Neural Computing and Applications (2022) 34:17175–17191
https://doi.org/10.1007/s00521-022-07365-4

ORIGINAL ARTICLE

Feature self-calibration network with global-local training strategy for multi-region deformable medical image registration

Zhiyuan Zheng[1] · Wenming Cao[1,2,3] · Deliang Lian[1] · Yi Luo[1]

Received: 13 July 2021 / Accepted: 27 April 2022 / Published online: 29 May 2022

Abstract
3D deformable medical image registration has important clinical significance. Deep learning-based methods have shown outstanding advantages in medical image registration. However, most current learning-based practices only have sound registration effects for large-area overall deformed medical images such as lung and liver CT scans. Medical images with multiple registration regions, such as brain MR scans, have become a significant shortcoming in the field of medical image registration. The current learning-based method uses only one displacement field to deform the entire moving image. For brain images with multiple registration areas, this method cannot guarantee accurate deformation of all regions, nor can it ensure the excellent maintenance of the topological structure of each part. Aiming to resolve this unresolved problem in the field of medical image registration, we propose a novel training strategy suitable for multi-region registration: global and local joint training strategy (*GoLo*). And then, we combine it with our proposed feature self-calibration network (*FSCN*) to solve the registration optimization problem of multi-region medical images. We conducted many quantitative and qualitative evaluation tests on the public brain MR scan data set(OASIS). The test results show that our method can further improve the accuracy of the registered image and ensure reversibility compared with the existing state-of-the-art methods.

Keywords Multi-region medical image registration · Global and local joint training strategy · Feature self-calibration network

1 Introduction

Deformable registration plays a vital role in the field of medical image analysis. It is sweltering for disease diagnosis, medical information fusion, tumor growth monitoring, image-guided surgical treatment, and radiotherapy planning. Deformable image registration is to find the nonlinear spatial correspondence between two images. The input image pair includes a fixed image and a moving image. After the deep neural network estimates the nonlinear transformation from the moving image to the fixed image, it is reflected in the form of a displacement field. The displacement field contains the mapping relationship of each voxel point in the moving image to the corresponding voxel point of the fixed image. The moving image aligns to the fixed image through the displacement field. There are many traditional algorithms applied in the field of medical image registration [1–6]. These algorithms generally have customized measurement of transformation quality, iteratively update a set of parameters according to the measurement consequences, and, finally, find optimal solutions in the updated parameters to achieve the best performance registration. Some traditional algorithms also introduce the concept of "image pyramid" [7, 8], which combines input images of different resolutions to constrain the transformation of the target image. Although traditional algorithms have shown excellent registration effects on

✉ Wenming Cao
wmcao@szu.edu.cn

Zhiyuan Zheng
zhengzhiyuan2020@email.szu.edu.cn

Deliang Lian
Liandl@szu.edu.cn

Yi Luo
1910434004@email.szu.edu.cn

[1] Guangdong Key Laboratory of Intelligent Information Processing and Shenzhen Key Laboratory of Media Security, Shenzhen University, Shenzhen 518060, China

[2] Video Processing and Communication Laboratory, Department of Electrical and Computer Engineering, University of Missouri, Columbia MO 65211, USA

[3] Peng Cheng Laboratory, shenzhen 5180601, China

some data sets, the required registration time is greatly affected by the degree of alignment between the image pairs before registration. Especially when the input image has an enormous dimension and a high resolution, the registration time required by the traditional algorithm is consuming, leading to the impractical use of conventional algorithms in clinical medicine.

Recently, image registration methods based on deep learning have received widespread attention. They can further improve the accuracy of registered images and achieve the fastest image registration, which is crucial in the field of time-sensitive medical image processing and is of great significance to the application of image registration in clinical medicine. Learning-based registration methods mainly include supervised registration and unsupervised registration. The difference between them is primarily whether the learning process uses the ground truth deformation field. The training data with the ground truth deformation field are hard to get, which significantly restricts the registration performance of the supervised registration method. Therefore, the registration method of unsupervised learning does not require prior information in the training process has attracted more attention. The learning-based deformable image registration method mainly uses a convolutional neural network (CNN) to predict the vector displacement field containing the mapping relationship between the input image pairs. The moving image aligns to the fixed image through the vector displacement field. The forward learning process is constrained by the similarity measure between the warped image and the fixed image, and the gradient can propagate back through the so-called warping operation [9].

The brain data image contains the measured heritable neuroanatomical features and is essential for developing neurological disease models. Heritable neuroanatomical characteristics are present in patients who have not developed the disease, and we can generally detect these features before the onset of the illness [10]. However, if the characteristics of the disease can change at any time, it is challenging to detect them in advance [11, 12]. Cerebrovascular diseases have been proven to be the leading cause of dementia. For example, white matter lesions (WML) can lead to a decline in predictive cognition and increase the risk of stroke and dementia.Specific disease characteristics are essential for pathological diagnosis and post-diagnosis treatment. To enable we can detect individual-specific disease characteristics for later treatment [13, 14], more and more methods based on deep learning have recently been applied in the brain imaging dataset.Using neuroimaging to check the cerebrovascular danger in the brain has become a hot issue in recent research. Then, the large-scale manual analysis is not only time-consuming and labor-intensive but also subjective and error-prone. Automatic and objective brain medical image analysis through deep learning tools is more efficient and robust.As an indispensable critical step in neuroimaging-related research, image registration plays a vital role in analyzing the changes in patients' diseases over time.

The registration analysis of brain data images generally aligns the brain images of the disease group and control groups' brain images. It compares the average information of all images in the template space.The iterative cascade network proposed by Zhao et al. [15] achieved the best performance in image registration in the previous learning-based method.They use the idea of the iterative cascade to gradually match the input images and realize the registration from coarse to fine.However, the iterative cascaded network consumes huge GPU memory. It has a significant effect only on large-area overall deformed medical images such as the liver.For more detailed brain images with multiple registration areas, its registration effect is minimal. Tony C et al. proposed a Laplacian pyramid network [16] to register brain data images in response to this problem, which achieved the best registration effect in the current brain data set.However,the Laplacian pyramid network proposes two methods to improve registration accuracy and ensure registration validity. It cannot guarantee registration accuracy and validity simultaneously.

We propose a feature self-calibration network based on the Laplacian image pyramid to solve the dilemma faced by those mentioned above current learning-based methods. And then, we combine it with the global and local joint training strategy, which can further improve the accuracy of the registration result while ensuring the differential homeomorphism of the registered image.The main contributions of this paper are as follows. We

- propose a novel global and local joint training strategy for brain medical images with multi-region complex registration.The original image is divided into patches of a specific size according to the number of registration regions to complete patch-level registration. And then, it cooperates with the initial image-level registration, finally realizing the overall registration from local to global and improving the accuracy of the registration result;
- combine the concept of feature pyramid and grouped convolution to propose a new feature self-calibration network. It can further extract features of input images in different scale spaces and collect global information of all spatial locations. And then, it uses output features in low-scale areas to guide and calibrate the output

features of the original scale space to prevent the loss of compelling features;
– further optimize the speed field superposition method of the original Laplacian pyramid network. Regarding the velocity field as a minimal displacement field, we use the displacement field superposition formula [15] to superimpose the velocity fields of different spaces.It can ensure better differential homeomorphism characteristics of the registered image.

2 Related work

2.1 A.Traditional algorithm

Previously, we usually used traditional algorithms to complete the registration of medical images such as fiRE [17], ANTS [5], and Elastix [6]. The conventional algorithm is to iteratively update the parameters by defining the quality of the registered image to find the final optimal transformation. Traditional algorithms solve the optimization problem in the deformation space [1, 3, 18–22] and these non-learning-based algorithms need to optimize the energy function of each input image pair. Therefore, it leads to the time-consuming registration process, making it impractical in clinical medical applications.

2.2 The method based on deep learning

Methods based on deep learning include supervised learning methods and unsupervised learning methods. The supervised learning method requires the ground truth deformation field as the comparison of the deformation field generated in the learning process [23–27] or requires an anatomical segmentation image to guide the registration process [28]. Traditional algorithm tools usually generate the ground truth deformation field, and professional doctors must manually generate high-quality anatomical segmentation images.The supervised learning method has significant advantages in registration accuracy, but the synthesized ground truth deformation field and the quality of the segmented image limit its registration performance.

More and more attention has been paid to unsupervised learning methods recently to eliminate the limitations of supervised learning registration. The unsupervised learning method does not require the ground truth deformation field and anatomical segmentation images. It only needs to input the image pair into the CNN, output the displacement vector field, and align the moving image to the fixed image through the displacement vector field. We measure the similarity between the distorted moving image and the fixed image as a loss function to achieve unsupervised learning of dense spatial mapping between input image pairs. In 2018, Balakrishnan et al. applied the unsupervised learning method to three-dimensional deformable medical image registration for the first time [29]. They defined the registration as a parameter function to optimize the parameters of a given input image pair. Given a new couple of images, we use the learned parameters to directly calculate the parameter function to register new image pairs rapidly.

