applied sciences

MDPI

Article

Reverse-Net: Few-Shot Learning with Reverse Teaching for Deformable Medical Image Registration

Xin Zhang [1], Tiejun Yang [2,3,4,*], Xiang Zhao [1] and Aolin Yang [1]

[1] School of Information Science and Engineering, Henan University of Technology, Zhengzhou 450001, China; learnerzx_haut@126.com (X.Z.); learnerzx@gmail.com (X.Z.); aolinyang_haut@126.com (A.Y.)
[2] School of Artificial Intelligence and Big Data, Henan University of Technology, Zhengzhou 450001, China
[3] Key Laboratory of Grain Information Processing and Control (HAUT), Ministry of Education, Zhengzhou 450001, China
[4] Henan Key Laboratory of Grain Photoelectric Detection and Control (HAUT), Zhengzhou 450001, China
* Correspondence: tjyanghlyu@126.com

Abstract: Multimodal medical image registration has an important role in monitoring tumor growth, radiotherapy, and disease diagnosis. Deep-learning-based methods have made great progress in the past few years. However, its success depends on large training datasets, and the performance of the model decreases due to overfitting and poor generalization when only limited data are available. In this paper, a multimodal medical image registration framework based on few-shot learning is proposed, named reverse-net, which can improve the accuracy and generalization ability of the network by using a few segmentation labels. Firstly, we used the border enhancement network to enhance the ROI (region of interest) boundaries of T1 images to provide high-quality data for the subsequent pixel alignment stage. Secondly, through a coarse registration network, the T1 image and T2 image were roughly aligned. Then, the pixel alignment network generated more smooth deformation fields. Finally, the reverse teaching network used the warped T1 segmentation labels and warped images generated by the deformation field to teach the border enhancement network more structural knowledge. The performance and generalizability of our model have been evaluated on publicly available brain datasets including the MRBrainS13DataNii-Pro, SRI24, CIT168, and OASIS datasets. Compared with VoxelMorph, the reverse-net obtained performance improvements of 4.36% in DSC on the publicly available MRBrainS13DataNii-Pro dataset. On the unseen dataset OASIS, the reverse-net obtained performance improvements of 4.2% in DSC compared with VoxelMorph, which shows that the model can obtain better generalizability. The promising performance on dataset CIT168 indicates that the model is practicable.

Keywords: multimodal registration; few-shot learning; generalizability; reverse teaching

Citation: Zhang, X.; Yang, T.; Zhao, X.; Yang, A. Reverse-Net: Few-Shot Learning with Reverse Teaching for Deformable Medical Image Registration. *Appl. Sci.* **2023**, *13*, 1040. https://doi.org/10.3390/app13021040

Academic Editors: Cecilia Di Ruberto, Andrea Loddo, Lorenzo Putzu, Alessandro Stefano and Albert Comelli

Received: 29 November 2022
Revised: 6 January 2023
Accepted: 8 January 2023
Published: 12 January 2023

1. Introduction

Medical image registration technology is the basis for accurate image fusion. Medical imaging techniques can be broadly divided into two categories: structural imaging and functional imaging. Structural imaging generally refers to X-rays, computed tomography (CT), and so on, which have high spatial resolution and can reflect the anatomical structure of tissues, while functional imaging generally refers to positron emission computed tomography (PET), single-photon emission computed tomography (SPECT), and so on, which have poor spatial resolution but can record the information of tissue metabolism and neural activity. In the process of diagnosis, it is often necessary to fuse the medical image information of different modalities to improve the accuracy of clinical diagnosis. However, due to different imaging devices, different acquisition moments, and the different acquisition perspectives of medical images of different modalities, the spatial location of images of the same object in different modalities is shifted. To obtain effective fusion

information, it is necessary to use image registration techniques to align the images of different modalities.

Deformable image registration can be divided into traditional methods and deep learning-based methods. Although conventional methods such as SyN [1], Demons [2], and optical flow [3] methods have achieved satisfactory performance after continuous parameter tuning, the higher time cost is not suitable for clinical treatment. Compared with traditional methods, deep-learning-based methods can provide fast registration [4] and improve registration accuracy through powerful feature extraction and learning capabilities.

Deep-learning-based methods can be divided into supervised methods and unsupervised methods. Supervised registration is achieved by minimizing the difference between the generated deformation field and the ground-truth using elaborate ground-truth. Salehi et al. [5] used mean square error to train regression neural networks for slice-to-volume registration and volume-to-volume registration scenarios. The model was applied to the registration between the T1 images of an infant brain and the T2 images of a fetus brain. Yan et al. [6] proposed an adversarial image registration model, in which the generator directly estimated the transformation parameter, and the discriminator determined whether the image pair was aligned. By introducing Euclidean distance and adversarial loss to construct the loss function, the model achieved better performance on the MR–TRUS dataset. Recent advances in using organ anatomy labels facilitate medical image registration. He et al. [7] proposed a model that utilizes the complementary nature of the joint model of medical image registration and segmentation to bring improvements in the complex scenes and few-shot situation. The diversity of the enhanced data is maintained by embedding a perturbation factor in the alignment to increase the activity of the deformation. The model finally achieved excellent results. Zhou et al. [8] provided a high-quality segmentation label for subsequent MR–CBCT multimodal registration using APA2Seg-Net. Hering et al. [9] proposed to align the surrounding structure by combining the complementary advantages of segmentation labels and local distance metrics. Compared with label-driven deep learning frameworks, its Dice scores were significantly improved. The above models achieve better performance, but the performance of the registration depends on the quality of the labels. Since segmentation labels are labeled by experienced medical experts, it is necessary to develop an unsupervised registration framework that does not require labels.

Unsupervised registration is achieved by reducing the difference between fixed and warped images. Tang et al. introduced adaptive convolution and adaptive transformer to enhance the ability of global semantic extraction. Compared with other fusion methods, satisfactory fusion results have been achieved [10]. In the field of multimodal registration, many methods are based on similarity metric training. Han et al. [11] proposed an unsupervised dual-channel registration network to resolve the deformation between preoperative MR and intraoperative CT. The model achieved excellent results in MR–CT registration. Since the above registration methods ignore the inverse consistency of the transform between a pair of images and the deformation field is only required to be locally smooth, the overlap of the deformation field cannot be completely avoided. Therefore, Zhang et al. [12] proposed to avoid folding in the transform by combining the traditional smoothness constraint with the anti-folding constraint. Mahapatra et al. [13] used GAN networks for multimodal retinal and cardiac image registration; the model outperformed the reverse deep learning-based B spline method by combining conditional and cyclic constraints [14]. For multimodal registration, it is difficult to correlate two modal images because the intensity of each modal image is different. Sun et al. [15] used image intensities and gradients to align each image pair for real-time registration. The above unsupervised registration methods do not require ground-truth and have attracted a lot of attention from researchers in recent years. However, unsupervised registration is easily misaligned on blurry structures and warped on task-independent backgrounds. Therefore, it is crucial to develop few-shot learning.

Few-shot learning trains a model from limited labeled data and reduces the need for data [16]. In medical image analysis, few-shot learning is urgently needed due to the lack of medical image data, and it has been successful in many scenarios. For example, He et al. [17] proposed few-shot learning deformable medical image registration framework. It achieved satisfactory results on three datasets using only five segmented labeled data. In addition, few-shot learning has been successful in medical image segmentation [18], object detection [19], and histopathology image classification [20].

Inspired by the above methods, we propose a framework for deformable image registration based on few-shot learning. By redesigning the network structure, our model achieves satisfactory registration performance. The main work is as follows:

1. We present a multimodal few-shot learning framework for deformable medical image registration. The border enhancement network is used to enhance the ROI boundaries of T1 images and provide quality data for subsequent pixel alignment network. The coarse registration network roughly aligns the T1 image with the T2 image, which can reduce image discrepancies. The pixel alignment network generates more smooth deformation field. The reverse teaching network teaches border enhancement network rich structural knowledge. By combining the four networks, the model can achieve better registration performance by using few segmentation labels.
2. We introduce reverse teaching which can improve the performance of the border enhancement network with only few segmentation labels. It utilizes the deformation field generated by the pixel alignment network to generate additional training data (warped images and warped segmentation label pairs), thus it reversely teaches the border enhancement network more structural knowledge. Therefore, the border enhancement network with better performance will provide clearer image of the ROI boundary for the pixel alignment network, finally improving the registration performance.
3. The model is trained on the IXI dataset and evaluated on the MRbrain (MRBrainS13DataNii-Pro), SRI24, CIT168, and OASIS datasets. The experimental results show that the model has better performance and generalization ability.

