Neurocomputing 501 (2022) 102–112

Contents lists available at ScienceDirect

Neurocomputing

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

Multi-strategy mutual learning network for deformable medical image registration

Zhiyuan Zheng [a,b], Wenming Cao [a,b,*], Ye Duan [c], Guitao Cao [d], Deliang Lian [a,b]

[a] Shenzhen Key Laboratory of Media Security, Shenzhen University, 518060, China
[b] Guangdong Multimedia Information Service Engineering Technology Research Center, Shenzhen 518060, China
[c] Computer Graphics and Image Understanding Laboratory, University of Missouri, Columbia, MO 65211, USA
[d] Shanghai Key Laboratory of Trustworthy Computing, East China Normal University, 200062, China

ARTICLE INFO

Article history:
Received 1 December 2021
Revised 12 April 2022
Accepted 6 June 2022
Available online 9 June 2022

Keyword:
Deformable medical image registration
Multiple training strategies
Mutual learning network

ABSTRACT

Deformable medical image registration plays a vital role in clinical diagnosis, monitoring treatment, and postoperative recovery. Nevertheless, the existing registration algorithms rely on a single network or training strategy to complete the registration task. Even the most advanced registration algorithms cannot capture sufficiently compelling feature information by a single network or strategy.This paper proposes a new unsupervised and deformable medical image registration network framework to get the compelling feature. The network uses 2D sub-images of 3D images as additional constraint information to supplement the training process. Therefore, the network's two-dimensional and three-dimensional data images can interactively learn each other's characteristic information. In addition, we propose to perform secondary registration on the concentrated part of the registration area according to the characteristics of the input image. It enables the complete image and critical regions to learn their characteristic information interactively. This paper uses the image pyramid to integrate the two mutual learning strategies, thus proposing a multi-strategy mutual learning network(MMLN) and conducting many evaluation tests on the public data set OASIS and LPBA40. The test results show that the network can achieve better registration performance than other learning-based methods and traditional algorithms.

1. Introduction

Deformable medical image registration has a wide range of application prospects in contemporary medical image analysis. It plays an essential role in the normalization of population studies, the dynamic interpretation of organ movement processes, and the construction of other image analysis algorithms. Image registration refers to the process of establishing spatial correspondence between image pairs, while "deformable" refers to the dense nonlinear spatial transformation between image pairs[1]. All voxels in the moving image align to the fixed image through the vector deformation field ϕ. The images are in the same geometric space after aligning, and we can directly compare the structure and changes between the two images. Traditional medical image registration achieves the registration optimization problem of each image pair by aligning the voxel points with similar appearance in the image pair and constraining the registration mapping. Therefore, the amount of its calculation is too large, and the registration time is too long, which leads to the traditional registration algorithm is not practical in actual clinical applications[2–7]. The registration algorithm based on deep learning uses an objective optimization function to solve the registration optimization problem of all image pairs in the training dataset. It seeks the best vector deformation field by updating and optimizing the similarity objective function parameters between image pairs. The best set of optimized parameters can quickly output the mapping of all voxel points in one volume to the corresponding voxel points in another volume. The most popular objective functions currently include cross-correlation (CC) and normalized cross-correlation (NCC), mutual information (MI), and mean square error (MSE). There are two general kinds of registration: rigid registration and non-rigid registration (deformable registration). Rigid registration means that it preserves the angle and edge length between pixels through rotation and translation operations and only performers linear transformation, but it does not change the image structure.

* Corresponding author at: Shenzhen Key Laboratory of Media Security, Shenzhen University, 518060, China.
E-mail address: wmcao@szu.edu.cn (W. Cao).

https://doi.org/10.1016/j.neucom.2022.06.020
0925-2312/

However, deformable registration refers to calculating the dense correspondence between two images to perform nonlinear distortion of the image. It has become the primary step of many medical image analysis tasks.

Due to their fast registration speed and high accuracy, medical image registration methods based on deep learning have recently become a research hotspot in image registration. In the past few years, we have witnessed the vigorous development of registration methods based on deep learning [8–13]. These methods have achieved the current state-of-the-art registration accuracy on many data sets. Then these methods almost all have more significant advantages in medical images with only one registration area, such as liver and lungs. However, the registration accuracy of complex medical images with multiple registration areas still has much room for improvement. Until 2020, Tony proposed the Laplacian Pyramid Network [14], the network first proposed to use the Laplacian image pyramid concept to design an image registration network framework. It uses multi-resolution input images to jointly constrain the registration process to look for the best deformation field, and the network has achieved the most advanced registration performance in the brain dataset. The Laplacian pyramid network uses a multi-resolution strategy to extract the feature information of the input image in different resolution spaces. It uses a pre-training model to feedback the adequate information extracted in the low-resolution space to the next high-resolution space to assist the registration process in high-resolution space. In each level of independent resolution space, they only use a single registration network to complete this level of registration. However, a single multi-resolution strategy cannot extract sufficient practical feature information, which still leaves room for improvement in the registration accuracy of brain datasets.

This paper proposes a new multi-strategy mutual learning network(MMLN) given this deficiency of the Laplace pyramid network. The main contributions of this work are as follows.

- Reducing the dimensionality of the three-dimensional image to obtain its two-dimensional sub-image and use the two-dimensional sub-image as additional constraint information for the training process. The registration process is divided into two branches, extracting the input pair's global and local feature information from the original three-dimensional images and their two-dimensional sub-images.
- Proposing the core enhancement strategy solves the registration problem of multi-region medical images such as brain medical images. We perform secondary registration on the centralized unit of the core region, extract the core features, force the network to focus on the easily ignored areas, and assist the registration process of the complete image pair.
- Using three layers of image pyramids with different resolutions and constructing two different registration branches in the same registration space to build the registration framework. We use the multi-resolution pyramid to integrate the multi-dimensional joint strategy and the core enhancement strategy. Adopting a new mutual learning network optimizes the way of supplementing additional parameters. In the same-level registration space, the input data of different levels in the two networks of different branches learns each other's feature information, finally achieving a higher-precision registration effect.

The rest of paper is organized as follows. Section 2 reviews the related work, Section 3 explains the idea of multi-strategy mutual learning network, Section 4 carries out the experiments on the two 3D registration benchmark datasets,Section 5 shows the results of experiments and the last Section 6 concludes the paper.

2. Related Work

2.1. Medical image registration based on non-learning

The most common non-learning-based traditional medical image registration algorithms such as Elastic[15,16], Fluid [17–21], and B-Spline[22] models are based on iterative numerical solutions of optimization problems. Most of them are optimized in the displacement vector field space. The Demons method proposed in 1998 is optimized in the space of the velocity vector field. The use of the velocity vector field has apparent advantages in ensuring the differential homeomorphism of the image during the transformation process. Demons calculate the optical flow and use a Gaussian filter to smooth the flow graph. After multiple iterations of optimization, we get the prediction result between each pair of images. In addition to Demons, the LDDMM algorithm that appeared in 2005 uses the derivation and implementation of the Euler–Lagrangian optimization algorithm to calculate the particle flow, solve the global variational problem, and estimate the metric value of the image. Subsequently, there are the region-specific differential metric mapping (RDMM), vector momentum parameterized static velocity field (VSVF), and symmetric image normalization (SYN)[23,24], and other variants. SYN[4] is currently the most widely used algorithm in medical image registration, describing asymmetric image normalization method based on Euler–Lagrangian optimization to maximize cross-correlation. All those mentioned above, non-learning-based traditional medical image registration algorithms do not require a training process, so they are very adaptable and robust when used in different image forms and organs. However, they are all based on iterative optimization., Resulting in too long registration time and low registration efficiency, so these methods are not practical in clinical medical applications.

2.2. Medical image registration based on learning

As deep learning shows more and more advantages in computer vision, some medical image registration algorithms based on deep learning have been proposed. The registration algorithm based on deep learning can improve the registration performance of the iterative method and use deep reinforcement learning to predict the transformation steps to obtain the best mapping matrix. It can also considerably shorten the registration time and improve the registration efficiency [8,13,25,26]. The spatial transformation network STN is the most representative of the first-generation deep learning-based registration algorithm, which realizes image registration by generating dense nonlinear deformation fields. Since then, STN began to use for image registration in various situations[27].