Unsupervised learning registration models the parameter function through CNN and uses a spatial transformation layer to reconstruct another image from one image. This method achieves the most advanced registration accuracy and the least registration time in 3D deformable medical image registration.In the same year, Balakrishnan et al. proposed a probabilistic generative model to guarantee the differential homeomorphism [30]. They performed variational reasoning on this model and proposed a new registration formula.It combines the original transformation layer with the diffeomorphism integration layer to learn the unsupervised registration process jointly. They defined the output of CNN as a static velocity field and used scaling and squaring to calculate the numerical integration of the velocity field in time [31], which ensured the reversibility of the final registration result and the maintenance of the topological structure. In 2019, Zhao et al. proposed the Volume Tweening Network (VTN) [32] based on the idea of the iterative cascade. VTN integrates the affine subnet and three deformable dense registration subnets into an overall network, which realizes end-to-end progressive registration from coarse to fine. Finally, it achieves the most advanced registration effect on a large-area overall deformed medical data set. In the same year, Zhao et al. further optimized the cascading method based on VTN and proposed a new iterative cascading network [15], which can customize several deformable dense registration subnetworks according to the registration task. Changing the position of the similarity loss function makes several deformable dense registrations sub-networks truly unite to complete the common registration task and further improve the accuracy of the registration result based on VTN. In 2020, Tony et al. proposed a Laplacian pyramid network [16] for multi-region registration medical images. They utilized the advantages of a multi-resolution strategy to improve the registration perfoemance while maintaining the non-linearity of the feature maps throughout the coarse-to-fine optimization scheme. The Laplacian image pyramid was used as the core idea to create A three-layer pyramid is created, and each layer integrates a detachable diffeomorphism module. Two registration methods are proposed, respectively, focusing on the accuracy and validity of the registration results.The Laplacian pyramid network has achieved the current state-of-the-art registration

performance on the brain data set. However, it cannot meet the accuracy and effectiveness of the registration results simultaneously but simultaneously uses two different registration modes to optimize them.Given the limitations of the Laplacian pyramid network, this paper proposes a more complex and robust design. For brain images that require multi-region complex registration, our network can simultaneously improve the accuracy and effectiveness of the registration results.

3 Method

3.1 Laplacian image pyramid

The idea of the Laplacian image pyramid has been applied in various fields of computer vision and has proved its significant advantages, such as the construction of high-resolution solutions and the robustness of training. As shown in Fig. 1, our network inherits the advantages of the LapIRN [16] and adopts a multi-resolution strategy. It consists of the LapIRN and builds three registration spaces with different resolution scales. The ratio of adjacent spaces is 1:2, and the final level of registration space is the original resolution size. To balance the weight of the loss produced by different resolution spaces, inspired by [16, 33, 34], we use coarse-to-fine training first, we use coarse-to-fine training and gradually embed the training data in low-resolution space into higher resolution registration space. To ensure that the training remains stable after replacing the registration space, we first freeze the training parameters in the upper-level space and unfreeze the learning parameters of the upper level after completing 2000 pieces of training at this level.According to the registration space scale, the same images are downsampled to different degrees and sent to the corresponding space as the input. We feed the low-dimensional feature information and velocity field information in the low-resolution space back to the corresponding position in the next-level registration space for information supplement. Each level of registration space outputs the velocity field of the corresponding scale, and we get the corresponding displacement field of the level after the integration operation. Except for the lowest resolution space of the first level, the input of each level of registration space and the image pair that needs to be registered will also upsample the output velocity field of the previous level as an auxiliary input.

3.2 Global and local joint training strategy

Brain data sets usually have dozens of regions that need to be registered. Most of the different regions are independent of each other and are not connected, which leads to the current deep learning-based registration methods that are still challenging when processing brain images. The previous registration model is limited to the complete image pair as the network input. However, the 3D brain image has a complex structure, and the network cannot reasonably extract the features of different registration regions within the complete image, resulting in the final registration performance is not ideal. The internal brain has many independent registration areas, hampering the registration performance, so we segmented the network input according to this feature of the brain image. Each input of the network only contained fewer registration regions, making the registration network more

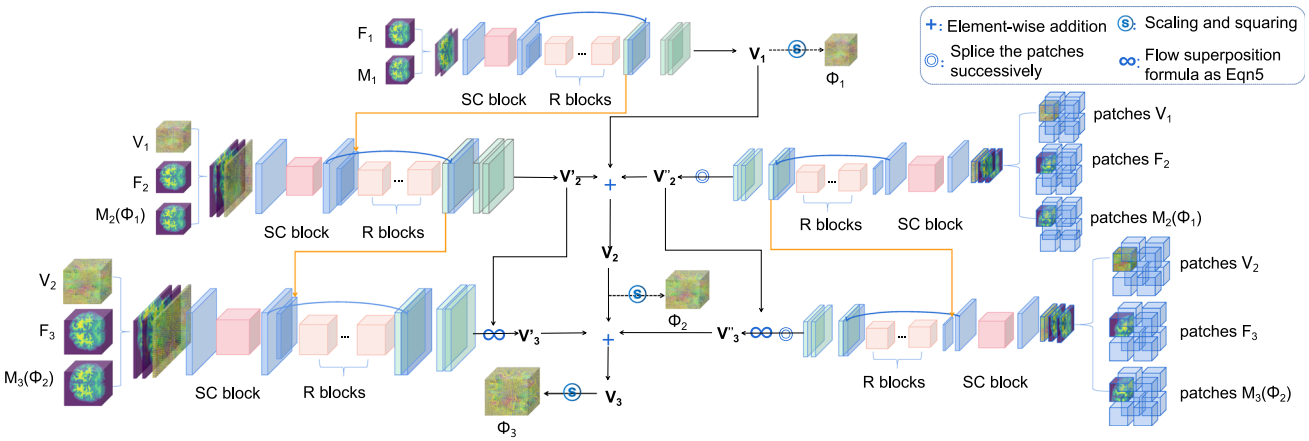


Fig. 1 The framework of feature self-calibration network based on global-local training strategy. The blue layer is the feature encoder, the red layer is the feature self-calibration module, the orange layer is the residual module, and the green layer is the feature decoder. The blue line is the jump connection, and the orange line is the gradual supplement of feature information. The dashed path only works during the training phase.In figure, + means the addition operation of the velocity field. Ⓢ means the numerical integration operation of the velocity field to obtain the deformation field. ◎ means that we successively spliced the velocity field patches to get the complete output speed field. ∞ means to use the velocity field superposition formula Eq. 5 to fuse the output velocity fields of different registration spaces

focused on the internal registration work of each region. However, a single block ignores the global characteristics of 3D brain images and the relationship between each registration region. Therefore, we combine the global image registration and propose a brand-new training strategy- a global and local joint training strategy (GoLo). In each level of registration space except the first level, we divide the registration process into two branches: the global registration branch and the local registration branch. On the local registration branch, we divide the input image pair into equal-sized image blocks and use the image blocks as the actual network input to complete the registration process for a specific registration area in a smaller range. The local network extracts finer local feature details, outputs the velocity field block for the image block, and finally stitches the velocity field block according to the block order to obtain the complete velocity field. However,it is unreasonable only to use blocks for training in medical image registration because we cannot accurately divide the blocks according to the registration area. The final image blocks may only contain the incomplete registration area, which restricts the network from learning the feature information of the connection area between image blocks. Therefore, the local registration branch must work together with the global registration branch. We use a complete image pair as the input of the global registration branch, extract the global features of the entire image pair, and finally generate a complete velocity field and displacement field. Then,the global velocity field output by the global branch supplements the velocity field output by the local branch. As shown in Fig. 1, in this paper, we divide the input of the second-level and third-level local branches into eight blocks, with the object of further preventing the boundary effect caused by the block.(The pixel points on the image boundary do not shift, resulting in the offset of the block boundary of the final stitched displacement field being 0.) When we divide the input image into blocks, each image block takes out two more rows of redundant pixels, and finally, when we spliced the velocity field, each velocity field block takes two fewer rows of pixels in each dimension. The global branch focuses on high-level information, and the local branch focuses on the detailed features of each specific registration area. To ensure that the final global velocity field completely matches the local velocity field during information fusion, the network structure of our global and local registration branches is the same.

3.3 Feature self-calibration network

A general CNN constructs a feature extraction module through a convolutional layer, uses a convolution kernel to sum all input data channels to calculate an output feature map, and repeats the convolution operation to output the feature map uniformly.Feature extraction in this way causes the receptive field of the spatial position in the feature transformation process to overly depend on the size of the convolution kernel so that it is impossible to extract enough distinguishing features [35, 36]. The Laplacian pyramid has proved its advantages in constructing high-resolution solutions and training robustness in many image processing fields, but the traditional pyramid structure is only applied to the image layer, ignoring the feature layer. In order to solve the difficulties faced by traditional registration networks in feature extraction, we combine the concepts of image pyramid and grouping convolution to propose a feature self-calibration network. A feature pyramid is constructed at the feature layer to replace the operation of repeated convolution in the traditional network to eliminate the dilemma faced in the feature extraction stage. We show the module structure in Fig. 2. This module merges richer information by establishing a correlation between channels around each spatial location, thereby helping the feature extraction module to extract more expressive features.The generally grouped convolution completes the feature transformation process synchronously and independently through multiple parallel branches and then cascades the output of all branches to get the final output of the grouped convolution. We improved the grouped convolution, combined with the multi-resolution strategy of the image pyramid. Finally, we proposed to perform different feature transformations in multiple spaces of different scales to achieve the purpose of efficiently collecting context information of all spatial positions. We send the initially extracted feature X to five parallel branches. The spatial scales of the five branches are different, and each branch is equipped with a convolution filter of the corresponding size. As shown in Fig. 2, we first divide the primary feature X into two equal-sized middle layer features X_1, X_2 after a convolution operation, and their number of channels is half of X. X_2 performs convolution feature transformation at the original spatial scale to obtain X_2'. The specific operation is as Eq. 1.

$$X_2' = F(X_2). \tag{1}$$

where the F is the convolution operation.X_1 is sent to the remaining four branches, respectively, and through different degrees of down-sampling in the three branches, the input features are further extracted in a smaller latent space. Then, the extracted features are up-sampled back to the initial resolution, and finally, we complement the information to obtain X_1''. The specific operation is as Eq. 2.

$$\begin{aligned} X_1'' =& X_1 + Up(F(Avgpool(X_1)_r)) \\ &+ Up(F(Avgpool(X1)_{2r})). \end{aligned} \tag{2}$$

where the $Avgpool(\Delta)_r$ represents average pooling to down-sample the input features by r times.We select r = 2