2. Related Work

2.1. Deep-Learning-Based Unimodal Medical Image Registration

Many researchers have introduced deep learning into the field of medical image registration and have proposed many models based on deep learning. Initially, in the deep iterative approach, deep learning is used to learn the similarity metric of two images. As deep learning continues to develop in the field of medical image registration, the unsupervised registration framework with the spatial transformation network (STN) [21] after the convolutional neural network was proposed by de Vos et al. [14] first. Subsequently, a large amount of STN-based unsupervised deformable image registration frameworks emerged. For example, Balakrishnan et al. [22] proposed VoxelMorph, which uses U-net [23] as the network backbone for unsupervised brain MR image registration. In a further development, Yang et al. [24] introduced the transformer to the field of medical image registration, enabling the model to acquire local information while obtaining rich global information. The model eventually achieved good results. Many transformer-based models have subsequently emerged, such as the better-known TransMorph [25]. In the same year, Yang et al. [26] introduced graph convolution to medical image registration again. The model aims to capture high-quality long-range dependencies and expand the receptive field. However, the above models are focused on unimodal registration and still need to be explored in the field of multimodal registration.

2.2. Deep-Learning-Based Multimodal Medical Image Registration

Compared to unimodal registration, multimodal registration can provide physicians with more comprehensive information. Since most similarity metrics are focused on the field of unimodal registration, there is a lack of research in the field of multimodal registra-

tion [27,28]. Therefore, Sloan et al. [29] proposed to incorporate the inverse consistency of the learned spatial transformations into the training network to add additional constraints. The model achieved better performance in T1 and T2 brain image registration compared with the MI-optimization-based image registration methods. Liu et al. [30] proposed a modally independent neighborhood descriptor of multi-dimensional tensor (tMIND) to measure the similarity between two images. The model outperformed MI-based [31] and MIND-based [32] multimodal registration methods. Chen et al. proposed a biogeography-based optimization algorithm with elite learning (BBO-EL). Firstly, the search space of the BBO algorithm is enhanced by proposing a hybrid full migration operator. Secondly, the accuracy and convergence speed of multimodal medical registration are greatly improved by improving the low bound and the up bound of the whole population's quality through some operations [33]. However, although the above methods propose corresponding solutions for multimodal registration, the similarity metric is still not good enough to evaluate the similarity between two images [34]. In order to use the well-developed unimodal registration similarity metric, Cao et al. [35] proposed to use an image synthesis method to synthesize pseudo-CT images and pseudo-MR images with the same anatomical structure. In this way, they converted the multimodal image registration task into a unimodal image registration task. However, although this method uses a unimodal registration similarity metric, the performance of the model is affected by the quality of the synthesized images [36].

3. Materials and Methods

The multimodal medical image registration framework based on few-shot learning is shown in Figure 1. It consists of boundary a enhancement network, a pixel alignment network, a coarse registration network, and a reverse teaching network. The boundary enhancement network is used to enhance the ROI boundaries of T1 images to provide high-quality data for the subsequent pixel alignment network and reduce registration distortion and misalignment caused by blurred ROI boundaries. The coarse registration network roughly aligns the T1 image with the T2 image, which can reduce the image discrepancy, effectively reduce the iteration time for subsequent neural network model training, and improve the efficiency of model training. The pixel alignment network generates more smooth deformation fields. Reverse teaching aligns the warped, segmented, labeled image with the warped image and maps the supervisory information in the warped, segmented, labeled image to the unlabeled image. Through the above method, rich structural knowledge is transmitted for the border enhancement network in reverse, which improves the performance of the border enhancement network. With the four networks mentioned above, we achieve deformable medical image registration with only a few segmentation labels. The detailed process of training is described in the subsequent sections.

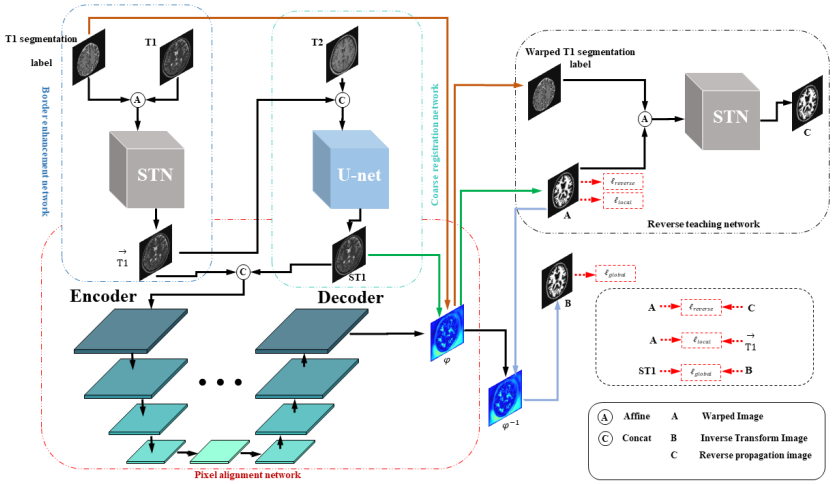

Figure 1. Overall architecture of our reverse-net.

3.1. Border Enhancement Network

The border enhancement network enhances the ROI boundaries of T1 images to provide quality data for the subsequent pixel alignment network, as shown in Figure 2. Let $x \in \mathcal{R}$ represent the sample count in the dataset, and $S \supset \{s^1, s^2, s^3 \ldots\ldots s^x\}$ and $T \supset \{t^1, t^2, t^3 \ldots\ldots t^x\}$ represent moving images and fixed images, respectively.

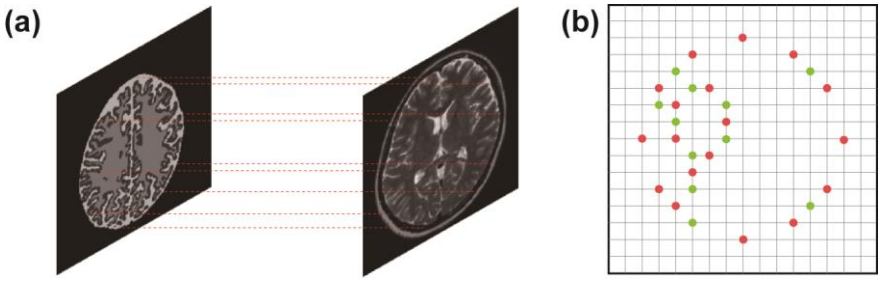

Figure 2. Process figure of the ROI border enhancement network. We have condensed the whole process into two figures. The figure (**a**) represents the correspondence between finding the T1 segmentation label and the T1 image. In this process it is assumed that the output pixels are defined on a regular network so that a mapping relationship between the input and output can be constituted. Finally, we can zero-fill the missing ROI boundaries. The red points in figure (**b**) represent the original contour points and the green points represent the fill points.

