Neural Computing and Applications (2021) 33:16471–16487
https://doi.org/10.1007/s00521-021-06243-9

ORIGINAL ARTICLE

A cascaded registration network RCINet with segmentation mask

Wenlan Zou[1] · Yi Luo[1] · Wenming Cao[1,2] · Zhiquan He[1] · Zhihai He[2]

Received: 7 December 2020 / Accepted: 17 June 2021 / Published online: 7 July 2021

Abstract

Traditional deformable registration methods achieve brilliant results and show strong theoretical support but are computational intensive since they optimize each image pair's objective function. Recently, supervised learning methods have facilitated fast registration. However, it requires ground truth and does not guarantee a diffeomorphism registration. This paper proposes a new unsupervised learning method Recursive Cascaded Network with a segmentation mask, for two-dimensional medical image registration. Different from the original cascaded network, the network framework into two parts. The first section obtains a pair of image rolls and uses the registration sub-network to predict the deformation vector field from the moving image to the fixed image. The second part introduces anatomical segmentation into the network during training, makes full use of the auxiliary information of the volume, adds an autoencoder to encode the anatomical segmentation, and incorporates it into the learning process of the model in the form of constraints. The local and global ideas are combined to ensure the deformation field's rationality and improve the distribution. The most important thing is that we propose a formula for calculating the cascaded network's deformation field used in the test stage to evaluate the relationship between the registration accuracy and the deformation field's effectiveness. Our experiments show that the system has a better registration effect and less information loss than the current state-of-the-art method. Simultaneously, the cascade method's accuracy is an improvement at a certain number of layers, and the increase in accuracy needs to sacrifice the effectiveness of the deformation field.

Keywords Image registration · Unsupervised learning · Segmentation mask · Cascaded network

1 Introduction

Deformation image registration is a fundamental task in many medical image processing and plays a critical role in medical image analysis. The process includes establishing accurate anatomical correspondences between the reference image and the floating image. The traditional registration approach solves the optimization problem by minimizing an objective function based on a predefined similarity metric. Unfortunately, solving pair optimization can be time-consuming, therefore, those methods are generally prolonged in practice [1].

With computer performance development, deep learning methods have drawn growing interest in terms of image registration. Most of these deep learning methods take a pair of images or their image patches as input and use convolutional neural networks to build prediction models of spatial transformation for image registration. Although these methods can quickly and accurately achieve image registration results, they require supervised information to

✉ Wenming Cao
wmcao@szu.edu.cn

Wenlan Zou
1900432054@email.szu.edu.cn

Yi Luo
1910434004@email.szu.edu.cn

Zhiquan He
zhiquan@szu.edu.cn

Zhihai He
hezhi@missouri.edu

[1] Guangdong Multimedia Information Service Engineering Technology Research Center, Shenzhen University, Shenzhen 518000, China

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

train the network module, such as dense displacement vector field (DVF) [2, 3], reference labels of segmentation [4]. However, accurate ground-truth is challenging to obtain, and the transformer does not guarantee the actual differential state properties.

Several recent works have studied the application of unsupervised learning to image registration to address the above problem. Bob et al. [5] developed a 2D end-to-end unsupervised registration network. Balakrishnan et al. [6] proposed VoxelMorph containing the spatial transformer network (STN), Jun et al. [7] implemented an Inverse-Consistent Deep Network combining an inverse-constraint and anti-folding constraint to prevent the deformation field folding. Both of them adopted regularization techniques to obtain spatially smooth and physically plausible transformations. However, smooth deformation fields do not guarantee realistic results: in some cases, the shaper deformation in the organ boundaries must preserve the anatomy.

In this paper, we propose a new recursive cascaded unsupervised registration network (RCINet), which learns from the limited recursive cascaded basic network established on the network to solve the above problems obtain the final deformation field. Specifically, we randomly extract a pair of images and their corresponding segmentation labels from the data set and use them to input the network's upper and lower parts. Both pairs of image volumes are recursively performed through the corresponding deformation field generated by each layer of the registration sub-network. Deformation, learn the network, and finally obtain a deformation field with sufficiently high registration accuracy and anatomical rationality to complete the registration purpose.

Unlike the previous work: (1) Since the parallel convolution kernel can extract and fuse information from images of different scales to better characterize these images and segmentation information, we have introduced inception module in the autoencoder (AED) and registration subnet. In the third part of the experiment, we can see that the primary network model can be implemented more than existing advanced methods with higher registration accuracy. (2) We introduce the local and global information of anatomical segmentation into the cascaded multilayer basic network training. The final deformation field can be decomposed into subtle and direct gradual changes, thereby reducing the difficulty of subnet learning, and at the same time, improving the unfolding of the deformation field. In the fourth part of the experimental chapter, as the cascade network gradually deepens, the unfolding of the deformation field is improving. (3) This paper proposes a formula for calculating the multilayer cascade deformation field. Two different test methods of config_1 verify the accuracy of the procedure. At the same time, comparing the two different configurations of the model config_1 and config_2, it can be seen that the network improves the registration accuracy and the folding voxel ratio at the cost of increasing the depth and reduces the reversibility and the registration speed.

The remaining sections are organized as follows: In Sect. 2, we give a related work on our methods. In Sect. 3, we present the formula for calculating the multilayer cascade deformation field and our new cascade registration framework. In Sect. 4, we conducted a six-part experiment. First, we evaluated the performance of the proposed network. Second, we observed the registration effect of cionfig_1 in the cascade 1 to 10. Third, the registration effect of config_1 and config_2 is compared with the effectiveness of the deformation field, and corresponding conclusions are drawn. Fourth, we conducted a detailed parameter study. In addition, we verified the accuracy of the multilayer mentioned above cascade formula. Finally, some failure case analysis is given to provide a reference for future work. In Sect. 5, we discuss the advantages and disadvantages of the proposed method. In Sect 6, we provide a brief conclusion.

2 Related work

2.1 Non-learning-based deformable image registration

There is a lot of work to be done in medical image registration. Traditionally, deformable registration is constructed by maximizing the similarity metric used to measure the source image's distance and the target image. Tools like ANTs [8] (affine and SyN for deformation) elastic [9] (affine and b-spline for deformation) have been proposed for image registration. Usually, the metric includes two parts: (1) the objective function measures the corresponding similarity between the input images. (2) the regularization term ensures the smoothness of the warped image.

Diffeomorphic transforms are anti-folded and reversible in image transformation [10, 11]. It is essential to choose a suitable differential symmetric registration algorithm [12]. Commonly used heteromorphic parameterization methods include distance metric mapping (LDDMM) [13–17], DARTEL [18], diffeomorphic demons [19] and standard symmetric normalization (SyN) [20]. The main problem faced by traditional registration methods is: for each image to be registered, iteratively optimizes the cost function from scratch, which severely limits registration speed and ignores the inherent registration mode shared between the same data [21]. It takes a long time to update each image's

parameters during the optimization process, making such methods less used in clinical applications.

2.2 Deformable image registration based on deep learning

With the development of deep learning, more and more researchers apply it to computer vision tasks, such as image recongnition, classification, and segmentation. In recent years, people have proposed image registration methods using deep neural networks, most of which rely on the ground truth field obtained through image pre-alignment, simulation transformation, or classic scan alignment. Some registrations use image similarity to help guide. Although supervised learning methods have achieved good performance, the great demand for ground truth alignment or synthetic data is a big problem. In contrast, RCINet is unsupervised and only utilizes segmentation mask information during training.

In earlier work, accurate and fast unsupervised learning-based image registration methods were the first to present in [5, 22]. They are all constructed by a convolutional neural network (CNN) and spatial transformation function [23], warping images mutually. However, these methods are mainly used on limited subsets of volumes, such as 3D sub-regions [22] or 2D slices [5], and only support small region [5] conversion, so they are more limited in data usage.

Some methods have recently proposed a segmentation driven cost function to register different image modalities, T1w, T2w MRI, and its slices and 3D ultrasound. There are [24, 25, 25], through the work of the authors, we can know that the loss function generated by the segmentation mapping can achieve more accurate cross-modal registration. Similar to this, we introduced two loss functions formed by the segmentation label in the registration work. One is the loss term of the floating mask of the anatomical segmentation and the mask's softmax function after the final deformation (ϕ_n). The second term is the Euclidean distance formed by encoding the two segmented masks by the ADE in the previous loss function. Besides, the advent of Google's inception makes everyone realize that improving convolutional neural networks' performance does not only depend on deepening the network's depth and width. Since the inception module connects multiple convolutions in parallel, it increases the network's adaptability to scales. We introduce the inception module into the registration network and the ADE so that the input image can be better extracted and fused at different scales, thereby improving the registration effect.

2.3 Medical image registration based on cascade thought

In the work of [26–29], the idea of a cascading network has appeared. In [28], a cascading recursive method is proposed for registration. Each sub-network input is the warped image. A fixed image of the previous network input of each sub-network performs a backpropagation to update all sub-networks parameters before it. In [26], the necessary modules in VTN and Voxelmorph are recursive cascaded. Unlike VTN, which performs registration in each sub-network and calculates each layer's similarity loss separately, its optimization goal in the last two images' similarity function. It gives all the sub-networks jointly learn advanced registration no backpropagation. The cascading idea proposed in [29] is to different cascade sub-networks to transform the image from global to local, while in [4], the idea of cyclic cascading appeared. After cascading once, the previous works' difference is that the deformed picture and the original floating image severe as the next cascade's input.

The above unsupervised cascading methods either directly make the deformation field smooth by using local smoothing or design the network structure to ensure the deformation field's smoothness. However, there is no standard to measure the deformation when designing the network framework differential homeomorphism of the posterior image. Especially for the [26] cascade method, although it has a witness that as the number of cascades increases, the accuracy of model registration is improved, the evaluation of image registration performance by the number of cascade layers is lacking. This paper proposes a related formula to evaluate the foldability of the deformation field in the registration task and observe the relationship between the image's registration accuracy under different layers and the Jacobian value experiments' change.