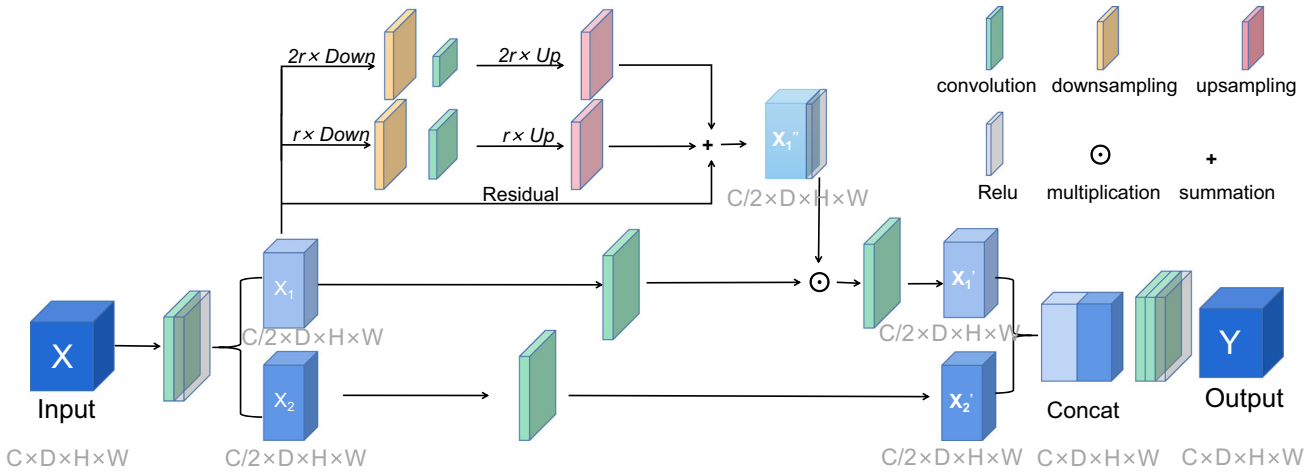


Fig. 2 The framework of feature self-calibration module. The feature self-calibration module includes the convolutional layer, the up-sampling layer, the down-sampling layer, and the activation layer, which is colored green, red, orange, and gray. The input feature X, the middle layer features X_1, X_2, $X1'$, X_1', X_2', and the shape of the output feature Y are marked below them, where C is the number of channels, D is the depth, H is the height, and W is the width

in this paper. $Up(\Delta)$ indicates that the input is up-sampled back to the initial resolution by bilinear interpolation, and X_1 is used as the correction weight in the form of residual. Finally,we perform the feature self-calibration process to obtain the output feature Y.

$$X_1' = F(Relu(X_1'') \times F(X_1)). \tag{3}$$

where the $Relu(\Delta)$ means using Relu to activate the input. The receptive field of the convolutional layer in the smaller latent space is large. Therefore, the X_1' obtained after feature conversion in the multi-scale space can be used as a reference to calibrate the feature extraction process of X_2 in the original feature space.

$$Y = Relu(F(F(Concat(X1', X2')))). \tag{4}$$

where the $Concat(\Delta)$ means that the input will be spliced on the channel. Compared with the original feature extraction module, the feature self-calibration module enables all spatial positions in the smaller latent space to adaptively consider the context around itself. However, it does not capture the global context information, which avoids the interference of too much irrelevant information when extracting feature information. Moreover, the convolutional layer in the feature self-calibration module can encode more significant and more delicate regions when extracting feature information [37].

The feature self-calibration network comprises a feature self-calibration module, feature encoder, residual module, and feature decoder. Figure 3 shows the specific structure. Each feature encoder contains two convolutional layers with different spans, the size of the convolution kernel is 3×3, and the convolution step size is 1 and 2, respectively.The feature decoder comprises a transposed

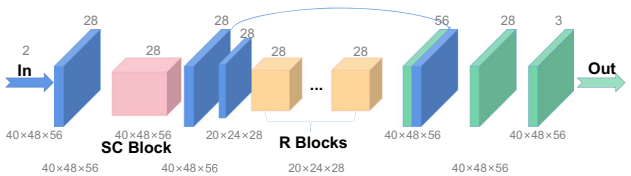


Fig. 3 The framework of feature self-calibration Network. The feature self-calibration network comprises a feature encoder, a feature decoder, a feature self-calibration module, and a residual module,-colored with blue, green, red, and orange, respectively. The blue line indicates a jump connection. Take the network of the first-level registration space as an example. The numbers of channels and sizes of the output by each module have been marked above and below, respectively

convolutional layer and two consecutive convolutional layers of equal size. We have added a jump connection part from the encoder output to the decoder input to prevent the disappearance of the underlying features in extracting features. The residual module contains ten identical convolutional layers, and each convolutional layer has a pre-activated structure, and we build a jump connection between every two convolutional layers. Figure 4 shows

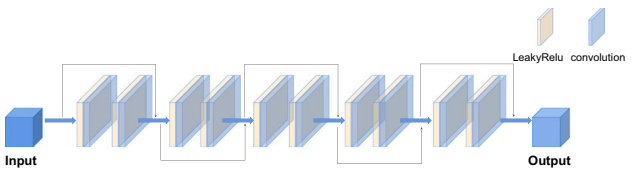


Fig. 4 The framework of the residual module. The residual module is composed of ten convolutional layers. Each convolutional layer has a pre-activated structure [38]. There is a jump connection between every two activated convolutional structures, and we use a black line to represent the operation. Figure 4 shows the specific components in detail

the specific structure.The network structure of FSCN on the global and local branches of each level of registration space is the same.

3.4 Velocity field superposition

When the original Laplacian pyramid network realizes the training process from coarse to fine, the velocity field generated in the low-resolution registration space is up-sampled step by step. Then, the velocity field generated in the current level of registration space is passed through a simple addition operation that realizes the supplement of velocity field information. However, a simple addition operation will result in the loss of a large amount of information when using the velocity field information generated by the low-resolution image, resulting in a waste of resources for the preliminary training work. Our network has chosen a more effective way to superimpose and fuse the velocity field information. Zhao et al. proposed a displacement field superposition formula in VTN. This formula can fuse multiple displacement field information.After using the displacement field to distort the moving image, we need interpolation to reconstruct the distorted image, and there will be a small amount of information loss in the process of interpolation and reconstruction. Therefore, the more cascaded registration networks, there will be more information loss eventually. The displacement field superposition formula reduces interpolation and reconstruction steps of the image between adjacent registration networks. When using the superposition formula to fuse the displacement field to distort the original moving image, only one interpolation reconstruction is required, thus reducing the information loss of the distorted image. Experiments have shown that it can achieve better effects than simple addition operations. We customize the FSCN output velocity field as a minimal displacement field. The pyramid network also has the characteristic of cascading network step-by-step registration, so it is also suitable for the displacement field superposition formula. We modified it to a form suitable for our network, as shown in Eq. 5.

$$V = V_2 \infty V_1 = V_2 + warp(V_1, V_2). \tag{5}$$

where V is the velocity field after superposition, V_1 is the velocity field output by the upper-level low-resolution space, and V_2 is the velocity field output by the registration space of this level. *Warp*(*image*, *flow*) is defined as the use of the flow field to distort the image and then completes the interpolation process.$\infty$is consistent with Fig. 1, indicating the information fusion operation of the velocity field.

4 Loss function

4.1 Similarity loss

We use the correlation coefficient as the similarity measure between the fixed map and the distorted moving image. The covariance between the two images (I_1, I_2) can be defined as:

$$\begin{aligned} Cov(I_1, I_2) = &\frac{1}{|\Omega|} * \sum_{x \in \Omega} I_1(x) I_2(x) \\ &- \frac{1}{|\Omega|^2} * \sum_{x \in \Omega} I_1(x) \sum_{x \in \Omega} I_2(x), \end{aligned} \tag{6}$$

where Ω represents the domain of all voxel points in the two images, and x represents each voxel point.We can calculate the covariance coefficient between images as the covariance between two vectors flattened from the images through element-wise operations. So the correlation coefficient can be defined as:

$$CorrCoef(I_1, I_2) = \frac{Cov(I_1, I_2)}{\sqrt{Cov(I_1, I_1) Cov(I_2, I_2)}}. \tag{7}$$

The correlation coefficient measures the degree of linear correlation between two images [32]. 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, the correlation coefficient should be non-negative.Therefore,we define the final correlation coefficient loss as:

$$L_{similarity} = 1 - CorrCoef(I_1, I_2). \tag{8}$$

4.2 Global regularization and local regularization

Most current learning-based methods apply a global regularization term to the deformation field to ensure its smoothness [29, 39, 40], such as the L2 norm on the spatial gradient. This paper also applies a regularization term to the output velocity field.As is shown in Eq. 9:

$$L_v = \frac{k}{2^{L-p}} \times \| \bigtriangledown v \|_2^2, \tag{9}$$

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

However, only applying a regularization term to the deformation field is not enough to ensure the preservation of the topological structure of the registered image. Therefore, on this basis, we additionally impose local orientation consistency constraints on the deformation field. Mathematically, the Jacobian determinant regularization loss function we chose is expressed as:

$$L_{jacc} = \frac{1}{N} \times \sum_{p \in \Omega} \sigma(-|J_\phi(p)|), \tag{10}$$

where N represents the number of all elements in the determinant.$\sigma(\bullet)$ is the activation function.In this paper, $\sigma(\bullet)$ means $max(0, \bullet)$, and it denotes the Jacobian matrix determinant at position p in the deformation field.Mathematically,its expression is as Eq. 11:

$$|J_\phi(p)| = \begin{Bmatrix} \frac{\partial \phi_x(p)}{\partial_x} & \frac{\partial \phi_x(p)}{\partial_y} & \frac{\partial \phi_x(p)}{\partial_z} \\ \frac{\partial \phi_y(p)}{\partial_x} & \frac{\partial \phi_y(p)}{\partial_y} & \frac{\partial \phi_y(p)}{\partial_z} \\ \frac{\partial \phi_z(p)}{\partial_x} & \frac{\partial \phi_z(p)}{\partial_y} & \frac{\partial \phi_z(p)}{\partial_z} \end{Bmatrix}. \tag{11}$$