We achieve the T1 image ROI border enhancement by the border enhancement network. When the ROI border is blurred and missing, we fill the zero to the corresponding image border by the border enhancement network [37]. The border enhancement network consists of five downsampling blocks and two fully connected layers. The detailed structure is shown in Figure 3. In this process, we need to find the correspondence between T1 segmentation labels (as shown in Figure 4) and T1 images. The whole transformation process is as follows:

$$\begin{pmatrix} x_i^s \\ y_i^s \end{pmatrix} = \begin{bmatrix} \theta_{11} & \theta_{12} & \theta_{13} \\ \theta_{21} & \theta_{22} & \theta_{23} \end{bmatrix} \begin{pmatrix} x_i^t \\ y_i^t \\ 1 \end{pmatrix} \tag{1}$$

where $\theta_{11}, \theta_{12}, \theta_{13}, \theta_{21}, \theta_{22}$, and θ_{23} represents the parameters required for the linear transformation, (x_i^s, y_i^s) represents the pixels sampling from the T1 segmentation label, and (x_i^t, y_i^t) represents the target coordinates of the output image ($\overrightarrow{\text{T1}}$). Since the linear transformed pixels are not necessarily integers, we use a bilinear sampler to compensate for the missing pixels. The formula is as follows:

$$V_{(e,f)}^c = \sum_{x^s}^{H} \sum_{y^s}^{W} U_{(x^s, y^s)}^c max(0, 1 - |x^t - e|) max(0, 1 - |y^t - f|) \tag{2}$$

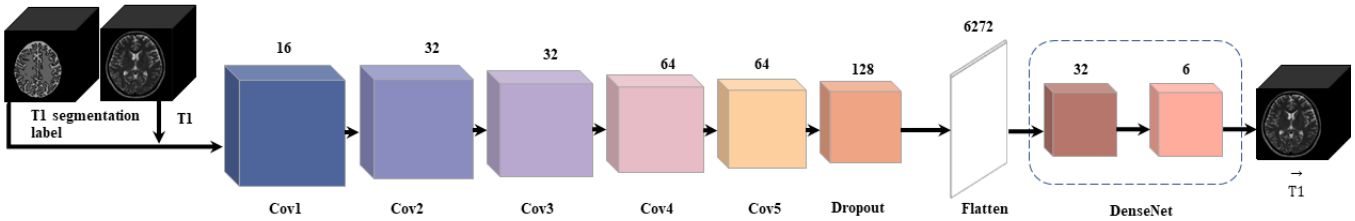

Figure 3. The detailed construction of our STN.

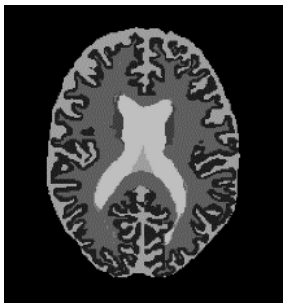

Figure 4. T1 segmentation label. A freehand spline drawing technique was used to segment all of the structures in the brain. The outline of each structure was delineated, starting with the innermost structures. By iteratively subtracting delineations to create holes, binary images were created for each structure. Segmentation was performed in a darkened room with optimal viewing conditions. The entire segmentation was inspected for correctness by an expert not involved in the segmentation procedure and corrections were made if needed. A third expert approved the entirety of final segmentation.

In channel C, $V^c_{(m,n)}$ is the output image at (e, f) and $U^c_{(x^s,y^s)}$ is the input image at (e, f). If (x^t,y^t) is close to (e, f), we use bilinear sampling to add the pixel at (x^s,y^s).

Through the above operation, we randomly select five segmentation labels to participate in the training to obtain the transformed prediction of the T1 segmentation label, which is represented as $S_A \supset \{S^1_A, S^2_A, S^3_A \ldots\ldots S^x_A\}$. Thus, we can obtain a T1 image with a clear ROI border ($\overrightarrow{\text{T1}}$).

3.2. Coarse Registration Network

To reduce the challenge of large differences in multimodal images, we designed a coarse registration network. Specifically, we pre-align the T1 and T2 images using the same U-net as the pixel alignment network. In this way, images with fewer differences are transmitted to the pixel alignment network, which reduces the learning difficulty of the pixel alignment network.

3.3. Pixel Alignment Network

3.3.1. Precision Registration

We obtain a spatial transformation through the pixel alignment network to align the ST1 image with the $\overrightarrow{\text{T1}}$ image. The spatial transformation determines the pixel-to-pixel correspondence [38].

$$\omega = argmin_\varphi \, reduse\left(\overrightarrow{\text{T1}}, ST1 \circ \varphi\right) + \lambda reg(\varphi) \tag{3}$$

where image similarity $reduse\left(\overrightarrow{\text{T1}}, ST1 \circ \varphi\right)$ can be described as a combination of mean square (MSE) error and structural similarity (SSIM). $reg(\varphi)$ is the regularization, whose objective is to ensure the smoothness of the deformation field. λ is the regularization parameter that balances similarity and smoothness.

We concatenated the ST1 and $\overrightarrow{\text{T1}}$ images generated by the border enhancement network and input them into the CNN to obtain the deformation field φ. We put the ST1 and segmented label images into the deformation field to obtain the warped image (image A) and the warped segmented label. Through this method, the warped segmented labeled images are rich in structural information, and they will be the quality data for

subsequent reverse teaching and learning. Our whole process can be expressed by the following formula:

$$w(t)' = w + \varphi(\mathrm{w}) \tag{4}$$

φ is the deformation field and $\varphi(\mathrm{w})$ indicates the offset distance. Each pixel w in the moving image after transformation can be defined as $w(t)'$.

Since the pixels are not necessarily integer after transformation, linear interpolation is used to avoid discontinuity in the transformed image. The formula is as follows:

$$S \circ \varphi(\mathrm{w}) = \sum_{q \in Z(\varphi(\mathrm{w}))} N(q) \prod_{d \in (x,y,z)} (1 - |\varphi_d(w) - q_d|) \tag{5}$$

where Z stands for the areas made up of nearby pixels. The projected results are smoother and more realistic as a result of this differentiable interpolation process.

3.3.2. Details of the Networks

The U-net used in this paper is modified from the famous U-net with an encoder-decoder symmetric architecture. In the process of feature extraction, the main role of the encoder is to extract the image features layer by layer, and the decoder recovers the image information layer by layer. The output feature map of the second convolution in each stage of the encoder is transmitted to the decoder through a skip connection and is channel-spliced with the output feature map of the upsampling layer in the corresponding stage of the decoder after clipping to achieve the fusion of shallow and deep information and provide more semantic information for the decoding process. As shown in Figure 5, each downsampling block is composed of two 3×3 convolutional layers and a 2×2 max pooling layer. In the last layer, we used two 3×3 convolutions with linear activations to obtain the final deformation field. After each convolution operation there is always batch normalization and leaky ReLU (unless otherwise stated). The channels of the U-net are set as 16, 32, 32, 64, 64, 64, 32, 32, and 16.

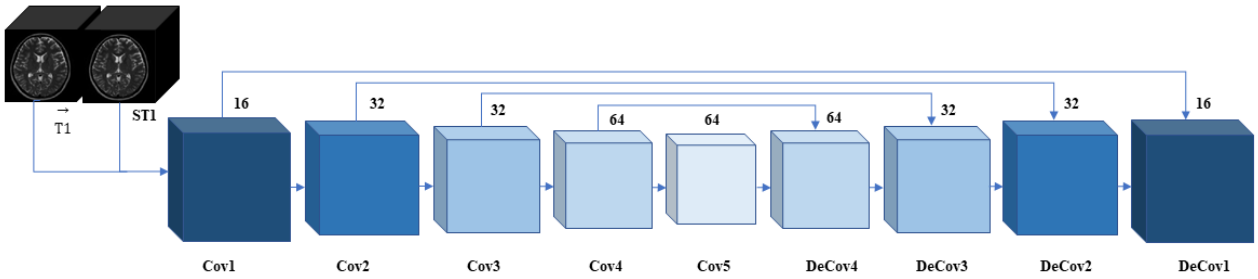

Figure 5. The detailed construction of our U-net.

3.3.3. Inverse Deformation Field

In most previous deep-learning-based deformable registration methods, researchers would divide the input image into a fixed image and a moving image, and only consider a single mapping from the fixed image to the moving image, ignoring the inverse transformation [39]. In our framework, we try to eliminate the division between fixed images and moving images. We divide the input image into S and T and have to implement not only φ_{ST} but also φ_{TS}. Inspired by previous inverse-consistent methods [40], we introduce the new inverse-consistent method, which does not require a complex network structure to achieve accurate registration results.

$$\varphi \xrightarrow{decompose} \varphi_x, \varphi_y \tag{6}$$

$$(\varphi_x, \varphi_y) \circ \varphi = \left(\varphi'_x, \varphi'_y\right) \xrightarrow{combine} \varphi' \tag{7}$$

$$\varphi' \times (-1) = \varphi^{-1} \tag{8}$$

where the above formula is the process of inverse deformation field formation. φ is the deformation field, φ_x and φ_y are the horizontal and vertical offsets, respectively, and $\circ$ is the transformation operator.

