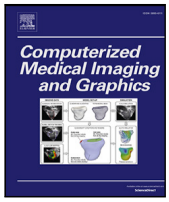
Computerized Medical Imaging and Graphics 104 (2023) 102184

Contents lists available at ScienceDirect

Computerized Medical Imaging and Graphics

journal homepage: www.elsevier.com/locate/compmedimag



Symmetric pyramid network for medical image inverse consistent diffeomorphic registration

Liutong Zhang, Guochen Ning, Lei Zhou, Hongen Liao *

Department of Biomedical Engineering, School of Medicine, Tsinghua University, Beijing, China

ARTICLE INFO

Keywords:
Symmetric diffeomorphic registration
Feature pyramid
Symmetric similarity
Inverse consistency

ABSTRACT

Over the past few years, deep learning-based image registration methods have achieved remarkable performance in medical image analysis. However, many existing methods struggle to ensure accurate registration while preserving the desired diffeomorphic properties and inverse consistency of the final deformation field. To address the problem, this paper presents a novel symmetric pyramid network for medical image inverse consistent diffeomorphic registration. Specifically, we first encode the multi-scale images to the feature pyramids via a shared-weights encoder network and then progressively conduct the feature-level diffeomorphic registration. The feature-level registration is implemented symmetrically to ensure inverse consistency. We independently carry out the forward and backward feature-level registration and average the estimated bidirectional velocity fields for more robust estimation. Finally, we employ symmetric multi-scale similarity loss to train the network. Experimental results on three public datasets, including Mindboggle101, CANDI, and OAI, show that our method significantly outperforms others, demonstrating that the proposed network can achieve accurate alignment and generate the deformation fields with expected properties. Our code will be available at https://github.com/zhangliutong/SPnet.

1. Introduction

Deformable image registration that serves as a crucial and fundamental problem in medical image analysis has received significant attention in the past few decades. The purpose of deformable image registration is to find the non-linear correspondences between a pair of images, allowing us to align images from different scan times, different devices, and different individuals to assist medical analysis and diagnosis. For instance, in brain image analysis, it is widely used to build brain atlas, compare brain activity, and monitor tumor growth (Chakravarty et al., 2006; Stefanescu et al., 2004). Traditional methods (Thirion, 1998; Rueckert et al., 1999; Glocker et al., 2008; Vercauteren et al., 2009; Beg et al., 2005; Avants et al., 2008) generally define a transformation model with a similarity measure and then iteratively optimize the model parameters. However, the optimization process is usually computationally intensive and time-consuming.

Recent deep learning-based image registration (DLIR) methods utilize the convolution neural network (CNN) to estimate the voxel-wise spatial correspondence and achieve fast and comparable registration results. Unsupervised methods (de Vos et al., 2017; Balakrishnan et al., 2018; Dalca et al., 2018; Kuang and Schmah, 2019) can train the network directly using the similarity measure without the ground truth deformation fields. Studies (Jiang et al., 2020; Hering et al., 2019, 2021; Fu et al., 2020; Li and Fan, 2020; Mok and Chung, 2020b) conduct multi-scale registration based on image pyramid to achieve more accurate alignment. However, these methods introduce a subnetwork at each scale and often train the networks separately or with the multi-stage training strategy. Notably, most of them cannot guarantee the diffeomorphic properties, such as invertibility and topology-preserving, limiting their clinical application.

In addition, the above approaches are formulated in asymmetrically, ignoring the inverse consistency, which plays a vital role in analyzing the variation of subtle anatomies (Gu et al., 2020). The inverse consistent registration implies that the forward and backward transformations estimated between an image pair should share the same pathway so that the composition of the bidirectional deformations should be identical. It should be noticed that the inverse consistency is not equal to invertibility, and diffeomorphic algorithms are not inverse consistent by default (Yang et al., 2008). Recent studies introduce the inverse consistent loss (Gu et al., 2020; Shen et al., 2019; Zhang et al., 2020) or adopt the "mean" shape (Mok and Chung, 2020a) to constrain the bidirectional registration results. Although these methods significantly reduce the inverse consistent error, it still has room for improvement.

* Corresponding author.
E-mail address: liao@tsinghua.edu.cn (H. Liao).

https://doi.org/10.1016/j.compmedimag.2023.102184
Received 25 March 2022; Received in revised form 31 December 2022; Accepted 3 January 2023
Available online 12 January 2023

To address the above challenges, we propose a symmetric pyramid network for medical image inverse consistent diffeomorphic registration. In our network, the multi-scale information of the input images is first encoded to the feature pyramids, respectively. Then we conduct the feature-level diffeomorphic registration progressively to achieve the coarse-to-fine alignment. The feature-level registration is implemented symmetrically. Specifically, we employ the shared-weights submodule to estimate the forward and backward feature registration velocity fields. The bidirectional velocity fields are then averaged for more robust estimation. Finally, the network is trained by the symmetric multi-scale similarity loss. Considering that our network is entirely symmetrical, we can generate the bidirectional results simultaneously, and the results are unrelated to the order of input images. To verify the performance of our proposed network, we implement the experiments on multiple datasets, including adult brain MRI, adolescent brain MRI, and knee MRI. The qualitative and quantitative results show that our network achieves more accurate alignment and maintains the desirable diffeomorphic properties and inverse consistency.

This paper is a significant extension of our previous work (Zhang et al., 2021) regarding the following aspects: The multi-scale image encoding is introduced to enhance the feature representation; We implement the registration in diffeomorphism space to ensure the diffeomorphic properties; The feature-level registration is conducted symmetrically to guarantee the inverse consistency; We employ more datasets, comparison methods, and evaluation metrics to assess the proposed network accurately and comprehensively. The main contributions can be summarized as follows:

- We propose a symmetric pyramid network for unsupervised medical image registration, aiming to achieve accurate alignment while maintaining the desirable diffeomorphic properties and inverse consistency.
- We encode multi-scale images to the feature pyramids and conduct the feature-level diffeomorphic registration in a coarse-to-fine manner.
- We implement the feature-level registration symmetrically by averaging the bidirectional velocity fields and introduce symmetric multi-scale similarity for optimization.
- We carry out experiments on three public datasets, and our method shows its superiority over others on multiple metrics, demonstrating the effectiveness and generalization of our network.

2. Related works

2.1. Classical deformable image registration

Classical deformable image registration methods often predefine a transformation model and iteratively optimize the model with a cost function. Some studies parameterized the problem based on the displacement field, such as Demons (Thirion, 1998), free form deformations with b-splines (Rueckert et al., 1999), and dense image registration with Markov Random Field (Glocker et al., 2008). Other studies conducted the registration in diffeomorphism space to guarantee the expected properties like continuity, differentiability, and topology-preserving. Diffeomorphic Demons (Vercauteren et al., 2009), Large Deformation Diffeomorphic Metric Matching (LDDMM) (Beg et al., 2005), and Symmetric image Normalization method (SyN) (Avants et al., 2008) are the common representatives of the diffeomorphic registration algorithms. Unfortunately, these traditional methods are time-consuming, limiting their clinical workflow application.

2.2. Deep learning-based image registration

Deep learning-based image registration methods employ CNN to directly predict the deformation field, which speeds up the registration process during the test phase. Nevertheless, supervised methods (Rohé et al., 2017; Sokooti et al., 2017; Eppenhof and Pluim, 2018) required the ground truth deformation fields to train the network. Although these methods achieve fast registration, the registration accuracy is severely limited to the quality of the generated ground truth deformation fields.

Recently, unsupervised methods adopted the differentiable similarity measure between the warped image and fixed image to optimize the network. de Vos et al. (2017) presented an end-to-end unsupervised 2D medical image registration network using cross-correlation as the similarity measure. Balakrishnan et al. (2018) further demonstrated a 3D registration network that introduced additional regularization to ensure the smoothness of the deformation field. Dalca et al. (2018) proposed a probabilistic diffeomorphic model. However, the above methods may fail in the case of complicated deformations because of their difficulty in dealing with the large deformations at the finest scale without initialized transformations. The study (de Vos et al., 2019) stacked multiple CNNs to perform the deformable image registration, but the multiple networks were trained separately. Zhao et al. (2019) developed an end-to-end recursive cascaded network for every subnetwork to learn a small deformation cooperatively. The network parameters and required memory increase linearly with the growth of cascades, which is highly inefficient. The studies (Jiang et al., 2020; Hering et al., 2019, 2021; Fu et al., 2020; Li and Fan, 2020; Mok and Chung, 2020b) employed the image pyramid to achieve coarse-to-fine registration. Nevertheless, these methods require a registration network at each scale, introducing lots of parameters. Moreover, most methods neglect the expected diffeomorphic properties and inverse consistency, resulting in their limitations in practical application.