There are two kinds of registration algorithm based on deep learning: supervised registration and unsupervised registration. The supervised registration algorithm uses the ground truth deformation field to evaluate the network's predicted deformation field. The segmented image corresponding to the training dataset can also impose additional segmented image constraints on the registration training process. The first generation of optical flow prediction network FlowNet[28] uses synthetic data for training and estimates pixel-level loss, which belongs to supervised registration. In 2019, Balakrishnan et al. used the segmented image corresponding to the training image to perform additional pixel-level constraints on the predicted deformation field in the 3D field for the first time[9].In 2020, Zheng et al. combined anatomical segmentation images with a cascade network, which achieves the current state-of-the-art registration performance in the liver dataset [29]. The method uses the segmentation of the training dataset

during the training process, so it also belongs to the supervised registration method. The supervised registration methods use the ground truth deformation field or segmented image information as additional constraints to achieve higher registration accuracy. However, obtaining such an excellent training dataset is very difficult, which makes the supervised registration impractical.

In contrast, the unsupervised registration method does not need to use the ground truth deformation field or segmented images during the training process. It only predicts the deformation field through the input pair of the fixed and moving images and uses the distortion operation to apply the deformation field to the moving image. The training process continuously minimizes the difference between the distorted and fixed images to achieve the optimal registration alignment. Most registration algorithms regard the registration task as an optimization problem and model it as:

$$\hat{\phi} = \text{argmin}_{\phi} S(M(\phi), F) + R(\phi), \tag{1}$$

where $\hat{\phi}$ represents the deformation field and F represents the fixed image. $M(\phi)$ represents the image obtained after using the deformation field to warp the moving image. $S(,)$ represents the similarity measure between the two images. $R(\phi)$ represents a regularization term that imposes geometric constraints on the deformation field.

Recently, unsupervised registration methods have gradually become a research hotspot in the field of registration and have made significant progress due to their easy acquisition of experimental data, less time-consuming registration, and high registration accuracy. In 2018, Balakrishnan et al. used an unsupervised registration method to customize the registration as a parameter function for the first time and model it through a convolutional neural network (VoxelMorph). It iteratively optimized the registration parameters and realized the 3D field the most advanced registration performance [8].VoxelMorph is a classic plug-and-play registration network. As a single registration network, it can quickly achieve high-performance registration, so it is often used as the primary registration subnet in many optimized registration frameworks, such as the cascade network proposed by Zhao et al. [30] and SYMNet proposed by Tony et al. [31]. VoxelMorph is used as the primary registration subnet to enrich the content of the registration framework. We can also regard the MMLN as a substantial extension of VoxelMorph, which is used as the primary 3D registration network after some internal adjustments, and then combined with three training strategies to build the final registration framework.In 2019, Zhao et al. used VoxelMorph as the basis for registering subnets. They combined it with the idea of iterative cascading and proposed an end-to-end iterative cascading network (VTN), which further improved the registration accuracy[32].In 2020, Zhao et al. improved the similarity constraint method of the original cascaded network based on VTN so that all registration subnets can achieve actual joint registration[30]. However, the above-unsupervised registration algorithms have significant advantages only on medical images with single registration regions such as the liver and lungs. They are not effective when applied to complex medical images with multi-region registration, such as the brain. These networks only use the original resolution image to extract features, reconstruct the deformation field, and measure the similarity between the original resolution distorted image and the fixed image. However, these methods ignore the truth that the gradient of the obtained similarity measure is usually very rough when the resolution is better. In the multi-resolution image registration framework, Hongming Li et al. gradually combined multiple complete convolutional networks in a profound self-supervision way to learn the spatial transformation process of the same pair of images at different resolutions to improve the registration accuracy of large deformation images[33].Meng Ye et al. estimated the motion field (INF) between any two consecutive T-MRI frames by a bi-directional neural network Generative diffeomorphic registration neural network. It provides a valuable solution for motion tracking and image registration in dynamic medical imaging Imaging[34].In 2020, Tony et al. proposed an unsupervised symmetric image registration method. They simultaneously estimated the forward and reverse transformations of the same pair of images, which significantly improved the similarity of images in the differential mapping space[31]. The recently proposed Vision Transformer (Vit) for image classification uses a purely self-attention-based model that learns long-range spatial relations to focus on the relevant parts of an image. Nevertheless, ViT emphasizes the low-resolution features because of the consecutive downsamplings, resulting in a lack of detailed localization information, making it unsuitable for image registration. Junyu Chen et al. proposed Vit-V-Net, which bridges Vit and ConvNet to provide volumetric medical image registration[35].And then,Junyu Chen et al. present TransMorph, a hybrid Transformer-ConvNet model for volumetric medical image registration. They also introduce three variants of TransMorph, with two diffeomorphic variants ensuring the topology-preserving deformations and a Bayesian variant producing a well-calibrated registration uncertainty estimate[36]. In 2020,Tony et al. first proposed the Laplacian Pyramid Registration Network[14] based on the Laplacian image pyramid structure. The network extracts image features from input images of different resolutions and then performs parameter embedding. It shows its strong registration advantage in brain datasets. However, Laplacian pyramid network only uses a single registration network to complete the spatial transformation of image pairs. It only uses a training strategy to build a registration framework, making the registration network unable to extract sufficient practical feature information from image pairs.This paper proposes a multi-strategy mutual learning network(MMLN) for the two problems of the Laplace network. It combines multiple training strategies and uses a mutual learning network to learn the characteristic input data information from each other. Therefore, MMLN realizes the actual fusion of the characteristic information at different levels and then reconstructs the optimal registration deformation field.

3. Method

3.1. Multi-resolution strategy

Inspired by traditional image pyramids, Laplacian Pyramid Network[14] adopts a new multi-1resolution strategy. In progressive registration from low resolution to high resolution, it adopts a pre-trained model in advance. We integrate the transformation parameters in the low-resolution registration space into the high-resolution registration space. It preserves the nonlinearity of feature mapping during the entire process of coarse-to-fine registration optimization.

This paper maintains the same multi-resolution strategy as LapIRN, using a three-layer pyramid structure. We build an image pyramid by downsampling, and perform trilinear interpolation on the input image to obtain $F_i \in (F_1, F_2, F_3), M_i \in (M_1, M_2, M_3)$, where F_3=F (original fixed image), M_3=M(Original moving image), the scale factor between adjacent layers of the pyramid is 0.5.F_1 and M_1 are used as the input of the first layer of the pyramid. The registration subnet extracts the nonlinear characteristics of the image pair in the lowest resolution space and outputs the three-channel dense vector velocity field V_1.We get the first level deformation field ϕ_1 after numerical integration. When the number of pyramid layers $i > 1$, we double upsample the V_{i-1} and ϕ_{i-1} output from the previous pyramid to obtain $V_{i-1'}$ and $\phi_{i-1'}$, and use $\phi_{i-1'}$ to warp M_i

to get $M_i(\phi_{i-1'})$. We superimpose $F_i, M_i(\phi_{i-1'})$, and $V_{i-1'}$ on the channel as the 5-channel input of this pyramid layer. And then, we add $V_{i-1'}$ to the output velocity field of this layer to get the actual velocity field of this layer. We show the specific regiatration framework in the Fig. 1. Three registration spaces with different resolutions learn sequentially. The two-layer registration space with lower resolution is equivalent to two pre-training sessions before the original resolution layer.Each layer of the pyramid uses a three-dimensional primary registration subnet with the same structure to complete the registration optimization task in the three-dimensional field of this layer. As shown in Fig. 2, we take the three-dimensional registration network in the first layer of registration space as an example to show the network's internal structure. The internal structure of the two-dimensional registration network is consistent with that of the three-dimensional registration network. In the second and third layers of registration space, we change the input channel of the network to 5 due to the superimposed output velocity field of the previous layer of registration space. Another structure is consistent with the network structure in the first layer of registration space.

