Applied Intelligence
https://doi.org/10.1007/s10489-022-04109-8

SEN-FCB: an unsupervised twinning neural network for image registration

Mingrui Ma[1,2] · Guixia Liu[1,2] · Lei Song[1,2] · Yuanbo Xu[1,2]

Accepted: 23 August 2022

Abstract

Medical image registration is a fundamental and vital task in medical image analysis. Deformable medical image registration generates a dense nonlinear transformation from the moving image to the fixed image. Current learning-based image registration methods utilize U-shaped networks, concatenate moving and fixed images as one input, and then impose a global regularization to ensure smooth deformation fields. However, existing deformable image registration approaches concatenate image pairs as one input to their model and may ignore independent anatomical relevance of the images. Moreover, the global regularization causes over/underconstraining, affecting their model registration accuracy and over/under enforcing the deformation field's smoothness. To address these two problems, we propose a twinning network, consisting of two subnetworks. The first subnetwork is the proposed separate encoding neural network (SEN) for predicting high-accuracy deformation fields, and the second subnetwork is a folding correction block (FCB) to correct the deformation fields to achieve folding reduction. Comparing our experimental results to the state-of-the-art displacement and diffeomorphic methods, the proposed method provides superior registration accuracy and reduces the folding numbers. Moreover, we utilize the FCB to correct the baselines' output deformation fields, proving that the FCB outperforms global regularization.

Keywords Deformable image registration · Deep learning · CNN · Folding correction

1 Introduction

Image registration is fundamental and crucial in many medical image analysis tasks. As a part of the medical image registration task, deformable registration aims to construct a dense and nonlinear transformation from a source image to a target image (denoted as moving image and fixed image) to represent the variations in anatomical shapes in images caused by factors including patient motion, organ motion, and disease development. For example, deformable registration enables researchers to compare the organ anatomical structure evolution of patients over time longitudinally, or the organ differences between individuals with disease and healthy individuals horizontally, which is critical for understanding the evolution of organ anatomical structures of a disease [1–3].

Recently, with the rapid development and superior performance of deep learning, deep learning has been widely applied in various medical imaging analysis tasks and has achieved remarkable success in many medical imaging applications. Especially in registration, unsupervised deep learning-based methods [4–8] have been proposed and demonstrated to achieve higher performance without ground-truth information for deformable medical image registration. These methods generally utilize a convolutional neural network (CNN) to estimate a deformation field from a pair of images. Then a spatial transform

Guixia Liu, Lei Song and Yuanbo Xu are contributed equally to this work.

✉ Guixia Liu
liugx@jlu.edu.cn

Mingrui Ma
mamr19@mails.jlu.edu.cn

Lei Song
songlei20@mails.jlu.edu.cn

Yuanbo Xu
yuanbox@jlu.edu.cn

[1] College of Computer Science and Technology, Jilin University, Changchun, 130012, Jilin, People's Republic of China

[2] Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education, Jilin University, Changchun, 130012, Jilin, People's Republic of China

Published online: 23 September 2022

network (STN) [9] is utilized to interpolate one image via the deformation field to the other. The average similarity metric score on the anatomical segmentations, achieves higher registration accuracy than conventional methods [10]. Most learning-based methods use UNet-like architectures and concatenate a pair of images as one 2-channel image input to their model, which fuses the features before extracting the independent tissue information contained in each resolution level [4, 5, 8, 8, 11, 12]. Nevertheless, these methods ignore the independent information of each image in an image pair. The recent dual-steam, also called two-stream research [13], states that modeling each image of the input image pair individually enhances the deep representation of their model.

Furthermore, medical images, which represent anatomical information in digital data, should have a realistic deformation field. Put differently, an image converted into another via a deformation field should retain its anatomical structure topology properties, which means that a deformation field should be smooth and have less folding within the transformation. Most unsupervised learning-based registration methods [4, 5, 8, 11] impose a global regularization on the gradient of an output displacement field to restrict the deformation to be smooth. However, the problem is that the regularization presumes that the deformations are in the same smoothness hypothesis, which causes over- or underconstraining for establishing anatomical correspondence. Learning-based diffeomorphic stationary velocity field methods [6, 12] provide diffeomorphic transformations to ensure the topology properties, i.e., to restrict the folding and even reduce the number of folding regions to zero. Nevertheless, studying the dynamic motion of organs requires discontinuous transformation [14], which is different than the continuous property of diffeomorphism. Thus, diffeomorphic methods perform poorly when registering the intrapatient organ (e.g., heart or lungs) images of different systolic cycles. Although the displacement-based approaches can generate discontinuous transformations, they also struggle with the annoying issue of folding.

To reduce folding in the deformation field and to ensure as much registration accuracy as possible, we divide the image registration task into two subtasks. The first is to compute highly accurate displacement fields using a global and coarse regularization function; the second is to use a model to find the folding and correct it to be smooth (i.e., reduce the number of folding regions). Here, we propose an unsupervised learning-based image registration method consisting of twinning networks, including a separate encoding network and a folding correction block. Our main contributions are as follows:

- We propose a novelty separate encoding network for unsupervised deformable image registration, which separately models the independent information of each image of an image pair to enhance the deep representation of the registration model.
- To further reduce the folding in a deformation field instead of using the global regularization, we propose a novelty folding correction block, a general module for 2D images and 3D volumes, which can learn folding features, recognize folding in displacement fields, and smooth the displacement fields.
- Quantitative and qualitative results demonstrate that the proposed twinning method outperforms the state-of-the-art displacement-based and velocity-based methods both in the registration accuracy and the number of folding regions in the transformation.
- We use folding a correction block to revises the state-of-the-art methods output displacement fields and correct them. The results prove that the folding correction block is more applicable and effective than global smooth regularization.

2 Related work

2.1 Conventional deformable approaches

Conventional deformable image registration methods usually employ a similarity function such as NCC [15, 16], MSE [17, 18], or NMI [19] to optimize the registration model iteratively to maximize the similarity between an input pair of images. These methods, including elastic-type models [15, 20], discrete methods [21, 22], and DRAMMS [23] establish the spatial correspondence of two images. These methods regularize the displacement field to be smoothed by using a regularization function or smoothing filter. In addition, several conventional studies use the diffeomorphic model to guarantee that the produced deformation field is differentiable, topology-preserving, and invertible [24, 25]. Diffeomorphic models such as LDDMM [26, 27] and SyN [10] are wildly used and recognized. However, these iterative methods are time-consuming and require a large number of computational resources to register an image pair.

2.2 Learning-based approaches

Many supervised learning-based methods have recently been proposed for deformable registration tasks. These methods usually utilize a CNN model to learn a dense correspondence between an input pair of images. Furthermore, most of these supervised methods [28–31] require images with a ground-truth deformation field or anatomical segmentation to supervise the learning process. Supervised methods have demonstrated outperformance in the image registration task. However, this ground-truth information requires complex annotations by experts or must be produced by conventional methods. Put differently, information for these approaches is difficult to obtain or is not appropriate as ground-truth information in practice.

In recent years, to avoid the difficulty of collecting supervised information, most learning-based approaches have focused on [4–8, 11, 12] unsupervised training. Unsupervised methods first compute a displacement field, then utilize an STN to warp the moving image to a fixed image, and then use a differentiable similarity function to learn the dense spatial transformations in the image pair. For example, [4] proposed an unsupervised UNet-like 3D volume registration method that employs NCC similarity and L_2 regularization constraint displacement field smoothness. However, computing smooth and topology-preserving transformations is still a challenge. To further avoid folding and obtain a topology-preserving warped image, stationary velocity fields are used in diffeomorphic approaches. Dalca et al. [6] proposed a probabilistic diffeomorphic registration method that offers uncertainty estimation within CNN and diffeomorphic integration models. Mok and Chung [12] proposed a diffeomorphic model to estimate both forward and inverse velocity fields simultaneously. Al Safadi and Song [32] proposed a meta-regularization method to learn regularization filters to generate a smoother displacement vector field. Kang et al. [13] employed the two-stream architecture, separately modeling the moving and fixed image to the bottom of the encoder, then restoring them to the upper resolution and fusing their feature maps in each level of the decoder.