2.3. Inverse consistent registration

The inverse consistency indicates that the inverse deformation field of the forward registration result equals the backward registration result and same for the reverse condition. Traditional methods (Christensen and Johnson, 2001; Tao et al., 2009) added an inverse consistent constraint in the optimized cost function. Inspired by this idea, recent studies (Gu et al., 2020; Shen et al., 2019; Zhang et al., 2020) employed the same network to generate the bidirectional deformation fields by swapping the order of input images. Then these methods introduced an inverse consistency loss on the composition of the bidirectional deformation fields. Zhang (2018) developed an inverse operation to solve the inverse deformation field of the current direction registration and used it to constrain the other direction. The proposed inverse operation, which does not meet the strict mathematical definition, generates pseudo-inverse. Kim et al. (2021) showed a cycle-consistent image registration method, but it required two different registration networks. Mok and Chung (2020a) followed the conventional methods (Avants et al., 2008; Yang et al., 2008; Lorenzi et al., 2013) to adopt the ‘mean’ shape to connect the bidirectional registration. Nevertheless, the single forward propagation of the network produces two-way registration results simultaneously, causing the registration results to be related to the order of input images. Although the above methods reduce the inverse consistent error compared with asymmetric algorithms, our method can show better inverse consistency.

3. Method

3.1. Method overview

Deformable image registration aims to find the optimal deformation field ϕ^* between the moving image I_m and fixed image I_f. The

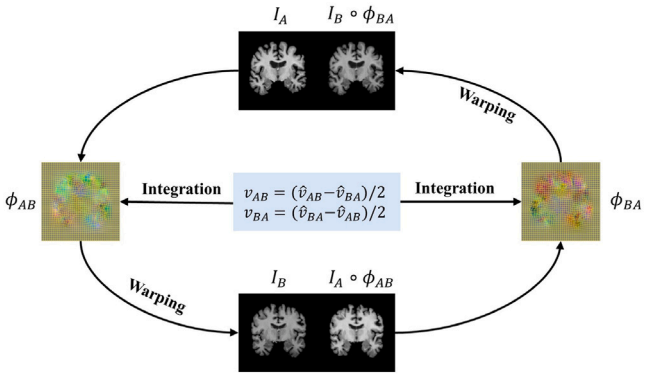


Fig. 1. Illustration of the proposed symmetric strategy. The forward and backward estimation are conducted, respectively. The bidirectional velocity fields are averaged for more robust estimation. The symmetric similarity is further introduced for optimization.

optimized process can be described as:

$$\phi^* = \arg\min_{\phi} \left[\mathcal{L}_{\text{sim}}(I_f, I_m \circ \phi) + \lambda_{\text{reg}}\ \mathcal{L}_{\text{smooth}}\ (\phi)\right] \tag{1}$$

where $I_m \circ \phi$ denotes the warped image, $\mathcal{L}_{\text{sim}}$ is the similarity measure loss, $\mathcal{L}_{\text{smooth}}$ is the regularization loss and λ_{reg} is the hyperparameter. DLIR methods often employ the CNN to directly estimate the displacement field u, and the final deformation field ϕ is generated by adding it to the identity transformation Id:

$$\phi(x) = Id + u(x) \tag{2}$$

Although this strategy is common and effective, it can hardly ensure the expected diffeomorphic properties, including topology-preservation and invertibility.

In this paper, we perform the diffeomorphic registration by estimating the stationary velocity field. The final deformation field ϕ is then defined as follows:

$$\frac{\partial \phi^{(t)}}{\partial t} = v\left(\phi^{(t)}\right), \phi^{(0)} = Id \tag{3}$$

where Id is identity transformation and t is time. The deformation field ϕ is obtained by integrating the stationary velocity field v over unit time. Besides, unlike most previous methods that only consider the single mapping from the moving image to the fixed image, we implement the diffeomorphic registration symmetrically to guarantee the inverse consistency of the bidirectional deformation fields. The proposed symmetric strategy is illustrated in Fig. 1. Specifically, we conduct the forward and backward registration respectively and average the bidirectional velocity fields for more robust estimation. We further introduce symmetric similarity for optimization. Let I_A and I_B denote the unaligned image volumes. ϕ_{AB} and v_{AB} represent the forward deformation and velocity field that register I_A to I_B while similar for ϕ_{BA} and v_{BA}. Our optimization process can be described as:

$$\begin{aligned}\phi^*_{AB}, \phi^*_{BA} = \arg\min_{\phi AB, \phi BA} \big[&\mathcal{L}_{\text{sim}}(I_B, I_A \circ \phi_{AB}) + \mathcal{L}_{\text{sim}}(I_A, I_B \circ \phi_{BA}) \\ &+\lambda_{\text{reg}}(\mathcal{L}_{\text{smooth}}\ (v_{AB}) + \mathcal{L}_{\text{smooth}}\ (v_{BA}))\big]\end{aligned} \tag{4}$$

Fig. 2 further presents the overall architecture of our proposed symmetric pyramid network. The unaligned images are input to a shared-weights encoder network to obtain two sets of feature pyramids. To achieve the coarse-to-fine registration, we progressively implement symmetric feature registration from the bottom to the top feature level.

3.2. Multi-scale image encoding

Given the unaligned image pair I_A and I_B, we adopt the shared-weights encoder network to acquire two sets of feature pyramids F^i_A and F^i_B $(i = 1, 2, 3, 4)$ as shown in Fig. 2(a). We employ two convolutional layers with $3 \times 3 \times 3$ kernels followed by LeakyReLU activation for each level of the encoder. The stride of the first convolutional layer in the two-layer structure is set to 2 to down-sample the extracted feature except for the top level. To compensate for the lack of detailed information in the down-sampling process, we also concatenate the feature extracted from the down-sampled images by an additional convolutional layer. The detailed architecture of the encoder is shown in Fig. 3. The encoding process can be described as:

$$\begin{aligned}F^i_A &= [C^i_1(F^{i-1}_A), C^i_2(I^i_A)], F^1_A = C^1_1(I^1_A) \\ F^i_B &= [C^i_1(F^{i-1}_B), C^i_2(I^i_B)], F^1_B = C^1_1(I^1_B)\end{aligned} \tag{5}$$

where $[*]$ denotes the concatenation operation at the channel dimension, C^i_1 and C^i_2 represent the two-layer structure and the additional convolutional layer at ith level, and I^i_A and I^i_B denote the down-sampled images. Finally, F^i_A and F^i_B are $1/2^{i-1}$ times of the initial size. The feature-level registration is then conducted progressively from the bottom to the top level in a coarse-to-fine manner. The network gradually achieves more accurate registration with the enrichment of detailed anatomical information in the feature.

3.3. Symmetric feature registration

3.3.1. Feature warping

Methods (Jiang et al., 2020; Hering et al., 2019, 2021; Fu et al., 2020; Li and Fan, 2020; Mok and Chung, 2020b) based on image pyramids usually use the registration results of the previous scale as the initialization of current scale. Specifically, the deformation field estimated in the previous stage is employed to warp the moving image of the current stage to reduce the correspondence distance in image space. Following this idea, we use the feature warping operation to reduce the matching distance in feature space. At the ith level, we up-sample the previous deformation fields ϕ^{i+1}_{AB} and ϕ^{i+1}_{BA} with a factor of 2 (denoted as $\hat{\phi}^{i+1}_{AB}$ and $\hat{\phi}^{i+1}_{BA}$) using trilinear interpolation, and then warp the corresponding feature in the forward and backward process, respectively. The whole process can be described as:

$$F^i_{WA} = F^i_A \circ \hat{\phi}^{i+1}_{AB}, F^i_{WB} = F^i_B \circ \hat{\phi}^{i+1}_{BA} \tag{6}$$

where $\circ$ represents the differentiable warping operation presented in Jaderberg et al. (2015), F^i_{WA} and F^i_{WB} denote the warped feature in the corresponding process.

3.3.2. Bidirectional velocity field estimation

After the feature warping operation, we employ the velocity field estimator to generate temporary velocity fields for the forward and backward registration. The estimator consists of three convolution layers. The first two layers, followed by LeakyReLU activation, are used for further feature extraction, and the last layer without activation is adopted to obtain the temporary velocity field. The feature map extracted by the first two layers is also up-sampled as an input of the next-level estimator to increase the global information. Consequently, the input of the ith level estimator contains the warped feature, the corresponding pyramid feature, the up-sampled global feature, and the up-sampled deformation field. These inputs are concatenated and fed to the estimator. The bidirectional estimation process can be written as:

$$\begin{aligned}\hat{v}^i_{AB} &= Est^i(F^i_{WA}, F^i_B, F^{i+1}_{FG}, \hat{\phi}^{i+1}_{AB}) \\ \hat{v}^i_{BA} &= Est^i(F^i_{WB}, F^i_A, F^{i+1}_{BG}, \hat{\phi}^{i+1}_{BA})\end{aligned} \tag{7}$$

where Est^i denotes the velocity field estimator at the ith level, F^{i+1}_{FG} and F^{i+1}_{BG} represent the forward and backward global feature up-sampled from the previous level, and $\hat{v}^i_{AB}$ and $\hat{v}^i_{BA}$ denote the estimated bidirectional temporary velocity fields. Especially, we just use the pyramid features as input for the bottom level. The estimator shares the same weights in the forward and backward process on the same levels other than different levels (see Fig. 2).