The main contributions of this paper are (1) adding different inception module blocks to the coding stage of the registration network and ADE, respectively. Extracting the information of different scales of the image in parallel through multiple convolution kernels, and performing information fusion to get the image more good representation, thereby improving the registration effect. (2) Using an unsupervised registration method, adding auxiliary information to the idea of the cascade, the pixel-level loss function L_{sc} and the low-dimensional representation encoded by the novel ICAE. The combination of the global information loss function L_{ac} makes it anatomically reasonable to retain the critical information of anatomical characteristics and deformation for anatomical segmentation. (3) Introduce the cascade idea into the registration

network framework and propose a computational cascade final deformation field formula of the network, combined with the Jacobian, to evaluate the degree of folding of the deformation field. At the same time, it can conclude that, in some cases, it is not that the more cascaded layers, the better, but the rationality of the deformation field structure.

3 Methods

Given a fixed image pair I_{f} and moving image I_{m}, they are both defined in the D-dimensional space Ω. Image registration's primary purpose is to construct a flow field prediction function G, taking I_{f} and I_{m} as inputs to predict a deformation field ϕ that aligns the two images: $\Omega \rightarrow \Omega$. In this work, we mainly discuss the registration task of $D = 2$.

For deformable image registration, an appropriate flow field should continuously transform and prevent folding. We recursively register warped images to achieve cascade processing. The deformed image $I_{\mathrm{m}}^{'}$ is the flow field and moving image composition, namely:

$$I_{\mathrm{m}}^{'} = I_{\mathrm{m}} \circ \phi \tag{1}$$

Theoretically:

$$G(I_{\mathrm{m}}, I_{\mathrm{f}}) = G_1(I_{\mathrm{m}} \circ \phi, I_{\mathrm{f}}) \circ \phi \tag{2}$$

where $\circ$ represents deforming the image with a deformation field, after this recursion, the moving image will be continuously warped so that the final prediction is divided into cascaded progressive refinement. A cascade is a flow field prediction function $g_{(\mathrm{f})}$, and the kth cascade predicts a flow field:

$$\phi_k = g_k(I_{\mathrm{m}}^{k-1}, I_{\mathrm{f}}) \tag{3}$$

Among them, I_{m}^{k-1} represents the first $k-1$ cascade warped moving images. Figure 1 describes the proposed framework. Assuming there are n cascades in total, there are n cascades in the training part, and the final output field is composed of all prediction fields, namely:

$$G(I_{\mathrm{m}}, I_{\mathrm{f}}) = \phi_n \circ \cdots \phi_1 \tag{4}$$

The formula obtains the final image:

$$I_{\mathrm{m}}^{n} = I_{\mathrm{m}} \circ G(I_{\mathrm{m}}, I_{\mathrm{f}}) = I_{\mathrm{m}} \circ \phi_n \circ \cdots \circ \phi_1 \tag{5}$$

3.1 Unsupervised image registration network

In our registration network RCINet, our impressive network is to learn part of the measurement or perform a particular type of calibration by cascading subnetworks. Unlike the previous cascading idea, we add the image's anatomical segmentation information to the registration network. The image registration realized by the network has more substantial anatomical rationality and a better registration effect.

Figure 1 is the registration diagram of our network framework. It consists of several cascaded registration subnetworks. After that, the moving image of each layer will be deformed. Four loss functions guide the unsupervised training of unsupervised network parameters. The deformation operation is performed in the sampler (also called the deformation field). We use the prediction field phi_1 to transform the previous deformed image moved $(M \circ \phi_{n-1})$ to get the next deformed image moved $(M \circ \phi_n)$ and also use the upper half of the deformation field $\phi_1, \phi_2 \cdots \phi_n$ is transformed. In addition to the general similarity loss function and deformation field constraint items for image registration. The composition of the loss function has a softmax cross-entropy loss function composed of final deformation segmentation and source segmentation. And the low-dimensionality obtained after the two are put into the encoder for encoding represents the composition of the Euclidean distance function.

The network framework is mainly composed of two parts. The first part is a network framework for image deformation. The framework is based on an UNet-like convolutional neural network (called ICUNet in Fig. 1), as shown in Fig. 2. Given a pair of sources for image F and target image M, ICUNet predicts the deformation field ϕ. After generating the corresponding displacement vector, it is input into a sampler similar to a space transformer, and the image is transformed by bilinear interpolation.

Figure 2 describes the CNN structure ICUNet we use to generate the prediction field. The network connects M and F into a 2-channel image as a single input. In our experiment, the input size is $32 \times 64 \times 64 \times 2$, but this frame does not limit the input image's size. In the encoding stage, we first use two 3×3 convolution kernels to perform convolution operations, use the activation function formed by the combination of Relu and Sigmoid to perform nonlinear transformations on the convolution output. Then, introduce it into inception module 1 (The basic structure of inception module 1 has four parts, the size is $1 \times 1, 3 \times 3, 5 \times 5$ convolution, and 3×3 maximum pooling layer). In the convolution operation, according to this, for four alternating operations, the difference is the size of the number of convolution channels, where *ap* represents the size of the average pooling, *dp* represents the size of the dropout. In the decoding stage, we alternate upsampling and connection jump connections. These jump connections will directly propagate the features learned in the encoding stage to the registration layer. The connection layer of the decoder operates on a finer spatial size. To achieve accurate registration, the second part uses the segmentation mask of the image as auxiliary information. It uses the displacement vector in the first part to transform the target

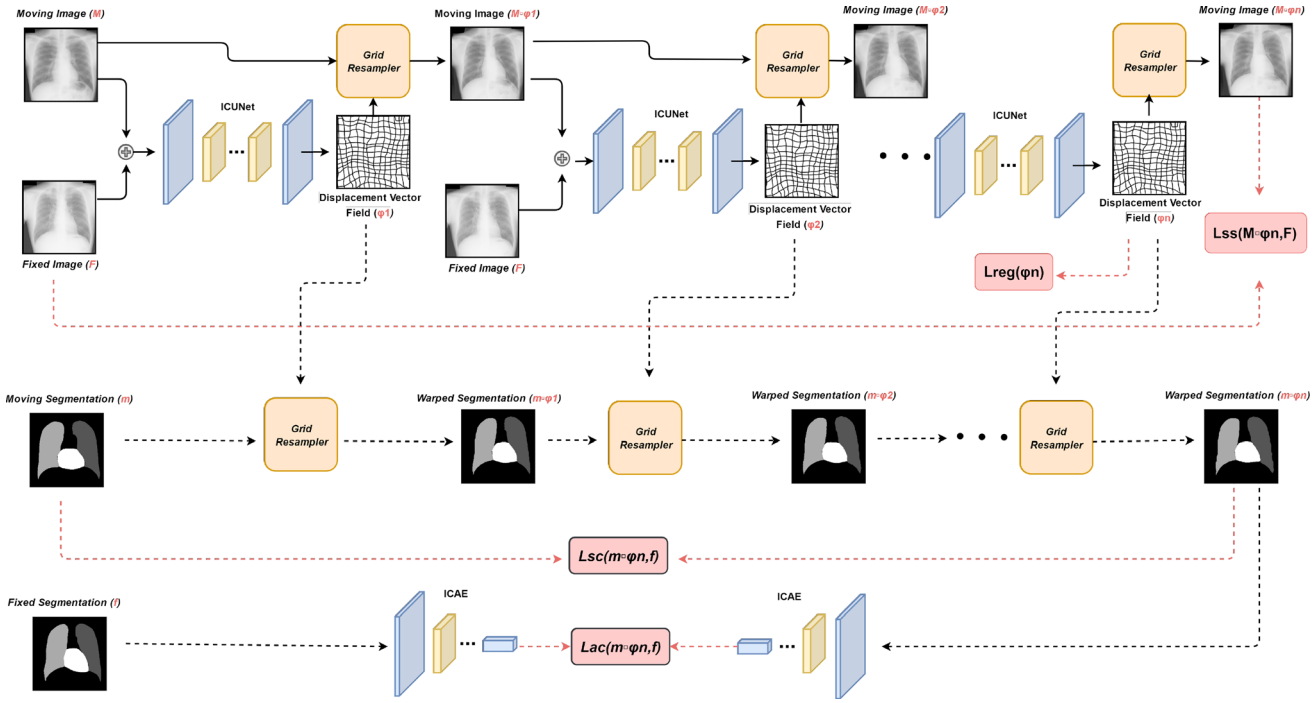


Fig. 1 Illustrates the overall structure and training method of the anatomical cascade RCINet. Each registration sub-network is responsible for finding the fixed image's deformation field and the current moving image. According to the deformation field, the moving image is deformed multiple times and sent to the next level cascade sub-network. The final level of image similarity item L_{ss}, and the local L_{sc} based on the final deformation-based anatomical segmentation and fixed anatomical segmentation combined with the global (L_{ac}) loss function to train the network. There is also a regularized loss term while ensuring the anatomical correspondence after registration and the deformation field's smoothness