In contrast to the abovementioned unsupervised learning-based approaches, we divide image registration into two subproblems and solve them step by step. Unlike most of these methods architectures, the proposed deformable image registration model fully considers the independent information of each image in the input pair and provide displacement fields with coarse regularization. Compared to the recent two-stream method [13], we introduce a separate encoding network that focuses on independent and combined hybrid encoding for the input image pair information. The correction model then learns to distinguish folded feature maps in the displacement field and regularize them to be smooth.

3 Methods

Let X and Y be two images defined in the spatial domain $\Omega = R^i$ $(i = 2, 3)$. Figure 1 illustrates the overall architecture of the proposed twinning deformable image registration neural network. Our proposed method is divided into two stages: one for image registration and another for correcting folding. In the image registration phase, SEN is the proposed convolutional neural network for deformable image registration, which computes the displacement fields between X and Y. $\mathcal{L}_{reg}$ is the registration loss function.

Many displacement field based approaches [4, 5, 7, 11] employ global regularization, which causes over-or under-constraining and affects their registration performance. Unlike the mentioned methods, we use a model to correct the output displacement fields to reduce the folding. In the folding correction phase, FCB is a proposed convolutional neural network for correcting the displacement field. After SEN training is finished in this coarse regularization way, we freeze the SEN parameters and output the predicted displacement field to FCB. $\mathcal{L}_{fc}$ is the folding correction loss.

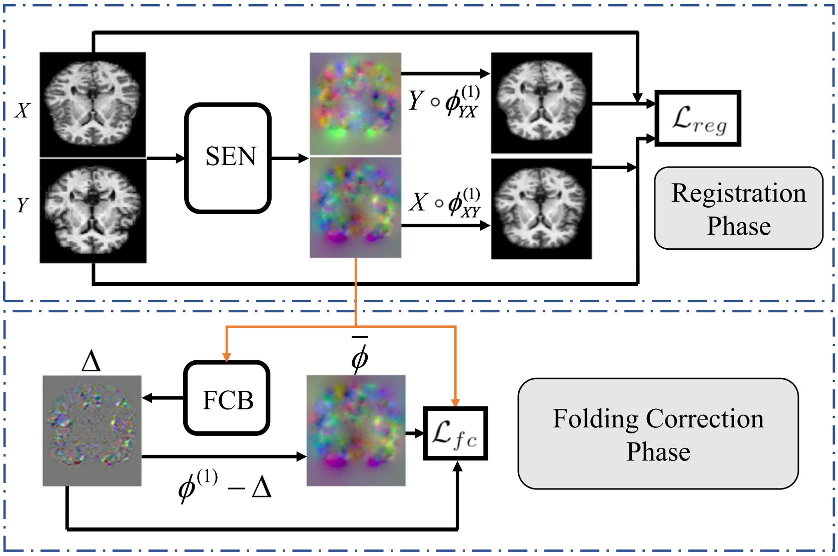


Fig. 1 Overview of the proposed twinning method for deformable image registration. $\circ$ denotes the spatial transform network. Δ is the residual factor. $\bar{\phi}$ represents the corrected deformation field

3.1 Separate encoding neural network

We take a deep unsupervised approach to learn the generation function for displacements, which the function denoted as $g_\theta(X, Y) = (\phi^{(1)}_{XY}, \phi^{(1)}_{YX})$. g_θ represents the separate encoding neural network (SEN) with its parameters θ. $\phi^{(1)}_{XY}$ and $\phi^{(1)}_{YX}$ are two output displacement fields, which are the direct and inverse displacement fields. The motivation for estimating the bidirectional deformation is to guarantee the existence of the inverse transformation [12, 33]. To guarantee the invertible property in registration, let $x \in X$ and $y \in Y$, we compute two directions by functions $\phi^{(1)}_{XY} = \phi^{(0)}_{Y\ X}(-\phi^{(0)}_{XY}(x))$ and $\phi^{(1)}_{YX} = \phi^{(0)}_{X\ Y}(-\phi^{(0)}_{YX}(y))$, where $\phi^{(0)}_{XY}$ and $\phi^{(0)}_{YX}$ are two output displacement fields from g_θ. Therefore, g_θ can be rewritten as $g_\theta(X, Y) = (\phi^{(0)}_{XY}(-\phi^{(0)}_{YX}(y)), \phi^{(0)}_{YX}(-\phi^{(0)}_{XY}(y)))$.

Architecture of SEN As shown in Fig. 2a, our proposed convolutional neural network consists of a 5-level hierarchical encoder-decoder with skip connections, which is similar to UNet [34]. Unlike formal U-shaped networks [4–7, 12] that concatenate *X*/fixed and *Y*/moving image volumes as a single 2-channel input, the proposed SEN is divided into three branches in the encoder. The first branch extracts feature maps for *X*, the second branch extracts feature maps for *Y*, and the third branch extracts feature maps for concatenated *XY* in each level encoder. The proposed SEN concatenates these three-branched feature maps at each level, then further computes the concatenated feature maps and downsamples these feature maps to the next resolution level encoder block as the input to the third branch. The blocks in the encoder consist of 3×3×3 kernel size convolutional layers with a stride of 1, followed by a rectified linear unit (ReLU) activation for computing the constant size feature maps in each resolution level. We apply 3×3×3 kernel size convolutional layers with a stride of 2 followed by a ReLU activation to downsample the feature maps in half until the lowest resolution level is reached. For each resolution level in the decoder, we apply 3 × 3 × 3 kernel size convolutional layers with a stride of 1 followed by ReLU activation and a 2 × 2 × 2 deconvolutional layer to upsample feature maps to twice their size and then concatenate them with the feature maps from the encoder through skip connections. To ensure that inverse transformation exists, we utilize two 3 × 3 × 3 convolutional layers with a stride of 1 followed by softsign activation (i.e., $softsign(x) = \frac{x}{1+|x|}$) to normalize the feature maps to $[-1, 1]$ to obtain direct $\phi^{(0)}_{XY}$ and inverse $\phi^{(0)}_{YX}$, and then each of them is multiplied by a constant c within the range $[-c, c]$ to obtain displacement fields.

3.2 Folding correction block

We freeze the parameters when our proposed SEN training is finished, and then we reuse the train set through the trained SEN to obtain the displacement fields. We take an unsupervised approach to learn the generation function $f_{\theta'}(\phi^{(1)}) = \Delta$ for correcting displacements. $f_{\theta'}$ is the proposed folding correction block (FCB) with its parameter $\phi^{(1)}$. Δ is a factor that is used to reduce the folding in the input displacement fields by the formula $\bar{\phi} = c \times \phi^{(1)} - \Delta$. This formula indicates that Δ includes folding location information and the magnitude of the displacement fields that need to be corrected. $\bar{\phi}$ is the corrected displacement field. Figure 3 shows some folding regions represented by the folding in the grid figure. We can observe that the FCB can recognize folding and correcting to smooth the transformation, i.e., the grid line without crossing.