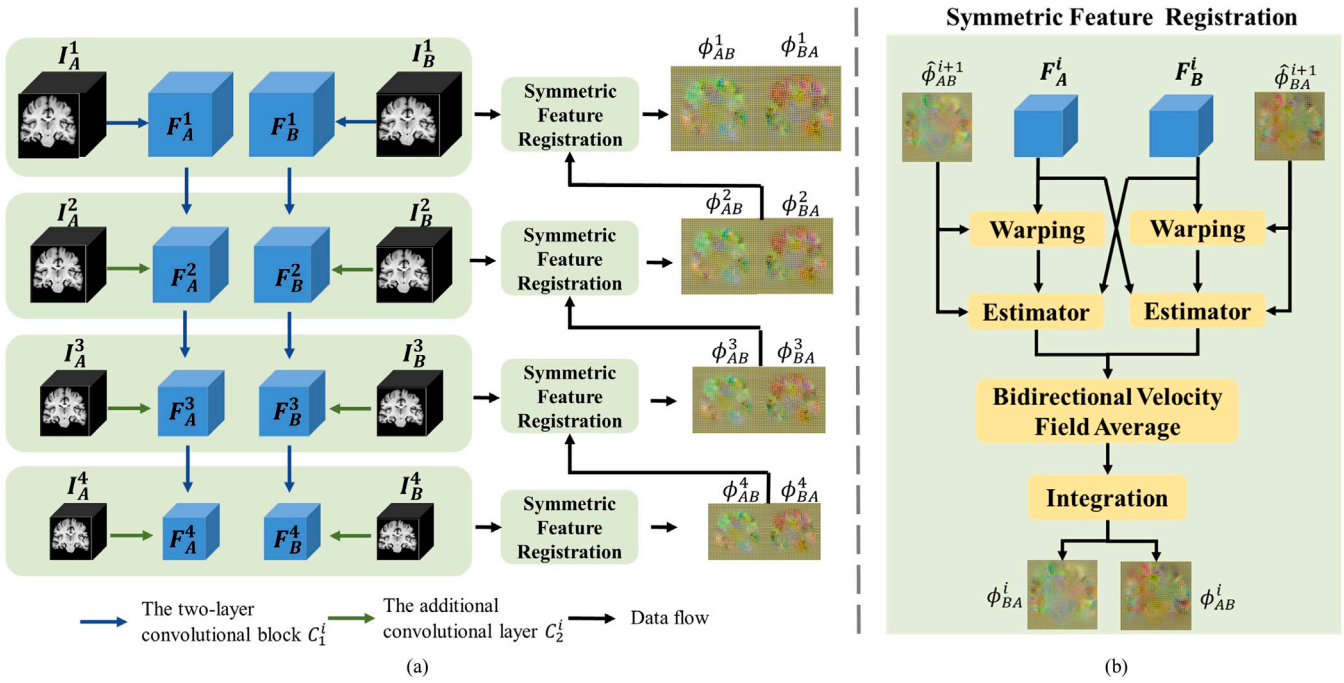


Fig. 2. Overview of (a) symmetric pyramid network and (b) symmetric feature registration. The input images are first encoded to the feature pyramids, respectively. Then we progressively conduct the feature-level registration from the bottom level to the top level. The feature-level registration is implemented symmetrically, and the estimated bidirectional velocity fields are averaged for more robust estimation.

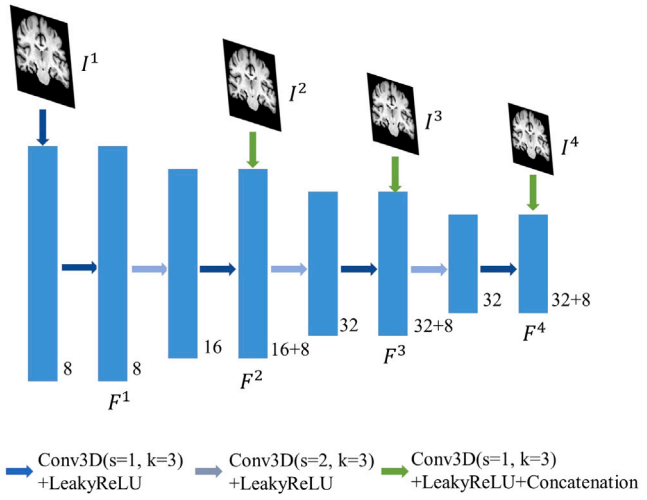


Fig. 3. Detailed architecture of the proposed multi-scale image encoder network. The light and dark blue arrows form the two-layer convolutional block C_1^i at each level. The green arrow represents the additional convolutional layer C_2^i.

3.3.3. Bidirectional velocity field average

Previous research (Dupuis et al., 1998) theoretically shows that the solution for the diffeomorphism system is unique and differentiable. The bidirectional stationary velocity fields are expected to be opposite under absolutely ideal registration (Lorenzi et al., 2013). With the bidirectional temporary velocity fields $\hat{v}^i_{AB}$ and $\hat{v}^i_{BA}$, we then consider averaging them for more robust estimation. The final velocity fields are calculated as follows:

$$v^i_{AB} = \frac{\hat{v}^i_{AB} - \hat{v}^i_{BA}}{2}, v^i_{BA} = \frac{\hat{v}^i_{BA} - \hat{v}^i_{AB}}{2} \tag{8}$$

where v^i_{AB} and v^i_{BA} are the final bidirectional velocity fields at the ith level.

3.3.4. Integration module

At each level, the final deformation field is defined through the ordinary differential equation (Eq. (3)). We integrate v^i_{AB} and v^i_{BA} to obtain ϕ^i_{AB} and ϕ^i_{BA}, respectively. The integration is computed using scaling and squaring (Lorenzi et al., 2013; Dalca et al., 2018). Specifically, considering v is a member of Lie algebra, $\phi^{(1)}$ is approximated to $exp(v)$ that is a member of Lie group. Starting from $\phi^{1/2^T}$, we adopt the recurrence $\phi^{1/2^{t-1}} = \phi^{1/2^t} \circ \phi^{1/2^t}$ to get $\phi^{(1)} = \phi^{1/2} \circ \phi^{1/2}$ where T is the time step.

3.4. Symmetric multi-scale similarity

After performing the symmetric feature registration from the bottom to the top level, we can obtain two sets of deformation field pyramids ϕ^i_{AB} and ϕ^i_{BA} $(i = 1, 2, 3, 4)$. Then we employ the symmetric multi-scale similarity measure to constrain the deformation fields. The similarity measure is implemented using normalized cross-correlation (NCC). Let I, J denote the input images. $\overline{I(p)}$ and $\overline{J(p)}$ are local means at position p while p_i denotes the positions within the local window centered at p. The NCC is then defined as:

$$NCC(I,J) = \frac{1}{|\Omega|} \sum_{p \in \Omega} \frac{\left(\sum_{p_i} \left(\boldsymbol{I}\left(p_i\right) - \overline{\boldsymbol{I}(\boldsymbol{p})}\right)\left(\boldsymbol{J}\left(p_i\right) - \overline{\boldsymbol{J}(\boldsymbol{p})}\right)\right)^2}{\sum_{p_i} \left(\boldsymbol{I}\left(p_i\right) - \overline{\boldsymbol{I}(\boldsymbol{p})}\right)^2 \sum_{p_i} \left(\boldsymbol{J}\left(p_i\right) - \overline{\boldsymbol{J}(\boldsymbol{p})}\right)^2} \tag{9}$$

To ensure the smoothness of the deformation fields, additional regularization items on the spatial gradients of the final velocity fields are also introduced at each level. The final complete loss is formulated as:

$$\begin{aligned} \mathcal{L} = \sum_{i=1}^{4} \frac{1}{2^i} \Big[& -NCC\left(I^i_A \circ \phi^i_{AB}, I^i_B\right) - NCC\left(I^i_B \circ \phi^i_{BA}, I^i_A\right) \\ & + \lambda \left(\left\| \nabla v^i_{AB} \right\|_2^2 + \left\| \nabla v^i_{BA} \right\|_2^2 \right) \Big] \end{aligned} \tag{10}$$

where λ is the regularization parameter. We also assign a lower weight to the higher level considering the lack of detailed information in the down-sampled images, and the window size of NCC is set to $(9, 7, 5, 3)$ from the first to the fourth level.