3.4. Reverse Teaching Network for Few-Shot Learning

Our reverse teaching warps some segmentation labels through a deformation field and transmits the structural information of the unlabeled images to the warped segmentation labels. Through the above operation, we can obtain additional training data (warped images and warped segmentation label pairs). Thus, these data are entered into the border enhancement network to teach the rich structural knowledge of the border enhancement network in reverse to achieve more accurate ROI boundary enhancement. The detailed process is as follows:

The reverse teaching warps several segmented, labeled images through the deformation field. Since the unlabeled images have rich structural knowledge, the warped, segmented, labeled images will have the structural knowledge of unlabeled images, which generates diverse pairs of warped images and warped segmented label images. Finally, the warped, segmented, labeled images will be input into our border enhancement network to teach our border enhancement network diverse structural knowledge, as shown in Figure 1. The warped image A and the warped segmented image are input to the border enhancement network to produce image C. Finally, the structure loss $\ell_{reverse}$ between image A and image C is calculated so that the border enhancement network will enhance more accurate ROI boundary information.

3.5. Objective Function

The objective function contains two components, including image similarity loss (ℓ_{sim}) and deformation field constraint conditions (ℓ_{reg}).

$$\ell(M, F, \varphi) = \ell_{sim}(\mathcal{H}(M, \varphi), F) + \omega \ell_{reg}(\varphi) \tag{9}$$

where F represents the fixed image, M represents the moving image, φ represents the deformation field, $\mathcal{H}$ represents the deformation process, ℓ_{sim} represents the similarity between moving and fixed images, ℓ_{reg} constrains the deformation field to make it smoother, and ω denotes the strength of the constrained deformation field.

where ℓ_{sim} consists of three losses, $\ell_{reverse}$, ℓ_{local}, and ℓ_{global}.

$$\begin{aligned} \ell_{sim}(\mathcal{H}(M, \varphi), F) = \alpha \ell_{local}\left(\overrightarrow{\mathrm{T1}}, ST1 * \varphi\right) + \beta \ell_{global}\left(ST1, A * \varphi^{-1}\right) \\ + \gamma \ell_{reverse}(A, C) \end{aligned} \tag{10}$$

ℓ_{local} consists of MSE and SSIM, which reflects the similarity between image $\overrightarrow{\mathrm{T1}}$ and image A.

$$\begin{aligned} \ell_{local}\left(\overrightarrow{\mathrm{T1}}, ST1 * \varphi\right) = \\ \xi_1 * \frac{1}{\Omega} \sum_{i \epsilon \Omega} ||ST1 * \varphi(i) - \overrightarrow{\mathrm{T1}}(i)||^2 + \psi_1 * \left(1 - \frac{\left(2\mu_{I_d}\mu_{I_f} + C_1\right)\left(2\sigma_{I_f I_d} + C_2\right)}{\left(\mu_{I_f}^2 + \mu_{I_d}^2 + C_1\right)\left(\sigma_{I_f}^2 + \sigma_{I_d}^2 + C_2\right)}\right) \end{aligned} \tag{11}$$

where ξ_1 and ψ_1 are used as the weight factors to prevent large deviations and Ω is the spatial domain. Where μ_{I_d} *and* μ_{I_f} represent the local means of image $\overrightarrow{\mathrm{T1}}$ and image A. σ_{I_f} and σ_{I_d} represent the standard deviations of images $\overrightarrow{\mathrm{T1}}$ and image A. C_1 and C_2 are minor constants required to prevent instability.

ℓ_{global} consists of PCC and SSIM, which reflects the similarity between image $ST1$ and image B.

$$\ell_{global}\left(ST1, A*\varphi^{-1}\right) = \xi_2 * \left(1 - \frac{\left(2\mu_{I_d}\mu_{I_f}+C_1\right)\left(2\sigma_{I_f I_d}+C_2\right)}{\left(\mu_{I_f}^2+\mu_{I_d}^2+C_1\right)\left(\sigma_{I_f}^2+\sigma_{I_d}^2+C_2\right)}\right) + \psi_2 * \left(1 - \frac{\sum_{i\in\Omega}\left(ST1(i)-\overline{ST1}\right)\left(A*\varphi^{-1}(i)-\overline{A*\varphi^{-1}}\right)}{\sqrt{\sum_{i\in\Omega}\left(ST1(i)-\overline{ST1}\right)}\sqrt{\sum_{i\in\Omega}A*\varphi^{-1}(i)-\overline{A*\varphi^{-1}}}}\right) \tag{12}$$

where ξ_2 and ψ_2 are used as the weight factors to prevent large deviations. Where μ_{I_d} *and* μ_{I_f} represent the local means of image ST1 and image B. σ_{I_f} and σ_{I_d} represent the standard deviations of image ST1 and image B. $\overline{A*\varphi^{-1}}$ and $\overline{ST1}$ stand for the average intensities.

$\ell_{reverse}$ consists of MI and MSE, which reflects the similarity between image A and image C.

$$\ell_{reverse}(A,C) = \xi_3 * \frac{1}{\Omega}||C-A||^2 + \psi_3 * \left(\iint p(A,C)\right) log\frac{p(A,C)}{p(A)p(C)}d_x d_y \tag{13}$$

where the adjustment factors ξ_3 and ψ_3 are used to balance the two losses. If A and C are independent, then $p(A,C)$ is equal to $p(A)p(C)$ and MI(A, C) will be 0, which means A and C are uncorrelated.

$$\ell_{reg} = \sum\nolimits_{x\in\Omega}\left|\left|\nabla h(x)\right|\right|^2 \tag{14}$$

The total losses are as follows:

$$\ell = \alpha\ell_{local}\left(\overrightarrow{\mathrm{T1}}, ST1*\varphi\right) + \beta\ell_{global}\left(ST1, A*\varphi^{-1}\right) + \gamma\ell_{reverse}(A,C) + \omega\ell_{reg}(\varphi) \tag{15}$$

4. Experiments

4.1. Dataset

We used the IXI dataset for training, which was acquired from a Philips 1.5T system (Philips, Amsterdam, The Netherlands), Philips 3T system(Philips, Amsterdam, The Netherlands), and GE 1.5T system (GE, Boston, MA, USA). We randomly selected 450 T1 and T2 images from the IXI dataset to participate in the training. Since the IXI dataset is a 3D dataset, each data format was 256 × 256 × 130, we crop them to 224 × 224 × 130, and 130 slices were acquired. We selected slices from 64 to 67 to participate in the training. The performance of the model was validated on the Mrbrain [41] dataset. MRbrain is composed of 20 images of normal subjects acquired on a 3T MRI scanner. We randomly selected eight 3D images and cropped each image from 240 × 240 × 48 to 224 × 224 × 48, so that 48 slices were acquired. In this experiment, we selected 22 to 27 slices for registration. Each subject is labeled as eight ROIs.

The generality of this experiment is verified on three datasets, OASIS [42], CIT168 [43], and SRI24 [44]. The OASIS dataset was acquired by 1.5T Siemens equipment, which consists of 271 T1 images of the normal human brain. We resampled from 192 × 160 to 224 × 224 and then selected 100 images to participate in the registration. Each subject was labeled as 36 ROIs. For the SRI24 and CIT168 datasets, we cropped them to 224 × 224, then selected two 3D images and sliced each 3D image into 10 slices to participate in the registration.

4.2. Implementation Details

Our proposed model was implemented on the Tensorflow (version 1.14.0) framework and trained on Intel i7-10700 CPU and GeForce RTX 2080ti GPU with 11 GB memory. We used a learning rate of 1×10^{-4} and the Adam optimizer to update the neural network weights in each network. Due to computer memory constraints, the batch size was set to

eight. When the best performance is reached, the experimental training will automatically end. The whole training process takes about 10 h.