The Jacobian determinant can analyze the local behavior of the deformation field. When the direction of the deformation field at point p reverses in its neighborhood, the Jacobian determinant at point p is negative. We use this feature to punish the local area with the negative Jacobian determinant to ensure the local direction consistency of the deformation field. The use of the global regularization term of the deformation field can further ensure the smoothness of the registration transformation process and the maintenance of the topological structure [41]. The final regularization term expression of the deformation field is:

$$L_{reg} = L_v + L_{jacc.} \tag{12}$$

5 Experiment

5.1 Data and preprocessing

We use 450 T1-weighted brain MR scans from the OASIS [42] dataset and 40 brain MR scans from LPBA40 [43] dataset to evaluate our method based on the brain atlas. Atlas-based registration is a standard application in multi-disciplinary image analysis.We can establish an anatomical correspondence can between the atlas and the moving image.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 dataset, including skull dissection, spatial normalization, and subcutaneous structure segmentation [44]. We use subcortical structure segmentation maps of 28 anatomical structures as the basis for our evaluation for the OASIS dataset.In the LPBA40 dataset, the MR scans in atlas space and its subcortical segmentation map of 56 structures, which experts manually delineate, are used in our experiments. To refine the registration area and improve the calculation efficiency, we reprocess all the preprocessed data images into $160 \times 192 \times 224$ sizes after cropping and scaling steps.To verify our network structure's robustness and universal applicability, we cross-validate the experimental data three times. Divide the dataset of OASIS into three equal batches. Each batch contains 150 brain images. Take one of the batches (150 images) as the test set in turn. We choose four from the test set as the atlas and use the rest of 146 pictures to align to the map in turn. We separate 250 images from the remaining two batches as the training set and 50 as the validation set. Besides, we use LPBA40 as an additional test set to test the registration method mentioned in the article across datasets. We use four arbitrarily selected maps from the LPBA40 dataset as the atlas and the remaining 36 as moving images. We use different registration methods to perform map-based registration to align the moving images in the test set to the fixed images in the map. For each registration method, $3 \times 4 \times 146 = 1752$ pairs of images were registered in the OASIS dataset, and $4 \times 36 = 144$ pairs of images were registered in the LPBA40 dataset.

5.2 Baseline method

We compare our method with SyN [3] and Elastic in traditional algorithms with the best registration performance. At the same time, we also selected the most advanced registration method based on unsupervised learning LapIRN [16] as our comparison object.We also select the plug-and-play initial registration network-*VoxelMorph* [29] and the latest conditional Laplacian Network- $CIR - DM$ [45] as the extra comparison object for the learning-based registration method. We have carefully adjusted the algorithm parameters for the two traditional registration algorithms to balance the registration effect and the registration time. We use their official default parameters for the three learning-based registration algorithms to train our data set from scratch.

5.3 Measurement

Several recent learning-based registration methods [15, 32] only use the dice score between the segmentation map of the warped moving image and the segmentation map of the fixed image to evaluate the performance of the registration method. However, using the dice score between the segmentation maps is not comprehensive as the only evaluation indicator. The maintenance of the topological structure of the medical image registration result and the reversibility of the displacement field predicted by the network also cannot be ignored for the image registration. Therefore, in addition to evaluating the accuracy of the registration result using the dice score between the

segmentation maps, we also use the percentage of voxels containing the non-positive Jacobian determinant ($|J_\phi|_{\leq 0}$) to evaluate the validity of the registration result (differential homeomorphism). In addition, we additionally use the volume change *TC* between the segmented images before and after the transformation to further evaluate the accuracy of the registration result. *TC* represents the topological changes of the anatomical structure in the image before and after registration. We express it by calculating the volume change of the corresponding registration area between the two images before and after registration. When $TC = 1$, it means that the image topology is maintained well during the registration process. When we cannot compare the registration performance of the two registration methods by dice and Jacobian scores, we use *TC* as an additional auxiliary evaluation index. *TC* verifies the maintenance of the topological structure, which is a secondary indicator of the Jacobian score ($|J_\phi|_{\leq 0}$).Intending to meet the timeliness of the registration method in clinical application, we also tested the average time required for each pair of image registration as a reference.

5.4 Implementation

Our proposed method($GoLo - FSCN$) is implemented based on PyTorch [46]. We use an Adam optimizer with a fixed learning rate (10^{-4}). To verify the effectiveness of our proposed global and local joint training strategy, the feature self-calibration network, and the new velocity field superposition method, we use the controlled variable method to compare the registration performance of the variant network with three innovation points added sequentially based on LapIRN. We train our network from scratch and select the model with the best performance on the validation set.

6 Results

6.1 Registration performance

Tables 1, 2 and 3 provide a comprehensive summary of the registration results. Among them, *Dice* represents the accuracy of the registration result (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 (the closer to 1, the better). *Time* represents the average registration time of each pair of images (in seconds). *Initial* means spatial normalization. The data in brackets are the standard deviations of each indicator.

Our method achieved high registration accuracy of 0.787, 0.788, and 0.792 in the three cross-tests on the OASIS data set. The registration accuracy has improved considerably compared with the traditional algorithm Syn, Elastic,and the three different learning-based methods.Compared with the current most advanced learning-based method-*LapIRN*, the folding percentage of non-zero Jacobian on the deformation field can reach 0.0017, 0.0043, and 0.0032 in our method, which has increased by 29%, 2%, and 6% compared with *LapIRN*, respectively.The anatomical structure change(TC) of $GoLo - FSCN$ is comparable to the anatomical structure change of *LapIRN*. The registration time of $GoLo - FSCN$ is slightly increased compared to *LapIRN*, but the increase is acceptable.Besides,we use the network model saved after training with the OASIS data set to test the corresponding cross-data set on the LPBA40 data set. It can be observed from the test results that our proposed network-$GoLo - FSCN$ also achieves the best registration performance.

Figure 5 randomly selects two pairs of registration images and uses different methods to register them to obtain the distorted image and mask. From the perspective of two-dimensional slices,observe the registration effect of each registration region on the two-dimensional plane. It can be observed from Fig. 5 that compared with the traditional method and *LapIRN*, the image registered by $GoLo - FSCN$ can achieve a better registration effect in large areas and small independent registration areas.

Figure 6 randomly takes two pairs of registered images and uses different methods to register them to obtain a distorted two-dimensional slice and a three-dimensional mask. From the overall registration effect of the brain from a three-dimensional perspective, we can observe that the overall shape of the distorted mask registered by $GoLo - FSCN$ is closer to the fixed atlas. Besides, the texture on the distorted mask is more precise and more detailed, and the texture is closer to that of the fixed atlas.

6.2 Contribution of velocity field superposition formula

Take the training data of batch3 as an example, using our velocity field superposition formula to optimize the velocity field of LapIRN, and then train the network model from scratch. Choose any four of the 150 test data as atlases and use the remaining 146 images to align to the four atlases, respectively. Figure 7 shows the dice score(*Dice*), Jacobian score($|J_\phi| \leq 0$), and topology change (*TC*) before and after registration,the average registration time of the four sets of atlas registration in the LapIRN

Table 1 Three cross-validation evaluations of ***OASIS*** (Taking batch1 of OASIS and LPBA40 as the test sets)

Method	Batch1 of OASIS				LPBA40			
	Dice	$\lvert J_\phi\rvert_{\leq 0}$	TC	Time(s)	Dice	$\lvert J_\phi\rvert_{\leq 0}$	TC	Time(s)
Initial	0.604 (0.020)	–	–	–	0.612 (0.011)	–	–	–
SyN	0.698 (0.015)	0.0000 (0.0000)	1.006 (0.011)	40.33	0.698 (0.015)	0.0000 (0.0000)	0.9985 (0.001)	29.98
Elastic	0.777 (0.018)	0.0028 (0.0011)	1.0104 (0.017)	56.47	0.699 (0.016)	0.0001 (0.0001)	0.9987 (0.001)	50.62
VoxelMorph	0.767 (0.005)	0.2044 (0.0012)	1.012 (0.012)	0.67	0.643 (0.013)	0.1957 (0.0022)	1.0014 (0.002)	0.89
LapIRN	0.783 (0.003)	0.0043 (0.0006)	1.008 (0.008)	0.88	0.699 (0.007)	0.0026 (0.0005)	0.9992 (0.008)	0.91
CIR-DM	0.761 (0.005)	0.0039 (0.0007)	0.995 (0.007)	0.87	0.691 (0.008)	0.0014 (0.0001)	1.0011 (0.001)	0.84
GoLo-FSCN (ours)	**0.787(0.003)**	**0.0017(0.0005)**	1.018 (0.016)	1.26	**0.696(0.009)**	**0.0008(0.0001)**	1.0002(0.001)	0.86

The experimental data measured by the registration method proposed in this paper.The overall performance of the boldface data is the best in the table

Table 2 Three cross-validation evaluations of ***OASIS***(Taking batch2 of OASIS and LPBA40 as the test sets)

Method	Batch2 of OASIS				LPBA40			
	Dice	$\lvert J_\phi\rvert_{\leq 0}$	TC	Time(s)	Dice	$\lvert J_\phi\rvert_{\leq 0}$	TC	Time(s)
Initial	0.603 (0.020)	–	–	–	0.612 (0.011)	–	–	–
SyN	0.758 (0.003)	0.0000 (0.0000)	1.005 (0.009)	40.37	0.698 (0.015)	0.0000 (0.0000)	0.9985 (0.001)	29.98
Elastic	0.779 (0.021)	0.0028 (0.0013)	1.010 (0.011)	56.51	0.699 (0.016)	0.0001 (0.0001)	0.9987 (0.001)	50.62
VoxelMorph	0.768 (0.010)	0.164 5(0.0028)	1.019 (0.013)	0.64	0.652 (0.013)	0.1537 (0.0039)	0.9989 (0.001)	0.91
LapIRN	0.780 (0.005)	0.0045 (0.0006)	1.014 (0.009)	0.85	0.700 (0.008)	0.0028 (0.0004)	1.0010 (0.001)	0.93
CIR-DM	0.763 (0.007)	0.0045 (0.0007)	1.0039 (0.007)	0.88	0.692 (0.009)	0.0015 (0.0001)	1.0013 (0.001)	0.83
GoLo-FSCN (ours)	**0.788 (0.005)**	**0.0043 (0.0005)**	1.033 (0.013)	1.17	**0.696 (0.009)**	**0.0007 (0.0001)**	1.0009 (0.001)	0.88