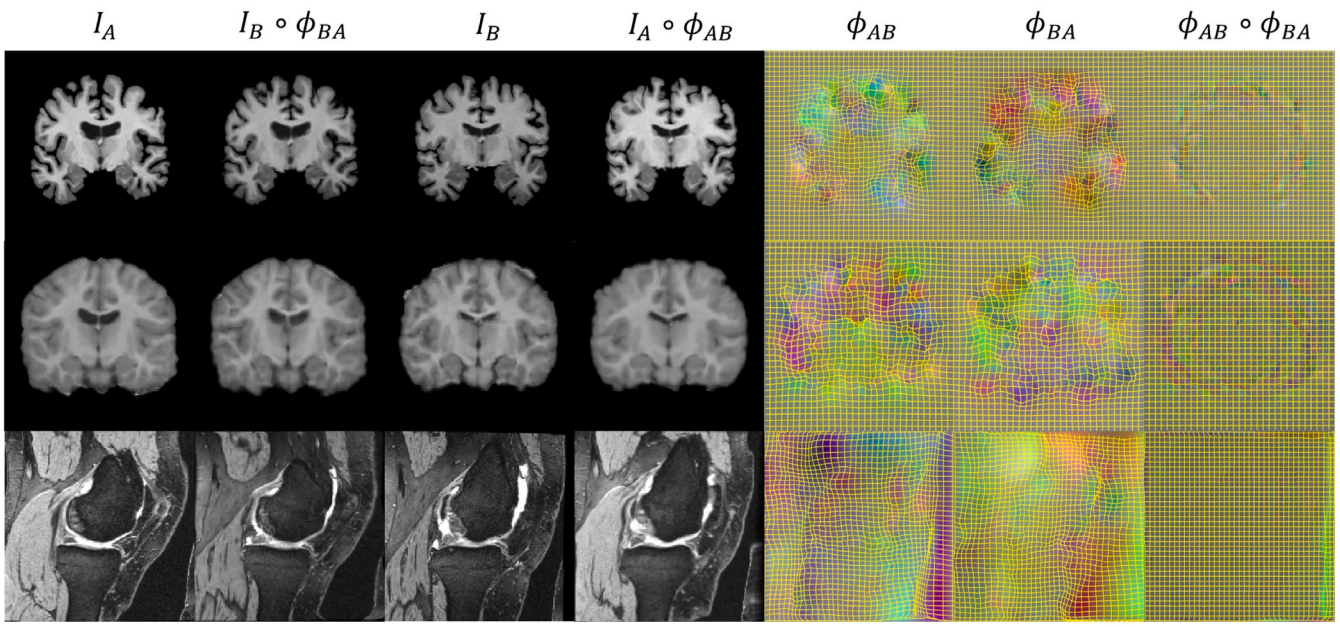


Fig. 4. Visualization of symmetrical registration results generated by our network. From top to bottom are the results on Mindbiggle101, CANDI, and OAI dataset. From left to right are the image I_A, the backward warped image $I_B \circ \phi_{BA}$, the image I_B, the forward warped image $I_A \circ \phi_{AB}$, the forward and backward deformation field and the composition of the bidirectional deformation fields.

4. Experiments

4.1. Datasets

4.1.1. Mindboggle101

MindBoggle101 (Klein and Tourville, 2012) consists of 101 skull-stripped T1-weighted brain MR scans, and parts of the data are annotated with 50 cortical regions. We followed the recent studies (Kuang and Schmah, 2019; Liu et al., 2019) to employ 42 images with 1742 pairs as the training set and 20 images with 380 pairs as the testing set. Considering the scans were affinely normalized, we center-cropped the image to $160 \times 192 \times 160$.

4.1.2. CANDI

CANDI (Kennedy et al., 2012) contains 103 child and adolescent brain MR images and the corresponding segmentation mask (28 anatomical labels). We randomly selected 30 images with 870 pairs as test data and used the others for training and validation. Standard preprocessing steps including, skull-stripping, affine normalization, and spacing resampling were conducted, and the images were then center-cropped to $128 \times 144 \times 128$.

4.1.3. OAI

OAI[1] provides labeled knee MR images with segmentations of femur and tibia as well as femoral and tibial cartilage (Ambellan et al., 2019). We randomly chose 50 scans with 2450 pairs as test data, and another 100 scans were employed for training and validation. The images were affinely normalized and then cropped to $160 \times 192 \times 192$.

4.2. Evaluation metrics and implementation

We employ the Dice score to evaluate the registration accuracy as follows:

$$Dice = \frac{2|S_A \circ \phi_{AB} \cap S_B|}{|S_A \circ \phi_{AB}| + |S_B|} \tag{11}$$

where S_A and S_B represent the segmentation masks of the input images. The Dice score ranges from $[0, 1]$, and a high Dice value indicates a better anatomical correspondence.

Jacobin matrix contains local information of the deformation fields. The Jacobin matrix $J_\phi(p)$ at position p is defined as:

$$J_\phi(p) = \begin{pmatrix} \frac{\partial \phi_x(p)}{\partial x} & \frac{\partial \phi_x(p)}{\partial y} & \frac{\partial \phi_x(p)}{\partial z} \\ \frac{\partial \phi_y(p)}{\partial x} & \frac{\partial \phi_y(p)}{\partial y} & \frac{\partial \phi_y(p)}{\partial z} \\ \frac{\partial \phi_z(p)}{\partial x} & \frac{\partial \phi_z(p)}{\partial y} & \frac{\partial \phi_z(p)}{\partial z} \end{pmatrix} \tag{12}$$

The determinant of the Jacobin matrix $|J_\phi(p)|$ can be adopted to evaluate the diffeomorphic property. Specifically, $|J_\phi(p)| > 1$ represents expansion at position p and $0 < |J_\phi(p)| < 1$ indicates shrink. $|J_\phi(p)| \leq 0$ means the appearance of folding and the loss of one-to-one mapping, which is undesirable in medical image registration. Consequently, we calculate the percentage of voxels with non-positive Jacobin determinant (Jab) to assess the quality of the deformation field in our experiments.

The inverse consistency error (ICE) measures the distance between the composition of the bidirectional deformation fields and the identity function. The ICE is formulated as:

$$ICE = \|\phi_{AB} \circ \phi_{BA} - Id\|_1 \tag{13}$$

Our proposed network was implemented using PyTorch (Paszke et al., 2019) and trained on an NVIDIA RTX GPU. We adopted the Adam (Kingma and Ba, 2014) optimizer with a learning rate of $1e^{-4}$. The batch size and regularization parameter λ was set to 1 for all three datasets. The time step T in the integration module was set to 7 following (Dalca et al., 2018). The feature channel of the two-layer structure in multi-scale image encoding was set to $(8, 16, 32, 32)$ from the first to the fourth level. The additional down-sampled image feature channel was set to 8.

4.3. Symmetric registration visualization

Fig. 4 first presents the symmetric diffeomorphic registration results of our network on the three datasets. The bidirectional warped images are similar to the corresponding target images, demonstrating that

[1] https://nda.nih.gov/oai

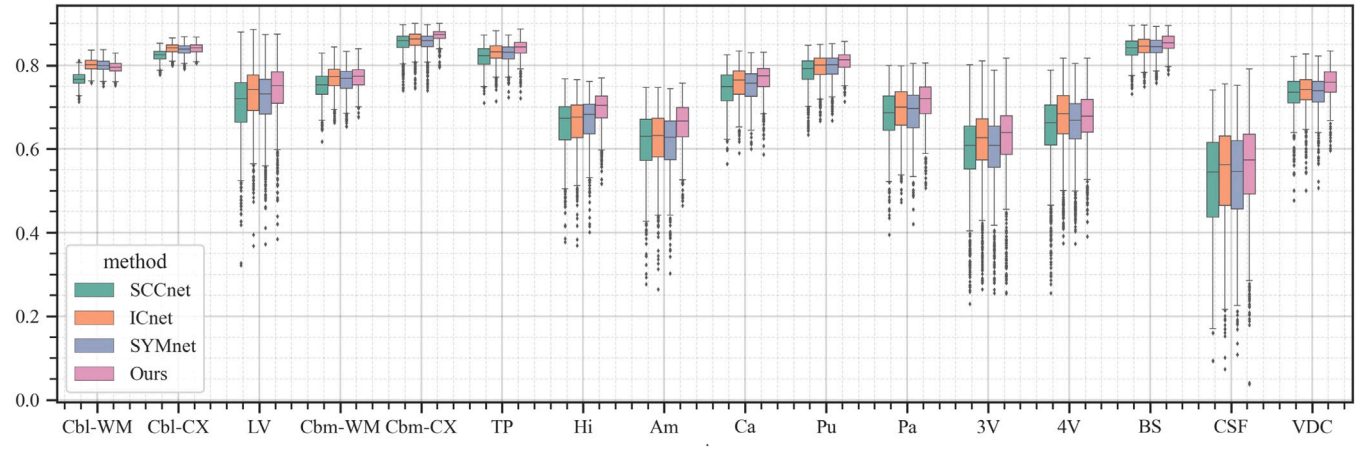


Fig. 5. Boxplots depicting the Dice accuracy of each anatomical structure in CANDI for SCCnet, ICnet, SYMnet, and our method. The left and right hemispheres of the brain are combined into one structure for visualization. The cerebral white matter (Cbl-WM), cerebral cortex (Cbl-CX), lateral vent (LV), cerebellum white matter (Cbm-WM), cerebellum cortex (Cbm-CX), thalamus proper (TP), hippocampus (Hi), amygdala (Am), caudate (Ca), putamen (Pu), Pallidum (Pa), 3rd vent (3V), 4th vent (4V), brain stem (BS), CSF (CSF), and ventralDC (VDC) are included. Our method outperforms the others on most anatomical structures.