3.2. Multi-dimensional joint strategy

In the three-dimensional field, the currently known image registration networks all use complete three-dimensional images as the network's input[8,32,30]. However, it is not enough to only extract the global feature information of the three-dimensional input to predict the deformation field, especially for complex medical images with many registration regions such as the brain. The global difference is more evident in three-dimensional space, and we can easily capture it. Therefore, feature extraction in three-dimensional space can capture global feature information, such as the size of the brain or the angle of the head. However, the local detail difference appears relatively weak in the three-dimensional space, so the local feature information extracted by the registration network is relatively rough. There are many independent registration areas in the brain, such as sulci, gyrus, and white matter. In the brain images of different patients or patients in different periods, the sulci or gyrus is more complex, and the anatomical differences of other registration regions are also complicated. These local differences information help predict the spatial mapping relationship of the brain's internal structure. These subtle anatomical differences are more evident at each slice level, so we cannot wholly rely on the global feature information in the three-dimensional space to complete the brain image registration task.

In order to extract richer local feature information and generate a more refined deformation field, we propose a new multi-dimensional joint strategy. We use the three-dimensional image as the input of the three-dimensional registration space and extract the global feature information of the three-dimensional

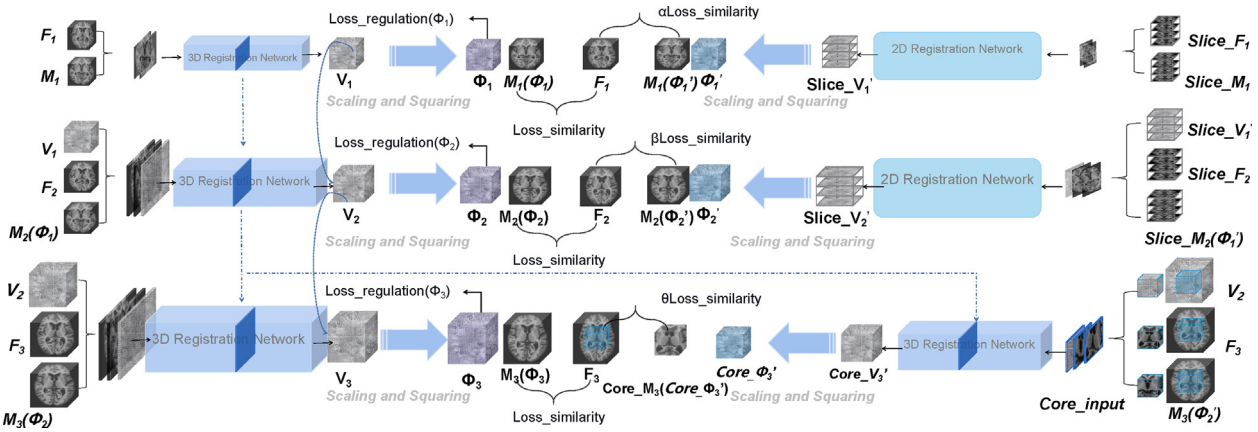


Fig. 1. Multi-strategy mutual learning network framework. The dark blue layer in the three-dimensional registration network of each registration space is the final feature map extracted by the network. We use the dotted line represents the gradual supplement of feature information between adjacent registration spaces. The blue arrow indicates the numerical integration of the static velocity field through scaling and squaring operations.

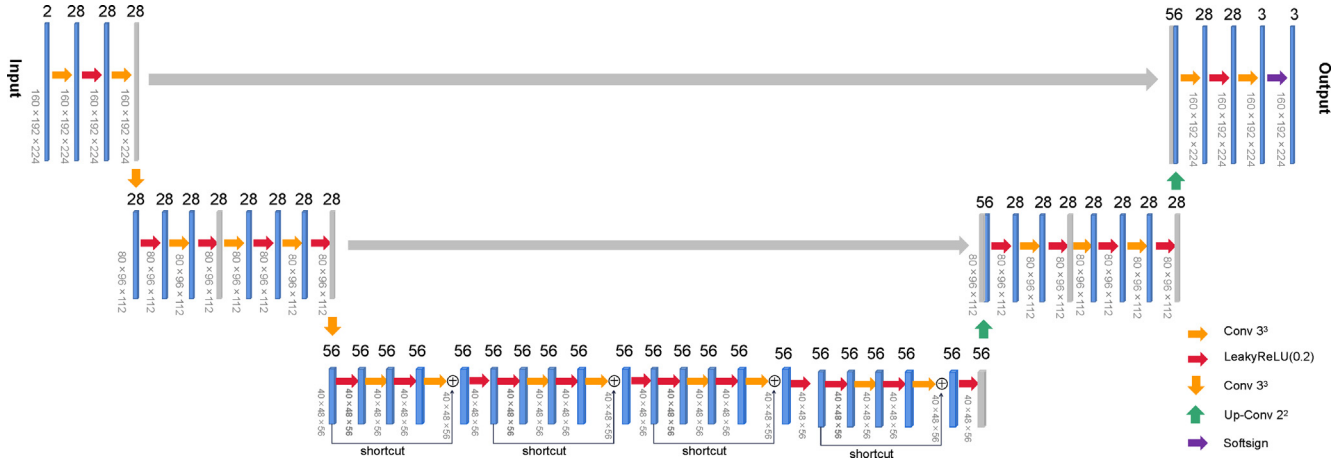


Fig. 2. The internal structure of the three-dimensional primary registration network. We mark the size of each feature map on its lower left, the corresponding channel number on the top of the feature map, and the gray arrow indicates a jump connection.

image through the three-dimensional registration network. At the same time, we use the two-dimensional as the input of the two-dimensional registration space and compare more subtle local differences between fixed and moving images from the slice level. The two-dimensional registration network can fine-tune the registration process of a three-dimensional registration network and refine the deformation field. We combine the three-dimensional registration network with the two-dimensional registration network to simultaneously manage global and local anatomical differences. It can extract richer feature information from multiple dimensions to make the predicted deformation field more refined.

3.3. Core strengthening strategy

Medical organs are all irregularly shaped, so there will always be a vacuum area in the corners no matter how to make the image size close to the organ size when acquiring medical organ images. These vacuum areas do not contain the voxel information of medical organs, so they cannot provide helpful feature information for predicting the deformation field. Except for the vacuum area, there are many independent and complex structures in the brain, and we need not register all structures. We need to align some regions of the brain according to the registration requirements. There are 56 labeled regions in this batch of brain data set, and the other labeled regions are redundant except for the selected 28 regions to be registered. The redundant regions that do not need to be registered will interfere with capturing adequate voxel information during the feature extraction stage of the network. The deformation field is the spatial mapping relationship of all corresponding voxel points between the fixed and moving images. The quality of the feature information extracted from the input image pair directly determines the quality of the predicted deformation field. Therefore, we propose the concept of feature reuse. As far as possible, we identify a region with the highest concentration of registration regions as the core region. The core region is not unique and may change according to the distribution of the selected registration regions. According to the distribution of 28 registration regions selected in this paper, we finally determine that the size of the core region is $80\times96\times112$, which is in the center of the original image. We remove the vacuum region and part of the redundant region and remain the core region. And then, we extract the features of the core region a second time to strengthen the core registration region while extracting features using complete image pairs. We must ensure the structural integrity of the essential registration regions, such as the brain sulcus and white matter, while dissecting the core region; otherwise, the incomplete white matter structure will harm the registration[37].

3.4. Mutual learning network

Considering that a single registration network (whether it is three-dimensional or two-dimensional, global or local) is relatively weak, predicting the deformation field through the feature information extracted by a single network is not precise. In recent years, the parameter embedding method based on the pre-training model has been used more frequently. A batch of high-quality model parameters are saved in advance in the form of a pre-training model, and high-quality network parameters are embedded in the learning process of the model during formal training to guide network learning.In this method, they perform the first-stage pre-training on the networks of different strategies in advance and then use the saved parameters as the prior guiding information in another strategy's network training process. The method can combine networks of different strategies, but it focuses more on network combination and optimization training in the second stage and provides more robust functions for the second stage network [38].In the first stage, the training of a single network loses the opportunity to use additional auxiliary prior information. The feature information of a single layer limits it, which affects the quality of the deformation field in the first stage, and thus cannot guarantee that those network parameters saved at the first layer are optimal. The parameter embedding method based on the pre-training model ignores that the registration networks of different strategies can promote each other in a unified way.