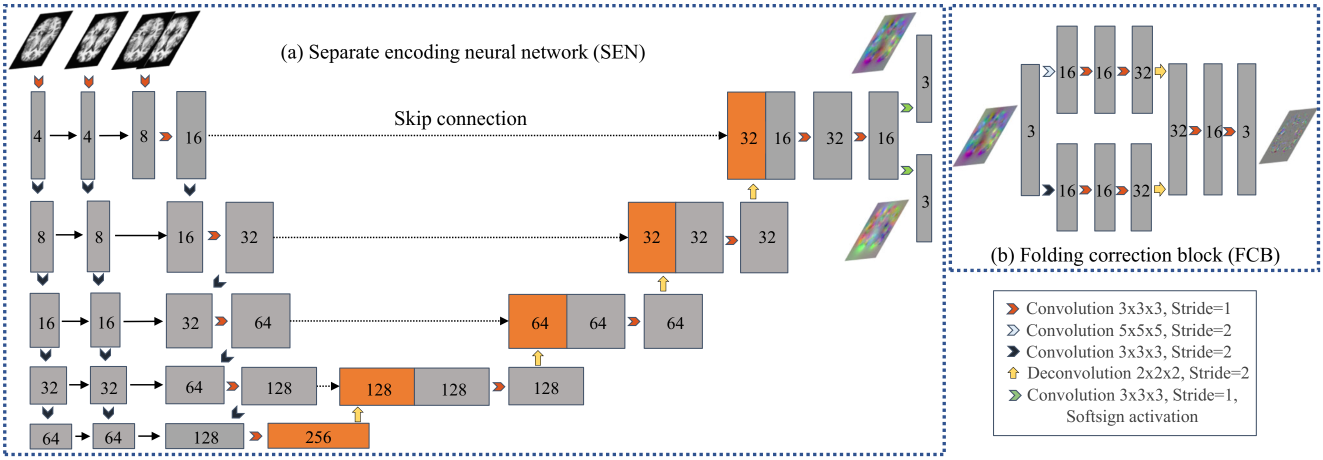

(a) Separate encoding neural network (SEN)
Skip connection
4
4
8
16
8
8
16
32
16
16
32
64
32
32
64
128
64
64
128
256
128
128
128
64
64
64
32
32
32
32
16
32
16
3
3
(b) Folding correction block (FCB)
3
16
16
32
16
16
32
32
16
3
Convolution 3x3x3, Stride=1
Convolution 5x5x5, Stride=2
Convolution 3x3x3, Stride=2
Deconvolution 2x2x2, Stride=2
Convolution 3x3x3, Stride=1,
Softsign activation

Fig. 2 Illustration of our two subnetworks. (a) illustrates our proposed fully connected network SEN architecture to predict the bidirectional deformation fields. The gray and orange blocks indicate the 3D feature maps from the encoder and decoder, respectively. (b) illustrates our proposed FCB architecture utilized to reduce the folding regions in the SEN predictions

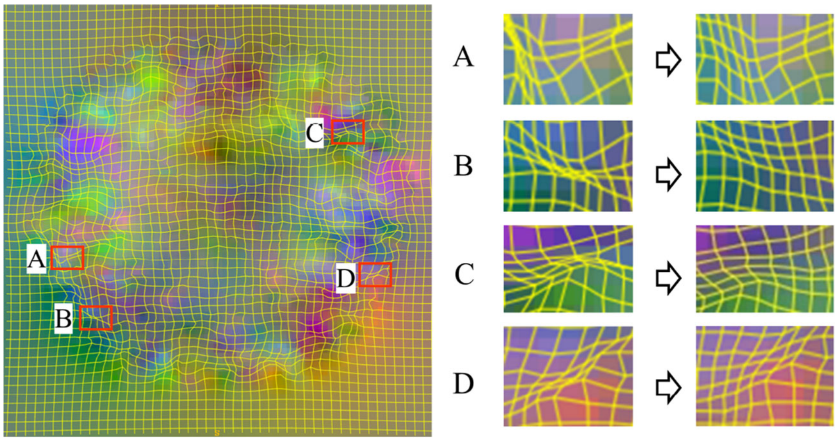


Fig. 3 An example of showing the FCB correcting the folding regions in a displacement field. The grid figure is the visualization of a displacement field. The red frames marked in the deformation field are parts of folding regions. We zoom the marked regions and then find the gridline crossing. The arrows indicate the corrected local displacement field

Architecture of FCB As shown in Fig. 2b, the proposed FCB consists of four 3×3×3 kernel size convolutional layers, each of them with a stride of 1 and followed by a ReLU activation except for the last layer. Convolutional layers with kernel sizes of 3×3×3 and 5×5×5 with different strides followed by ReLU activation downsample the input displacement fields to the same size. The deconvolutional layers upsample the 1/2 resolution feature maps to the shaped of the original input displacement fields. Then, the last layer outputs the residual factor Δ.

3.3 Loss functions

3.3.1 Registration loss function

We employ the registration loss function to penalize the displacement field $\mathcal{L}_{reg} = \mathcal{L}_{sim}(\cdot) + \mathcal{L}_{smooth}(\cdot)$, which is divided into the similarity loss function and the regularization loss function. Each of these two loss functions is pairwise, both consisting of bidirection losses. We use normalized cross-correlation (NCC) and mean square error (MSE) as the similarity loss functions to measure the similarity between the warped image and the fixed image. To measure the similarity between warped X and Y and warped Y and X, the similarity loss function is formulated as

$$\mathcal{L}_{sim}(X, Y) = \mathcal{L}_{sim}(X \circ (\phi_{XY}^{(1)}), Y) + \mathcal{L}_{sim}(Y \circ (\phi_{YX}^{(1)}), X), \tag{1}$$

where $X \circ (\phi_{XY}^{(1)})$ and $Y \circ (\phi_{YX}^{(1)})$ represent image X warped toward Y via the displacement field $\phi_{XY}^{(1)}$ and image Y warped toward X via the displacement field $\phi_{YX}^{(1)}$ respectively. $\mathcal{L}_{sim}$ is NCC when $\Omega = R^3$ and $\mathcal{L}_{sim}$ is MSE when $\Omega = R^2$. A higher NCC value or a smaller MSE value indicates a better alignment.

We enforce the deformation field coarse smoothness using an L_2 regularization loss function with ∇, which denotes the spatial gradient using differences with neighboring positions. Thus, $\mathcal{L}_{smooth}$ can be defined as follows:

$$\mathcal{L}_{smo}(\phi_{XY}^{(1)}, \phi_{YX}^{(1)}) = \sum_{x\in\Omega} (\|grad(\phi_{XY}^{(1)})\|^2 + \|grad(\phi_{YX}^{(1)})\|^2). \tag{2}$$

Therefore, the registration loss function of our first network can be written as follows:

$$\mathcal{L}_{reg}(X, Y) = \mathcal{L}_{sim}(X, Y) + \lambda_1 \mathcal{L}_{smo}(\phi_{XY}^{(1)}, \phi_{YX}^{(1)}), \tag{3}$$

where λ_1 is a hyperparameter that balances the accuracy of the network predictions and the coarse smoothness of the output displacement fields.

3.3.2 Folding correction loss function

We propose a folding correction loss function $\mathcal{L}_{fc} = \mathcal{L}_{sim_2}(\cdot) + \mathcal{L}_{Jdet}(\cdot) + \mathcal{L}_{enc}(\cdot)$ consisting of three terms, including the deformation field similarity loss $\mathcal{L}_{sim_2}$, the Jacobian determinant regularization loss $\mathcal{L}_{Jdet}$ and the regularization $\mathcal{L}_{enc}$. In this section, $\mathcal{L}_{sim_2}$ is an MSE similarity function, which is used to measure the similarity between $\bar{\phi}$ and $\phi^{(1)}$. $\mathcal{L}_{sim_2}$ can be formulated as follows:

$$\mathcal{L}_{sim}(\phi^{(1)}, \bar{\phi}) = MSE(\phi^{(1)}, \bar{\phi}), \tag{4}$$

where $\phi^{(1)}$ is the input displacement field and $\bar{\phi}$ is the corrected displacement field.