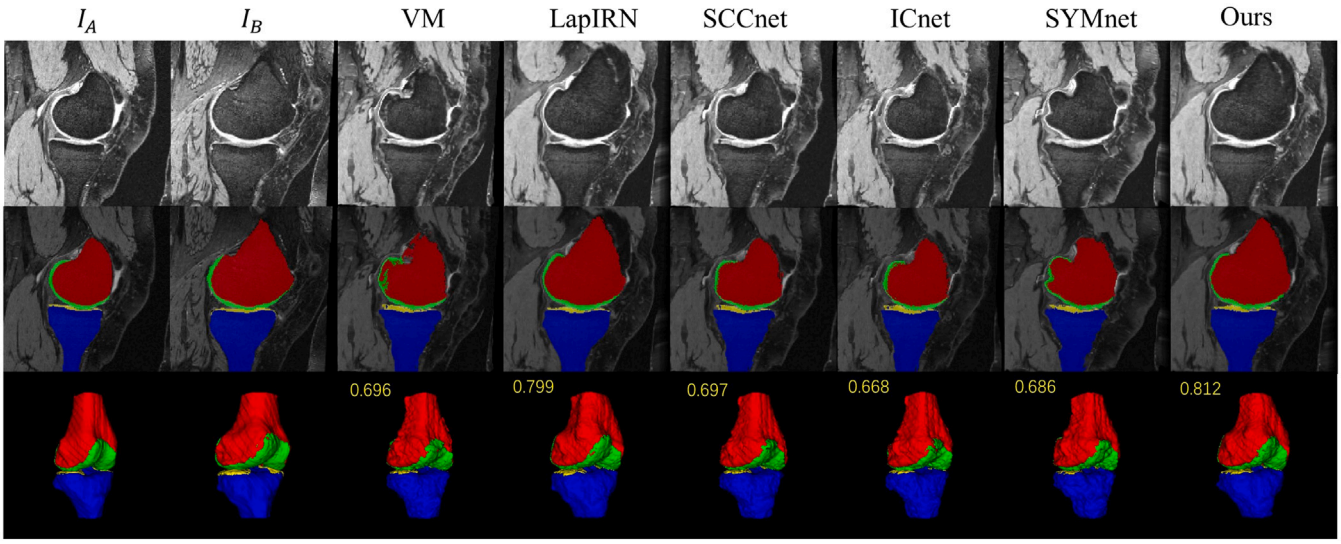


Fig. 6. Qualitative comparison of registration results generated by different methods. From top to bottom are the image, 2D, and 3D segmentation mask. The red and blue areas represent bone, and the green and yellow areas represent cartilage. The registration accuracy is marked in the upper left corner of the third row.

our network achieves the forward and backward registration simultaneously and accurately. Besides, the composition of the bidirectional deformation fields shown in the last column is close to identity transformation, indicating that the bidirectional deformation fields generated by our method are inverse consistent. This result qualitatively illustrates that our network achieves accurate alignment while maintaining inverse consistency.

4.4. Comparison with other methods

4.4.1. Baseline settings

We further compared our method with eight other approaches: SyN (Avants et al., 2008), VM (Balakrishnan et al., 2018), DIF-VM (Dalca et al., 2018), PMRnet (Liu et al., 2019), LapIRN (Mok and Chung, 2020b), RCnet (Zhao et al., 2019), CFWnet (Zhang et al., 2021), SCCnet (Gu et al., 2020), ICnet (Zhang, 2018), and SYMnet (Mok and Chung, 2020a). SyN was a traditional iteration-based symmetric diffeomorphic method, which was inverse consistent. We used the official implementation in ANTs (Avants et al., 2009) for SyN. VM adopted the "U" shape network to directly estimate the deformation field between the input images, and DIF-VM was a probabilistic diffeomorphic variant of VM. PMRnet developed a multi-scale feature-level probabilistic model. LapIRN conducted the multi-scale diffeomorphic registration based on the image pyramids. RCnet and CFWnet are cascade-based registration algorithms. SCCnet, ICnet, and SYMnet were inverse consistent registration methods. More specifically, SCCnet introduced the inverse consistency loss on the composition of the bidirectional deformation fields. ICnet developed an inverse consistency constraint based on the proposed inverse network, and SYMnet presented the 'mean' shape to constrain the bidirectional registration results. For a fair comparison, we used the public code and fine-tuned the hyperparameters on each dataset. Moreover, we followed (Liu et al., 2019) to report the Dice on five large regions grouped from the initial cortical regions on Mindboggle101 and showed the Dice of bone and cartilage on OAI conforming to Xu and Niethammer (2019). The average Dice of all anatomical labels was also reported on the three datasets.

4.4.2. Registration accuracy on Mindboggle101

Table 1 summarizes the registration accuracy on the Mindboggle101. It is clear that our method outperforms the others on all five grouped anatomical regions. We also achieve 0.587 average Dice and

Table 1
Quantitative comparison of Dice accuracy on Mindboggle101 dataset mean(std).

Method	Region					Average
	Frontal	Parietal	Occipital	Temporal	Cingulate	
SyN (Avants et al., 2008)	0.520(0.022)	0.452(0.032)	0.402(0.039)	0.544(0.021)	0.631(0.032)	0.503(0.015)
VM (Balakrishnan et al., 2018)	0.583(0.024)	0.523(0.031)	0.438(0.050)	0.581(0.022)	0.672(0.036)	0.554(0.017)
DIF-VM (Dalca et al., 2018)	0.566(0.023)	0.509(0.031)	0.429(0.047)	0.563(0.023)	0.648(0.034)	0.538(0.016)
PMRnet (Liu et al., 2019)	0.588(0.024)	0.529(0.032)	0.452(0.047)	0.591(0.022)	0.667(0.034)	0.562(0.016)
LapIRN (Mok and Chung, 2020b)	0.575(0.024)	0.502(0.031)	0.433(0.045)	0.576(0.021)	0.669(0.033)	0.545(0.015)
RCnet (Zhao et al., 2019)	0.605(0.025)	0.539(0.032)	0.466(0.049)	0.595(0.066)	0.688(0.033)	0.576(0.016)
CFWnet (Zhang et al., 2021)	0.595(0.023)	0.531(0.032)	0.463(0.046)	0.595(0.023)	0.676(0.034)	0.567(0.016)
SCCnet (Gu et al., 2020)	0.551(0.025)	0.485(0.031)	0.402(0.050)	0.555(0.023)	0.651(0.036)	0.523(0.017)
ICnet (Zhang, 2018)	0.578(0.025)	0.530(0.030)	0.437(0.051)	0.570(0.023)	0.657(0.038)	0.550(0.018)
SYMnet (Mok and Chung, 2020a)	0.580(0.025)	0.517(0.032)	0.434(0.050)	0.579(0.023)	0.667(0.035)	0.551(0.017)
Ours	**0.614(0.023)**	**0.553(0.031)**	**0.485(0.043)**	**0.614(0.020)**	**0.689(0.034)**	**0.587(0.014)**

Table 2
Quantitative comparison of Dice accuracy on CANDI and OAI dataset mean(std).

Method	CANDI	OAI		
	Average	Bone	Cartilage	Average
SyN (Avants et al., 2008)	0.720(0.023)	0.909(0.054)	0.595(0.132)	0.752(0.090)
VM (Balakrishnan et al., 2018)	0.744(0.023)	0.911(0.044)	0.574(0.042)	0.742(0.052)
DIF-VM (Dalca et al., 2018)	0.704(0.033)	0.914(0.034)	0.549(0.063)	0.731(0.041)
PMRnet (Liu et al., 2019)	0.719(0.027)	0.935(0.020)	0.631(0.053)	0.783(0.031)
LapIRN (Mok and Chung, 2020b)	0.751(0.022)	0.940(0.027)	0.670(0.059)	0.805(0.038)
RCnet (Zhao et al., 2019)	0.758(0.021)	0.942(0.027	0.653(0.057)	0.798(0.038)
CFWnet (Zhang et al., 2021)	0.755(0.023)	0.947(0.019)	0.661(0.054)	0.804(0.033)
SCCnet (Gu et al., 2020)	0.731(0.024)	0.903(0.050)	0.609(0.071)	0.756(0.054)
ICnet (Zhang, 2018)	0.745(0.023)	0.906(0.047)	0.573(0.072)	0.740(0.053)
SYMnet (Mok and Chung, 2020a)	0.741(0.023)	0.923(0.033)	0.590(0.061)	0.756(0.042)
Ours	**0.758(0.021)**	**0.953(0.010)**	**0.677(0.049)**	**0.815(0.027)**

Table 3
Quantitative comparison of Jab and ICE on the three datasets mean(std).

Method	Mindboggle101		CANDI		OAI	
	Jab(%)	ICE	Jab(%)	ICE	Jab(%)	ICE
VM (Balakrishnan et al., 2018)	1.086(0.117)	0.531(0.034)	0.447(0.010)	0.271(0.021)	4.528(1.335)	0.565(0.108)
DIF-VM (Dalca et al., 2018)	0.044(0.012)	0.645(0.036)	0.020(0.010)	0.366(0.021)	0.839(0.442)	1.887(0.364)
PMRnet (Liu et al., 2019)	**0.000(0.000)**	0.942(0.068)	0.000(0.001)	0.504(0.036)	**0.001(0.002)**	1.567(0.410)
LapIRN (Mok and Chung, 2020b)	0.295(0.117)	0.786(0.054)	0.023(0.029)	0.460(0.041)	0.227(0.216)	3.425(0.831)
RCnet (Zhao et al., 2019)	0.004(0.001)	0.422(0.014)	0.000(0.001)	0.504(0.036)	0.002(0.001)	2.021(0.282)
CFWnet (Zhang et al., 2021)	0.002(0.001)	0.925(0.018)	0.000(0.000)	0.750(0.016)	0.002(0.002)	2.287(0.238)
SCCnet (Gu et al., 2020)	0.095(0.015)	0.113(0.003)	0.032(0.015)	0.081(0.003)	1.750(0.812)	0.152(0.008)
ICnet (Zhang, 2018)	1.717(0.145)	0.266(0.011)	0.691(0.132)	0.167(0.008)	3.683(0.660)	0.570(0.045)
SYMnet (Mok and Chung, 2020a)	0.010(0.002)	0.065(0.002)	0.007(0.002)	0.065(0.002)	0.052(0.036)	0.155(0.021)
Ours	0.002(0.001)	**0.046(0.002)**	**0.000(0.000)**	**0.031(0.001)**	0.005(0.012)	**0.120(0.027)**

obtain 8.4%, 3.3%, 3.9%, 2.5%, 4.3%, 1.1%, 2.0%, 6.4%, 3.7%, and 3.6% improvement compared with SyN (Avants et al., 2008), VM (Balakrishnan et al., 2018), DIF-VM (Dalca et al., 2018), PMRnet (Liu et al., 2019), LapIRN (Mok and Chung, 2020b), RCnet (Zhao et al., 2019), CFWnet (Zhang et al., 2021), SCCnet (Gu et al., 2020), ICnet (Zhang, 2018), and SYMnet (Mok and Chung, 2020a).