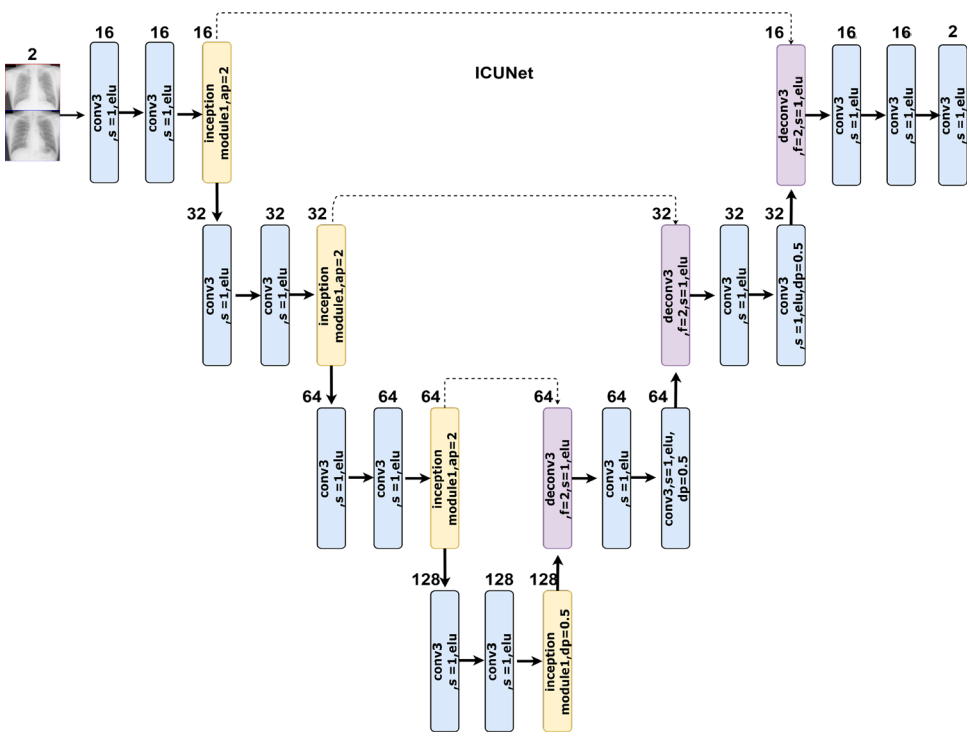


Fig. 2 The ICUNet network we use to predict the registration network's deformation field similarly applies inception modul 1 to UNet

anatomical segmentation to achieve accurate registration. We also transform it into an n-cascaded and finally get the warped anatomical segmentation. Besides, we used the noise reduction ICAE to learn the low-dimensional representation of anatomical segmentation (Vincent et al. 2010) to combine the global loss function with the local loss function and improve the anatomical rationality the registration.

Figure 3 depicts the ICAE encoder used for anatomical segmentation low-dimensional feature learning. Its encoding stage is mainly composed of 3×3 convolution and inception module 2 (Inception module 2 converts the 5×5 convolution kernel of inception module 1 into two 3×3 convolution kernels series connection) alternately, replacing the 5×5 in inception module 1 with two. The 3×3 convolution can combine more nonlinear features, which helps learn the mapping from the input space X to the novel low-dimensional representation h, which retains essential information. These neural networks usually follow a codec architecture, where the encoding $h = \text{enc}(X)$ from the middle fully connected layer. This code contains a large amount of information to decode the original input through the decoding stage $X = \text{dec}(\text{enc}(X))$ and is forced to store important information (used to reconstruct the original anatomical mask) and convert it into a learned representation. The registration effectiveness of the registration network is further improved.

3.2 Loss function

To ensure our model is trained in an unsupervised manner, we design a four-loss function for the RICNet, which contains L_{ss}, L_{reg}, and L_{sc}, L_{ac}. The first item measures the (dis)similarity between the moving images warped by the last spatial transformation and the fixed image. The regularization loss is to encourage the flow field smoothness and to prevent unrealistic or overfitting. The last two items leverage anatomical segmentations at training time to improve the registration effect and anatomical rationality.

3.2.1 Normalization cross-correlation L_{ss}:

We use the normalized cross-correlation function commonly used in single-modal registration to construct L_{ss}, where I_1 and I_2 represent the pixel value of the fixed image and the pixel value of the final deformed image, respectively:

$$\text{NCC}[I_1, I_2] = \frac{\sum I_1 \cdot I_2}{\sqrt{(\sum I_1 \cdot I_1) \times (\sum I_2 \cdot I_2)}} \tag{6}$$

A higher NCC indicates a more similar between the input images. The Normalization cross-correlation is defined as:

$$L_{ss}(M \circ \phi_n, F) = -\text{NCC}[M \circ \phi_n, F] \tag{7}$$

3.2.2 Total variation loss (smooth item) L_{reg}

For the final dense flow field, we regularize it with the subsequent loss that encourages realism physically

$$L_{reg}(\phi_n) = \frac{1}{2|\Omega|} \sum_x \sum_{i=1}^{2} (g(x + e_i) - g(x))^2 \tag{8}$$

where $e_{1,2}$ form the natural basis of $\mathbb{R}^2$. This varies from the initial definition of the improvement in total variation loss [33], but the formula's loss term is better than most regularization.

3.2.3 Segmentation mask similarity L_{sc}:

The loss function formed by the fixed segmentation mask f and the moving segmentation mask m: this section will describe the segmentation information that RCINet is used in training but not used in the test. The loss function is formed by the fixed segmentation mask f and the moving segmentation mask m. The segmentation label is obtained by segmenting the image. The deformed segmentation and the fixed segmentation constitute two corresponding formulas. These two formulas can improve the image's smoothness and help improve the registration's robustness and accuracy.

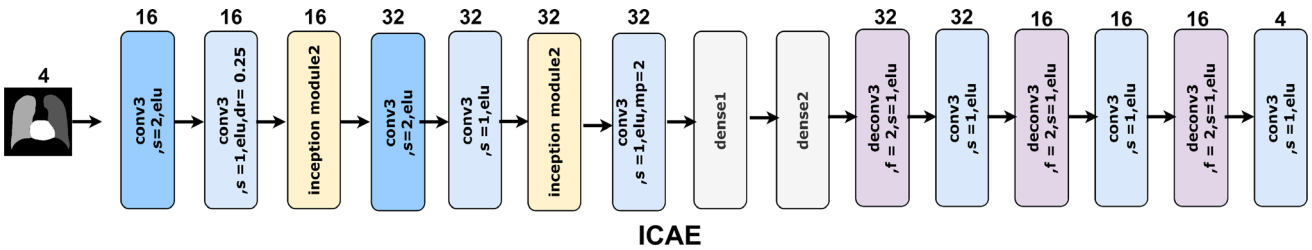


Fig. 3 The encoder that encodes the segmentation mask. Inception module 2 is added in the encoding stage to extract the low-dimensional feature representation better

To enhance the anatomical environment provided to the network, we considered a common strategy, that is, combining the loss mentioned above based on similarity strength with segmentation and including the anatomical segmentation into the loss function. We finally obtain moved $(m \circ \phi_n)$ by cascading the sub-registration sub-network and get the loss function L_{sc}. This loss function quantifies the alignment between the segmentation mask moved $(m \circ \phi_n)$ and the source segmentation f. And the size of each segmentation mask is the same as the corresponding size. Among them, L_{sc} is the classical softmax cross-entropy defined at the pixel level, so the first term composed of anatomical segmentation is:

$$L_{sc} = \text{softmax\_cross\_entroy}(m^{\circ}\phi_n, f) \tag{9}$$

3.2.4 Anatomical segmentation encoding function L_{ac}

The loss function L_{sc} is only defined at the local pixel-level prediction. Therefore, it cannot be guaranteed that the deformation source and the target source can be segmented in anatomy on the whole good registration. Besides, since the segmentation markers used in the data in this paper represent anatomical structures such as lungs and hearts, even if there are high differences in the anatomical structures of different patients, human organs still maintain a high degree of regularity. We use this regularity to restrict the registration process and design a loss function to analyze a global anatomical segmentation. The deformation source segmentation and target segmentation are anatomically reasonable.

To this end, we introduced the autoencoder ICAE to learn the low-dimensional representation of anatomical segmentation. Autoencoder is a convolutional neural network that can retain important information about the input. These neural network structures are usually encoding and decoding systems. The encoding includes Important information. We extract the code $c = \text{enc}(x)$ from the middle fully connected layer. ICAE stores a large amount of the original anatomical segmentation used to reconstruct the original anatomical segmentation into the learned representation, which is to learn important information representing the overall anatomical characteristics of the organ, such as shape and topological structure. In other words, the loss function combines it with pixel-level loss to make the deformation of anatomical segmentation more realistic and reasonable. L_{ac} is defined as the Euclidean distance between the code c generated by the deformation source $(m \circ \phi_n)$ and the corresponding target segmentation f:

$$L_{ac}(m \circ \phi_n, f) = \|\text{AE}(m \circ \phi_n, f) - \text{AE}(f)\|_2^2 \tag{10}$$

It should be noted that both Euclidean distance formula (10) are differentiable, so L_{ac} is the differentiable loss term. Considering both global and local RICNet models, the final loss function is:

$$\begin{aligned}\mathcal{L}(M, F, m, f, \phi) = \mathcal{L}_{ss}(M \circ \phi_n, F) + \lambda_r L_{reg}(\phi_n) \\ + \lambda_{sc}\mathcal{L}_{sc}(m \circ \phi_n, f) + \lambda_{ac}\mathcal{L}_{ac}(m \circ \phi_n, f)\end{aligned} \tag{11}$$

Among them, λ_r, λ_{sc}, λ_{ac} represents the weight of the last three loss functions, and the weighting factor method is used to control the influence of the term L_{sc} on the loss function. The combination of L_{sc} and L_{ac} better reflects the consistency of anatomical shape changes.