We utilize the Jacobian determinant in the second term in the proposed folding correction loss function because it is positive when the displacement field is smooth. Put differently, we can say that the Jacobian determinant is folding sensitive. The definition of the Jacobian matrix can be written as follows:

$$J_{\bar{\phi}}(p) = \begin{Vmatrix} \frac{\partial \bar{\phi}_x(p)}{\partial x} & \frac{\partial \bar{\phi}_x(p)}{\partial y} & \frac{\partial \bar{\phi}_x(p)}{\partial z} \\ \frac{\partial \bar{\phi}_y(p)}{\partial x} & \frac{\partial \bar{\phi}_y(p)}{\partial y} & \frac{\partial \bar{\phi}_y(p)}{\partial z} \\ \frac{\partial \bar{\phi}_z(p)}{\partial x} & \frac{\partial \bar{\phi}_z(p)}{\partial y} & \frac{\partial \bar{\phi}_z(p)}{\partial z} \end{Vmatrix}. \tag{5}$$

$J_{\bar{\phi}}(p)$ denotes the Jacobian determinant metric over deformation field $\bar{\phi}$ at position p.

To measure the degree of a folding region in the deformation field, we utilize Jacobian determinant regularization in [12] and give a smooth formulation. The Jacobian determinant regularization is written as follows:

$$\mathcal{L}_{Jdet} = \ln\left(\frac{1}{N}\sum_{p\in\Omega} ReLU(-|J_{\bar{\phi}}(p)|)\right), \tag{6}$$

where N is the total number of elements in $|J_{\bar{\phi}}|$. ReLU is the linear activation function that maintains the values when $J_{\bar{\phi}}(p) \leq 0$ and sets the values to zero when $J_{\bar{\phi}}(p) > 0$.

Aiming at balancing the contribution of $\mathcal{L}_{sim}$ and $\mathcal{L}_{Jdet}$, we use a variant L_2 regularization on the spatial gradient of Δ to encourage its change. Thus, $\mathcal{L}_{enc}$ is defined as follows:

$$\mathcal{L}_{enc} = \ln\sum_{x\in\Omega}(\|grad(\Delta)\|^2) \tag{7}$$

These two functions (4) and (6) in $\mathcal{L}_{fc}$ enforce the adjustment to the local regions with negative Jacobian determinants in the deformation field $\phi^{(1)}$. In contrast, the local regions with positive Jacobian determinants are not corrected. This adjustment is made under the premise that the adjusted deformation field $\bar{\phi}$ is similar to the original deformation field $\phi^{(1)}$. Put differently, the adjusted local region deformation field maintains the constraints and magnitude constrains in the neighborhood. We balance the contributions of these two terms with weight λ_2 multiplied by $J_{\bar{\phi}}(p)$. Therefore, $\mathcal{L}_{fc}$ can be written as $\mathcal{L}_{fc} = \mathcal{L}_{sim} + \lambda_2\mathcal{L}_{Jdet} + \lambda_3\mathcal{L}_{enc}$.

4 Experiments

4.1 Datasets

The first dataset is the EchoNet-Dynamic dataset [35]. This dataset is composed of echocardiogram videos and human expert annotations for the left cardiac ventricle of each subject. We select 1276 image pairs representing end-systole and end-diastole at two separate times in each video, which are annotated by human experts. We use the end-systole phase image as Y and the end-diastole phase image as X. The selected image pairs are randomly divided into 920 for training, 100 for validation, and 256 for evaluatation for each method.

The second dataset is OASIS [36] preprocessed in [4], which consists of a cross-sectional collection of T1-weighted MRI scans from 416 subjects aged 18 to 96, as one of our experimental datasets. These raw MRI scans with shapes of 256 × 256 × 256 and $1\,mm \times 1\,mm \times 1\,mm$ resolution are preprocessed by using FreeSurfer [37], resulting in shapes that are 160 × 224 × 192. We resample these scans into 96 × 112 × 96. We randomly select 270 MRI scans from the dataset, and the scans are divided into 200, 36, and 34 scans for training, validation and testing, respectively. We randomly select 4 and 6 MRI volumes from our validation and testing set as fixed, and the remainder denotes the moving image volumes. We perform a registration task by aligning the moving image volumes to each fixed image. To compare with the other methods, we use X as the fixed image and Y as the moving image volumes in our method, and we register a total of 120 fixed/moving image volume pairs for each method.

4.2 Measurement

Because the ground-truth nonlinear deformation field is challenging to obtain, we evaluate registration performance with the Dice similarity coefficient metric and Jacobian determinant ($|J_\phi|$). For example, we first warp each moving brain MRI volume to each atlas to obtain the deformation field. Then, we warp the anatomical segmentation maps belonging to each moving image to align with the anatomical segmentation maps belonging to each fixed image by using the predicted deformation fields. We evaluate the overlap of the segmentation maps using the percentage of Dice metrics (higher is better). Then, compute $|J_\phi|$ on each displacement field and count the number of pixels with nonpositive Jacobian determinants (i.e., $|J_\phi| \leq 0$, lower is better).

4.2.1 Dice

Dice is a metric for measuring the overlap of anatomical segmentation maps between the warped moving image and the fixed image. In our experiments, for brain MRI, 36 anatomical structures are used for analysis. For cardiac ultrasound images, only the left ventricle annotation is used for analysis. The Dice values ranged from [0, 1], and a high Dice metric indicates a high registration accuracy.

4.2.2 Jacobian determinant

The Jacobian metric is defined in (5). In our experiments, we compute the Jacobian determinant of each displacement field and count the number of pixels or voxels with nonpositive Jacobian determinants (i.e., $|J_\phi| \leq 0$).

4.2.3 Baseline methods and implementation

We compare our proposed method to three unsupervised deep learning-based deformable registration methods. The first and second baseline methods are VoxelMorph (VM) [4] and Vit-V-Net (VVN) [11], both of which predict a displacement field and then utilize global regularization to restrict the displacement fields smoothing. VM employs a UNet and outputs the displacement field directly. VVN is a transformer-based method, that introduces transformer into image registration. The third baseline method is SYMNet (SN) [12], which predicts diffeomorphic transformations. For these methods, we use their official online implementation. We train VM, VVN, and SN and follow the recommended parameter settings in [4, 11, 12]. The proposed method is implemented based on PyTorch. We adopt the Adam [38] optimizer with a learning rate of 0.0001 for SEN and FCB. We train our method and baseline methods on an RTX 3080 GPU. What needs to noted is that we first train our proposed SEN, then we freeze the SEN parameter to compute displacement fields between each image pair. Finally, FCB uses these displacement fields to learn its parameters. The FCB mentioned below are trained based on the SEN predictions. For different datasets, λ_1, λ_2, and λ_3 have different settings, and the specific settings are shown in Section 4, the experimental results.

4.3 Experimental results

4.3.1 Validation on the cardiac dataset

We first evaluate displacement-based methods (VM, VVN) with global regularization $\lambda_1 = (0.04, 0.05)$ and our proposed method on the cardiac dataset. We utilize MSE as the loss function and train these methods for 160,000 iterations. We tune the hyperparameter $\lambda_1 = 0.05$ for the coarse regularization in our method. We set λ_2 and λ_3 to (40, -1).

Analysis and discussion The first part of Table 1 shows the registration results on 256 cardiac ultrasound image pairs. We can observe that our proposed single method SEN outperforms the other two baseline methods on the average Dice metric. Our proposed twinning method, denoted as SEN+FCB, outperforms the others both on the Dice metric and the number of nonpositive $|J_\phi|$. Figure 4 shows a registration result, including displacement field computed by the SEN, the displacement field corrected by the FCB, and the final warped image.