4.4.3. Registration accuracy on CANDI

The comparison results on the CANDI are shown in Table 2. The proposed method obtains 0.758 average Dice. Fig. 5 further presents a detailed anatomical comparison. Our method shows its advances over the other three inverse consistent methods on most anatomical areas and achieves 2.7%, 1.3%, and 1.7% Dice improvement in comparison to SCCnet (Gu et al., 2020), ICnet (Zhang, 2018), and SYMnet (Mok and Chung, 2020a).

4.4.4. Registration accuracy on OAI

Table 2 sums up the results on OAI. The proposed network demonstrates superior registration accuracy compared with methods (Avants et al., 2008; Balakrishnan et al., 2018; Dalca et al., 2018; Liu et al., 2019; Mok and Chung, 2020b; Zhao et al., 2019; Zhang et al., 2021; Gu et al., 2020; Zhang, 2018; Mok and Chung, 2020a) obtaining 6.3%, 7.3%, 8.4%, 3.2%, 1.0%, 1.7%, 1.1%, 5.9%, 7.5%, 5.9% in Dice. We can also observe that the LapIRN (Mok and Chung, 2020b) based on image pyramids achieves comparable registration accuracy, but the approach requires a multi-stage training strategy. Furthermore, the parameter amount of LapIRN is $0.92M$, which is far beyond the VM ($0.39M$) and our method ($0.41M$). Our network achieves superior performance with a relatively small number of parameters, revealing its parameter efficiency. Fig. 6 further shows the qualitative registration results generated by different methods.

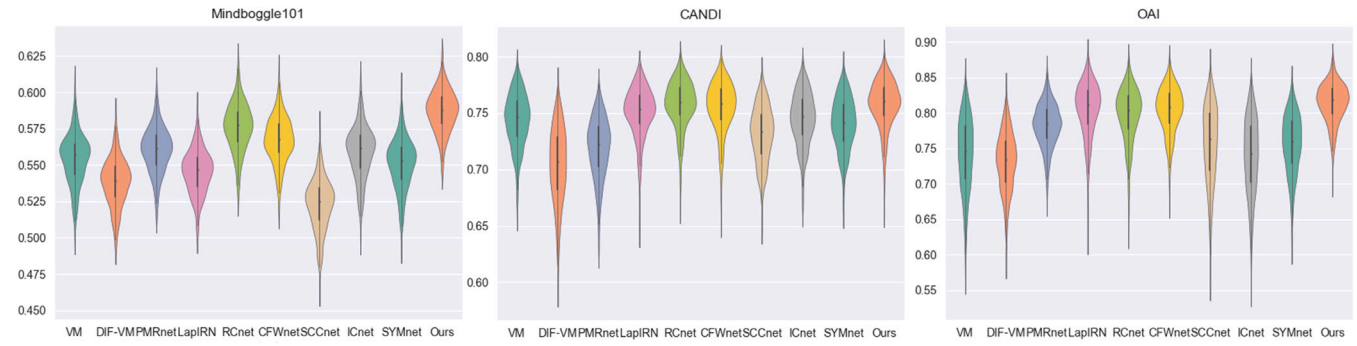


Fig. 7. Registration accuracy distributions of DLIR methods on the three datasets. Our method outperforms the others on all three datasets.

Table 4
Quantitative results of the ablation study on Mindboggle101 dataset mean(std).

Method	Region					Average	Jab(%)	ICE
	Frontal	Parietal	Occipital	Temporal	Cingulate			
w\o SVF	0.599(0.024)	0.534(0.032)	0.465(0.048)	0.601(0.021)	0.692(0.034)	0.573(0.016)	0.223(0.036)	0.142(0.004)
w\o MIE	0.605(0.023)	0.543(0.032)	0.469(0.045)	0.604(0.020)	0.683(0.033)	0.577(0.015)	0.003(0.001)	0.046(0.002)
w\o FW	0.590(0.024)	0.526(0.032)	0.448(0.047)	0.591(0.021)	0.677(0.034)	0.562(0.016)	0.003(0.001)	0.046(0.002)
w\o SR	0.598(0.023)	0.534(0.031)	0.465(0.046)	0.599(0.022)	0.680(0.034)	0.571(0.015)	0.001(0.001)	0.720(0.014)
Ours	0.614(0.023)	0.553(0.031)	0.485(0.043)	0.614(0.020)	0.689(0.034)	0.587(0.014)	0.002(0.001)	0.046(0.002)
SS	0.599(0.023)	0.532(0.031)	0.465(0.045)	0.599(0.022)	0.678(0.034)	0.570(0.016)	0.002(0.001)	0.787(0.016)
SS+BE	0.600(0.023)	0.536(0.031)	0.468(0.044)	0.601(0.020)	0.682(0.033)	0.573(0.015)	0.001(0.000)	0.425(0.013)
SS+BE+BA	0.614(0.023)	0.553(0.031)	0.485(0.043)	0.614(0.020)	0.689(0.034)	0.587(0.014)	0.002(0.001)	0.046(0.002)
n=2	0.577(0.025)	0.514(0.031)	0.433(0.049)	0.578(0.023)	0.669(0.035)	0.549(0.017)	0.003(0.001)	0.043(0.002)
n=3	0.603(0.024)	0.539(0.032)	0.464(0.047)	0.602(0.021)	0.684(0.034)	0.573(0.015)	0.003(0.001)	0.045(0.002)
n=4	0.614(0.023)	0.553(0.031)	0.485(0.043)	0.614(0.020)	0.689(0.034)	0.587(0.014)	0.002(0.001)	0.046(0.002)

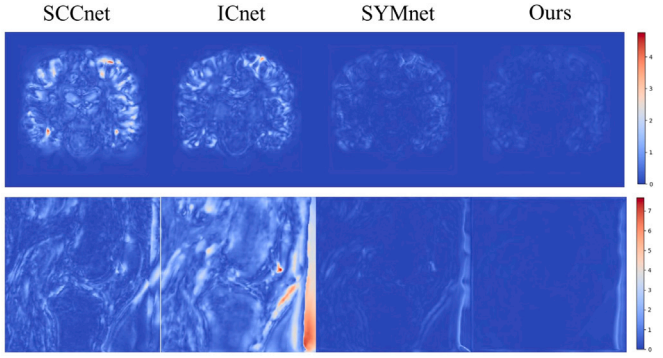


Fig. 8. Qualitative ICE comparison of the three inverse consistent methods and ours. Our method shows lower ICE.

Fig. 7 summarizes the performance of DLIR methods on all three datasets. Our network consistently outperforms the others, which implies that our method achieves more accurate alignment by progressively conducting the symmetric feature registration. The comparison of Jab and ICE are shown in Table 3. The proposed network performs better diffeomorphic properties with the percentage of voxels with non-positive Jacobin determinant no more than 0.005% on all three datasets. Especially, there are only several voxels with the appearance of folding on the CANDI dataset. Although the three inverse consistent methods significantly reduce the inverse consistency error compared with the other asymmetric approaches, our approach shows the best inverse consistency on all datasets. Fig. 8 further shows a qualitative ICE comparison. Additionally, it should be noticed that our approach achieves excellent performance on all metrics simultaneously, distinct from the LapIRN (Mok and Chung, 2020b), PMRnet (Liu et al., 2019) and SYMnet (Mok and Chung, 2020a) showing competitive Dice, Jab or ICE. This result demonstrates that the proposed network implements the registration accurately and generates the deformation fields with better diffeomorphic properties and inverse consistency.

4.5. Ablation study

We further conducted several ablation studies on the Mindboggle101 to verify the contributions of different parts. The ablation experiments include the following three aspects.

4.5.1. Effectiveness of major components

We evaluated four major components, including stationary velocity field (SVF), multi-scale image encoding (MIE), feature warping (FW), and symmetric registration (SR). The baseline models are summarized as follows:

- w/o SVF: It conducts the registration by directly estimating the displacement field and the bidirectional deformation fields are constrained with inverse consistency loss.
- w/o MIE: It adopts the initial images to generate the feature pyramids without down-sampled images embedding.
- w/o FW: It excludes the feature warping operation of the symmetric feature registration.
- w/o SR: It only conducts the forward registration without any symmetric modules.