The experimental data measured by the registration method proposed in this paper.The overall performance of the boldface data is the best in the table

Table 3 Three cross-validation evaluations of ***OASIS***(Taking batch3 of OASIS and LPBA40 as the test sets)

Method	Batch3 of OASIS				LPBA40			
	Dice	$\lvert J_\phi\rvert_{\leq 0}$	TC	Time(s)	Dice	$\lvert J_\phi\rvert_{\leq 0}$	TC	Time(s)
Initial	0.611 (0.019)	–	–	–	0.612 (0.011)	–	–	–
SyN	0.761 (0.021)	0.0000 (0.0000)	1.006 (0.009)	42.05	0.698 (0.015)	0.0000 (0.0000)	0.9985 (0.001)	29.98
Elastic	0.782 (0.019)	0.0029 (0.0012)	1.010 (0.013)	56.43	0.699 (0.016)	0.0001 (0.0001)	0.9987 (0.001)	50.62
VoxelMorph	0.774 (0.007)	0.1935 (0.0021)	1.005 (0.011)	0.68	0.646 (0.012)	0.1797 (0.0035)	1.0010 (0.002)	0.89
LapIRN	0.783 (0.002)	0.0043 (0.0006)	1.008 (0.008)	0.88	0.700 (0.007)	0.0027 (0.0008)	1.0004 (0.001)	0 .91
CIR-DM	0.765 (0.005)	0.0045 (0.0007)	0.999 (0.007)	0.87	0.692 (0.008)	0.0017 (0.0001)	1.0011 (0.001)	0.85
GoLo-FSCN (ours)	**0.793 (0.002)**	**0.0032 (0.0005)**	1.012 (0.022)	1.16	**0.697 (0.008)**	**0.0008 (0.0001)**	1.0007 (0.001)	0.88

The experimental data measured by the registration method proposed in this paper.The overall performance of the boldface data is the best in the table

network and the optimized velocity field network. Figure 7 comprehensively compares the registration performance of the two networks, which directly highlights the ability of our velocity field superimposition method. In the map-based registration of four sets of different fixed maps, the dice score, *TC* and *Time* of the registered image after optimizing the flow field superimposition method are comparable to the LapIRN network, and the Jacobian score is further optimized, that is, while ensuring the registration accuracy, it further improves the differential homeomorphism characteristics of the registered image.

From the traditional LapIRN and the network that changes the speed field superposition mode, choose four pairs of flow fields for comparison.It can be observed from

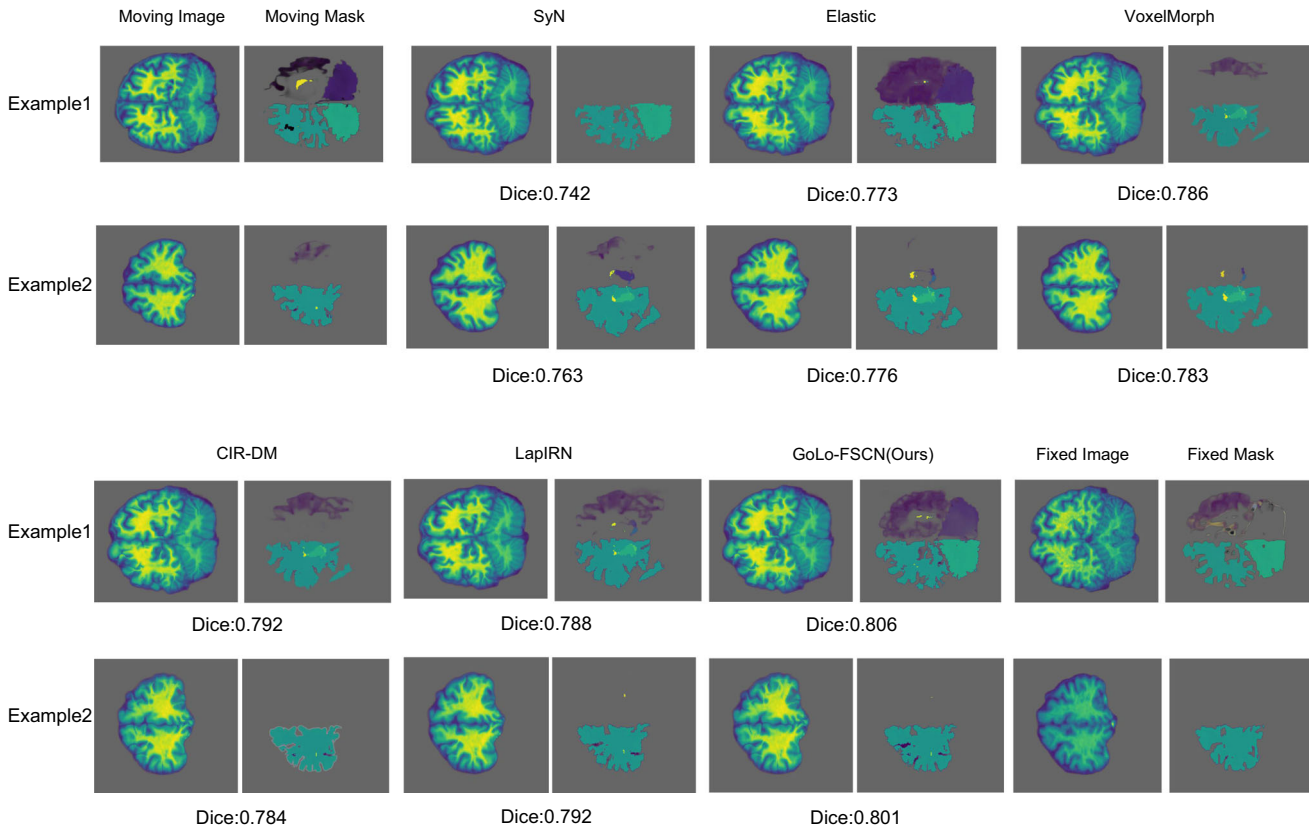


Fig. 5 2D visualization results of registration results

Fig. 8 that the LapIRN output flow field has a stronger sense of sawtooth, and the registration details are rough.However, the output flow field after the optimized velocity field superposition method is smooth,and the flow field at the details has not folded so that the registered image can maintain a better topology structure and have better reversibility.

6.3 Contribution of feature self-calibration network

Add the self-calibration module to the optimized flow field network and start training again from the beginning and then save the optimal model. Four sets of atlas registration are adopted and compared with the network registration effect after optimizing the flow field. It can be seen intuitively from Fig. 9 that the self-calibration module does not affect the effectiveness of the flow field superposition method and can further reduce the Jacobian score under the premise of ensuring the accuracy of the registration, that is, further optimizing the registration validity. Although the registration time of the four sets of atlas registration is increased compared with that without the self-calibration module added, the increased minimization time is acceptable compared with the improved performance.The three-dimensional mask image is convenient for us to observe the overall structure and surface texture of the registration area and intuitively feel the anatomical structure of the registration image to judge the effectiveness of the network output image.From the registration area in the red circle in the brain mask image in Fig. 10, it is evident that after adding the self-calibration module to the network, the distorted mask output by the network is closer to the fixed mask, and the registration performance of the network in the detail part has been further improved.It can be proved that the self-calibration module we proposed is feasible, and the further extraction and fusion of spatial feature information of different scales and the correction of original spatial feature information by low-dimensional spatial feature information can further optimize the validity of registration.

6.4 Contribution of global and local joint registration strategy

Figure 11 can intuitively see that after adding a global and local joint training strategy to the network,the registration effectiveness is improved to a certain extent. Besides, the registration accuracy of the network can be significantly improved. The segmentation area with a small area has