3.3 Calculation of cascade deformation field

This article mainly focuses on 2D image registration, and we implement the network in a recursive cascade network. But when training more cascade layers, the calculation of the deformation field is needed; there are specific formulas to calculate it. Given the above situation, in this section, we will extend the work on the calculation of the two-layer cascaded deformation field in [28] to transform the image with the deformation field of each layer and the final number of layers in the test phase. Use the transformation of each layer of the image and the final deformation field to transform the image. We use S to represent the displacement field, ϕ to represent the deformation field, use $\circ$ to represent the transformation using the deformation field, $*$ to represent the transformation using the displacement field, there are the following calculation formulas:

$$I \circ \phi = \{I(x) \circ \phi(x), X \in \mathbb{R}^2\} \tag{12}$$

Theoretically:

$$\phi(x) = x + S(x) \tag{13}$$

$$I * S = I \circ \phi(x) = \{I(x + S(x)), X \in \mathbb{R}^2\} \tag{14}$$

Then, the deformed image obtained by cascading two layers can be expressed as:

$$\begin{aligned} & I * S_1 * S_2 \\ & = \{I(x) * S_1(x), x \in \mathbb{R}^2\} * S_2 \\ & = \{I(x + S_1(x)), x \in \mathbb{R}^2\} * S_2 \\ & = \{I(x + S_1(x)) * S_2(x + S_1(x)), x \in \mathbb{R}^2\} \\ & = \{I(x + S_1(x) + S_2(x + S_1(x))), x \in \mathbb{R}^2\} \\ & = \{I(x + S_1(x) + S_2(x) * S_1(x)), x \in \mathbb{R}^2\} \\ & = \{I(x + S_1(x) + S_2(x) * S_1(x)), x \in \mathbb{R}^2\} \end{aligned} \tag{15}$$

so we can simplify the formula to:

$$I * S_1 * S_2 = I * (S_1 + S_2 * S_1) \tag{16}$$

The displacement field of double cascades is represented by S_n, representing the nth cascades' predicted displacement field. Assuming for n cascades in total, the final displacement field is a composition that:

$$\begin{aligned} f_1 &= S_1 \\ f_2 &= f_1 + S_2 * f_1 \\ &\cdots \\ f_n &= f_{n-1} + S_n * f_{n-1} \end{aligned} \tag{17}$$

4 Experiment

In this section, we evaluate the proposed registration model in 2D chest X-ray image registration. We first introduced the evaluation indicators of the experiment in the second part. The third part compares the registration method in this paper with the other three data sets in turn. The fourth and fifth parts introduce the influence of changing the deformation field constraints on the cascade effect. The sixth part performed a detailed parametric study. The seventh part uses the output deformation field of the last layer to transform the image and compares the effect of changing layer by layer. Finally, some failed case analysis was provided for the benefit of future work propositions.

4.1 Experiment content setting

a. Original experimental data: Our experimental image database is mainly divided into three: the Japanese Society of Radiology (JSRT) database [30], Montgomery County X-ray database, and Shenzhen Hospital X-ray database [31, 32] 2D chest X-rays between subjects from healthy and sick people. JSRT is a public database containing 247 chest X-ray images with and without lung segments (154 images with nodes and 93 images without nodes); the image pixel is 2048 × 2048. The Montgomery set contains 138 PA X-ray images, 80 and 58 images with or without tuberculosis, respectively; the image pixels are 4020 × 4892 or 4892 × 4020. The Shenzhen set contains X-ray images of 662 cases of different sizes with or without tuberculosis (326 normal and 336 pathological images).

   The segmentation mask is obtained by manually segmenting the three types of data sets. JSRT provides manual lung and heart segmentation for each picture. Manual lung segmentation is performed on the Montgomery and Shenzhen data sets, the Montgomery and Shenzhen data sets are manually segmented. The image and segmentation are preprocessed to obtain a square image with the same spatial resolution. These segmentation masks introduce segmentation auxiliary information into the registration problem.

b. Data preprocessing: Thanks to Mansilla et al. [34] for generously providing experimental datasets. We give the image preprocessing to do a simple instruction. An image was taken as a reference image and resized by filling its shortest side with a background color to make it square. Then, all the images of each dataset were registered against this image, taken as a reference image, through a similarity transform using SimpleElastix, finally obtaining images of 4892 × 4892 pixels in the Montgomery set and 4892 × 4892 [34].

c. Parameter Setting: We use TensorFlow to implement our registration method, use two-dimensional linear interpolation to deform the image in the spatial conversion layer, and set the ADAM optimizer's learning rate to 10^{-3}. At the same time, our experiment allows small batches to drop randomly. In our experiment, by performing Sects. 4.4–4.6, we set $\lambda_r = 5 \times 10^{-5}$, $\lambda_{sc} = 1$, and $\lambda_{ac} = 10^{-1}$, the default number of training iterations is 18,000.

4.2 The matric of evaluation

Due to the different samples in the training set, validation set, and test set, the registration accuracy may differ. Therefore, in the experiment, we use the fivefold cross-validation method to register the learning-based method. In the data set of each fold, the distribution of training set, validation set, and test set are 60%, 20%, and 20%, respectively. We set the image size in the training image to 64 × 64 and the test image to 256 × 256. To evaluate the effectiveness of the image registration algorithm, we used four common indicators in the literature. These indicators quantify the configuration the consistency between the source anatomical segmentation and the target mask after deformation:

1. Dice similarity coefficient [35]: a measure of segmentation similarity, which is used to quantify the performance of registration based on the segmentation of anatomical structure. The dice score of two registration A, B is formulated as:

$$\mathrm{Dice}(A, B) = 2 \cdot \frac{|A \cap B|}{|A| + |B|} \tag{18}$$

   Dice ranges from 0 (no overlap) to 1 (complete overlap). And A, B represent the voxels in the segmentation.

2. Hausdorff's distance: a measure of the maximum distance between all surface points of two segmentation volumes. It is defined as:

$$d_{\mathrm{H}}(X,Y) = \max|d_{XY}, d_{YX}| = \max\left\{\max_{x\in X}\min_{y\in Y} d\left(x,y\mid,\max_{y\in Y}\min_{x\in X}(x,y)\right)\right\} \tag{19}$$

Where x and y are points of lesion segmentations X and Y, respectively, and it is measured in millimeters, and a smaller value indicates higher accuracy.

3. Average symmetric surface distance: ASSD is a measure of all Euclidean distances between two image volumes. Given the average surface distance(ASD):

$$\mathrm{ASD}(X,Y) = \sum_{x\in X}\min_{yCY} d(x,y)/|X| \tag{20}$$

Where $d(x, y)$ is a 2D matric consisting of the Euclidean distances between the two image volumes X and Y, ASSD is given as:

$$\mathrm{ASSD}(X,Y) = \{\mathrm{ASD}(X,Y) + \mathrm{ASD}(Y,X)\}/2 \tag{21}$$

Similar to HD, the ASSD is measured in millimeters, and a smaller value indicates higher accuracy.

4. Jacobian: the evaluation metric is the Jacobian determinant $\left|J_\phi(x)\right|$

$$\left|J_\phi(x)\right| = |\nabla\phi(x)| = \begin{vmatrix} \frac{\partial\phi_1(x)}{\partial x_1} & \frac{\partial\phi_1(x)}{\partial x_2} \\ \frac{\partial\phi_2(x)}{\partial x_1} & \frac{\partial\phi_2(x)}{\partial x_2} \end{vmatrix} \tag{22}$$

Use the volume overlap of anatomical segments to evaluate our method and calculate the number of image deformation field overlap; a smaller value indicates higher effectiveness.

5. Time: Comparing the registration time of different methods, we use seconds as the unit; the shorter the registration time, the faster the registration speed.

4.3 Compare with other methods

In this experiment, we use SimpleElastix, a non-learning method, as the first comparison baseline. SimpleElastix is a classic medical image registration toolbox, which is considered the most advanced and is listed as one of the most popular image registration software packages. The second is to compare the state-of-the-art ACRegNet [34] method, which has brilliant effectiveness in the current 2D image registration method. And the third method RICNet(-cas1)no∗seg, which is the RICNet, does not consider segmentation-aware loss function during training. Our model is compared with the three approaches on the three data, and the following results can be obtained, as shown in Tables 1, 2, and 3.

Tables 1, 2, and 3 gives a comprehensive summary of the results. The Dice scores of our proposed method RICNet on the three data sets are 0.5 to 0.8 percentage points higher than the existing state-of-the-art methods. The values of HD and ASSD are both there is a significant decrease. Observing the Jacobian value, we can know that the deformation field's unfolding has been significantly improved. Comparing the time, we can understand that the registration time of the RICNet basic model is almost the same as that of the ACRegNet model. The increase in the number of cascading layers makes the registration time slightly increase, but compared with the registration tool SimpleElastix, the registration method based on learning, there are advantages. We can conclude that our model is superior to all current baseline methods and has a more plausible and smooth deformation field.

4.4 Cascade method with significant weight of constraints

In our experiment in the previous section, our model fixed the constraint term coefficient to $\lambda_r = 5 \times 10^{-5}$, which is called config_1. In this experiment, we will observe the change of config_1 as the number of cascade layers increases. It is worth noting that our registration sub-networks are all the same. Experiments are conducted in 3 data sets with fivefold cross-validation experiments. Training is performed on layers 1 to 10 of the cascade. Training starts from 17,500 times. This model is saved every 100 times, and five models are saved. In the test phase, respectively, calculate the values of the four metrics.

In Fig. 4, we have, respectively, given box plots of various metrics obtained from the fivefold cross-validation test on three data sets with config_1 cascaded from 1 to 10 layers. We can see that the value of the config_1 measurement index will fluctuate. Still, from the general trend, Dice gradually rises as the number of cascade layers increases, HD and ASSD decrease slightly, but JAC has a significant decrease. We can conclude that config_1 has a specific improvement in registration effectiveness as the number of cascade layers deepens.