Table 1 Comparison of cardiac ultrasound image results

Method	Dice (%)	$\lvert J_\phi\rvert \leq 0$
Affine Only	75.51 ± 7.00	–
VM ($\lambda_1 = 0.04$)	88.49 ± 4.52	45.39 ± 51.50
VM ($\lambda_1 = 0.05$)	88.14 ± 4.78	36.47 ± 45.65
VVN ($\lambda_1 = 0.04$)	88.97 ± 4.33	61.79 ± 64.84
VVN ($\lambda_1 = 0.05$)	88.22 ± 4.79	39.84 ± 46.70
SEN	**89.32** ± **4.14**	59.27 ± 44.36
SEN + FCB	**89.35** ± **4.16**	37.00 ± 34.85
VM ($\lambda_1 = 0.04$) + FCB	88.58 ± 4.49	21.14 ± 30.65
VVN ($\lambda_1 = 0.04$) + FCB	89.14 ± 4.28	24.73 ± 35.12

Affine only: the results from preprocessing only

The bold entries are the highlighted results that prove our methods outperform the baseline methods

To illustrate that the FCB correction outperforms global regularization, we use the FCB to correct the output displacement fields of VM ($\lambda_1 = 0.04$) and VVN ($\lambda_1 = 0.04$), which indicates VM and VVN are trained with the hyperparameter setting of the global regularization as $\lambda_1 = 0.04$. Then we compare the correction results to the output displacement fields of VM ($\lambda_1 = 0.05$) and VVN ($\lambda_1 = 0.05$), which are trained with the hyperparameter setting of the global regularization as $\lambda_1 = 0.05$. Compared to VM, SN, and VVN, our proposed SEN and SEN+FCB achieve the best Dice metrics. On the average nonpositive $|J_\phi|$ metric, the results of VM are slightly higher than our proposed SEN+FCB. Comparing the nonpositive $|J_\phi|$ standard deviations of all three methods, our proposed SEN+FCB is the lowest among all approaches, which indicates that our method is robust in predicting deformation fields. The second part of Table 1 shows the global regularization and the correction results. It is worth noting that the Dice values are improved while the number of nonpositive $|J_\phi|$ is significantly reduced when the FCB is utilized for VM, VVN, and SEN. This demonstrates that the correction by using the FCB is more effective than using a global regularization. Figure 7 shows the deformation fields of the cardiac images predicted by each method and the warped image with the overlaid segmentation map.

4.3.2 Validation on brain dataset

We evaluate VM, VVN, SN, and our proposed twinning method on 3D brain MRI volumes. For VM, SN, and the proposed SEN, we use NCC as the loss function. We find that the Dice metric is the best when VM is trained with smooth hyperparameter $\lambda_1 = 3$ and VVN is trained with $\lambda_1 = 0.02$ on this 3D brain dataset. We use the recommended global regularization hyperparameters in [12] for SN. SN employs the explicit Jacobian loss term with the hyperparameter λ_o to achieve folding reduction. We use $\lambda_o = (0, 1000)$, which is the recommended setting in [12] to restrict the folding, and then compare it to our proposed twinning method results. We tune the hyperparameter $\lambda_1 = 2$ for the coarse regularization in our method. We set λ_2 and λ_3 to (50000, -0.01).

Analysis and discussion The first part of Table 2 shows the experimental results on brain MRI volumes. We find that for SN, the Dice metrics changes too much while the folding reduction is insufficient when the hyperparameter λ_o of the

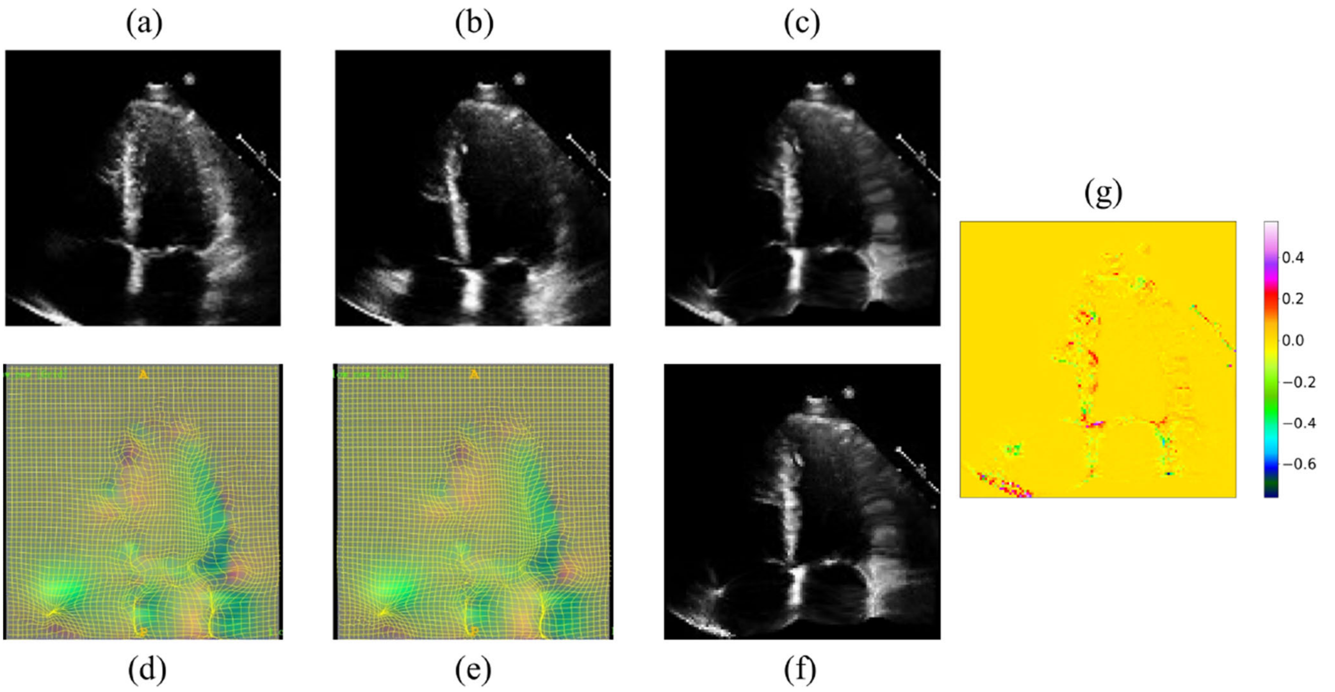


Fig. 4 A sample result of registering two cardiac ultrasound images pair. (a) Fixed image. (b) Moving image. (c) Warped moving image by displacement field. (d) Displacement field. (e) Corrected displacement field. (f) Warped moving image by corrected displacement field. (g) Difference representation between warped image and corrected warped image. Colorbar in (g) represents the normalized difference

explicit Jacobian loss term changed (i.e., 0 → 1000). This result motivates us to design the folding correction block by utilizing the smooth Jacobian loss term. The single SEN achieves the best results on Dice metrics and leads the other methods by almost 1-2%. The proposed twinning method outperforms the others both on Dice and nonpositive $|J_\phi|$ metrics except VVN with the smooth hyperparameter $\lambda_1 = 0.02$. When we adjust λ_1 to 0.05 to make the number of nonpositive $|J_\phi|$ of VVN's result similar to ours, VVN does not perform well on the Dice metric. This is due to the worse performance of the model when the smooth hyperparameter is larger, as stated in [6, 12]. Figure 5 shows a registration result, including the displacement field computed by the SEN, the displacement field corrected by the FCB, and the final transformed brain image. The boxplot in Fig. 6 shows the comparison results for each anatomical structure.