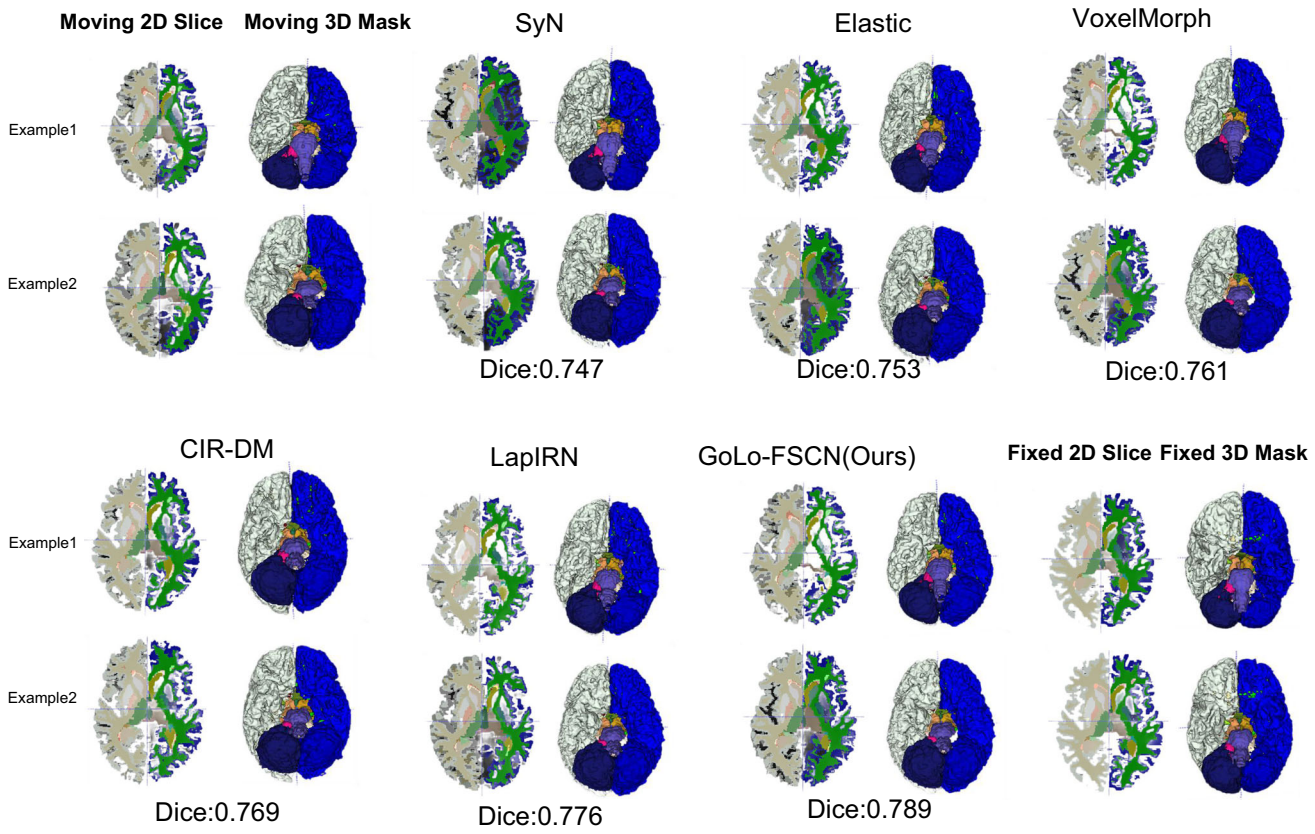
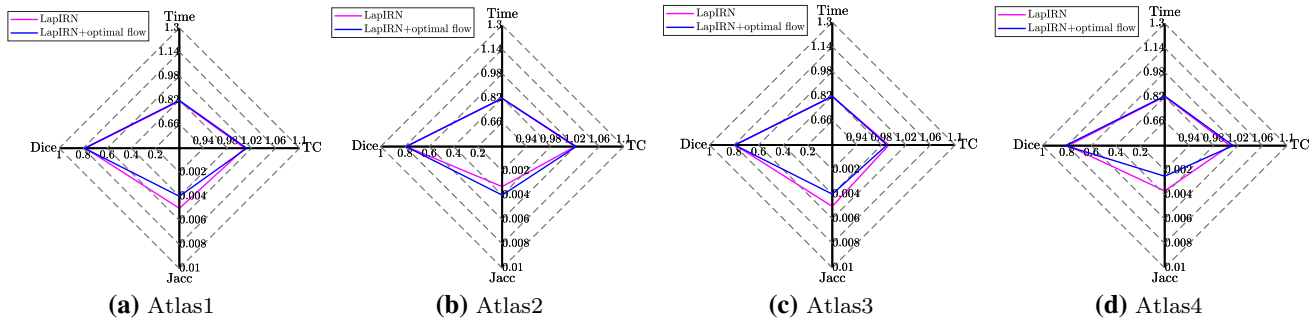


Fig. 6 3D visualization of the registration results



(a) Atlas1 **(b)** Atlas2 **(c)** Atlas3 **(d)** Atlas4

Fig. 7 Comparing the results based on the registration effect of the flow field optimization network and LapIRN. Choose four fixed atlases arbitrarily to align the moving image to the fixed atlas. It shows the performance comparison of the two networks in the registration of four sets of atlases. **a–d** are the configurations of the four different sets of fixed atlases. Among them, *Dice* indicates the accuracy of registration (the higher, the better), and *Jacc* indicates the effectiveness of registration ($|J_\phi|_{\leq 0}$) (the lower, the better). We use *TC* as an additional reference index for registration validity (The closer to 1, the better). *Time* is the average registration duration of a single pair of images (In seconds, the smaller, the better)

always been a problem in the registration field. Because the registration area is small, if the flow field is not smooth enough, these areas will easily disappear during the distortion process, leading to the destruction of the anatomical structure. Figure 12 comprehensively analyzes the registration performance of the global and local joint training strategy in different registration areas through the box plot.It can be found from Fig. 12 that compared with the network using the traditional training method, the network, after adopting the global-local joint training strategy, has better registration performance in the eight large-area registration regions in (*a*). Moreover, the registration effect of the eight small-area registration regions in (*b*) can also be significantly improved,proving the effectiveness of our

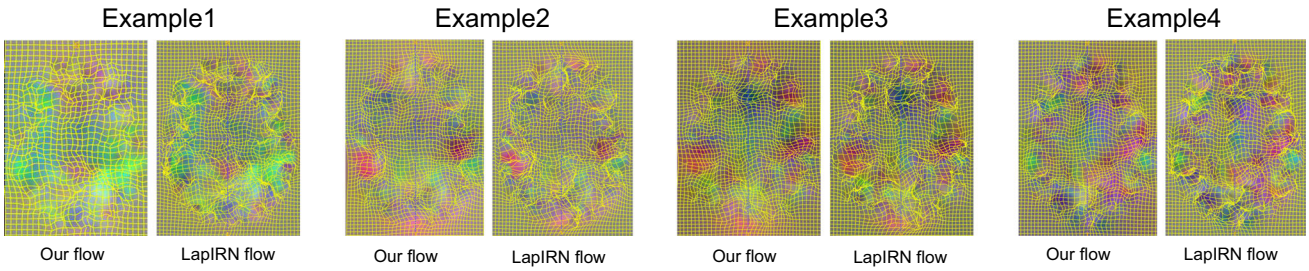
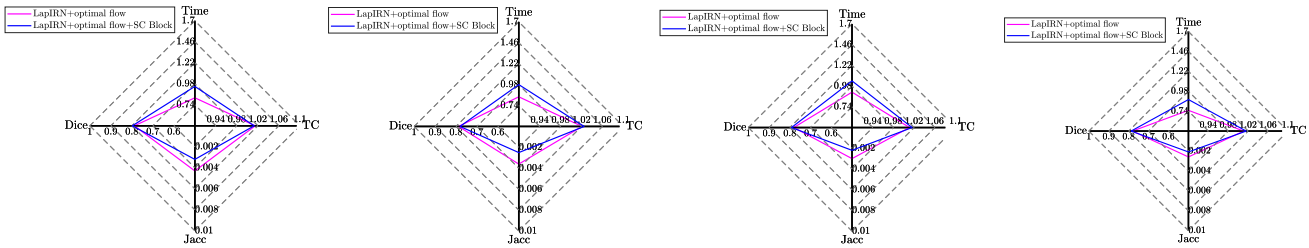


Fig. 8 The visualization of the flow field comparison. Choose any four sets of output flow fields from the four sets of atlas registration and compare the difference between the original network flow field and the flow field after modifying the superposition method of the velocity field



(a) Atlas1 **(b)** Atlas2 **(c)** Atlas3 **(d)** Atlas4

Fig. 9 Add a self-calibration module to the network after the flow field is optimized and compare its registration performance with the network that only optimizes the flow field. It shows the performance comparison of the two networks in the registration of four sets of atlases. **a**–**d** are the configurations of the four different sets of fixed atlases. Among them, *Dice* indicates the accuracy of registration (the higher, the better), and *Jacc* indicates the effectiveness of registration ($|J_\phi|_{\leq 0}$) (the lower,the better). We use *TC* as an additional reference index for registration validity (The closer to 1, the better). *Time* is the average registration duration of a single pair of images (In seconds, the smaller, the better)

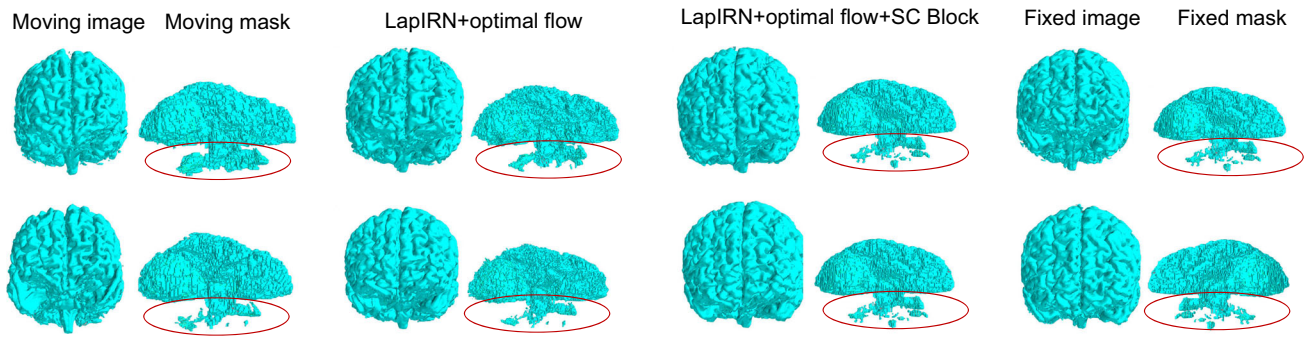
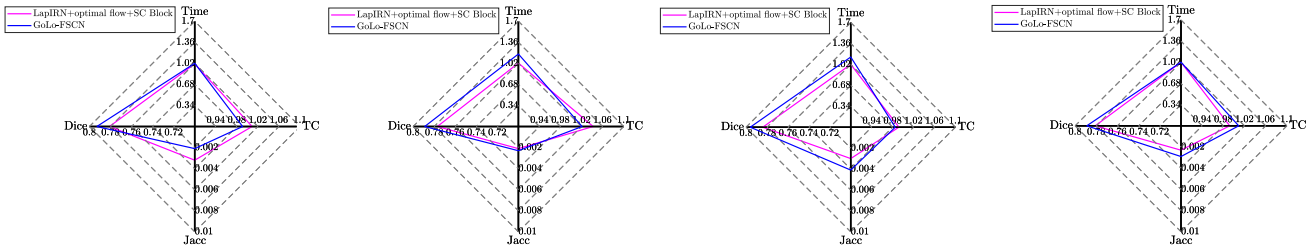


Fig. 10 Two groups are arbitrarily selected from the four groups of map-based registration to show the three-dimensional visualization effect of the two network registration performances.The first and second columns are three-dimensional moving image and moving mask,the third and fourth columns are the distorted image and masks output by the network after optimizing the flow field, and the fifth and sixth columns are the output of the distorted image and mask by the network after optimizing the flow field and adding the self-calibration module. The seventh and eighth columns are three-dimensional fixed image and fixed mask.The red circle encircles the detailed registration area of the mask

proposed global and local joint training strategy for medical images with complex registration multiple regions such as the brain. For medical images with multi-region registration, the global registration branch ensures the registration of the overall structure, and the local registration branch ensures the complex registration of the internal structure of each region. The combination of the two can achieve the most advanced performance in the current brain registration field.