4.3. Evaluation and Validation

We use three metrics to evaluate the registration performance of the model, the percentage of non-positive values in the determinant of the Jacobian matrix over the deformation field, the DSC, and precision.

$$\mathrm{Dice}(F, M) = \frac{2TP}{2TP + FP + FN} \times 100\% \tag{16}$$

$$\mathrm{Precision}(F, M) = \frac{TP}{TP + FP} \times 100\% \tag{17}$$

$$\mathrm{Recall}(F, M) = \frac{TP}{TP + FN} \tag{18}$$

Ground Truth	Prediction	
	True	False
True	TP (true positive)	FN (false negative)
False	FP (false positive)	

TP, FP, and FN represent the number of pixels in the different cases, respectively. We compute these scores for each case separately and report the average results.

$$\mathrm{Det.Jac}(i, j, k) = \begin{vmatrix} \frac{\partial i}{\partial x} & \frac{\partial j}{\partial x} & \frac{\partial k}{\partial x} \\ \frac{\partial i}{\partial y} & \frac{\partial j}{\partial y} & \frac{\partial k}{\partial y} \end{vmatrix} \tag{19}$$

$$\mathrm{Npr.Jac} = \frac{\sum \sigma(\mathrm{Det}(J(i, j, k)) \leq 0)}{n} \tag{20}$$

$\mathrm{Det}\,(J) = 1$ indicates that no volume behavior has occurred, where Det.Jac represents Jacobian determinant. Det (J) > 1 indicates that the point is expanding. 0 < Det (J) < 1 indicates that the point is contracting. Det (J) ≤ 0 indicates a folding phenomenon. σ is a linear activation function. n represents the total number of elements.

5. Results

In the registration task, we compared our model with the traditional method SyN [1] and three deep-learning-based methods, VoxelMorph [22], VM-diff (VoxelMorph-diff) [45], and Cof-net [46]. Pre-net is the model trained by removing a reverse teaching network on the base of reverse-net and using 40 segmentation labels. Un-reverse-net is the model without labels involved in training (it participates in training with T2 images instead of segmented label images). We validated the performance and generalization ability of the model with four brain MR datasets.

5.1. Border Enhancement Experiment

Our reverse teaching network utilizes the deformation field to generate additional training data (warped images and warped segmentation label pairs), which can reversely teach rich structural knowledge to the border enhancement network. The above operation makes the border enhancement network enhance more accurate ROI boundaries and provide quality data for the subsequent pixel alignment network. As shown in Figure 6, the ROI boundary enhancement results show a significant visual improvement, as reflected by finer and clearer borders with fewer blurred structures. The enlarged area of the MRI in the brain showed complete border enhancement areas with clear borders. Our ROI border enhancement network enhanced the ROI boundary information of the image and effectively suppressed the misalignment due to boundary blurring.

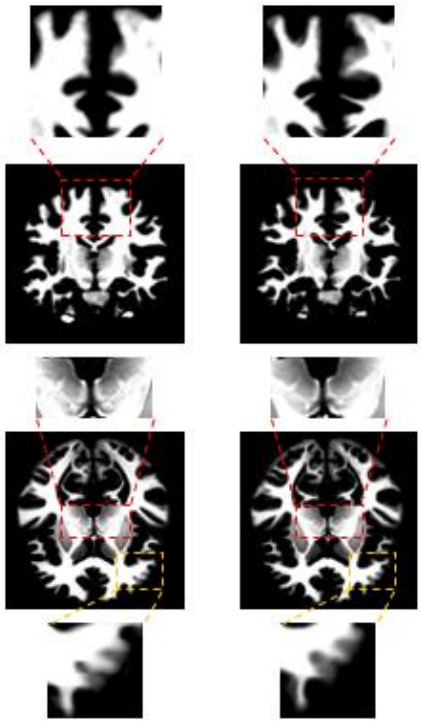

Figure 6. ROI boundary enhancement results on CIT168 and OASIS.

5.2. Ablation Experiment

We devised ablation experiments to validate the contribution of reverse teaching networks. We conducted experiments using brain MR data to evaluate the registration performance of the model using the average DSC, the percentage of non-positive values in the determinant of the Jacobi matrix of the deformation field, and the average alignment accuracy per image.

The results of the ablation experiments are shown in Table 1. The structural information of different unlabeled images is passed to the warped, labeled images through the deformation field formed by the pixel alignment network, and image A and the warped, labeled images with rich structural knowledge will provide this rich structural knowledge to the reverse teaching network. The above operation can provide quality data for the subsequent pixel alignment network. Compared with Pre-net, the structural information brings 0.76% DSC and 4.47% precision improvement to our model. Compared with Cof-net, the structural information brings 3.5% DSC and 3.82% precision improvement to our model. In addition, as shown in Figure 7. Compared with the other two models, our model is richer in detail. The above results show that the reverse teaching network improves the registration performance.

Table 1. Quantitative results of ablation experiments. The numbers following the ± indicate the standard deviation.

Method	DSC (%)	Precision (%)	Det.Jac (%)
Cof-net [46]	81.70 ± 3.71	90.59 ± 3.05	0.006 ± 0.002
Pre-net	84.44 ± 0.67	89.94 ± 4.27	**0.002 ± 0.001**
Un-reverse-net	83.29 ± 0.15	89.45 ± 1.23	0.004 ± 0.012
Reverse-net	**85.20 ± 1.18**	**94.41 ± 0.89**	0.004 ± 0.001

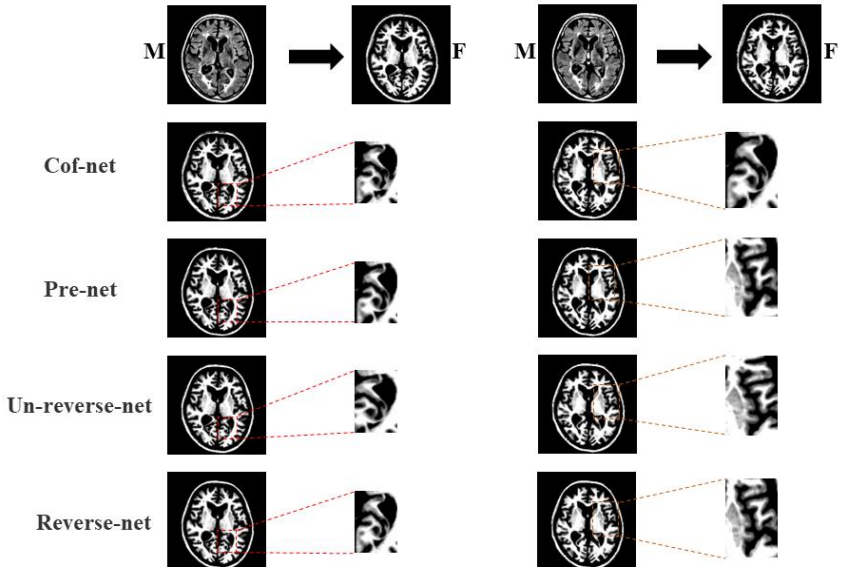

Figure 7. Results of ablation experiments on MRbrain.

In addition, we investigated the effect of different numbers of segmented labels on the registration performance. As shown in Figure 8. When having one segmented label, the model decreases 2.33% on DSC compared with Pre-net. When having three labels, the model improves 0.41% on DSC compared with Pre-net. This shows that our reverse teaching network can improve the registration performance of the model. As the number of labels increases, our model is further enhanced. When the number of labels is increased to five, the reverse-net improves 0.76% on DSC compared with Pre-net. This further shows that our network has improved the ability of the border enhancement network sensing boundary through reverse teaching.