4.5 Cascade method with a small weight of constraints

In the previous section, we experimented with cascade config_1 with a more significant constraint item weight. We can find that as the number of cascade layers deepens, the increase in Dice at a certain number of layers will not

Table 1 performance comparison of different methods in ***JSRT*** data

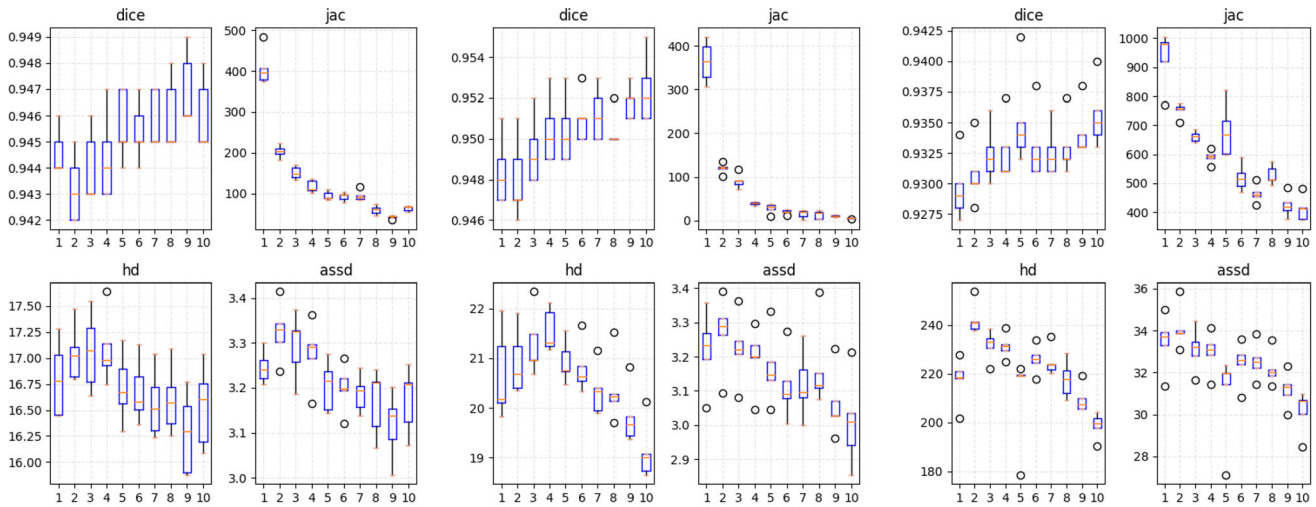
Method	Dice	HD	ASSD	Time	JAC
SimpleElastix	0.871 (0.062)	26.655 (11.898)	5.282 (2.481)	1576.4	
ACRegNet	0.942 (0.019)	17.229 (7.553)	3.360 (1.139)	4.9	399 (663)
RICNet$(cas1)_{no*seg}$	0.844 (0.073)	33.330 (19.036)	9.010 (4.538)	5.1	624 (806)
RICNet$(cas1)_{ours}$	0.945 (0.018)	16.945 (8.920)	3.184 (1.136)	5.8	386 (579)
RICNet$(cas9)_{ours}$	***0.947*** (***0.019***)	***16.233*** (***7.541***)	***3.071*** (***1.143***)	***25.4***	***35*** (***146***)

Table 2 performance comparison of different methods in ***Montgomery*** data

Method	Dice	HD	ASSD	Time	JAC
SimpleElastix	0.887 (0.087)	27.278 (21.178)	4.272 (3.766)	1575.8	
ACRegNet	0.945 (0.034)	20.685 (18.726)	3.330 (2.571)	3.3	453(762)
RICNet$(cas1)_{no*seg}$	0.883 (0.064)	34.450 (34.554)	7.091 (5.815)	3.2	383 (660)
RICNet$(cas1)_{ours}$	0.951 (0.027)	18.488 (17.055)	2.930 (2.222)	4.3	415 (737)
RICNet$(cas9)_{ours}$	***0.953*** (***0.028***)	***18.291*** (***17.425***)	***2.888*** (***2.289***)	***16.0***	***8*** (***51***)

Table 3 performance comparison of different methods in ***Shenzhen*** data

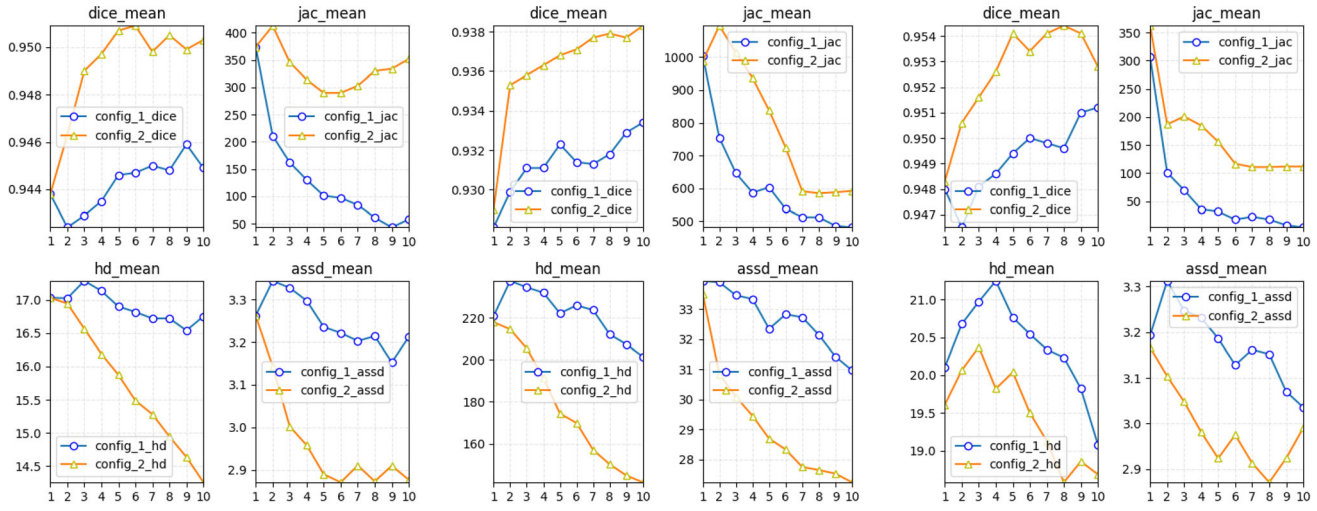
Method	Dice	HD	ASSD	Time	JAC
SimpleElastix	0.916 (0.037)	344.818 (247.637)	47.364 (23.984)	1630.4	
ACRegNet	0.927(0.036)	220.593 (142.494)	34.108 (18.598)	8.7	1045 (1200)
RICNet$(cas1)_{no*seg}$	0.836 (0.0721)	375.640 (208.336)	75.76 (36.022)	8.7	563(641.9)
RICNet$(cas1)_{ours}$	0.929 (0.035)	219.169 (142.463)	33.248 (17.687)	9.2	1048 (1205)
RICNet$(cas9)_{ours}$	***0.934*** (***0.034***)	***196.287*** (***130.963***)	***30.074*** (***16.393***)	***45.5***	***460*** (***729***)



(a) JSRT **(b)** Montgomery **(c)** Shenzhen

Fig. 4 The registration performance of config_1 under different layers in the three data sets

be particularly significant. So, we intend to set the constraint term loss coefficient $\lambda_r = 5 \times 10^{-5}$ divided by the number of cascade layers to reduce its weight, which we called config_2 and observe the changes in other metrics. In other words, we compare and analyze config_1 and config_2. Both are carried out in the same experimental



(a) JSRT **(b)** Montgomery **(c)** Shenzhen

Fig. 5 Comparison of the registration effect of config_1 and config_2 cascading 1 to 10 layers on three data sets

environment, and other hyperparameter settings and testing methods are also the same.

Figure 5 shows the average curve of the registration effect of config_1 and config_2 in each measurement index. Comparing the graphs, we can see that as the number of cascading layers increases, the Dice of config_2 is significantly improved than config_1. The HD and ASSD are reduced considerably. But we can see that the JAC decline is not significant, indicating the deformation field unfolding config_2 is significantly lower than config_1.

To intuitively observe the difference in the registration effect of config_1 and config_2, we randomly select a pair of pictures and their segmentation masks as input to the model during the three data test phases and save the registered images of the two models. We then put the contour edge of the deformed segmentation mask into the registered image to visualize the two models' registration effect. The two configs are only visually compared in cascades 1, 3, 5, 7, and 9 layers. Which means its effectiveness is declining. The results show that a deeper network improves the registration accuracy at the expense of increasing the proportion of folded voxels and reducing the reversibility.

Figure 6 shows the registration effect of config_1 and config_2 on three data sets. *M* represents the source image, *F* represents the target image, and cas n represents the number of cascaded layers. We can see that the difference in the registration effect of the warped images of the two models in the first three layers is not very obvious. However, with the increasing of layers, the smoothness of the deformation field of the config_2 is significantly lower than config_1. The registration image has jagged edges, and the Jacobian value gradually rises. The information loss of the deformation field is more serious, which means its effectiveness is declining. The results show that a deeper network improves the registration accuracy by increasing the folded voxels' proportion and reducing the reversibility.

4.6 A detailed parametric study

In this section, we mainly performed the weighting factors for the loss functions were chosen through grid search using the validation fold. From the experiment in Sect. 4.5, we can know that while pursuing registration accuracy, we need to consider the effectiveness of the deformation field. So we fix $\lambda_r = 5 \times 10^{-5}$ and set the value ranges of λ_{sc} and λ_{ac} to $[1e-3, 1]$ and $[1e-4, 1e-1]$, respectively. In other words, we tested only a discrete set of values for each parameter and trained a model with different combinations of them. The optimal values were obtained from the model that achieved the minimum registration error in the validation.

In the experiment, we set the discrete set value of λ_{sc} to 1e−3, 1e−2, 5e−2, 1e−1, 2e−1, 5e−1, 1, and the discrete value of λ_{ac} to 1e−4, 1e−3, 5e−3, 1e−2, 3e−2, 6e−2, 1e−1, which together form 49 different combinations. We observe the registration effect of each discrete parameter combination of the cascade one layer in RICNet in the three data sets, and fit each data in different 3D parameter surface maps through the obtained scattered point values.

Figure 7 shows the effect of parameter experiments using the cross-validated grid search method under the premise of ensuring the effectiveness of the deformation field. We can know that $\lambda_r = 5 \times 10^{-5}$, $\lambda_{sc} = 1$, and $\lambda_{ac} = 1 \times 10^{-1}$ can obtain the best model for the RICNet network within the specified range.