The second part of Table 2 shows the FCB correcting the folding of the other three methods. The results demonstrate that FCB effectively reduces the number of nonpositive $|J_\phi|$ while sacrificing some registration accuracy. FCB reduces almost 85-90% nonpositive $|J_\phi|$ for VM, VVN, and our proposed SEN while reducing 65% nonpositive $|J_\phi|$ for SN. We attribute this gap in the percentage of reducing results to the deformation field generation form: one is based on the displacement field, and the other is based on the velocity field. The results of SN + FCB prove that the use of the additional convolutional block with the Jacobian loss term outperforms than the single network with the explicit Jacobian loss term. Compared to the experimental results on the cardiac dataset, the Dice metrics are reduced after being corrected by the FCB. This is because of the different number of anatomical labels for each subject for evaluation: one label for each ultrasound cardiac image and 36 labels for each brain MRI volume. Overall, compared to SN, FCB can correct the displacement field more effectively and maintain the registration accuracy well. We give each

Table 2 Comparison of brain MRI scans results

Method	Dice(%)	$\|J_\phi\| \leq 0$
Affine Only	56.51 ± 6.32	–
VM ($\lambda_1 = 3$)	72.24 ± 3.00	1066.22 ± 800.86
VM ($\lambda_1 = 5$)	71.74 ± 3.25	212.66 ± 243.01
VVN ($\lambda_1 = 0.02$)	73.00 ± 2.65	1636.18 ± 712.95
VVN ($\lambda_1 = 0.05$)	72.43 ± 2.75	228.641 ± 137.66
SN ($\lambda_o = 0$)	71.92 ± 2.81	1038.65 ± 270.93
SN ($\lambda_o = 1000$)	71.51 ± 2.87	993.62 ± 247.62
SEN (Ours)	**73.32 ± 2.65**	1069.94 ± 227.81
SEN + FCB (Ours)	**72.80 ± 2.83**	**155.39 ± 69.10**
VM ($\lambda_1 = 3$) + FCB	71.74 ± 3.06	98.90 ± 108.68
VVN ($\lambda_1 = 0.02$) + FCB	72.46 ± 2.68	251.48 ± 147.64
SN ($\lambda_o = 0$) + FCB	71.70 ± 2.80	341.33 ± 138.47

Affine only: the results from preprocessing only

The bold entries are the highlighted results that prove our methods outperform the baseline methods

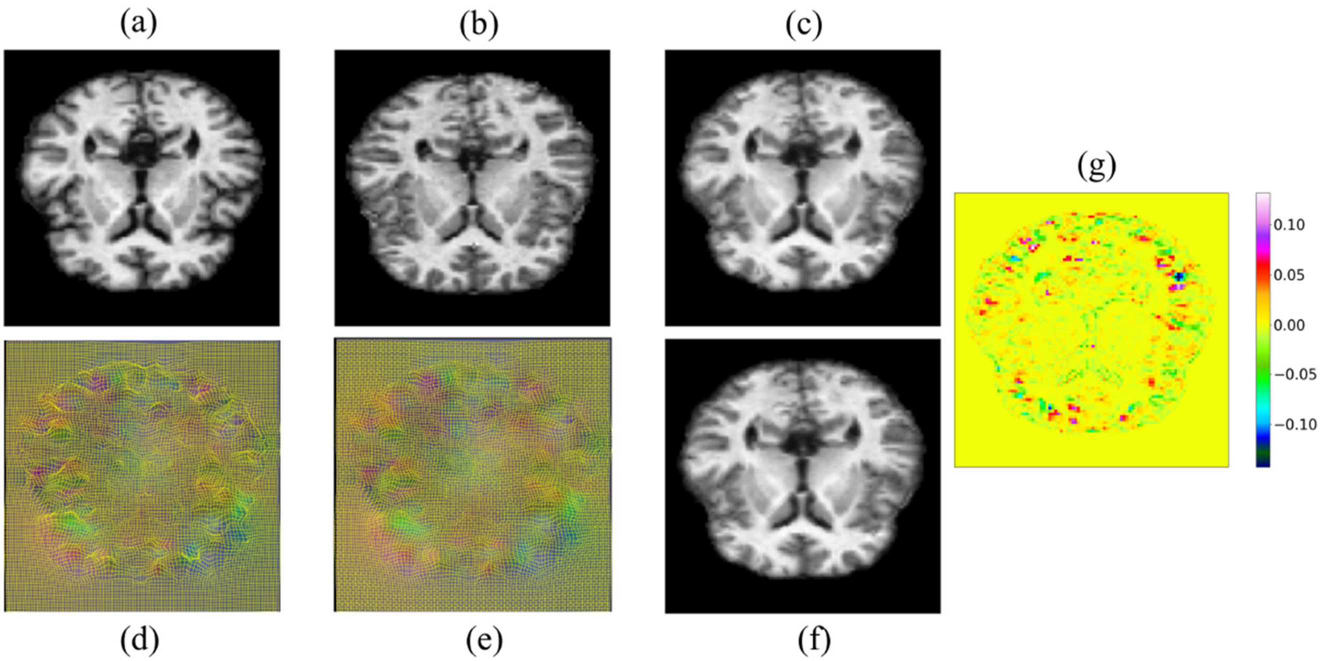


Fig. 5 A sample slice of a result of registering two brain MRI volumes pair. (a) Fixed image, (b) Moving image, (c) Warped moving image by displacement field, (d) Displacement field, (e) Corrected displacement field, (f) Warped moving image by corrected displacement field. (g) Difference representation between warped image and corrected warped image. Colorbar in (g) represents the normalized difference

method's output deformation field of the brain images and the warped images in Fig. 7.

4.3.3 Runtime analysis

We register each pair of images for the nonlinear deformable registration task using an NVIDIA RTX 3080 GPU. We measure the execution time for VM, VVN, SN, SEN, and SEN+FCB. Figure 8 shows the average runtime of our proposed methods and these baseline methods. The results show that our method is faster than both of these baseline methods for registering a pair of images. Furthermore, it is worth noting that utilizing the FCB to correct folding in a deformation field does not significantly increase the registration method runtime.

4.3.4 Ablation study

To demonstrate the effectiveness of our proposed SEN, we remove the separate encoding of each image, leaving

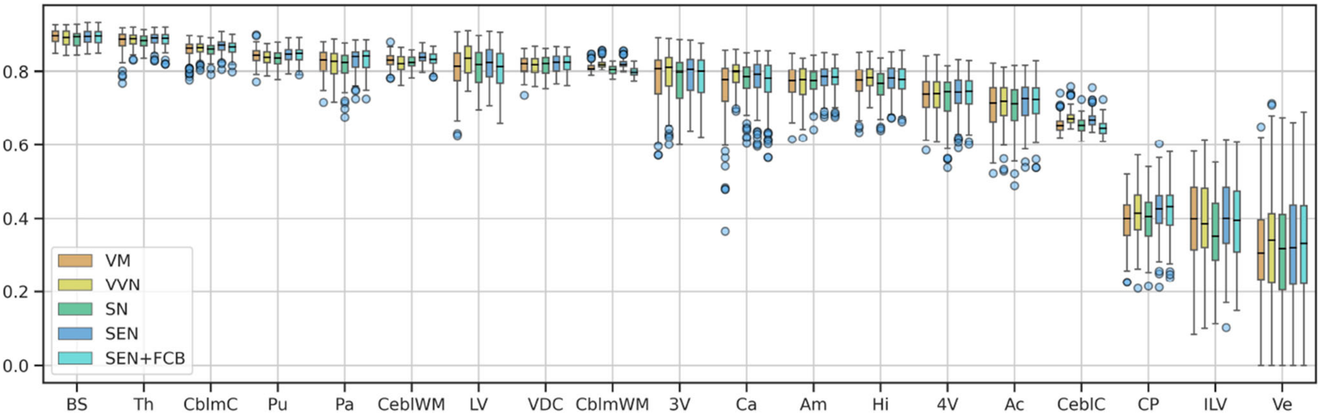


Fig. 6 A boxplot illustrating the Dice value of each anatomical structure segmentation for VM, VVN, SN, and our proposed twinning method. We averaged the Dice values of the left and right brain hemispheres and combined them into one structure for visualization