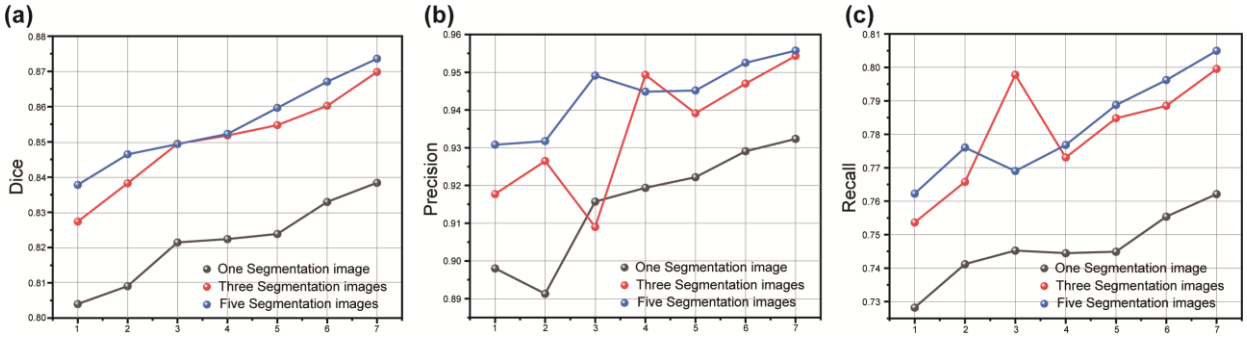

Figure 8. Registration ability of different numbers of segmented labels on the MRbrain dataset ((**a**–**c**) are the line graphs formed by the model under the three evaluation metrics of Dice, Preciaion, Recall respectively).

Finally, we evaluated the effect of the model without segmented label participation. The results of quantitative analysis are shown in Table 1. Compared to Cof-net, our model improves 1.59% on DSC and decreases 0.002 on Det.Jac. Compared to VoxelMorph, our model improves 2.45% on DSC and decreases 0.002 on Det.Jac. The above results show the superiority of our model design. When labels are involved, the model results are even better.

5.3. Comparative Experiments

We compared our method with four state-of-the-art registration methods, including symmetric normalization (SyN) and three deep-learning-based methods, VoxelMorph, VM-diff, and Cof-net. On the same datasets, we trained deep learning methods with rec-

ommended hyper-parameters from scratch. We quantitatively evaluated the performance from four perspectives: Precision, DSC, Runtime, and Det.Jac.

Table 2 summarizes the experimental results on the MRbrain dataset. It can be seen that our method obtains better mean DSC scores and lower running times while achieving the lowest number of folds. In the MRbrain dataset, compared with VoxelMorph, the DSC score and Precision score of our model increased by 4.36% and 3.31%, respectively. The Det.Jac indexes of our model are decreased from 0.006% to 0.004%. Compared with Cof-net, our model improves 3.5% in the DSC metric and 3.82% in Precision, and Det.Jac decreases from 0.006% to 0.004%. At the same time, our model has a lower average processing time per slice compared to Voxelmorph, VM-diff, and Cof-net. In addition, it can also be seen from the boxplot in Figure 9 that our model achieves the best results among the three evaluation metrics.

Table 2. Quantitative results on the MRbrain dataset. Bold indicates the highest score.

Method	DSC (%)	Precision (%)	Runtime (s)/Slice	Det.Jac (%)
SyN [1]	79.48 ± 0.78	84.07 ± 0.86	2.034	0.071 ± 0.007
Voxelmorph [22]	80.84 ± 0.78	91.10 ± 1.18	0.558 ± 0.017	0.006 ± 0.002
VM-diff [45]	79.59 ± 2.31	86.68 ± 2.52	0.423 ± 0.011	0.013 ± 0.005
Cof-net [46]	81.70 ± 3.71	90.59 ± 3.05	0.367 ± 0.013	0.006 ± 0.002
Reverse-net	**85.20 ± 1.18**	**94.41 ± 0.89**	**0.359 ± 0.011**	**0.004 ± 0.001**

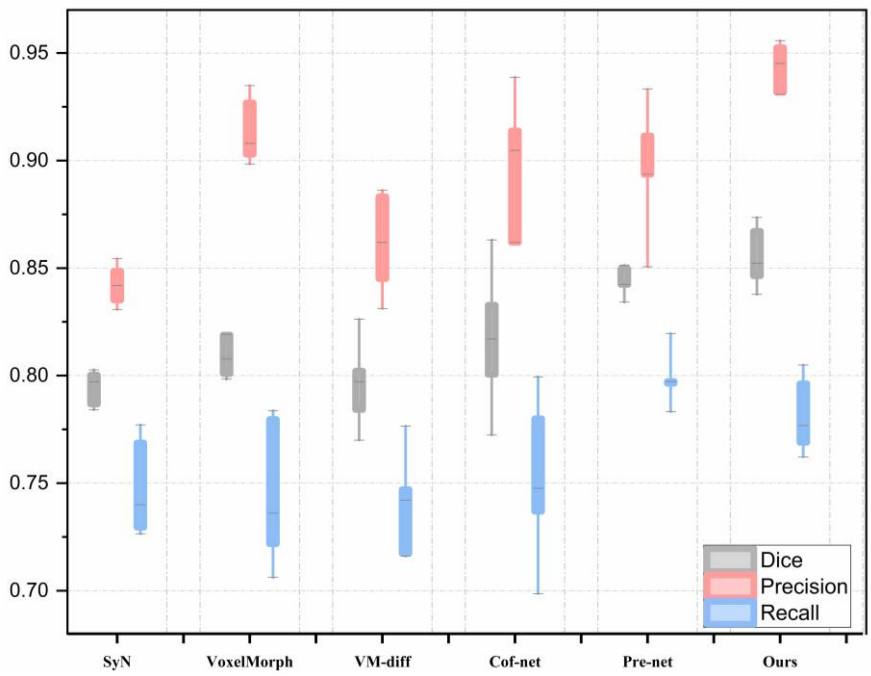

Figure 9. Boxplots of the results of different methods on the MRbrain dataset.

Figure 10 shows the visualization of the images from the perspective of axial slices. The columns marked with W indicate the warped images. φ indicates the deformation field visualization figure. The red color in the visualized image indicates the horizontal transformation and the green color indicates the vertical transformation. The higher the red or green signal, the larger the transformation. As shown in Figure 10, typical registration results from the MRbrain dataset show that the warped image generated by our method is most like the fixed image; our method can achieve the better registration among several methods.

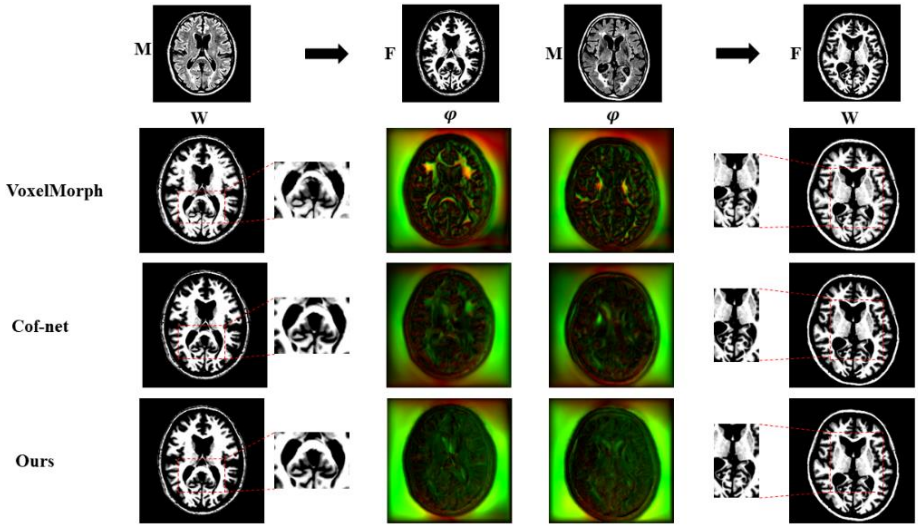

Figure 10. Experimental results on the MRbrain dataset.

5.4. Evaluation and Analysis on CIT168, SRI24, and OASIS

In order to evaluate the generalizability of the network, many experiments were performed on images from different scanners. It is worth noting that the proposed method uses the same model as the comparative experimental model. The experiments were performed on the CIT168, SRI24, and OASIS datasets. As shown in Table 3, compared to VoxelMorph, the proposed method improved 7.3%, 4.7%, and 4.2% on the CIT168, SRI24 and OASIS datasets, respectively. As shown in Figure 11, our model shows more details on the CIT168, SRI24, and OASIS datasets. The above results show that our method achieves the best registration among several methods.