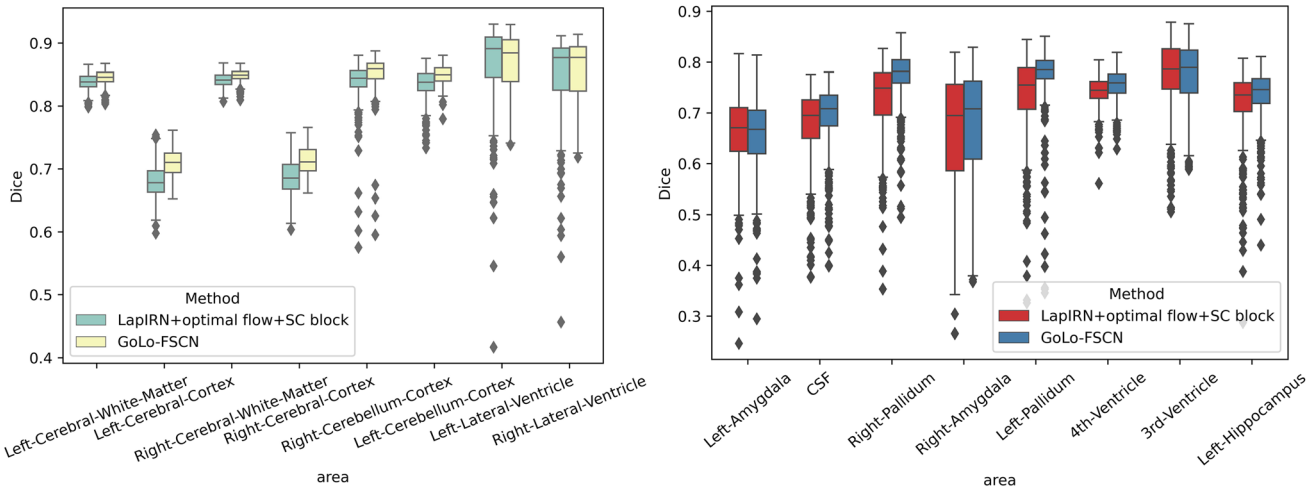
6.5 Optimization experiment of global and local joint training strategy

Regarding the local block content in the global and local joint training strategy, we initially adopted the standard blocks; we equally divided the original three-dimensional image into eight small blocks of the same size. We registered each block in turn, and the eight blocks were spliced finally after being registered to get a complete velocity



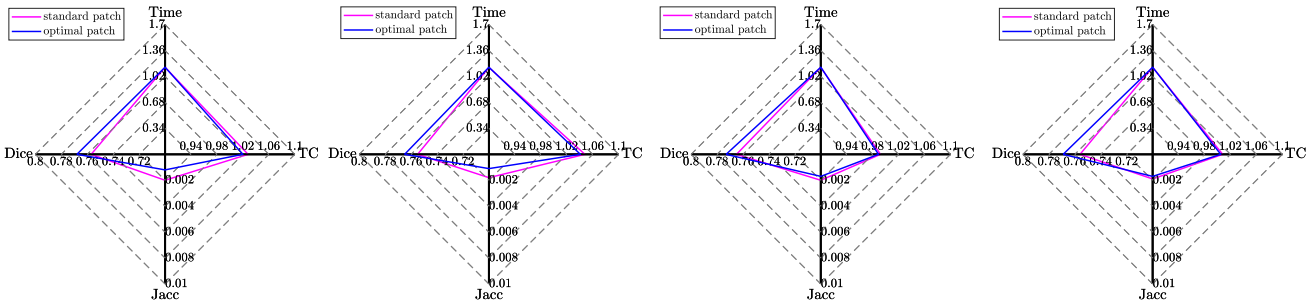
(a) Atlas1 **(b)** Atlas2 **(c)** Atlas3 **(d)** Atlas4

Fig. 11 Perform the same four sets of atlas registration on GoLo-FSCN, compare the registration performance with the network using traditional training methods and draw Fig. 11 to verify the ability of the global and local joint training strategy. Figure 11 shows the performance comparison of the two networks in the registration of four sets of atlases. **a**–**d** are the configurations of the four different sets of fixed atlases. Among them, *Dice* indicates the accuracy of registration(the higher,the better), and *Jacc* indicates the effectiveness of registration ($|J_\phi|_{\leq 0}$) (the lower,the better). We use *TC* as an additional reference index for registration validity (The closer to 1, the better). *Time* is the average registration duration of a single pair of images (In seconds, the smaller, the better)



(a) Big registration areas **(b)** Small registration areas

Fig. 12 The box plot analyzes the registration effect of each brain region by two networks in the four sets of atlas-based registration.We select the eight regions with the most extensive area and the eight regions with the smallest area from the 28 segmented regions of the brain.And then, we used box plots to compare the registration performance of the two networks in regions with different areas. **a** is a comparison of the registration performance of the two networks in eight large-area registration areas, and **b** is a comparison of the registration performance of the two networks in eight small-area registration areas



(a) Atlas1 **(b)** Atlas2 **(c)** Atlas3 **(d)** Atlas4

Fig. 13 Comparison of the effect of standard block network and redundant block network.It shows the performance comparison of the two networks in the registration of four sets of atlases. **a**–**d** are the configurations of the four different sets of fixed atlases. Among them, *Dice* indicates the accuracy of registration (the higher, the better), and *Jacc* indicates the effectiveness of registration ($|J_\phi|_{\leq 0}$) (the lower,the better). We use *TC* as an additional reference index for registration validity(The closer to 1, the better). *Time* is the average registration duration of a single pair of images (In seconds, the smaller, the better)

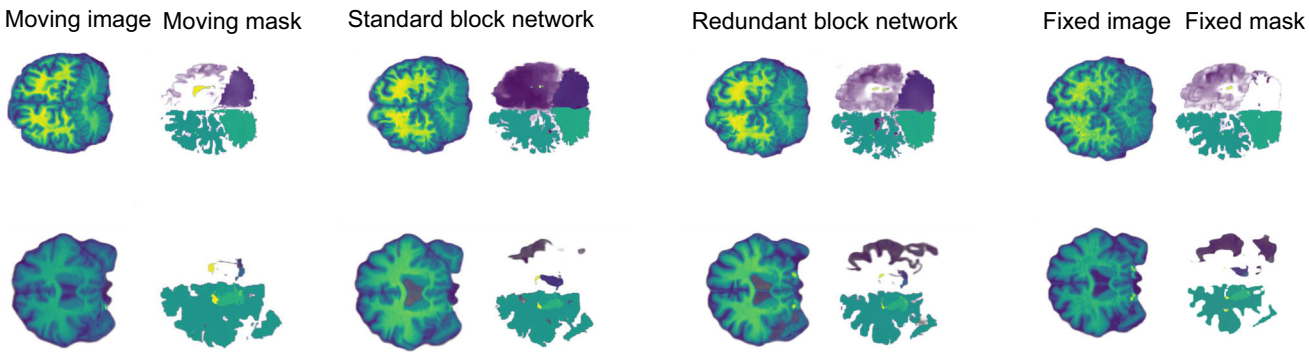


Fig. 14 Two groups are selected arbitrarily from the four groups of map-based registration to show the two-dimensional visualization effect of the two network registration performances.The first and second columns are two-dimensional moving images and moving masks, the third and fourth columns are the distorted images and masks output by the standard block network, and the fifth and sixth columns are distorted image and mask output by the redundant block network, and the seventh and eighth columns are two-dimensional fixed images and fixed masks

field. However, we found that if we equally divided the original image into eight small blocks during the registration process, each small block cannot produce the flow field at the edge, and the final spliced flow field is blank between the small blocks. The voxel point at the splicing location has no offset; that is, a boundary effect will occur.Given the defects of this standard block, we design a redundant block; when we divided the original image into blocks,each standard block is given two additional rows of voxel points in each dimension, and we cut off the extra rows voxels when splicing.The redundant block could eliminate the boundary effect caused by the standard block.We separately trained the standard and redundant blocks network from scratch to verify whether our thinking was correct and then performed four sets of atlas registration tests. We show our comparison results in Fig. 13.

As shown in Fig. 13, compared with the standard block network, the registration performance of the redundant block network has been significantly improved in accuracy and effectiveness, and the registration time is still comparable with that of the standard block network.The two-dimensional slice image helps us understand the internal structure of the registered image.From the slicing effect of the two sets of registered images shown in Fig. 14, we can observe that the distorted mask of the redundant block network is not only closer to the fixed mask in the small registration area at the center but also the registration in the external large-area registration area is also smoother, which can prove that our thinking is correct. It is feasible to use redundant blocks to eliminate the influence of boundary effects.