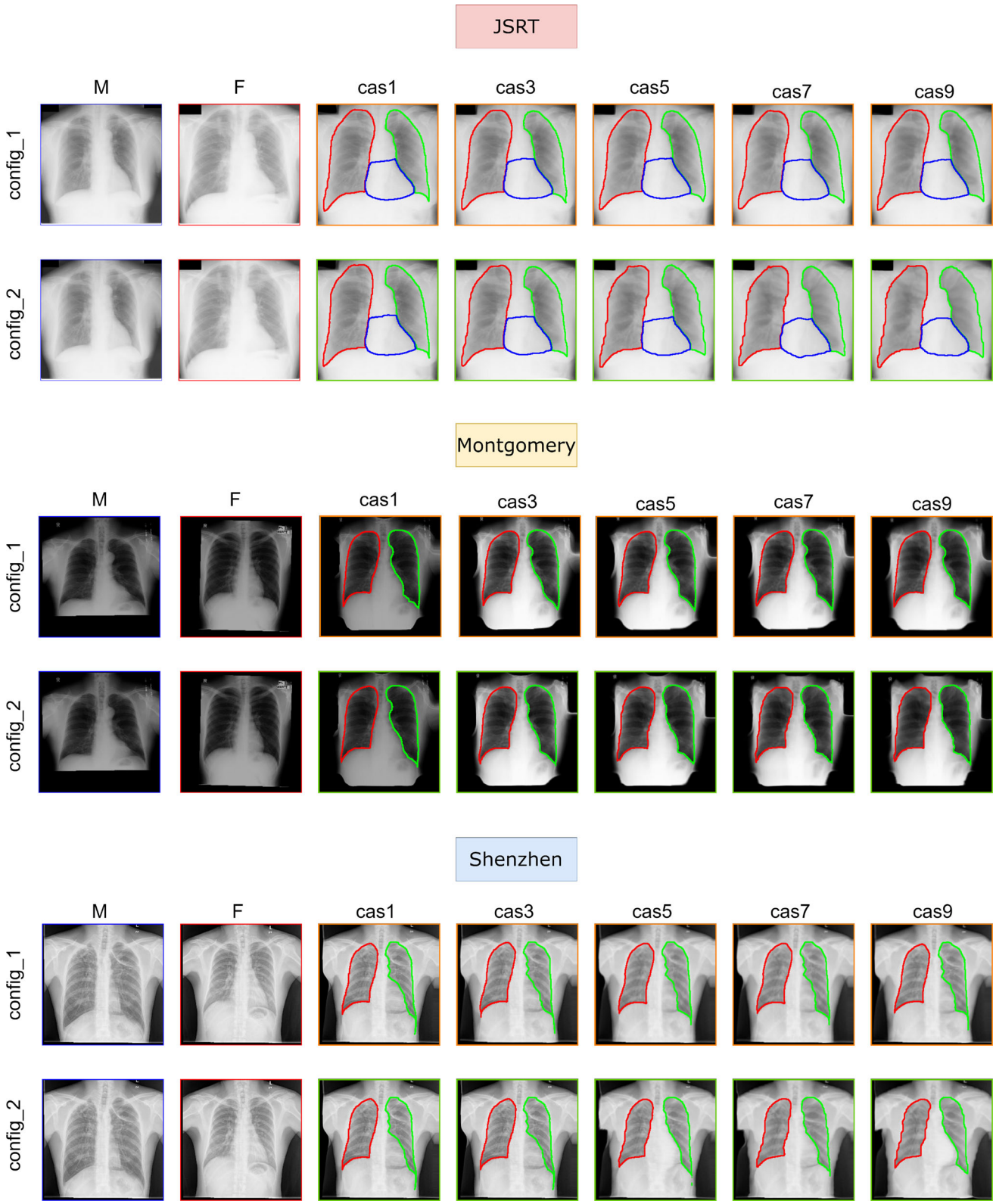


Fig. 6 Visualization of the registration effect of config_1 and config_2 at the same cascade layer number

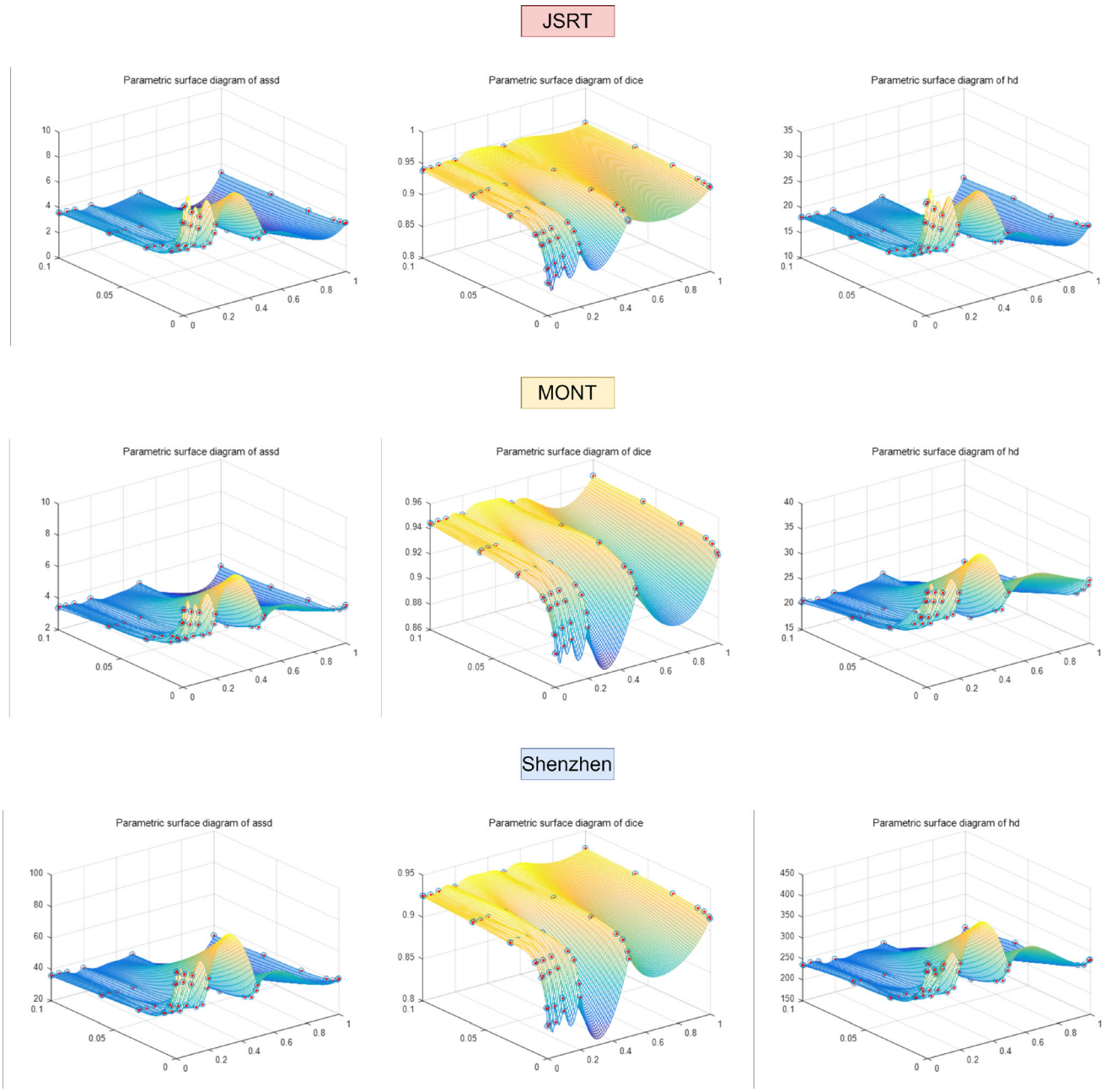
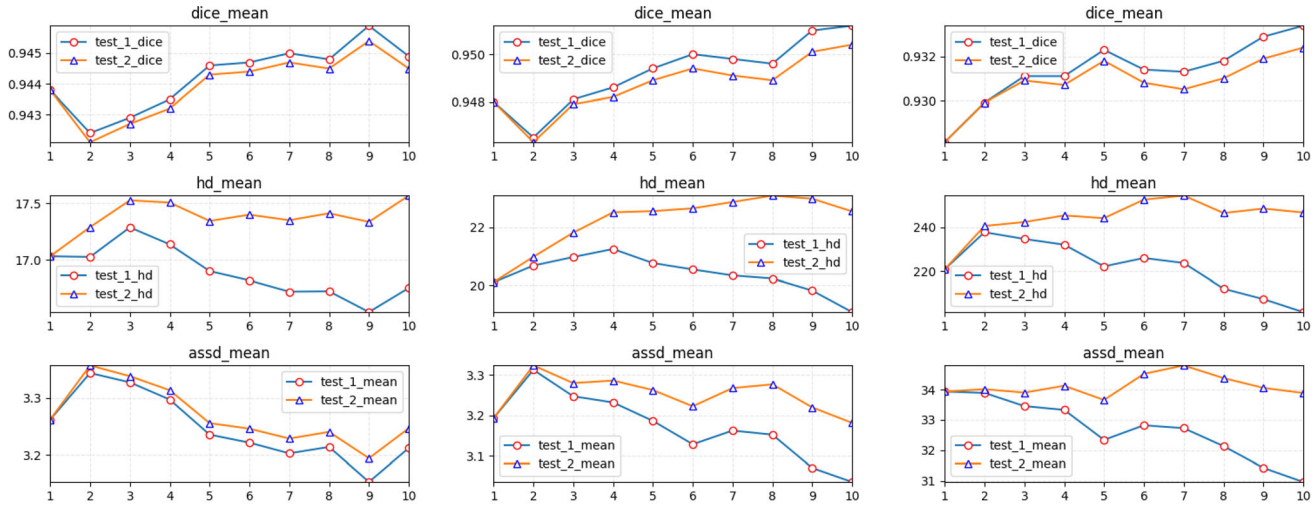


Fig. 7 The 3D parameter fitting surface diagram of each measurement index formed by combining sc and ac at different discrete values

4.7 The effectiveness of the final deformation field formula

In formula (17), we propose a calculation formula for the multilayer cascade deformation field. In this experiment, we will use config_1 to perform two different test methods to verify the formula. The first test method is to transform the input picture through a network layer to change the mobile image, called test_1. The other is to use the formula (17) to obtain deformation fields corresponding to each layer cascade directly deform the moving image, called test_2. And observe and analyze the registration effect diagram of the two methods.