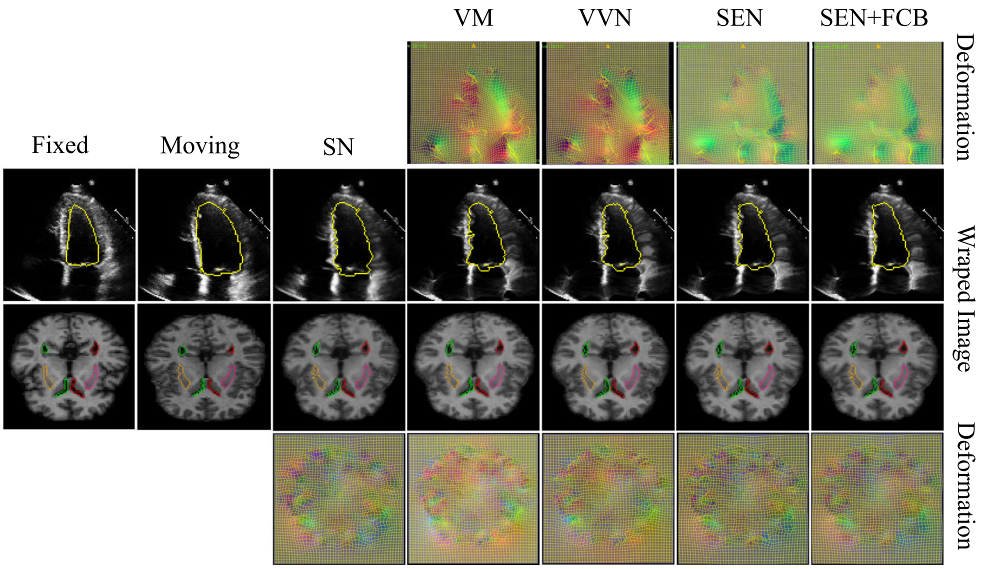


Fig. 7 The view of the fixed/moving image slices and each baseline method's deformation fields and the warped images with the overlaid segmentation maps

only the concatenated encoding branch. Then, we doubled the number of concatenated branch channels to double to keep the number of channels in each level unchanged. As shown in Fig. 2a, this network with the separate encoding branch removed degenerates to an ordinary U-shaped architecture, which is denoted as SEN-1. We apply SEN-1 and SEN to the cardiac and brain datasets. We train SEN and SEN-1 with $\lambda_1 = (0.01, 0.05, 0.1)$ on the cardiac dataset and $\lambda_1 = (2, 4)$ on the brain dataset. We evaluate these two methods on the testing set of the cardiac and brain datasets. Then, the two-direction output displacement field is utilized to warp X and Y. The average Dice metric on the two warped anatomical segmentation maps indicates the registration performance. Table 3 shows that the SEN consistently outperforms the SEN-1 on all hyperparameter settings. This demonstrates that separate encoding for each image enhances the registration accuracy.

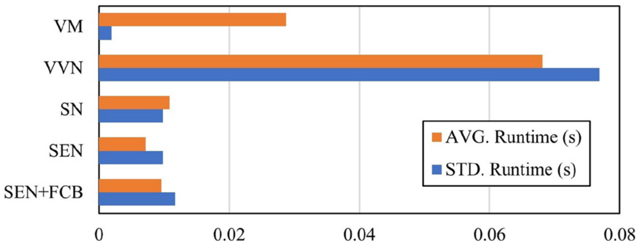


Fig. 8 The bar chart of runtime for each method to register a pair of images. The orange bars are the average runtime. The blue bars are the standard deviations of runtime

5 Conclusion

In this work, we introduce a twinning network for learning-based deformable image registration, which consists of two subnetworks. We utilize the proposed SEN to compute the high-accuracy symmetric displacement fields. Then, we utilize the proposed FCB to correct folding in the output displacement field from SEN. We validate our proposed twinning method on 2D ultrasound cardiac images and 3D brain MRI scans. Compared with three other unsupervised learning-based methods, the experimental results demonstrate that our twinning method achieves high registration accuracy on Dice metrics and reduces the number of nonpositive Jacobian determinants in the predicted

Table 3 Ablation comparison between SEN and SEN-1

Method	Data	λ_1	Dice %	$\lvert J_\phi \rvert$
SEN-1	Cardiac	0.01	88.59 ± 3.85	275.15 ± 124.82
		0.05	89.67 ± 4.96	48.84 ± 38.72
		0.1	89.32 ± 4.24	13.20 ± 15.58
	Brain	2	72.97 ± 2.69	1203.42 ± 230.91
		4	72.48 ± 2.69	132.84 ± 38.69
SEN	Cardiac	0.01	**88.82** ± **3.83**	281.63 ± 126.37
		0.05	**89.84** ± **3.72**	48.30 ± 39.00
		0.1	**89.71** ± **4.10**	19.04 ± 21.43
	Brain	2	**73.24** ± **2.68**	1082.90 ± 207.34
		4	**72.54** ± **2.90**	149.53 ± 40.69

The bold entries are the highlighted results that prove our methods outperform the baseline methods displacement fields compared to baseline methods. Furthermore, the experimental results on FCB correcting displacement fields of the baseline methods demonstrate that FCB outperforms global regularization on folding reduction. The ablation study shows that separate encoding improves the registration performance.

Acknowledgments This work is supported by The National Nature Science Foundation of China (Grant Nos. 61772226 and 61862056), The Natural Science Foundation of Jilin Province (Grant number No. 20210204133YY), Key Laboratory for Symbol Computation and Knowledge Engineering of the National Education Ministry of China, Jilin University.

Author Contributions All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Lei Song, Mingrui Ma and Guixia Liu. The first draft of the manuscript was written by Lei Song and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding This work was supported by the National Nature Science Foundation of China [grant number 61772226, 61862056]; Science and Technology Development Program of Jilin Province [grant number 20210204133YY]; The Natural Science Foundation of Jilin Province (Grant number No. 20200201159JC); Key Laboratory for Symbol Computation and Knowledge Engineering of the National Education Ministry of China,Jilin University.

Availability of data and materials The data is a public data set and it can be obtained https://www.oasis-brains.org/. We preprocessed the data.

Code Availability Our code is available at https://github.com/MingR-Ma/SEN-FCB.

Declarations

Conflict of Interests The authors have no conflicts of interest to declare that are relevant to the content of this article.