7 Conclusion

This paper proposes a novel global and local joint training strategy, features a self-calibration network for multi-region medical image registration, and combines the two to create "GoLo-FSCN" based on the Laplacian image pyramid. This method can ensure the differential homeomorphism of the registration result and further improve the registration accuracy. We use the large-scale public brain data set OASIS to evaluate our method and then use the three-time cross-validation method to test the universality and robustness of our network and compare our method with the classic traditional registration algorithm and the most advanced learning-based registration method.Comprehensive experimental results show that our method is better than other methods of topological structure preservation, deformation field quality of the registration results, and registration accuracy. Finally, although the registration time has increased slightly, the increase is almost negligible in clinical medicine, and the registration performance is acceptable compared with the improvement of our registration performance.

Acknowledgements This work was supported in part by the National Natural Science Foundation of China under Grant 61771322, Grant 61871186, Grant 61971290, and in part by the Fundamental Research Foundation of Shenzhen under Grant JCYJ20190808160815125.

Declarations

Conflict of interest We wish to draw the attention of the Editor to the following facts, which may be considered potential conflicts of interest, and to significant financial contributions to this work. We confirm that the manuscript has been read and approved by all named

authors and that there are no other persons who satisfied the criteria for authorship but are not listed. We further confirm that all have approved the order of authors listed in the manuscript. We confirm that we have given due consideration to the protection of intellectual property associated with this work and that there are no impediments to publication, including the timing of publication, concerning intellectual property. In so doing, we confirm that we have followed the regulations of our institutions concerning intellectual property. We understand that the Corresponding Author is the sole contact for the Editorial process (including Editorial Manager and direct communications with the office). He is responsible for communicating with the other authors about progress, submissions of revisions, and final approval of proofs. We confirm that we have provided a current, correct email address accessible by the Corresponding Author.

References

1. Ashburner J (2007) A fast diffeomorphic image registration algorithm. Neuroimage 38(1):95–113
2. Ashburner J, Friston KJ (2000) Voxel-based morphometry-the methods. Neuroimage 11(6):805–821
3. Brian BA, Charles LE, Murray G, James C (2008) Gee. Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. Med image Anal, 12(1):26–41,
4. Huang X, Paragios N, Metaxas DN (2006) Shape registration in implicit spaces using information theory and free form deformations. IEEE Trans Pattern Anal Mach Intell 28(8):1303–1318
5. Avants BB, Tustison N, Song G et al (2009) Advanced normalization tools (ants). Insight J 2(365):1–35
6. Klein S, Staring M, Murphy K, Viergever MA, Pluim JPW (2009) Elastix: a toolbox for intensity-based medical image registration. IEEE Trans Med Imag 29(1):196–205
7. Thirion J-P (1998) Image matching as a diffusion process: an analogy with maxwell's demons. Med Image Anal 2(3):243–260
8. Vercauteren T, Pennec X, Perchant A, Ayache N (2009) Diffeomorphic demons: efficient non-parametric image registration. Neuroimage 45(1):S61–S72
9. Jaderberg M, Simonyan K, Zisserman A et al (2015) Spatial transformer networks. Adv Neural Inf Process Syst 28:2017–2025
10. Cullen H, Krishnan ML, Selzam S, Ball G, Visconti A, Saxena A, Counsell SJ, Hajnal J, Breen G, Plomin R et al (2019) Polygenic risk for neuropsychiatric disease and vulnerability to abnormal deep grey matter development. Sci Rep 9(1):1–8
11. Iqbal K, Flory M, Khatoon S, Soininen H, Pirttila T, Lehtovirta M, Alafuzoff I, Blennow K, Andreasen N, Vanmechelen E et al (2005) Subgroups of alzheimer's disease based on cerebrospinal fluid molecular markers. Ann Neurol: official J Am Neurol Assoc Child Neurol Soc 58(5):748–757
12. Ross CA, Margolis RL, Reading SAJ, Pletnikov M, Joseph T (2006) Coyle. Neurobiol Schizophrenia. Neuron 52(1):139–153
13. Glasser MF, Coalson TS, Robinson EC, Hacker CD, Harwell J, Yacoub E, Ugurbil K, Andersson J, Beckmann CF, Jenkinson M et al (2016) A multi-modal parcellation of human cerebral cortex. Nature 536(7615):171–178
14. Adrian VD, Marianne R, John G, Mert RS (2019) Learning conditional deformable templates with convolutional networks. arXiv preprint arXiv:1908.02738
15. Shengyu Z, Yue D, Eric IC, Yan X et al (2019) Recursive cascaded networks for unsupervised medical image registration. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 10600–10610,
16. Mok Tony CW, Albert CSC (2020) Large deformation diffeomorphic image registration with laplacian pyramid networks. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 211–221. Springer
17. Jan M(2009) FAIR: flexible algorithms for image registration. SIAM,
18. Bajcsy R, Kovačič S (1989) Multiresolution elastic matching. Comput Vis, Graphics, Image Process 46(1):1–21
19. Faisal Beg M, Miller MI, Trouvé A, Younes L (2005) Computing large deformation metric mappings via geodesic flows of diffeomorphisms. Int J Comput Vision 61(2):139–157
20. Adrian VD, Andreea B, Natalia SR, Polina G (2016) Patch-based discrete registration of clinical brain images. In: International Workshop on Patch-based Techniques in Medical Imaging, pp. 60–67. Springer,
21. Glocker B, Komodakis N, Tziritas G, Navab N, Paragios N (2008) Dense image registration through mrfs and efficient linear programming. Med Image Anal 12(6):731–741
22. Thomas Yeo BT, Sabuncu MR, Vercauteren T, Holt DJ, Amunts K, Zilles K, Golland P, Fischl B (2010) Learning task-optimal registration cost functions for localizing cytoarchitecture and function in the cerebral cortex. IEEE Trans Med Imag 29(7):1424–1441
23. Xiaohuan C, Jianhua Y, Jun Z, Dong N, Minjeong K, Qian W, Dinggang S (2017) Deformable image registration based on similarity-steered cnn regression. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 300–308. Springer
24. Julian K, Tommaso M, Hervé D, Li Z, Florin CG, Shun M, Andreas KM, Nicholas A, Rui L, Ali K(2017) Robust non-rigid registration through agent-based action learning. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 344–352. Springer,
25. Marc-Michel R, Manasi D, Tobias H, Maxime S, Xavier P (2017) Svf-net: Learning deformable image registration using shape matching. In: International conference on medical image computing and computer-assisted intervention, pp. 266–274. Springer,
26. Hessam S, Bob De V, Floris B, Boudewijn PF, Lelieveldt I, Marius S (2017) Nonrigid image registration using multi-scale 3d convolutional neural networks. In: International conference on medical image computing and computer-assisted intervention, pp. 232–239. Springer,
27. Yang X, Kwitt R, Styner M, Niethammer M (2017) Quicksilver: Fast predictive image registration-a deep learning approach. Neuroimage 158:378–396
28. Mansilla L, Milone DH, Ferrante E (2020) Learning deformable registration of medical images with anatomical constraints. Neural Netw 124:269–279
29. Guha B, Amy Z, Mert RS, John G, Adrian VD (2018) An unsupervised learning model for deformable medical image registration. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 9252–9260,
30. Adrian VD, Guha B, John G, Mert RS (2018) Unsupervised learning for fast probabilistic diffeomorphic registration. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 29–738. Springer,
31. Vincent A, Olivier C, Xavier P, Nicholas A (2006) A log-euclidean framework for statistics on diffeomorphisms. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 924–931. Springer,
32. Zhao S, Lau T, Ji L, Eric Chao C, Yan X (2019) Unsupervised 3d end-to-end medical image registration with volume tweening network. IEEE journal of biomedical and health informatics. 24(5):1394–1404

33. Tero K, Timo A, Samuli L, Jaakko L (2017) Progressive growing of gans for improved quality, stability, and variation. arXiv preprint arXiv:1710.10196
34. Ting-Chun W, Ming-Yu L, Jun-Yan Z, Andrew T, Jan K, Bryan C (2018) High-resolution image synthesis and semantic manipulation with conditional gans. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 8798–8807
35. Hengshuang Z, Jianping S, Xiaojuan Q, Xiaogang W, Jiaya J (2017) Pyramid scene parsing network. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2881–2890
36. Bolei Z, Aditya K, Agata L, Aude O, Antonio T (2016) Learning deep features for discriminative localization. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2921–2929
37. Jiang-Jiang L, Qibin Hou, Ming-Ming C, Changhu W, Jiashi F (2020) Improving convolutional networks with self-calibrated convolutions. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 10096–10105,
38. Kaiming H, Xiangyu Z, Shaoqing R, Jian S (2016) Identity mappings in deep residual networks. In: European conference on computer vision, pp. 630–645. Springer
39. de Vos BD, Berendsen FF, Viergever MA, Sokooti H, Staring M, Išgum I (2019) A deep learning framework for unsupervised affine and deformable image registration. Med Image Anal 52:128–143
40. Boah K, Jieun K, June-Goo L, Dong Hwan K, Seong Ho P, Jong Chul Y (2019)Unsupervised deformable image registration using cycle-consistent cnn. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 166–174. Springer,
41. Mok Tony CW, Albert C (2020) Fast symmetric diffeomorphic image registration with convolutional neural networks. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 4644–4653,
42. Marcus DS, Wang TH, Parker J, Csernansky JG, Morris JC, Buckner RL (2007) Open access series of imaging studies (oasis): cross-sectional mri data in young, middle aged, nondemented, and demented older adults. J Cogn Neurosci 19(9):1498–1507
43. Shattuck DW, Mirza M, Adisetiyo V, Hojatkashani C, Salamon G, Narr KL, Poldrack RA, Bilder RM, Toga AW (2008) Construction of a 3d probabilistic atlas of human cortical structures. Neuroimage 39(3):1064–1080
44. Fischl B (2012) Freesurfer. Neuroimage 62(2):774–781
45. Mok Tony CW, Albert C (2021) Conditional deformable image registration with convolutional neural network. pp. 35–45
46. Paszke A, Gross S, Chintala S, Chanan G, Yang E, DeVito Z (2017) Zeming Lin. Luca Antiga, and Adam Lerer. Automatic differentiation in pytorch, Alban Desmaison