Given the shortcomings mentioned above based on the pre-training model, we propose a new mutual compatible learning network, which realizes the true sense of multi-strategy joint training. As shown in Fig. 1, we use a multi-dimensional joint training strategy in the first two layers of registration space and use the three-dimensional medical image as the input of the three-dimensional registration space to output an entire three-dimensional flow field. We successively use the corresponding two-dimensional sub-images of the three-dimensional image as the input of the two-dimensional registration network. Finally, we splice the output two-dimensional flow field to obtain an entire three-dimensional flow field. The output flow field of the two-dimensional and three-dimensional network respectively distorts the moving image in the registration space of this layer and then measures the similarity with the fixed image. We superimpose the two similarity measurement results with hyperparameters(α, β) as the similarity loss function of the registration space of this layer.

Through this method, we can realize the compatibility of the two-dimensional view and the three-dimensional view. The three-dimensional registration network extracts the global difference of the registered organ from the three-dimensional image, and the two-dimensional registration network captures the local feature information of the internal details of the organ from the two-dimensional image. More importantly, the mutual learning process of two registration networks in the same registration space further promotes mutual collaborative learning and the invisible exploration of their helpful information.We adopt the core strengthening strategy in the third-level registration space, using the complete 3D image and the core region as the input of the two registration branches. We use the secondary registration of the core region to obtain an enhanced deformation field focused on the core region and use it to distort the corresponding moving region.The similarity measurement result of the core region pair combines with the hyperparameter(θ) and then is added to the similarity measurement result of the complete image pair to get the similarity loss of the registration space of this layer. Starting from the first level of registration space, we up-sampled the final feature map extracted from the 3D registration network and then fed it back to the 3D registration network of the next level of registration space to supplement the feature information. The mutual learning network further promotes cooperation between different strategies. Moreover, the mutual learning network corrects the "biased feature information" captured by a single registration network so that the registration space can extract more prosperous and fairer feature differences, thereby outputting more precise and detailed deformation fields.

3.5. Loss function

3.5.1. Similarity loss

In the two registration branches of each layer of registration space, we use the correlation coefficient as the similarity measurement function to test the similarity between the moving image and the fixed image after spatial transformation. We define the correlation coefficient between two images (I_1, I_2) as:

$$CorrCoef(I_1, I_2) = \frac{Cov(I_1, I_2)}{\sqrt{Cov(I_1, I_1)Cov(I_2, I_2)}}. \quad (2)$$

$Cov(I_1, I_2)$ is the covariance between I_1 and I_2 defined as

$$Cov(I_1, I_2) = \frac{1}{|\Omega|} * \sum_{x\in\Omega} I_1(x)I_2(x) - \frac{1}{|\Omega|^2} * \sum_{x\in\Omega} I_1(x) \sum_{x\in\Omega} I_2(x), \quad (3)$$

where Ω is the set of image coordinate indexes. The correlation coefficient can measure the degree of linear correlation between two images. It is more robust to use it to measure the similarity. Usually, the value range of the correlation coefficient is (-1, 1), but because the images taken are all authentic images, they are correlated coefficients should all be non-negative, and define the final similarity loss[32] as:

$$L_{similarity} = 1 - CorrCoef(I_1, I_2). \quad (4)$$

3.5.2. Regularization loss of deformation field

We only perform parameter embedding in the 3D registration network, and we only use the complete deformation field predicted by the 3D registration network during testing. Therefore, when we impose regularization constraints on the deformation field, we only use the deformation field output by the whole 3D registration network. We use regularization loss to prevent the deformation field from being unrealistic or over-fitting and ensure the deformation field's smoothness [8,13,39].In this paper, we select the *L2* norm on the spatial gradient to impose regularization constraints on the velocity field, as shown in Eqn. 5:

$$L_v = \frac{k}{2^{L-p}} \times \| \bigtriangledown v\|_2^2, \quad (5)$$

where $p \in (1, L)$, represents the number of pyramid levels, and k is the regularization parameter.

3.5.3. Total loss of the network

We carry out the network training of each layer of registration space, respectively, so the total loss function of each layer of space only includes the similarity measurement loss of this space and the deformation field regularization term. We show the total loss function of the first-level registration space in Eqn. 6:

$$Loss_{(1)} = L_{v(1)} + L_{similarity}(F_1, M_1(\phi_1)) + \alpha L_{similarity}(F_1, M_1(\phi_1')), \quad (6)$$

where ϕ_1 and ϕ_1' are the output deformation fields of the three-dimensional and two-dimensional registration network in the first-level registration space, respectively. We show the total loss function of the second-level registration space in Eqn. 7:

$$Loss_{(2)} = L_{v(2)} + L_{similarity}(F_2, M_2(\phi_2)) + \beta L_{similarity}(F_2, M_2(\phi_2')), \quad (7)$$

where ϕ_2 and ϕ_2' are the output deformation fields of the three-dimensional and two-dimensional registration network in the second-level registration space, respectively. We show the total loss function of the third-level registration space in Eqn. 8:

$$Loss_{(3)} = L_{v(3)} + L_{similarity}(F_3, M_3(\phi_3)) + \theta L_{similarity}(CoreF_3, CoreM_3(Core\phi_3')), \quad (8)$$

where ϕ_3 and $Core\phi_3'$ are the deformation fields output by the registration network of the complete branch and the core branch in the third-level registration space,respectively.This paper sets α, β, θ to 0.25, 0.5, 0.25, respectively.

4. Experiments

4.1. Data and preprocessing

We use 450 T1-weighted brain MR scans from the OASIS[40] dataset and 40 brain MR scans from LPBA40[41] to evaluate our method based on the atlas. Atlas-based registration is a standard application in multidisciplinary image analysis, which can establish an anatomical correspondence between the atlas and moving images. The OASIS dataset contains data images of subjects ranging from 18 to 96 years old. Among them, 100 subjects have mild to moderate Alzheimer's disease. We perform standard preprocessing steps on the data set, including skull stripping, spatial normalization, and subcutaneous structure segmentation[42].In the OASIS data set, we use 28 subcortical structure segmentation of anatomical structures as the basis for our evaluation. In the LPBA40 dataset, we use 56 subcortical structure segmentation of 56 anatomical structures hand-drawn by experts as the basis for evaluation. In order to refine the registration area and improve computational efficiency, we further crop all data images to the size of 160×192×224. This article randomly divides the OASIS data set into 255, 20, and 150 volumes for training, verification, and testing. We use the LPBA40 data set as a test set for cross-data set verification. We randomly select five volumes from the two test sets as the fixed atlas of OASIS and LPBA40 and align the remaining 145 and 35 volumes as moving images to the fixed atlas. The test results in this paper are from the mean values of 725 and 175 test combinations of OASIS and LPBA40.

4.2. Baseline method

We compare our method with two traditional registration algorithms, SyN[4] and Elastic[15]. At the same time, in the learning-based method, we also selected the most commonly used VoxelMorph[9] network and the LapIRN[14] and its variant network-CIR-DM[43], and the newest ViT-V-Net[35] and TransMorph[36] as our comparison objects. VoxelMorph, LapIRN, CIR-DM, ViT-V-Net and TransMorph use the hyperparameters set in the original author's public code to train our data set from scratch. We have carefully adjusted the algorithm parameters for the two traditional registration algorithms to balance the registration effect and the registration time.

4.3. Test

We use the segmented images corresponding to the fixed image and the distorted moving image to calculate their dice scores to evaluate the degree of similarity between the two images. The maintenance of the topological structure of the medical image registration result and the reversibility of the displacement field predicted by the network are also significant for image registration. Therefore, we also use the percentage of voxels in non-positive Jacobian determinants in the displacement field($|J_\phi|_{\leq 0}$) to evaluate the validity of the registration results (differential homeomorphism characteristics).In addition, we use the volume ratio (*TC*) of the two segmented images before and after the spatial transformation as an auxiliary evaluation index for the characteristics of the differential homeomorphism. We also tested the average time required for each pair of images to be registered.

4.4. Implementation

Our proposed method (MMLN) is implemented based on PyTorch[44]. We use an ADAM optimizer with a fixed learning rate (10^{-4}).In order to verify the effectiveness of our proposed multi-dimensional joint strategy, core strengthening strategy, and mutual learning network, we use the controlled variable method to perform ablation experiments on MMLN to verify the registration performance of the three innovative points. We train our network from scratch and select a model with the best overall performance on the validation set.