The results are shown in Table 4. We can find that all components have improved the registration accuracy. The percentage of voxels with non-positive Jacobin determinant and the inverse consistency error highly decrease by estimating the stationary velocity field. The multi-scale image encoding obtains a gain of 1.0% Dice, showing that enhancing the representational power of the feature pyramid can achieve more accurate alignment. The feature warping operation without introducing any parameters achieves a 2.5% improvement, revealing the importance of reducing the feature matching distance. Moreover, the symmetric registration also promotes registration accuracy with a relative improvement of 1.6% and highly reduces the inverse consistency error.

Table 5
Quantitative results of different backbones on Mindboggle101 dataset mean(std).

Method	Region					Average	Jab(%)	ICE
	Frontal	Parietal	Occipital	Temporal	Cingulate			
Baseline	0.605(0.023)	0.543(0.032)	0.469(0.045)	0.604(0.020)	0.683(0.033)	0.577(0.015)	0.003(0.001)	0.046(0.002)
ResNet	0.609(0.023)	0.543(0.032)	0.474(0.043)	0.606(0.020)	0.685(0.034)	0.579(0.015)	0.002(0.001)	0.046(0.002)
Inception	0.616(0.023)	0.553(0.032)	0.483(0.044)	0.614(0.020)	0.689(0.034)	0.587(0.015)	0.002(0.001)	0.046(0.002)
MIE	0.614(0.023)	0.553(0.031)	0.485(0.043)	0.614(0.020)	0.689(0.034)	0.587(0.014)	0.002(0.001)	0.046(0.002)

Table 6
Quantitative results of cross dataset brain registration mean(std).

Train	Test					
	Mindboggle101			CANDI		
	Dice	Jab(%)	ICE	Dice	Jab(%)	ICE
Mindboggle101	0.587(0.014)	0.002(0.001)	0.046(0.002)	0.752(0.022)	0.000(0.000)	0.031(0.002)
CANDI	0.562(0.016)	0.006(0.002)	0.047(0.002)	0.758(0.021)	0.000(0.000)	0.031(0.001)

4.5.2. Effectiveness of symmetric strategy

The proposed symmetric registration has been demonstrated effective. We then explored the detailed contributions of the symmetric modules, including symmetric similarity (SS), bidirectional velocity field estimation (BE), and bidirectional velocity field average (BA). The models are constructed as follows:

- SS: It just implements the forward feature registration but with symmetric similarity. The inverse deformation field is acquired by integrating the negative velocity field.
- SS+BE: It carries out the forward and backward feature registration separately and employs symmetric similarity for optimization.
- SS+BE+BA: It represents our proposed symmetric strategy that with all three modules.

The results are also presented in Table 4. It shows that simply using symmetric similarity or conducting the bidirectional velocity field estimation cannot significantly improve the registration performance. Nevertheless, the registration accuracy and inverse consistency are highly enhanced by introducing the bidirectional velocity field average, demonstrating that we can get a more robust estimation by averaging the bidirectional velocity fields.

4.5.3. Effectiveness of pyramid layer number

We further explored the effect of the number of feature pyramids. The pyramid layer number is set to 2, 3, and 4. We obtain 2.4% and 1.4% Dice improvements, respectively. The results indicate that as the number of pyramid layer increases, the bottom level can provide better initialization for the top level, leading to more accurate alignment.

4.6. Compatible with different backbones

To validate that the proposed symmetric feature registration is compatible with different backbones, we compared several variants on the Mindboggle101 dataset. The baseline represents the encoder that directly employs successive convolutional layers at each level. The ResNet denotes the encoder that uses the residual module (He et al., 2016) at each level. The Inception denotes the encoder uses the inception module (Szegedy et al., 2015). The MIE is the proposed multi-scale image encoding. The results are shown in Table 5. We can find that all the models can achieve promising registration performance, demonstrating the compatibility of the proposed symmetric feature registration. The ResNet with residual connection achieves similar performance with the baseline. The Inception using multiple different-size kernels to enhance the feature representation can further improve the registration accuracy. The proposed MIE also promotes accuracy and is more parameter efficient compared with Inception. Additionally, the strong backbone HRNet (Wang et al., 2020) is also more likely to enhance the registration performance, but the network requiring large memory is not suitable for 3D medical image registration.

4.7. Registration generalization validation

We also conducted cross dataset brain registration to verify the generalization ability of our network. The Mindboggle101 includes adult brain MRI, while the CANDI consists of child and adolescent brain MRI. There are apparent domain gaps between the two datasets, as shown in Fig. 4. We employ one of them as training set and test the model on the other dataset. The results are shown in Table 6. The transfer model from Mindboggle101 to CANDI shows close registration performance compared with the original model trained on the same dataset. The registration accuracy of the transfer model from CANDI to Mindboggle101 drops a little, but it still outperforms several DLIR methods. These results indicate that our network has good generalization ability.

5. Discussion and conclusion

This paper presents a novel symmetric pyramid network for medical image inverse consistent diffeomorphic registration. Different from previous multi-scale registration methods that train an individual registration network at each scale, we first encode the multi-scale images to the feature pyramids, then progressively conduct the feature level diffeomorphic registration from the bottom to the top level and train the network in an end-to-end manner. Fig. 9 shows the multi-scale registration results of our method, revealing that the network gradually achieves more accurate alignment. The feature warping operation employs the previous level deformation field to warp the current level feature. Fig. 10 further visualizes the feature cosine similarity map before and after the operation. We can observe that the warped feature pair show higher similarity than the initial feature pair, demonstrating this operation can reduce the feature-matching distance.

To guarantee inverse consistency, Most of the existing methods are based on inverse consistency loss or “mean” shape. We independently conduct the forward and backward feature level registration and average the bidirectional velocity fields for more robust estimation. It should be noticed that the bidirectional velocity field average improves the registration accuracy and highly reduces the inverse consistent error. Additionally, our approach is entirely symmetrical, and the results are unrelated to the order of input images.

The proposed symmetric pyramid network still has several limitations. First, compared with VM (Balakrishnan et al., 2018), the proposed network has a similar number of parameters but takes twice the GPU memory in the training stage. This is mainly because our network performs registration at multi-scale feature levels and implements bidirectional registration simultaneously. Furthermore, the current model is still limited to single-modality registration. In the future, we will consider extending the network to multi-modality registration by adopting the encoder to extract modality-independent features and employing multi-modality similarity measures.

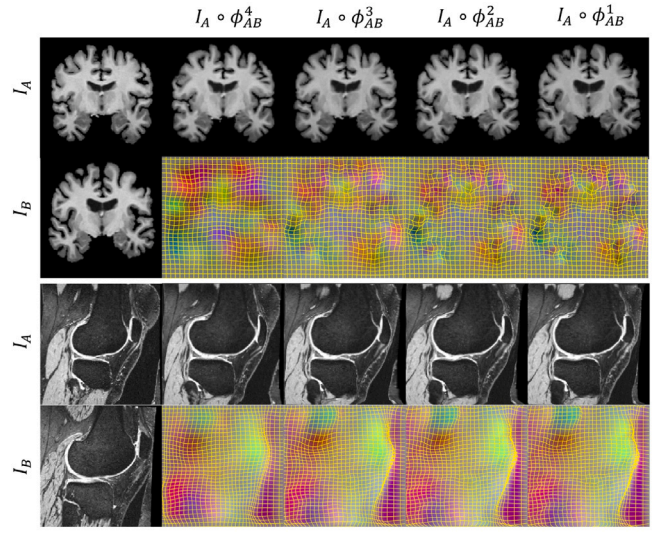


Fig. 9. Visualization of the multi-scale registration results. The network achieves more accurate alignment from the bottom to the top level.

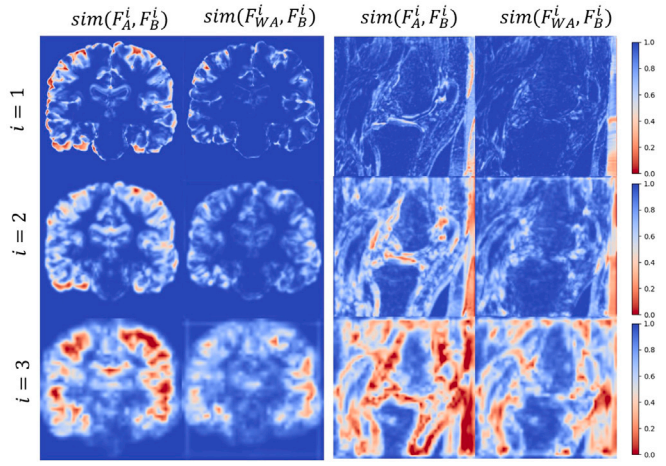


Fig. 10. Visualization of the feature cosine similarity map before and after the feature warping operation. The warped feature pair show higher similarity.

To sum up, we propose a general framework that can achieve accurate registration while maintaining the diffeomorphic properties and inverse consistency. We conduct the experiments on three different kinds of benchmark datasets. The results show that our method significantly outperforms other methods on multiple metrics, indicating the effectiveness and generalization of the proposed network. Our method shows great potential in medical research and clinical workflow.

