learning approach. *Neuroimage* 158, 378–396. doi: 10.1016/j.neuroimage.2017.07.008
- Zhao, S., Dong, Y., Chang, E. I., Xu, Y., et al. (2019a). “Recursive cascaded networks for unsupervised medical image registration,” in *Proceedings of the IEEE International Conference on Computer Vision* (Seoul), 10600–10610.
- Zhao, S., Lau, T., Luo, J., Eric, I., Chang, C., and Xu, Y. (2019b). Unsupervised 3d end-to-end medical image registration with volume tweening network. *IEEE J. Biomed. Health Inform.* 24, 1394–1404. doi: 10.1109/JBHI.2019.2951024

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Cao, Zhu, Rao, Qin, Lin, Dou, Ni and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.