5. Result

5.1. Registration performance

Table 1 compares the registration performance of two traditional registration algorithms and four learning-based registration methods. We can observe from Table 1 that the traditional algorithm has apparent advantages in registration validity. That is, it can achieve the lowest Jacobian score($|J_\phi|_{\leq 0}$). However, its registration time is too long, which affects its practical application ability. The four learning-based methods can complete the registration task in a short time. Among them, the MMLN proposed in the paper can achieve the best results in registration accuracy (Dice) and registration effectiveness ($|J_\phi|_{\leq 0}$), which proves the advantages of MMLN in the field of registration.

Table 2 summarizes the cross-data-set verification results for the six registration algorithms on LPBA40. Since the LPBA40 data set has more regions that need to be registered, it is more complicated than OASIS, so the performance of each registration algorithm is reduced to a certain extent. However, the MMLN proposed in this paper can still achieve the best registration effect in the six registration methods when the registration performance is reduced. The results in Table 2 prove the robustness and versatility of the MMLN network, and its advantages in the registration field are not limited to a specific data set and certain registration areas.

As shown in Fig. 3, the three-dimensional contour of the mask after registration using MMLN is closer to the fixed mask. Moreover, from the two-dimensional slices in the coronal section, cross-section, and sagittal section, we can also observe that the detailed texture of the mask section after the MMLN registration is more similar to the texture of the corresponding slice the fixed image. In the three professional slices, the area inside the red circle of the image after registration using MMLN is closer to the corresponding area of the fixed map. It can also be seen intuitively from the box plot in Fig. 4 that in most of the registration areas, the registration accuracy of MMLN is higher than that of LapIRN.

5.2. Contribution of the core strengthening strategy

In order to verify the effectiveness of the core strengthening strategy proposed in this paper, we compare the registration performance of the original LapIRN and the registration performance of the network added to the core enhancement strategy. The input image resolution in the first-level and second-level registration spaces is relatively small, making it challenging to eliminate redundant regions. Therefore, we only use the core strengthening strategy in the third-level registration space. We keep the same network parameters as the original LapIRN, train the network after adding the core enhancement strategy from the beginning and verify the registration performance of the two networks in four sets of atlas-based test sets. As shown in Table 3, we compare the performance of two networks in the OASIS data set. When we ensure that the two networks achieve the same registration similarity, the network, after adding the core strengthening strategy, can achieve a lower Jacobian score($|J_\phi|_{\leq 0}$) and an anatomical structure change score(TC) closer to 1.

Compared with the original LapIRN, the Jacobian score of the network after adding the core strengthening strategy is reduced by 33%. We use the visualization of the flow field of the two net-

Table 1
Comparison of registration performance between our proposed MMLN and other five registration algorithms on the OASIS data set. Among them, Dice represents the similarity between the images after registration (the bigger, the better); $|J_\phi|_{\leq 0}$ represents the percentage of folded voxels in the deformation field (the smaller, the better); TC represents the topological change of the anatomical structure (closer to 1 the better); Time represents the average registration time of each pair of images (in seconds). Initial: spatial normalization. We add the standard deviation of each indicator in parentheses. Using VoxelMorph as the baseline in the learning-based registration method, we supplemented the improvement in registration accuracy and registration effectiveness of the other three learning-based registration networks compared with VoxelMorph.

Method	OASIS						
	Dice	%Dice	$\|J_\phi\| \leq 0$	$\%\|J_\phi\| \leq 0$	TC	Time(s)	GPU memory(training)
Initial	0.611(0.019)	-	-	-	-	-	-
SyN	0.761(0.021)	-	0.0000(0.0000)	-	1.006(0.009)	42.05	-
Elastic	0.782(0.019)	-	0.0029(0.0012)	-	1.010(0.013)	56.43	-
VoxelMorph	0.772(0.007)	-	0.1935(0.0021)	-	1.005(0.013)	0.68	9.970 GiB
LapIRN	0.783(0.002)	+1.4%	0.0043(0.0006)	−97.8%	1.008(0.008)	0.88	15.48 GiB
CIR-DM	0.765(0.005)	−0.9%	0.0045(0.0007)	−97.7%	0.999(0.007)	0.87	17.76 GiB
ViT-V-Net	0.796(0.004)	+3.1%	0.0032(0.0008)	−98.3%	1.047(0.006)	0.21	10.09 GiB
TransMorph	0.799(0.003)	+3.5%	0.0021(0.0006)	−98.9%	1.045(0.009)	0.34	12.92 GiB
MMLN(Ours)	**0.791(0.002)**	**+2.5**%	**0.0016(0.0006)**	−**99.2**%	1.015(0.007)	0.90	16.61 GiB

Table 2
Comparison of the cross-data-set registration performance of our proposed MMLN and the other five registration algorithms on the data set LPBA40.Among them, Dice represents the similarity between the images after registration (the bigger, the better); $|J_\phi|_{\leq 0}$ represents the percentage of folded voxels in the deformation field (the smaller, the better); TC represents the topological change of the anatomical structure (closer to 1 the better); Time represents the average registration time of each pair of images (in seconds). Initial: spatial normalization. We add the standard deviation of each indicator in parentheses. Using VoxelMorph as the baseline in the learning-based registration method, we supplemented the improvement in registration accuracy and registration effectiveness of the other three learning-based registration networks compared with VoxelMorph.

Method	LPBA40					
	Dice	%Dice	$\|J_\phi\| \leq 0$	$\%\|J_\phi\| \leq 0$	TC	Time(s)
Initial	0.612(0.011)	-	-	-	-	-
SyN	0.698(0.015)	-	0.0000(0.000)	-	0.9985(0.001)	28.98
Elastic	0.699(0.016)	-	0.0001(0.0001)	-	0.9987(0.001)	50.62
VoxelMorph	0.646(0.012)	-	0.1797(0.0035)	-	1.0010(0.002)	0.89
LapIRN	0.700(0.007)	+8.4%	0.0027(0.0008)	−98.5%	1.0004(0.001)	0.91
CIR-DM	0.692(0.008)	+7.1%	0.0017(0.0001)	−99.1%	1.0011(0.001)	0.85
ViT-V-Net	0.648(0.007)	+0.3%	0.0034(0.0002)	−98.1%	0.9835(0.001)	0.23
TransMorph	0.661(0.006)	+2.3%	0.0021(0.0008)	−98.8%	0.9846(0.005)	0.35
MMLN(Ours)	**0.705(0.006)**	**+9.1**%	**0.0014(0.0002)**	−**99.2**%	1.0006(0.001)	0.88