Table 3. Quantitative results on OASIS, CIT168, and SRI24 datasets. Bold indicates the highest DSC score.

Dataset \ Method	SyN [1]	VoxelMorph [22]	VM-Diff [45]	Reverse-Net
OASIS [42]	67.7 ± 2.9	74.9 ± 13.7	75.2 ± 13.9	**79.1 ± 1.3**
SRI24 [44]	70.3 ± 5.6	77.9 ± 2.4	78.6 ± 2.9	**82.6 ± 0.6**
CIT168 [43]	67.8 ± 2.6	73.4 ± 3.9	73.2 ± 3.1	**80.5 ± 1.9**

Figure 11. Experimental results on CIT168, SRI24, and OASIS.

6. Discussion

For supervised registration, the number of labels limits the registration performance. Unsupervised registration is easily disturbed by outliers and task-irrelevant areas due to the constraint of missing labels, which can lead to eventual misalignment and distortion. To solve these problems, we propose a few-shot deformable image registration method based on reverse learning, which can improve registration performance effectively. For few-shot learning, the reverse teaching network utilizes the deformation field generated by the pixel alignment network to generate additional training data from different unlabeled images and teaches the border enhancement network rich structural knowledge for more accurate enhancement of T1 image ROI boundaries. Therefore, the above operation can provide high-quality data for the subsequent pixel alignment network and improve the registration performance of the model. The registration results in Figure 6 also show the ability of the border enhancement network to enhance the ROI boundaries,;our experiments on brain MRI images show that our network has better registration performance and effective texture preservation. Compared with Cof-net, we achieved 3.5% improvement in DSC while reducing the processing time per slice, which indicates that our framework has better potential for future clinical applications.

Reverse-net has improved deformable medical image registration with only a few segmentation labels, but the number of segmented labels becomes a challenge. In order to fully utilize the segmentation labels, we introduced the reverse teaching network and redesigned the loss function. Experiments in which different numbers of segmentation labels are used are conducted on the MRbrain dataset. It was shown that competitive registration accuracy could be obtained by using only five segmentation labels.

Although methods based on few-shot learning are efficient, the smoothness of the deformation field is insufficient. In order to enhance the smoothness, various joint loss functions were designed and introduced, and suitable hyperparameters were selected after extensive experiments. As can be seen in Figure 10, the smoothness of the deformation field is better with our method.

By redesigning the network structure and loss function, our model shows better performance on brain datasets. However, MR–CT multimodal registration based on few-shot learning is still a challenging task in medical image deformation registration, so future research applying our reverse-net for MR–CT multimodal registration tasks is significant.

7. Conclusions

In this paper, we propose a deformable medical image registration model based on few-shot learning, which can provide doctors with a reliable diagnosis and reduce misdiagnosis rates. Different from other methods, we introduced the reverse teaching network, which can reduce the number of segmentation labels while transmitting rich structural knowledge for the border enhancement network. The above operations can make the ROI boundaries of T1 images clearer which provides better quality data for the subsequent pixel alignment network and improve the registration performance of the model. The smoothness of the deformation field was improved by introducing a new loss function and redesigning the individual weights of the loss function. The method obtained satisfactory registration results on the brain dataset. In addition, the experimental results on the unseen data also show that our model has better generalizability. MR–CT registration remains a challenging task in the field of medical image registration due to the lack of MR–CT data. Therefore, in the future, we will apply the reverse-net to MR–CT multimodal registration task based on few-shot learning.

Author Contributions: X.Z. (Xin Zhang) devised the project, performed the experiments, and drafted the manuscript. T.Y. provided critical revision of the manuscript for important intellectual content and technical and material support. X.Z. (Xiang Zhao) and A.Y. contributed to the design of this study and the revision of the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the key specialized research and development program of Henan Province (Grant No. 202102210170), the Open Fund Project of Key Laboratory of Grain Information Processing & Control (Grant No. KFJJ2021101), and the Innovative Funds Plan of Henan University of Technology (Grant No. 2021ZKCJ14).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are publicly available. It can be used as long as the corresponding article is cited.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Avants, B.B.; Epstein, C.L.; Grossman, M.; Gee, J.C. Symmetric diffeomorphic image registration with CrossCorrelation: Evaluating automated labeling of elderly and neurodegenerative brain. *Med. Image Anal.* **2008**, *1*, 26–41. [CrossRef] [PubMed]
2. Thirion, J.-P. Image matching as a diffusion process: An analogy with Maxwell's demons. *Med. Image Anal.* **1998**, *2*, 243–260. [CrossRef] [PubMed]
3. Yang, D.; Li, H.; Low, D.A.; Deasy, J.O. A fast inverse consistent deformable image registration method based on symmetric optical flow computation. *Phys. Med. Biol.* **2008**, *53*, 6143–6165. [CrossRef] [PubMed]
4. Zhou, S.K.; Le, H.N.; Luu, K. Deep reinforcement learning in medical imaging: A literature review. *Med. Image Anal.* **2021**, *73*, 102193. [CrossRef] [PubMed]
5. Salehi, S.S.M.; Khan, S.; Erdogmus, D.; Gholipour, A. Real-time deep pose estimation with geodesic loss for Image-to-Template rigid registration. *IEEE Trans. Med. Imaging* **2019**, *38*, 470–481. [CrossRef]
6. Yan, P.; Xu, S.; Rastinehad, A.R.; Wood, B.J. Adversarial image registration with application for MR and TRUS image fusion. In *International Workshop on Machine Learning in Medical Imaging*; Springer: Cham, Switzerland, 2018; pp. 197–204.
7. He, Y.; Li, T.; Yang, G.; Kong, Y. L Deep complementary joint model for complex scene registration and few-shot segmentation on medical images. In Proceedings of the European Conference on Computer Vision, Glasgow, UK, 23–28 August 2020; pp. 770–786.
8. Zhou, B.; Augenfeld, Z.; Chapiro, J.; Zhou, S.K.; Liu, C.; Duncan, J.S. Anatomy-guided multimodal registration by learning segmentation without ground truth: Application to intraprocedural CBCT/MR liver segmentation and registration. *Med. Image Anal.* **2021**, *71*, 102041–102051. [CrossRef]
9. Hering, A.; Kuckertz, S.; Heldmann, S.; Heinrich, M.P. Enhancing Label-Driven Deep Deformable Image Registration with Local Distance Metrics for State-of-the-Art Cardiac Motion Tracking. In *Bildverarbeitung für die Medizin*; Springer: Wiesbaden, Germany, 2019; pp. 309–314.
10. Tang, W.; He, F.; Liu, Y.; Duan, Y. MATR: Multimodal Medical Image Fusion via Multiscale Adaptive Transformer. *IEEE Trans Image Process.* **2022**, *31*, 5134–5149. [CrossRef]
11. Han, R.; Jones, C.K.; Lee, J. Deformable MR-CT image registration using an unsupervised dual-channel network for neurosurgical guidance. *Med. Image Anal.* **2022**, *75*, 102292. [CrossRef]
12. Zhang, J. Inverse-Consistent deep networks for unsupervised deformable image registration. *arXiv* **2018**, arXiv:1809.03443.
13. Mahapatra, D.; Antony, B.; Sedai, S.; Garnavi, R. Deformable medical image registration using generative adversarial networks. In Proceedings of the IEEE International Symposium on Biomedical Imaging, Washington, DC, USA, 4–7 April 2018.
14. De Vos, B.D.; Berendsen, F.F.; Viergever, M.A.; Staring, M.; Isgum, I. End-to-End Unsupervised Deformable Image Registration with a Convolutional Neural Network. In Proceedings of the Computer Vision and Pattern Recognition, Québec City, QC, Canada, 14 September 2017; pp. 204–212.
15. Sun, L.; Zhang, S. Deformable MRI-Ultrasound Registration Using 3D Convolutional Neural Network. In Proceedings of the Simulation, Image Processing, and Ultrasound Systems for Assisted Diagnosis and Navigation, Granada, Spain, 16–20 September 2018; pp. 152–158.
16. Wang, Y.; Yao, Q.; Kwok, J.T.; Ni, L.M. Generalizing from a Few Examples: A Survey on Few-Shot Learning. *ACM Comput. Surv.* **2021**, *1*, 1–34. [CrossRef]
17. Medela, A.; Picon, A.; Saratxaga, C.L.; Belar, O.; Cabezón, V.; Cicchi, R.; Bilbao, R.; Glover, B. Few Shot Learning in histopathological images:Reducing the need of labeled data on biological datasets. In Proceedings of the International Symposium on Biomedical Imaging, Venice, Italy, 8–11 April 2019.
18. Zhao, A.; Balakrishnan, G.; Durand, F.; Guttag, J.V.; Dalca, A.V. Data augmentation using learned transformations for One-Shot medical image segmentation. In Proceedings of the Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–20 June 2019.
19. Puch, S.; Sanchez, I.; Rowe, M. Few-shot Learning with deep triplet networks for brain imaging modality recognition. In Proceedings of the Medical Image Computing and Computer Assisted Intervention Society, Shenzhen, China, 13–17 October 2019.
20. He, Y.; Li, T.; Ge, R.; Yang, J.; Kong, Y.; Zhu, J.; Shu, H.; Yang, G.; Li, S. Few-Shot Learning for Deformable Medical Image Registration With Perception-Correspondence Decoupling and Reverse Teaching. *IEEE J. Biomed. Health Inform.* **2022**, *26*, 1177–1187. [CrossRef] [PubMed]