Figure 8 shows the registration results of config_1 using two different test methods on three data sets. Observing the picture, we can see that the Dice of test_1 is slightly higher than that of test_2, and the distance between HD and ASSD is also slightly smaller than the latter, and the registration effect is relatively good. However, the changing trend of the two registration measures is the same in each layer, and the value is not much different. It shows that the registration formula is effective.



(a) JSRT **(b)** Montgomery **(c)** Shenzhen

Fig. 8 Comparison of registration effect of test_1 and test_2 cascaded 1 to 10 layers

Figure 9 shows the registration effect images of test_1 and test_2 at the 1, 5, and 9 layers of the cascade. The one with the green frame is test_1, and the one with the purple frame is test_ 2. We can see that the sharpness of the deformed image of test_1 is much lower than that of test_2. In test_1, it can be concluded due to multiple linear interpolation operations on the image, the registration accuracy can be improved. Still, the clarity is slightly lower than in test_2.

4.8 Failure case analysis

Although our work has achieved some results, we found some relatively failed cases in the experiment. That is, there will be points in the test where the measurement

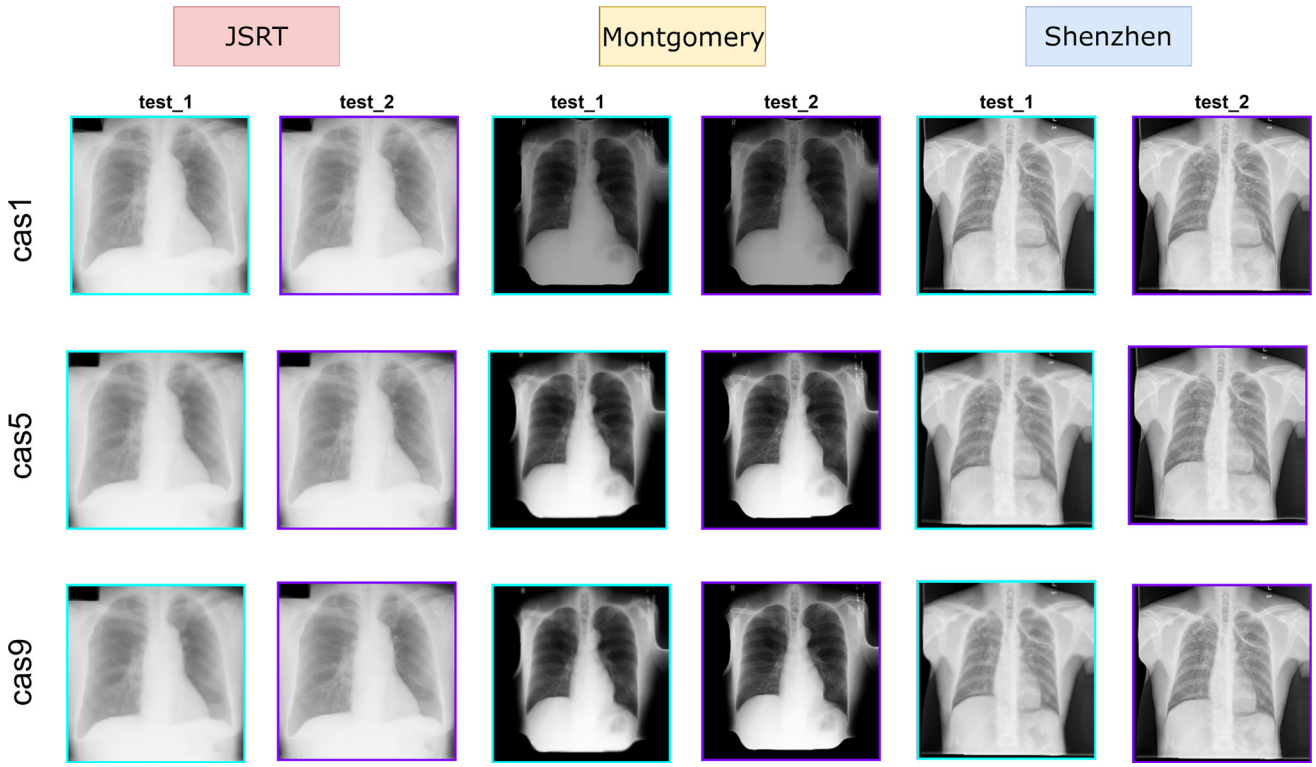


Fig. 9 Two different test methods of model1 are compared on three data sets, cascaded 1, 5, and 9 layers after deformation

standard deviates from the mean. Specifically, for some image pairs with too large structural differences, our model will significantly differ in the registration effect after swapping the registration positions.

We use the RICNet (cas9) model to transform the image. The registration from the large part of the thoracic cavity to the small one is called warped_1, and vice versa; it is called warped_2. During the experiment, we found that the registration measure dice of warped_1 will be significantly lower than warped_2. Figure 10 shows some registered pictures and their segmentation.

Based on the above cases, in future work, we can start to conceive a bidirectionally deformed symmetric registration network or add the relevant loss function of the inverse registration transformation to the network. It is used to overcome the problem of poor registration effect due to significant differences in image structure.

5 Discussion

In terms of dice scores, the network RCINet shows better registration effectiveness than existing methods and does not limit the registration image's size. The inception module is added to the encoding stage of the registration sub-network and ADE. The network provides adaptability to different receptive fields. Simultaneously, the auxiliary information is quoted in the training process. The designed

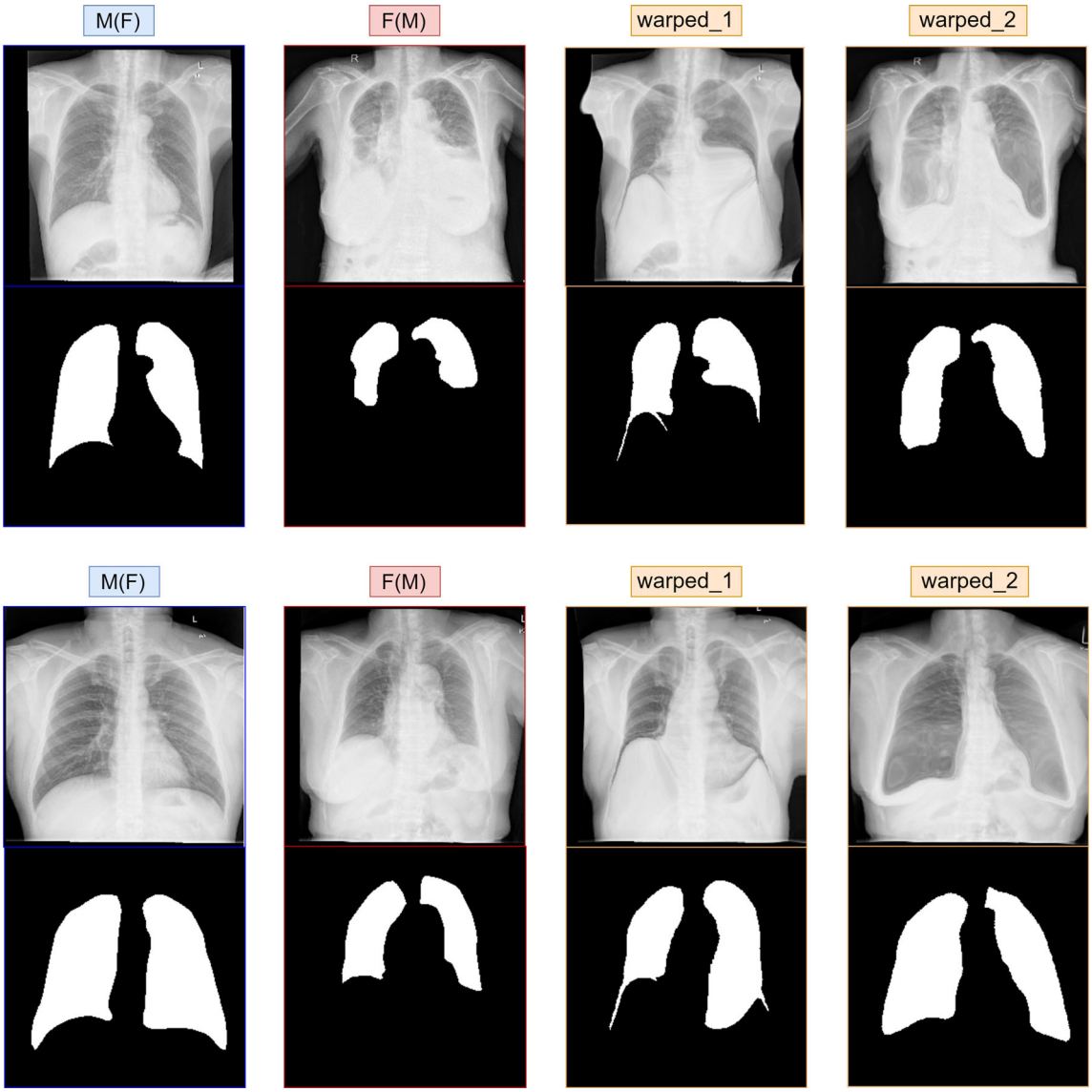


Fig. 10 The registration effect map of the failed cases of the Shenzhen dataset

loss function combines global and local ideas to make the deformation of anatomical segmentation more real and reasonable. The most important thing is that we compared the cascade method of changing the weight of the constraint items to get the approximate relationship between the effectiveness of the deformation field and the registration accuracy and concluded that the increase in the registration accuracy is accompanied by the decrease in the efficacy of the deformation field. Finally, we verified the multilayer cascade deformation field formula, compared and analyzed the two test methods' registration effects.