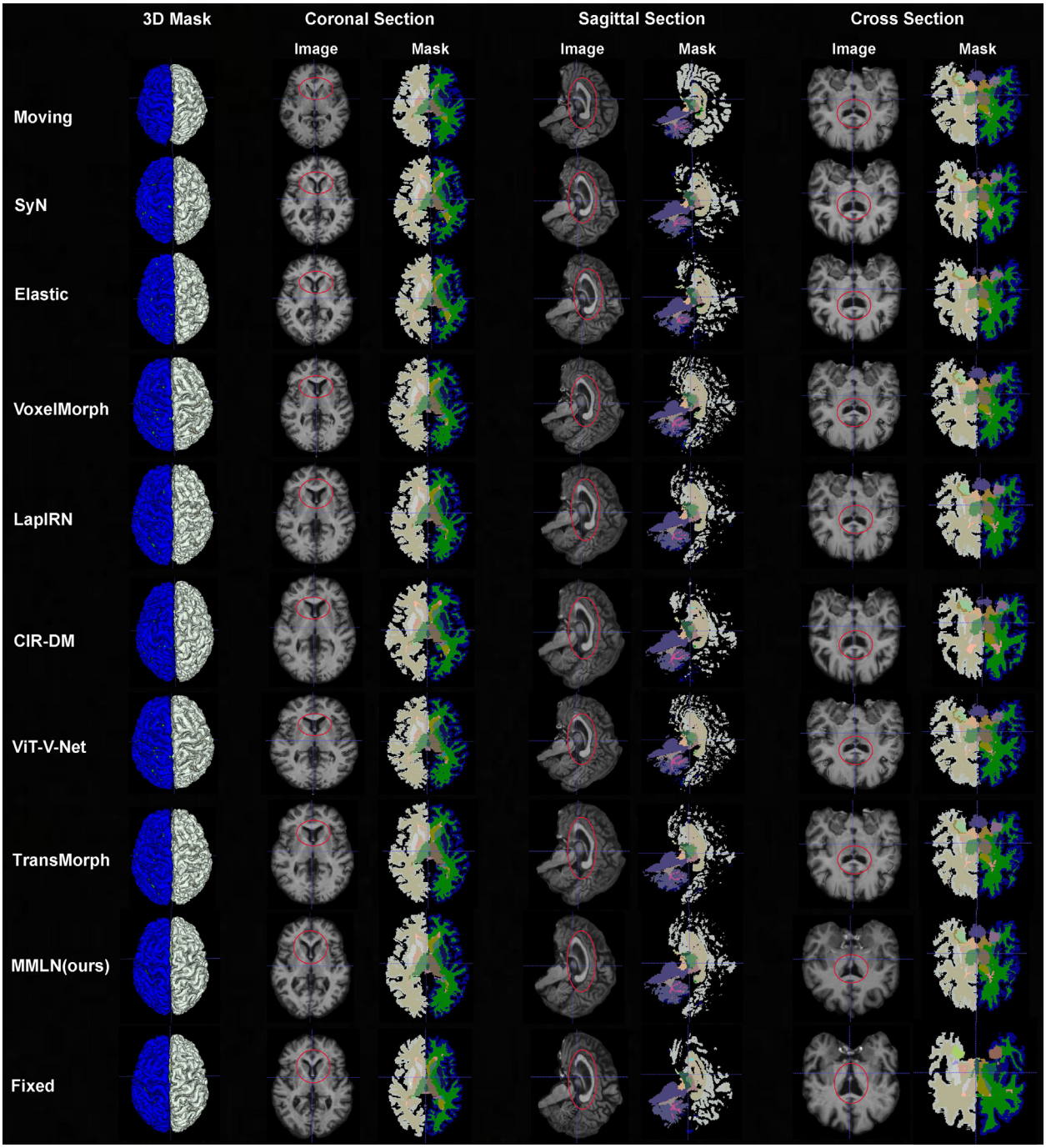


Fig. 3. Visualized results of the registration effect. We randomly select a pair of registration images in the test set of OASIS and use six registration algorithms to perform registration experiments on them. We display the registered images and masks in the three-dimensional view and three professional medical sections, respectively.We use red circles to circle the main registration areas in the three professional slices.The area inside the red circle shows that the image registered using MMLN (ours) is more similar to the fixed image in the main registration area.

works to show the significance of reducing the Jacobian score ($|J_\phi|_{\leq 0}$) by 33%. As shown in Fig. 5, we can observe from the cut surface of the three different dimensions that the flow field output by the network added the core strengthening strategy is very smooth. The flow field output by the original LapIRN has much pixel folding in the area that needs to be densely displaced, which will cause the information loss of the distorted image. The addition of the core strengthening strategy can significantly reduce the pixel folding of the original flow field, thereby improving the registration validity of the network.

5.3. Contribution of the multi-dimensional joint strategy

Using two-dimensional sub-images to assist in three-dimensional image registration can encourage the network to learn more detailed feature information. However, the extraction of two-dimensional sub-images and sequential registration tasks will also consume much time simultaneously. We cannot ignore the registration time in clinical medical applications. Therefore, we choose to adopt a two-dimensional joint strategy in the first and second levels of registration space. The spatial resolution of the first two

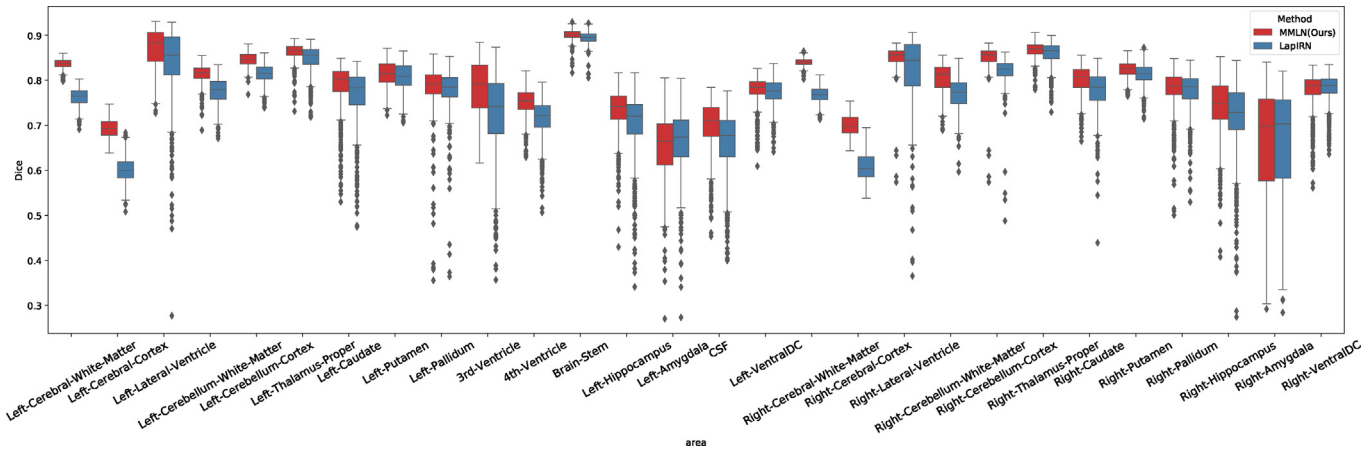


Fig. 4. Comparison of registration performance between LapIRN and MMLN in each registration area. In the OASIS test set, we plot the registration accuracy box plots of LapIRN and MMLN in 28 different registration regions.

Table 3
Comparison of registration performance between traditional LapIRN and LapIRN with the core strengthening strategy added.

Method	OASIS			
	Dice	$\lvert J_\phi\rvert_{\leq 0}$	TC	Time (s)
LapIRN	0.783 (0.002)	0.0043 (0.0006)	1.008 (0.008)	0.88
LapIRN with core strengthening	0.782 (0.002)	0.0029 (0.0005)	1.006 (0.009)	0.90

Table 4
Performance comparison of multi-dimensional federation strategies verified on the OASIS data set.

Method	OASIS			
	Dice	$\lvert J_\phi\rvert_{\leq 0}$	TC	Time (s)
LapIRN with core strengthening	0.782 (0.002)	0.0029 (0.0005)	1.006 (0.009)	0.90
MMLN(Ours)	0.779 (0.001)	0.0003 (0.0005)	1.016 (0.008)	0.93

levels of registration is relatively small, and the number of two-dimensional sub-images is small, so the time consumed is relatively more minor. In addition, when testing, we only need to use the registration network in the third layer of registration space to predict the required deformation field and only use the first two layers as auxiliary parameters of the third layer during training. Therefore, using the dimensional joint strategy in the first two layers can minimize the loss of time.

As shown in Table 4, we compared the registration performance of the two networks with and without a multi-dimensional joint strategy on the OASIS data set. We randomly select four maps in the test set as the fixed atlas, and the remaining ones are used as moving images and aligned to the fixed atlas in turn. Table 4 records the mean value of all registration results in the four batches of atlas-based registration. Compared with the network without a multi-dimensional joint strategy, after adopting the multi-dimensional joint strategy in the first two layers, the network slightly decreases the dice score, but its Jacobian score ($|J_\phi|_{\leq 0}$) is optimized by 90%. By adjusting the weight of the loss function later, we can transfer the optimization effect on the Jacobian score to the dice score.