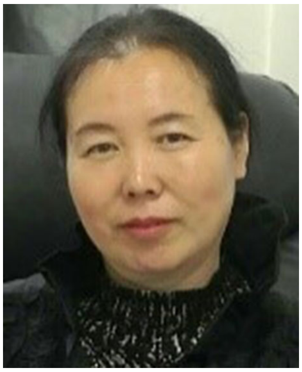
References

1. Veiga C, Janssens G, Teng C-L, Baudier T, Hotoiu L, McClelland JR, Royle G, Lin L, Yin L, Metz J, Solberg TD, Tochner Z, Simone CB, McDonough J, Kevin Teo B-K (2016) First clinical investigation of cone beam computed tomography and deformable registration for adaptive proton therapy for lung cancer. Int J Rad Oncol Biol Phys 95(1):549–559
2. Nakao M, Kobayashi K, Tokuno J, Chen-Yoshikawa T, Date H, Matsuda T (2021) Deformation analysis of surface and bronchial structures in intraoperative pneumothorax using deformable mesh registration. Med Image Anal 102181:73
3. Alvarez P, Rouzé S, Miga MI, Payan Y, Dillenseger J-L, Chabanas M (2021) A hybrid, image-based and biomechanics-based registration approach to markerless intraoperative nodule localization during video-assisted thoracoscopic surgery. Med Image Anal 101983:69
4. Balakrishnan G, Zhao A, Sabuncu MR, Guttag J, Dalca AV (2018) An unsupervised learning model for deformable medical image registration. In: 2018 IEEE/CVF conference on computer vision and pattern recognition
5. Kim B, Dong HK, Park SH, Kim J, Ye JC (2021) Cyclemorph: Cycle consistent unsupervised deformable image registration. Med Image Anal 71(1):102036
6. Dalca AV, Balakrishnan G, Guttag J, Sabuncu MR (2019) Unsupervised learning of probabilistic diffeomorphic registration for images and surfaces. Med Image Anal 57:226–236
7. Zhao S, Dong Y, Chang EI-C, Xu Y (2019) Recursive cascaded networks for unsupervised medical image registration. In: Proceedings of the IEEE/CVF international conference on computer vision (ICCV)
8. Wang J, Zhang M (2020) Deepflash: an efficient network for learning-based medical image registration. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR)
9. Jaderberg M, Simonyan K, Zisserman A, Kavukcuoglu K (2015) Spatial transformer networks Advances in neural information processing systems 28: Annual conference on neural information processing systems 2015, december 7-12, 2015, montreal, quebec, canada, pp 2017–2025
10. Avants BB, Epstein CL, Grossman M, Gee JC (2008) Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. Med Image Anal 12(1):26–41
11. Chen J, He Y, Frey EC, Li Y, Du Y (2021) ViT-V-Net vision transformer for unsupervised volumetric medical image registration
12. Mok T, Chung A (2020) Fast symmetric diffeomorphic image registration with convolutional neural networks. In: 2020 IEEE/CVF Conference on computer vision and pattern recognition (CVPR)
13. Kang M, Hu X, Huang W, Scott MR, Reyes M (2022) Dual-stream pyramid registration network. Med Image Anal 78:102379
14. Jud C, Möri N, Bitterli B, Cattin PC (2016) Bilateral regularization in reproducing kernel hilbert spaces for discontinuity preserving image registration. In: International workshop on machine learning in medical imaging, Springer, pp 10–17
15. Shen D, Davatzikos C (2002) Hammer: hierarchical attribute matching mechanism for elastic registration. IEEE Trans Med Imaging 21(11):1421–1439
16. Ardekani BA, Guckemus S, Bachman A, Hoptman MJ, Wojtaszek M, Nierenberg J (2005) Quantitative comparison of algorithms for inter-subject registration of 3d volumetric brain mri scans. J Neurosci Methods 142(1):67–76
17. Woods RP, Grafton ST, Holmes CJ, Cherry SR, Mazziotta JC (1998) Automated image registration: i. general methods and intrasubject, intramodality validation. J Comput Assist Tomogr 22(1):139
18. Hellier P, Ashburner J, Corouge I, Barillot C, Friston K (2002) Inter-subject registration of functional and anatomical data using spm. Springer, Berlin, pp 590–597
19. Maes F, Collignon A, Vandermeulen D, Marchal G, Suetens P (1997) Multimodality image registration by maximization of mutual information. IEEE Trans Med Imaging 16(2):187–198
20. Davatzikos C (1997) Spatial transformation and registration of brain images using elastically deformable models. Comp Vision Image Underst (CVIU) 66(2):207
21. 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
22. Dalca AV, Bobu A, Rost NS, Golland P (2016) Patch-based discrete registration of clinical brain images. In: International workshop on patch-based techniques in medical imaging
23. Ou Y, Sotiras A, Paragios N, Davatzikos C (2011) Dramms: deformable registration via attribute matching and mutual-saliency weighting. Med Image Anal 15(4):622–639
24. Beg MF, Miller MI, Trouvé A, Younes L (2005) Computing large deformation metric mappings via geodesic flows of diffeomorphisms. Int J Comput Vis 61(2):139–157
25. Vercauteren T, Pennec X, Perchant A, Ayache N (2009) Diffeomorphic demons: efficient non-parametric image registration. Neuroimage 45(1):61–72

26. Zhang M, Liao R, Dalca AV, Turk EA, Golland P (2017) Frequency diffeomorphisms for efficient image registration. In: International conference on information processing in medical imaging
27. Ashburner J (2007) A fast diffeomorphic image registration algorithm. Neuroimage 38(1):95–113
28. Cao X, Yang J, Zhang J, Nie D, Kim M, Wang Q, Shen D (2017) Deformable image registration based on similarity-steered cnn regression. In: International conference on medical image computing and computer-assisted intervention, Springer, pp 300–308
29. Cao X, Yang J, Zhang J, Wang Q, Yap P-T, Shen D (2018) Deformable image registration using a cue-aware deep regression network. IEEE Trans Biomed Eng 65(9):1900–1911
30. Liao R, Miao S, de Tournemire P, Grbic S, Kamen A, Mansi T, Comaniciu D (2017) An artificial agent for robust image registration. In: Proceedings of the thirty-first AAAI conference on artificial intelligence, pp 4168–4175
31. Miao S, Piat S, Fischer PW, Tuysuzoglu A, Mewes PW, Mansi T, Liao R (2018) Dilated FCN for multi-agent 2d/3d medical image registration. In: Proceedings of the thirty-second AAAI conference on artificial intelligence, pp 4694–4701
32. Al Safadi E, Song X (2021) Learning-based image registration with meta-regularization. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 10928–10937
33. Christensen GE, Johnson HJ (2001) Consistent image registration. IEEE Trans Med Imaging 20(7):568–582
34. Ronneberger O, Fischer P, Brox T (2015) U-net: convolutional networks for biomedical image segmentation. In: Navab N, Hornegger J, Wells WM, Frangi AF (eds) Medical image computing and computer-assisted intervention – MICCAI 2015, Springer, pp 234–241
35. Ouyang D, He B, Ghorbani A, Yuan N, Ebinger J, Langlotz CP, Heidenreich PA, Harrington RA, Liang DH, Ashley EA, et al. (2020) Video-based ai for beat-to-beat assessment of cardiac function. Nature 580(7802):252–256
36. Marcus DS, Wang TH, Parker J, Csernansky JG, Morris JC, Buckner RL (2007) Open access series of imaging studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. J Cogn Neurosci 19(9):1498–1507
37. Fischl B (2012) Freesurfer. Neuroimage 62(2):774–781. 20 YEARS OF fMRI
38. Kingma DP, Ba J (2017) Adam: a method for stochastic optimization

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Mingrui Ma is a PhD student in the Key Laboratory for Symbol Computation and Knowledge Engineering of the National Education Ministry of China, College of Computer Science, Jilin University, Changchun, China. His research interest is medical image analysis.



Guixia Liu received her B.S, M.S, and PhD degree in the College of Computer Science and Technology, Jilin University, Changchun, China. She is a research professor in the College of Computer Science and Technology, Jilin University, Changchun, China. Her research interests include machine learning, deep learning, medical image processing and analysis, and biological big data mining and analysis. She has published several papers in journals, BIB, PR and AI.

Lei Song is a master's student in the Key Laboratory for Symbol Computation and Knowledge Engineering of the National Education Ministry of China, College of Computer Science, Jilin University, Changchun, China. His research interest is medical image analysis.

Yuanbo Xu received his B.S, M.S, and PhD degree in the College of Computer Science and Technology, Jilin University, Changchun, China. He is an associate research professor in the College of Computer Science and Technology, Jilin University, Changchun, China. He is also a visiting scholar in Rutgers, the state university of New Jersery, under the supervision of Prof Hui Xiong. His research interests include applications of data mining, recommender system, mobile computing, and medical image analysis. He has published several papers in journals, TKDE, TKDD, TNNLS, TMC and in conferences INFOCOM, CIKM, ICDM, IWQoS, MASS.