The limitation of this work is that as the number of cascading layers deepens, GPU performance requirements are higher, and the time it takes for image registration becomes longer.

6 Conclusion

This work proposes a novel cascade method with anatomical segmentation, which introduces anatomical segmentation into the cascade network for training. And we presented a new registration sub-network and auto-encoder with inception to generate corresponding global and local loss functions to train the network. We evaluated our model using a 2D liver CT image dataset and compared our method with classic registration methods and the most advanced methods based on unsupervised learning. Comprehensive experimental results show that this method is superior to traditional methods and learning-based methods in terms of accuracy and quality of deformation field registration

In future work, since we can know that config_1 and config_2 still have a massive difference in registration accuracy, we plan to find a balance between the registration accuracy and the deformation field's effectiveness. We will further study conceiving a bidirectionally deformed symmetric registration network or adding the relevant loss function of the inverse registration transformation to the network to improve our method.

Appendix

In the process of deriving formula (15) in Sect. 3.3, we found that the following steps can calculate the inverse transformation S^{-1} of the displacement field:

$$\begin{aligned} & I * S * S^{-1} \\ & = I(x + S(x)) * S^{-1} \\ & = I\big(x + S(x) + S^{-1}(x + S(x))\big) \\ & = I(x) \end{aligned} \tag{23}$$

We can obtain that:

$$\begin{aligned} & S(x) + S^{-1}(x + S(x)) = 0 \\ & S^{-1}(x + S(x)) = -S(x) \\ & S^{-1}(x) = -S(x - S(x)) \\ & S^{-1}(x) = -S * (-S) \end{aligned} \tag{24}$$

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 as 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 the order of authors listed in the manuscript has been approved by all of us. 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, with respect to 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 which is accessible by the Corresponding Author.

References

1. Fan J, Cao X, Xue Z, Yap P-T, Shen D (2018) Adversarial similarity network for evaluating image alignment in deep learning based registration. In: International conference on medical image computing and computer-assisted intervention. Springer, pp. 739–746
2. Sokooti H, De Vos B, Berendsen F, Lelieveldt BPF, Išgum I, Staring M (2017) Nonrigid image registration using multi-scale 3d convolutional neural networks. In: International conference on medical image computing and computer-assisted intervention. Springer, pp 232–239
3. Rohé M-M, Datar M, Heimann T, Sermesant M, Pennec X (2017) Svf-net: learning deformable image registration using shape matching. In: International conference on medical image computing and computer-assisted intervention. Springer, pp 266–274

4. Lu Z, Yang G, Hua T, Hu L, Kong Y, Tang L, Zhu X, Dillenseger J-L, Shu H, Coatrieux J-L (2019) Unsupervised three-dimensional image registration using a cycle convolutional neural network. In: 2019 IEEE international conference on image processing (ICIP). IEEE, pp 2174–2178
5. de Vos BD, Berendsen FF, Viergever MA, Staring M, Išgum I (2017) End-to-end unsupervised deformable image registration with a convolutional neural network. In: Deep learning in medical image analysis and multimodal learning for clinical decision support. Springer, pp 204–212
6. Balakrishnan G, Zhao A, Sabuncu MR, Guttag J, Dalca AV (2019) Voxelmorph: a learning framework for deformable medical image registration. IEEE Trans Med Imaging 38(8):1788–1800
7. Zhang J (2018) Inverse-consistent deep networks for unsupervised deformable image registration. arXiv preprint arXiv:1809.03443
8. Avants BB, Tustison N, Song G (2009) Advanced normalization tools (ants). Insight J 2(365):1–35
9. Klein S, Staring M, Murphy K, Viergever MA, Pluim JPW (2009) Elastix: a toolbox for intensity-based medical image registration. IEEE Trans Med Imaging 29(1):196–205
10. Hernandez M, Bossa MN, Olmos S (2009) Registration of anatomical images using paths of diffeomorphisms parameterized with stationary vector field flows. Int J Comput Vis 85(3):291–306
11. Joshi SC, Miller MI (2000) Landmark matching via large deformation diffeomorphisms. IEEE Trans Image Process 9(8):1357–1370
12. Miller MI, Beg MF, Ceritoglu C, Stark C (2005) Increasing the power of functional maps of the medial temporal lobe by using large deformation diffeomorphic metric mapping. Proc Natl Acad Sci 102(27):9685–9690
13. Beg MF, Miller MI, Trouvé A, Younes L (2005) Computing large deformation metric mappings via geodesic flows of diffeomorphisms. Int J Comput Vis 61(2):139–157
14. Zhang M, Liao R, Dalca AV, Turk EA, Luo J, Grant PE, Golland P (2017) Frequency diffeomorphisms for efficient image registration. In: International conference on information processing in medical imaging. Springer, pp 559–570
15. Cao Y, Miller MI, Winslow RL, Younes L (2005) Large deformation diffeomorphic metric mapping of vector fields. IEEE Trans Med Imaging 24(9):1216–1230
16. Ceritoglu C, Oishi K, Li X, Chou M-C, Younes L, Albert M, Lyketsos C, van Zijl PCM, Miller MI, Mori S (2009) Multi-contrast large deformation diffeomorphic metric mapping for diffusion tensor imaging. Neuroimage 47(2):618–627
17. Oishi K, Faria A, Jiang H, Li X, Akhter K, Zhang J, Hsu JT, Miller MI, van Zijl PCM, Albert M et al (2009) Atlas-based whole brain white matter analysis using large deformation diffeomorphic metric mapping: application to normal elderly and alzheimer's disease participants. Neuroimage 46(2):486–499
18. Ashburner J (2007) A fast diffeomorphic image registration algorithm. Neuroimage 38(1):95–113
19. Vercauteren T, Pennec X, Perchant A, Ayache N (2009) Diffeomorphic demons: Efficient non-parametric image registration. NeuroImage 45(1):S61–S72
20. Balakrishnan G, Zhao A, Sabuncu MR, Guttag J, Dalca AV (2018) An unsupervised learning model for deformable medical image registration. In: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR). IEEE, pp 9252–9260
21. Shan S, Yan W, Guo X, Chang EI, Fan Y, Xu Y, et al (2017) Unsupervised end-to-end learning for deformable medical image registration. arXiv preprint arXiv:171108608
22. Li H, Fan Y (2017) Non-rigid image registration using fully convolutional networks with deep self-supervision. arXiv preprint arXiv:170900799
23. Jaderberg M, Simonyan K, Zisserman A, et al (2015) Spatial transformer networks. In: Advances in neural information processing systems (NIPS), pp 2017–2025
24. Hu Y, Modat M, Gibson E, Li W, Ghavami N, Bonmati E, Wang G, Bandula S, Moore CM, Emberton M et al (2018) Weakly-supervised convolutional neural networks for multimodal image registration. Med Image Analysis 49:1–13
25. Hu Y, Modat M, Gibson E, Ghavami N, Bonmati E, Moore CM, Emberton M, Noble JA, Barratt DC, Vercauteren T (2018) Label-driven weakly-supervised learning for multimodal deformable image registration. In: 2018 IEEE 15th international symposium on biomedical imaging (ISBI). IEEE, pp 1070–1074
26. Zhao S, Dong Y, Chang EI, Xu Y et al (2019) Recursive cascaded networks for unsupervised medical image registration. In: Proceedings of the IEEE international conference on computer vision, pp 10600–10610
27. Ali S, Rittscher J (2019) Conv2warp: An unsupervised deformable image registration with continuous convolution and warping. In: International workshop on machine learning in medical imaging. Springer, pp 489–497
28. Zhao S, Lau T, Luo J, Eric I, Chang C, Xu Y (2019) Unsupervised 3d end-to-end medical image registration with volume tweening network. IEEE J Biomed Health Informatics 24(5):1394–1404
29. Cheng Z, Guo K, Wu C, Shen J, Qu L (2019) U-net cascaded with dilated convolution for medical image registration. In: 2019 Chinese automation congress (CAC). IEEE, pp 3647–3651
30. Junji S, Shigehiko K, Junpei I, Tsuneo M, Takeshi K, Ken-ichi K, Mitate M, Hiroshi F, Yoshie K, Kunio D (2000) Development of a digital image database for chest radiographs with and without a lung nodule: receiver operating characteristic analysis of radiologists' detection of pulmonary nodules. Am J Roentgenol 174(1):71–74
31. Candemir S, Jaeger S, Palaniappan K, Musco JP, Singh RK, Xue Z, Karargyris A, Antani S, Thoma G, McDonald CJ (2013) Lung segmentation in chest radiographs using anatomical atlases with nonrigid registration. IEEE Trans Med Imag 33(2):577–590
32. Jaeger S, Karargyris A, Candemir S, Folio L, Siegelman J, Callaghan F, Xue Z, Palaniappan K, Singh RK, Antani S et al (2013) Automatic tuberculosis screening using chest radiographs. IEEE Trans Med Imag 33(2):233–245
33. Rudin LI, Osher S, Fatemi E (1992) Nonlinear total variation based noise removal algorithms. Physica D Nonlinear Phenomena 60(1–4):259–268
34. Mansilla L, Milone DH, Ferrante E (2020) Learning deformable registration of medical images with anatomical constraints. Neural Netw 124:269–279
35. Chechik G, Shalit U, Sharma V, Bengio S (2009) An online algorithm for large scale image similarity learning. In: Advances in neural information processing systems (NIPS), pp 306–314