Fig. 6 shows the registration effect of four different atlases, respectively. We can comprehensively compare the four performance indicators of each group of registered images through the radar chart. As shown in Fig. 6, in the registration batches of different atlases, the multi-dimensional joint strategy can optimize the Jacobian score to a large extent without making too many sacrifices on the other three indicators. This result is consistent with Table 4,

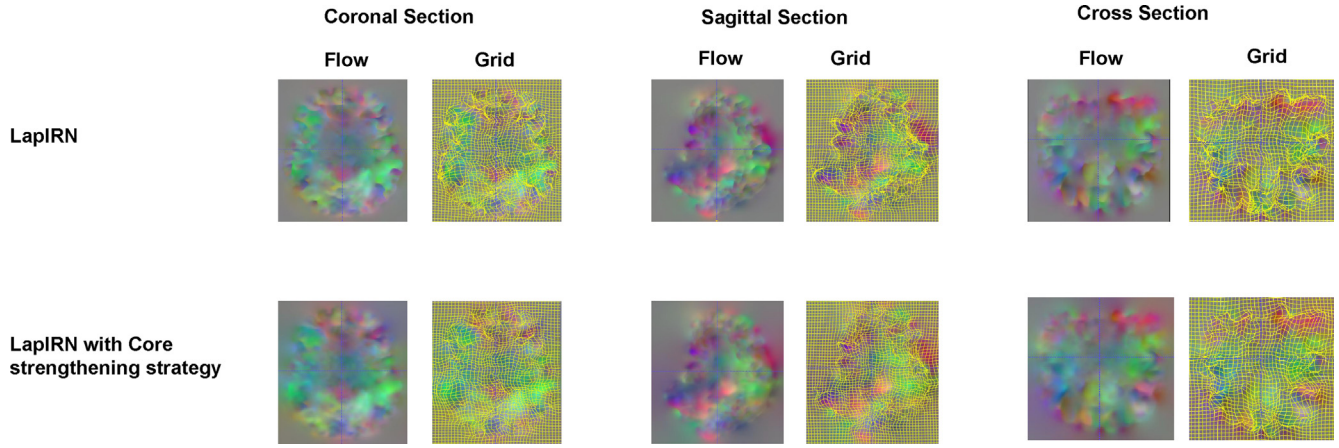


Fig. 5. The flow field visualization comparison between traditional LapIRN and LapIRN with core strengthening strategy. We randomly take out the deformation fields of the two networks on the same pair of registered images from the test set and show the flow fields of the two networks in three different dimensions: cross-section, coronal section, and sagittal section.The flow field of the network with the core strengthening strategy is smoother compared with traditional LapIRN in the three sections.

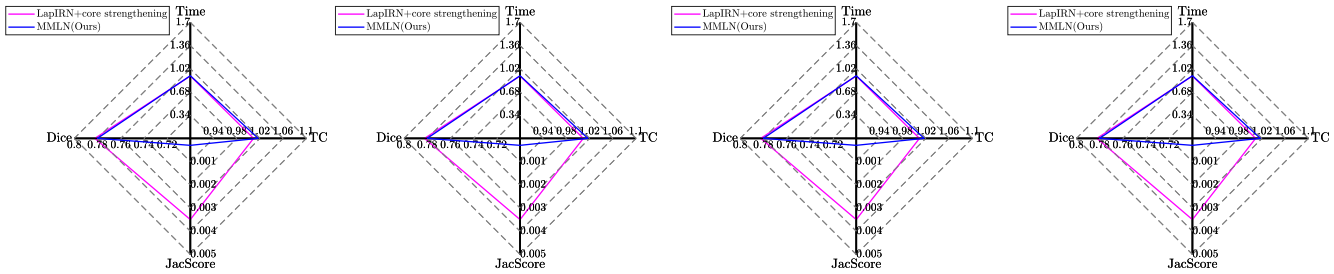


Fig. 6. Comparison of four batches of atlas-based registration effects. The JacScore represents the Jacobian score ($|J_\phi|_{\leq 0}$). Dice stands for registration image similarity. We use *TC* to represent the topological structure change of the registered image before and after registration. Time represents the average registration time of each pair of images.

which proves that the multi-dimensional joint strategy is not only effective for a specific batch of registration images but has strong versatility.

6. Conclusion

This paper proposes a multi-dimensional joint training strategy, a core strengthening training strategy, and a novel mutual learning mode. We conducted a lot of testing and verification of our network on the public data set OASIS. Then, we used another public data set, LPBA40, to verify the network model trained on OASIS across data sets to evaluate further the general applicability of our network Sexuality and robustness. We use the similarity metric indicator-Dice, the effectiveness indicator-"Jacobi score" and "TC" to conduct a comprehensive evaluation of the registration performance of our network. The experimental results prove that the multi-strategy mutual learning network(MMLN) can simultaneously improve registration accuracy and registration validity. The multi-strategy mentioned in the paper includes the multi-resolution strategy, multi-dimensional joint strategy, and core strengthening strategy. If there are other better training strategies in the future, the three training strategies in the paper can be replaced or supplemented to improve the registration performance further.

CRediT authorship contribution statement

Zhiyuan Zheng: Conceptualization, Methodology, Software, Formal analysis, Writing - original draft, Writing - review & editing. **Wenming Cao:** Supervision, Project administration. **Ye Duan:** Validation. **Guitao Cao:** Investigation. **Deliang Lian:** Visualization.

Declaration of Competing Interest

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

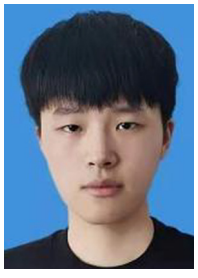
Acknowledgment

This work is supported by National Natural Science Foundation of China under grant 61771322 and 61871186, and the Fundamental Research Foundation of Shenzhen under Grant JCYJ20190808160815125.

References

[1] A. Sotiras, C. Davatzikos, N. Paragios, Deformable medical image registration: A survey, IEEE transactions on medical imaging 32 (7) (2013) 1153–1190.
[2] J. Ashburner, A fast diffeomorphic image registration algorithm, Neuroimage 38 (1) (2007) 95–113.
[3] J. Ashburner, K.J. Friston, Voxel-based morphometry–the methods, Neuroimage 11 (6) (2000) 805–821.
[4] B.B. Avants, C.L. Epstein, M. Grossman, J.C. Gee, Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain, Medical image analysis 12 (1) (2008) 26–41.
[5] X. Huang, N. Paragios, D.N. Metaxas, Shape registration in implicit spaces using information theory and free form deformations, IEEE transactions on pattern analysis and machine intelligence 28 (8) (2006) 1303–1318.
[6] B.B. Avants, N. Tustison, G. Song, et al., Advanced normalization tools(ants), Insight j 2 (365) (2009) 1–35.
[7] S. Klein, M. Staring, K. Murphy, M.A. Viergever, J.P. Pluim, elastix: A toolbox for intensity-based medical image registration, IEEE transactions on medical imaging 29 (1) (2009) 196–205.
[8] G. Balakrishnan, A. Zhao, M.R. Sabuncu, J. Guttag, A.V. Dalca, An unsupervised learning model for deformable medical image registration, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 9252–9260.
[9] G. Balakrishnan, A. Zhao, M.R. Sabuncu, J. Guttag, A.V. Dalca, Voxelmorph: a learning framework for deformable medical image registration, IEEE transactions on medical imaging 38 (8) (2019) 1788–1800.
[10] A.V. Dalca, G. Balakrishnan, J. Guttag, M.R. Sabuncu, Unsupervised learning for fast probabilistic diffeomorphic registration, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2018, pp. 729–738.
[11] M.-M. Rohé, M. Datar, T. Heimann, M. Sermesant, X. Pennec, Svf-net:learning deformable image registration using shape matching, in: International conference on medical image computing and computer-assisted intervention, Springer, 2017, pp. 266–274..
[12] B.D. de Vos, F.F. Berendsen, M.A. Viergever, M. Staring, I. Išgum, End-to-end unsupervised deformable image registration with a convolutional neural network, in: Deep learning in medical image analysis and multimodal learning for clinical decision support, Springer, 2017, pp. 204–212..
[13] B.D. de Vos, F.F. Berendsen, M.A. Viergever, H. Sokooti, M. Staring, I. Išgum, A deep learning framework for unsupervised affine and deformable image registration, Medical image analysis 52 (2019) 128–143.
[14] T.C. Mok, A.C. Chung, Large deformation diffeomorphic image registration with laplacian pyramid networks, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2020, pp. 211–221.
[15] R. Bajcsy, S. Kovačič, Multiresolution elastic matching, Computer vision, graphics, and image processing 46 (1) (1989) 1–21.
[16] D. Shen, C. Davatzikos, Hammer: hierarchical attribute matching mechanism for elastic registration, IEEE transactions on medical imaging 21 (11) (2002) 1421–1439.
[17] M.F. Beg, M.I. Miller, A. Trouvé, L. Younes, Computing large deformation metric mappings via geodesic flows of diffeomorphisms, International journal of computer vision 61 (2) (2005) 139–157.
[18] G.L. Hart, C. Zach, M. Niethammer, An optimal control approach for deformable registration, in: 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, IEEE, 2009, pp. 9–16.
[19] T. Vercauteren, X. Pennec, A. Perchant, N. Ayache, Diffeomorphic demons: Efficient non-parametric image registration, NeuroImage 45 (1) (2009) S61–S72.
[20] Z. Chen, H. Jin, Z. Lin, S. Cohen, Y. Wu, Large displacement optical flow from nearest neighbor fields, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp. 2443–2450.
[21] J. Wulff, M.J. Black, Efficient sparse-to-dense optical flow estimation using a learned basis and layers, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 120–130.
[22] D. Rueckert, L.I. Sonoda, C. Hayes, D.L. Hill, M.O. Leach, D.J. Hawkes, Nonrigid registration using free-form deformations:application to breast mr images, IEEE transactions on medical imaging 18 (8) (1999) 712–721.
[23] Z. Shen, F.-X. Vialard, M. Niethammer, Region-specific diffeomorphic metric mapping, arXiv preprint arXiv:1906.00139..
[24] Z. Shen, X. Han, Z. Xu, M. Niethammer, Networks for joint affine and non-parametric image registration, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 4224–4233.