CRediT authorship contribution statement

Liutong Zhang: Conceptualization, Methodology, Writing – original draft, Writing – review & editing. **Guochen Ning:** Validation, Writing – review & editing. **Lei Zhou:** Validation, Visualization. **Hongen Liao:** Funding acquisition, Methodology, Supervision, Writing – review & editing, Project administration.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

The authors acknowledge supports from National Natural Science Foundation of China (82027807, 62201315, U22A2051), Beijing Municipal Natural Science Foundation, China (7212202), and National Key Research and Development Program of China (2022YFC2405200).

References

Ambellan, F., Tack, A., Ehlke, M., Zachow, S., 2019. Automated segmentation of knee bone and cartilage combining statistical shape knowledge and convolutional neural networks: Data from the osteoarthritis initiative. Med. Image Anal. 52, 109–118.

Avants, B.B., Epstein, C.L., Grossman, M., Gee, J.C., 2008. Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. Med. Image Anal. 12 (1), 26–41.

Avants, B.B., Tustison, N., Song, G., et al., 2009. Advanced normalization tools (ANTS). Insight J. 2 (365), 1–35.

Balakrishnan, G., Zhao, A., Sabuncu, M.R., Guttag, J., 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. pp. 9252–9260.

Beg, M.F., Miller, M.I., Trouvé, A., Younes, L., 2005. Computing large deformation metric mappings via geodesic flows of diffeomorphisms. Int. J. Comput. Vis. 61 (2), 139–157.

Chakravarty, M.M., Bertrand, G., Hodge, C.P., Sadikot, A.F., Collins, D.L., 2006. The creation of a brain atlas for image guided neurosurgery using serial histological data. Neuroimage 30 (2), 359–376.

Christensen, G.E., Johnson, H.J., 2001. Consistent image registration. IEEE Trans. Med. Imaging 20 (7), 568–582.

Dalca, A.V., Balakrishnan, G., Guttag, J., Sabuncu, M.R., 2018. Unsupervised learning for fast probabilistic diffeomorphic registration. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 729–738.

de Vos, B.D., Berendsen, F.F., Viergever, M.A., Sokooti, H., Staring, M., Išgum, I., 2019. A deep learning framework for unsupervised affine and deformable image registration. Med. Image Anal. 52, 128–143.

de Vos, B.D., Berendsen, F.F., Viergever, M.A., Staring, M., Išgum, I., 2017. End-to-end unsupervised deformable image registration with a convolutional neural network. In: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. Springer, pp. 204–212.

Dupuis, P., Grenander, U., Miller, M.I., 1998. Variational problems on flows of diffeomorphisms for image matching. Quart. Appl. Math. 587–600.

Eppenhof, K.A., Pluim, J.P., 2018. Pulmonary CT registration through supervised learning with convolutional neural networks. IEEE Trans. Med. Imaging 38 (5), 1097–1105.

Fu, Y., Lei, Y., Wang, T., Higgins, K., Bradley, J.D., Curran, W.J., Liu, T., Yang, X., 2020. LungRegNet: An unsupervised deformable image registration method for 4D-CT lung. Med. Phys. 47 (4), 1763–1774.

Glocker, B., Komodakis, N., Tziritas, G., Navab, N., Paragios, N., 2008. Dense image registration through MRFs and efficient linear programming. Med. Image Anal. 12 (6), 731–741.

Gu, D., Cao, X., Ma, S., Chen, L., Liu, G., Shen, D., Xue, Z., 2020. Pair-wise and group-wise deformation consistency in deep registration network. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 171–180.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 770–778.

Hering, A., van Ginneken, B., Heldmann, S., 2019. Mlvirnet: Multilevel variational image registration network. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 257–265.

Hering, A., Häger, S., Moltz, J., Lessmann, N., Heldmann, S., van Ginneken, B., 2021. CNN-based lung CT registration with multiple anatomical constraints. Med. Image Anal. 102139.

Jaderberg, M., Simonyan, K., Zisserman, A., et al., 2015. Spatial transformer networks. Adv. Neural Inf. Process. Syst. 28, 2017–2025.

Jiang, Z., Yin, F.-F., Ge, Y., Ren, L., 2020. A multi-scale framework with unsupervised joint training of convolutional neural networks for pulmonary deformable image registration. Phys. Med. Biol. 65 (1), 015011.

Kennedy, D.N., Haselgrove, C., Hodge, S.M., Rane, P.S., Makris, N., Frazier, J.A., 2012. CANDIShare: a resource for pediatric neuroimaging data. Springer.

Kim, B., Kim, D.H., Park, S.H., Kim, J., Lee, J.-G., Ye, J.C., 2021. CycleMorph: Cycle consistent unsupervised deformable image registration. Med. Image Anal. 71, 102036.

Kingma, D.P., Ba, J., 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

Klein, A., Tourville, J., 2012. 101 Labeled brain images and a consistent human cortical labeling protocol. Front. Neurosci. 6, 171.

Kuang, D., Schmah, T., 2019. Faim–a convnet method for unsupervised 3d medical image registration. In: International Workshop on Machine Learning in Medical Imaging. Springer, pp. 646–654.

Li, H., Fan, Y., 2020. MDReg-Net: Multi-resolution diffeomorphic image registration using fully convolutional networks with deep self-supervision. arXiv preprint arXiv:2010.01465.

Liu, L., Hu, X., Zhu, L., Heng, P.-A., 2019. Probabilistic multilayer regularization network for unsupervised 3D brain image registration. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 346–354.

Lorenzi, M., Ayache, N., Frisoni, G.B., Pennec, X., (ADNI, A.D.N.I., et al., 2013. LCC-Demons: A robust and accurate symmetric diffeomorphic registration algorithm. NeuroImage 81, 470–483.

Mok, T.C., Chung, A., 2020a. Fast symmetric diffeomorphic image registration with convolutional neural networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4644–4653.

Mok, T.C., Chung, A., 2020b. Large deformation diffeomorphic image registration with Laplacian pyramid networks. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 211–221.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al., 2019. Pytorch: An imperative style, high-performance deep learning library. Adv. Neural Inf. Process. Syst. 32, 8026–8037.

Rohé, M.-M., Datar, M., Heimann, T., Sermesant, M., Pennec, X., 2017. SVF-net: Learning deformable image registration using shape matching. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 266–274.

Rueckert, D., Sonoda, L.I., Hayes, C., Hill, D.L., Leach, M.O., Hawkes, D.J., 1999. Nonrigid registration using free-form deformations: Application to breast MR images. IEEE Trans. Med. Imaging 18 (8), 712–721.

Shen, Z., Han, X., Xu, Z., Niethammer, M., 2019. Networks for joint affine and non-parametric image registration. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4224–4233.

Sokooti, H., De Vos, B., Berendsen, F., Lelieveldt, B.P., Išgum, I., Staring, M., 2017. Nonrigid image registration using multi-scale 3D convolutional neural networks. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 232–239.

Stefanescu, R., Pennec, X., Ayache, N., 2004. Grid powered nonlinear image registration with locally adaptive regularization. Med. Image Anal. 8 (3), 325–342.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., 2015. Going deeper with convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1–9.

Tao, G., He, R., Datta, S., Narayana, P.A., 2009. Symmetric inverse consistent nonlinear registration driven by mutual information. Comput. Methods Programs Biomed. 95 (2), 105–115.

Thirion, J.-P., 1998. Image matching as a diffusion process: An analogy with Maxwell's demons. Med. Image Anal. 2 (3), 243–260.

Vercauteren, T., Pennec, X., Perchant, A., Ayache, N., 2009. Diffeomorphic demons: Efficient non-parametric image registration. NeuroImage 45 (1), S61–S72.

Wang, J., Sun, K., Cheng, T., Jiang, B., Deng, C., Zhao, Y., Liu, D., Mu, Y., Tan, M., Wang, X., et al., 2020. Deep high-resolution representation learning for visual recognition. IEEE Trans. Pattern Anal. Mach. Intell. 43 (10), 3349–3364.

Xu, Z., Niethammer, M., 2019. DeepAtlas: Joint semi-supervised learning of image registration and segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 420–429.

Yang, D., Li, H., Low, D.A., Deasy, J.O., El Naqa, I., 2008. A fast inverse consistent deformable image registration method based on symmetric optical flow computation. Phys. Med. Biol. 53 (21), 6143.

Zhang, J., 2018. Inverse-consistent deep networks for unsupervised deformable image registration. arXiv preprint arXiv:1809.03443.

Zhang, Y., Pei, Y., Guo, Y., Ma, G., Xu, T., Zha, H., 2020. Fully convolutional network for consistent voxel-wise correspondence. In: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 34, no. 07. pp. 12935–12942.

Zhang, L., Zhou, L., Li, R., Wang, X., Han, B., Liao, H., 2021. Cascaded feature warping network for unsupervised medical image registration. In: 2021 IEEE 18th International Symposium on Biomedical Imaging. ISBI, IEEE, pp. 913–916.

Zhao, S., Dong, Y., Chang, E.I., Xu, Y., et al., 2019. Recursive cascaded networks for unsupervised medical image registration. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 10600–10610.