21. Jaderberg, M.; Simonyan, K.; Zisserman, A.; Kavukcuoglu, K. Spatial transformer networks. In Proceedings of the Conference on Computer Vision and Pattern Recognition, Montreal, QC, Canada, 27–30 June 2016.
22. Balakrishnan, G.; Zhao, A.; Sabuncu, M.R.; Guttag, J.; Dalca, A.V. VoxelMorph: A learning framework for deformable medical image registration. *IEEE T. Med. Imaging* **2019**, *38*, 1788–1800. [CrossRef]
23. Ronneberger, O.F.P.; Brox, T. U-Net: Convolutional networks for biomedical image segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015.
24. Yang, T.; Bai, X.; Cui, X.; Gong, Y.; Li, L. TransDIR: Deformable imaging registration network based on transformer to improve the feature extraction ability. *Med. Phys.* **2022**, *49*, 952–965. [CrossRef]
25. Chen, J.; Du, Y.; He, Y.; Segars, W.P.; Li, Y.; Frey, E.C. TransMorph: Transformer for unsupervised medical image registration. *Med. Image Anal.* **2022**, *82*, 102615–102649. [CrossRef]
26. Yang, T.; Bai, X.; Cui, X.; Gong, Y.; Li, L. GraformerDIR: Graph convolution transformer for deformable image registration. *Comput. Biol. Med.* **2022**, *147*, 105799–105808. [CrossRef] [PubMed]
27. Li, Y.; Sixou, B.; Peyrin, F. A Review of the Deep Learning Methods for Medical Images Super Resolution Problems. *IRBM* **2021**, *42*, 120–133. [CrossRef]
28. Belderrar, A.; Hazzab, A. Real-time estimation of hospital discharge using fuzzy radial basis function network and electronic health record data. *Int. J. Med. Inform.* **2021**, *13*, 75–84. [CrossRef]
29. Siebert, J.P.; Goatman, K.A.; Sloan, J.M. Learning Rigid Image Registration—Utilizing Convolutional Neural Networks for Medical Image Registration. In Proceedings of the International Conference on on Bioimaging, Funchal, Portugal, 19–21 January 2018.
30. Liu, Q.; Leung, H. Tensor-based descriptor for image registration via unsupervised network. In Proceedings of the International Conference on Information Fusion, Xi'an, China, 10–13 July 2017.
31. Myronenko, A.; Song, X. Intensity-based image registration by minimizing residual complexity. *IEEE Trans. Med. Imaging* **2010**, *29*, 1882–1891. [CrossRef]
32. Heinrich, M.P.; Jenkinson, M.; Bhushan, M.; Matin, T.; Gleeson, F.V.; Brady, S.M.; Schnabel, J.A. MIND: Modality independent neighbourhood descriptor for multi-modal deformable registration. *Med. Image Anal.* **2012**, *16*, 1423–1435. [CrossRef]
33. Chen, Y.; He, F.; Li, H.; Zhang, D.; Wu, Y. A full migration BBO algorithm with enhanced population quality bounds for multimodal biomedical image registration. *Appl. Soft Comput.* **2020**, *93*, 106335–106343. [CrossRef]
34. Rubeaux, M.; Nunes, J.C.; Albera, L.; Garreau, M. Medical image registration using Edgeworth-based approximation of Mutual Information. *IRBM* **2014**, *35*, 139–148. [CrossRef]
35. Cao, X.; Yang, J.; Gao, Y.; Guo, Y.; Wu, G.; Shen, D. Dual-core steered non-rigid registration for multi-modal images via bi-directional image synthesis. *Med. Image Anal.* **2017**, *41*, 18–31. [CrossRef] [PubMed]
36. Song, Y.; He, F.; Duan, Y.; Liang, Y.; Yan, X. A Kernel Correlation-Based Approach to Adaptively Acquire Local Features for Learning 3D Point Clouds. *Comput.-Aided Design* **2022**, *146*, 103196–103207. [CrossRef]
37. Zheng, Z.; Zheng, L.; Yang, Y. Pedestrian alignment network for large-scale person re-Identification. *IEEE Trans. Circuits Syst. Video Technol.* **2019**, *29*, 3037–3045. [CrossRef]
38. Maintz, J.; Viergever, M. An overview of medical image registration methods. *Med. Image Anal.* **1996**, *12*, 1–22.
39. Balakrishnan, G.; Zhao, A.; Sabuncu, M.R.; Dalca, A.V.; Guttag, J. An unsupervised learning model for deformable medical image registration. In Proceedings of the Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018.
40. Mok, T.C.W.; Chung, A.C.S. Fast symmetric diffeomorphic image registration with convolutional neural networks. In Proceedings of the Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 16–20 June 2020.
41. Mendrik, A.M.; Vincken, K.L.; Kuijf, H.J.; Breeuwer, M.; Bouvy, W.H.; de Bresser, J.; Alansary, A.; de Bruijne, M.; Carass, A.; El-Baz, A.; et al. MRBrainS Challenge: Online Evaluation Framework for Brain Image Segmentation in 3T MRI Scans. *Comput. Intell. Neurosci.* **2015**, *2015*, 813696–813712. [CrossRef]
42. Marcus, D.S.; Wang, T.H.; Parker, J.; Csernansky, J.G.; Morris, J.C.; Buckner, R.L. Open Access Series of Imaging Studies (OASIS): Cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. *J. Cogn. Neurosci.* **2007**, *19*, 1498–1507. [CrossRef]
43. Pauli, W.M.; Nili, A.N.; Tyszka, J.M. A high-resolution probabilistic in vivo atlas of human subcortical brain nuclei. *Sci. Data* **2018**, *5*, 180063. [CrossRef]
44. Rohlfing, T.; Zahr, N.M.; Sullivan, E.V.; Pfefferbaum, A. The SRI24 multichannel atlas of normal adult human brain structure. *Hum. Brain Mapp.* **2010**, *31*, 798–819. [CrossRef]
45. Dalca, A.V.; Balakrishnan, G.; Guttag, J.; Sabuncu, M.R. Unsupervised learning of probabilistic diffeomorphic registration for images and surfaces. *Med. Image Anal.* **2019**, *57*, 226–236. [CrossRef]
46. Huang, W.; Yang, H.; Liu, X.; Li, C.; Zhang, I.; Wang, R.; Zheng, H. A coarse-to-fine deformable transformation framework for unsupervised multi-contrast MR image registration with dual consistency constraint. *IEEE Trans. Med. Imaging* **2021**, *40*, 2589–2599. [CrossRef] [PubMed]