[25] H. Li, Y. Fan, Non-rigid image registration using self-supervised fully convolutional networks without training data, in: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), IEEE, 2018, pp. 1075–1078.
[26] G. Haskins, U. Kruger, P. Yan, Deep learning in medical image registration: a survey, Machine Vision and Applications 31 (1) (2020) 1–18.
[27] M. Jaderberg, K. Simonyan, A. Zisserman, et al., Spatial transformer networks, Advances in neural information processing systems 28 (2015) 2017–2025.
[28] A. Dosovitskiy, P. Fischer, E. Ilg, P. Hausser, C. Hazirbas, V. Golkov, P. Van Der Smagt, D. Cremers, T. Brox, Flownet: Learning optical flow with convolutional networks, in: Proceedings of the IEEE international conference on computer vision, 2015, pp. 2758–2766.
[29] Z. Zheng, W. Cao, Z. He, Y. Luo, Progressive anatomically constrained deep neural network for 3d deformable medical image registration, Neurocomputing 465 (2021) 417–427.
[30] S. Zhao, Y. Dong, E.I. Chang, Y. Xu, et al., Recursive cascaded networks for unsupervised medical image registration, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 10600–10610.
[31] T.C. Mok, A. Chung, Fast symmetric diffeomorphic image registration with convolutional neural networks, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020, pp. 4644–4653.
[32] S. Zhao, T. Lau, J. Luo, I. Eric, C. Chang, Y. Xu, Unsupervised 3d end-to-end medical image registration with volume tweening network, IEEE journal of biomedical and health informatics 24 (5) (2019) 1394–1404.
[33] Y. Li, H. Li, Y. Fan, Acenet: Anatomical context-encoding network for neuroanatomy segmentation, Medical Image Analysis 70 (2021) 101991.
[34] M. Ye, M. Kanski, D. Yang, Q. Chang, Z. Yan, Q. Huang, L. Axel, D. Metaxas, Deeptag: An unsupervised deep learning method for motion tracking on cardiac tagging magnetic resonance images, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 7261–7271.
[35] J. Chen, Y. He, E.C. Frey, Y. Li, Y. Du, Vit-v-net: Vision transformer for unsupervised volumetric medical image registration, arXiv preprint arXiv:2104.06468..
[36] J. Chen, Y. Du, Y. He, W.P. Segars, Y. Li, E.C. Frey, Transmorph: Transformer for unsupervised medical image registration, arXiv preprint arXiv:2111.10480..
[37] M. Sdika, D. Pelletier, Nonrigid registration of multiple sclerosis brain images using lesion inpainting for morphometry or lesion mapping Tech. rep., Wiley Online Library, 2009.
[38] B. Liu, Y. Cao, Y. Lin, Q. Li, Z. Zhang, M. Long, H. Hu, Negative margin matters: Understanding margin in few-shot classification, in: European Conference on Computer Vision, Springer, 2020, pp. 438–455.
[39] B. Kim, J. Kim, J.-G. Lee, D.H. Kim, S.H. Park, J.C. Ye, Unsupervised deformable image registration using cycle-consistent cnn, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2019, pp. 166–174.
[40] D.S. Marcus, T.H. Wang, J. Parker, J.G. Csernansky, J.C. Morris, R.L. Buckner, Open access series of imaging studies (oasis): cross-sectional mri data in young, middle aged, nondemented, and demented older adults, Journal of cognitive neuroscience 19 (9) (2007) 1498–1507.
[41] D.W. Shattuck, M. Mirza, V. Adisetiyo, C. Hojatkashani, G. Salamon, K.L. Narr, R. A. Poldrack, R.M. Bilder, A.W. Toga, Construction of a 3d probabilistic atlas of human cortical structures, Neuroimage 39 (3) (2008) 1064–1080.
[42] B. Fischl, Freesurfer, Neuroimage 62 (2) (2012) 774–781.
[43] T.C. Mok, A. Chung, Conditional deformable image registration with convolutional neural network, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2021, pp. 35–45.
[44] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, A. Lerer, Automatic differentiation in pytorch..



Zhiyuan Zheng received the B.Eng. Degree in electronic information engineering from Hubei University in 2020. He is currently pursuing the master's degree in electronic information engineering with Shenzhen University, Shenzhen, China. He current research interests include deep learning, medical image analysis, and medical image registration.

Wenming Cao received the M.S. degree from the System Science Institute, China Science Academy, Beijing, China, in 1991, and the Ph.D. degree from the School of Automation, Southeast University, Nanjing, China, in 2003. From 2005 to 2007, he was a ostdoctoral Researcher with the Institute of Semiconductors, Chinese Academy of Sciences, Beijing, China. He is currently a Professor with Shenzhen University, Shenzhen, China. He has authored or coauthored more than 80 publications in top-tier conferences and journals. His research interests include pattern recognition, image processing, and visual tracking.

Ye Duan is an associate professor in the Department of Electrical Engineering and Computer Science at the University of Missouri. He is the director of the Computer Graphics and Image Understanding Lab and the director of the Cognitive Internet of Things (CIoT) for Intelligent Communities - Industry Supported Consortium (ISC). His research interests include computer graphics, computer vision, machine learning, and biomedical imaging, with a special focus on 3D deep learning. He has received funding from NSF, NIH, DoD, DoE, ARL, NGA, Honeywell, Bosch, etc.

Prof. Guitao Cao obtained her Ph.D. in 2006 from Shanghai Jiao Tong University with a focus on pattern recognition. She is currently a professor of Software Engineering Institute, East China Normal University (ECNU), Shanghai. ECNU is the top tier university in China with a high rank (Level A) in Software Engineering in China. She was also a visiting researcher with University of Missouri Columbia. She has published decades of peer reviewed papers in top venues including IEEE Transactions on Cybernetics, IEEE Transactions on Multimedia, and IEEE Transactions on Biomedical Engineering. Prof. Cao is also the Principal Investigator for many research funding with major sponsors including the National Science Foundation of China, Ministry of Industry and Information Technology of the People's Republic of China, and Science Foundation of Shanghai. Her research interests include pattern recognition, image processing and machine learning.

Delliang Lian received his MASTER's degree in physics from the Graduate University of Chinese Academy of Sciences and his doctor's degree in microelectronics from Beijing Normal University. From 1996 to 1997, he was engaged in developing photoelectric devices at Hong Kong Baptist University. Since 1999, he has been teaching in the School of Information Engineering, Shenzhen University and was promoted to associate professor in 2002.