ISSN 2616-5775 Vol. 6, Issue 9: 102-109, DOI: 10.25236/AJCIS.2023.060915

Large Deformation Features Guided Network for Medical Image Registration

Haoxuan Sun[1,2], Xiaogang Du[1,2]

[1]Shaanxi Joint Laboratory of Artificial Intelligence, Shaanxi University of Science and Technology, Xi'an, 710021, China
[2]The School of Electronic Information and Artificial Intelligence, Shaanxi University of Science and Technology, Xi'an, 710021, China

Abstract: *Aiming at the problem of the loss of detail information in medical image registration, which leads to poor registration results in large deformation regions, a large deformation features guided network(LDGNet) for medical image registration is proposed. LDGNet mainly includes two contributions: first, a large deformation feature enhancement module is designed at the encoding and decoding connection to enable the network to enhance the extraction of large deformation features. Secondly, a large deformation feature guidance module is designed at the skip connection, which can help fully fuse the large deformation features from the encoded feature map, and effectively improve the registration accuracy of the network in large deformation regions. Registration experiments on the brain dataset IXI show that LDGNet achieves higher registration accuracy compared with current popular medical image registration methods.*

Keywords: *deep learning, deformable registration, unsupervised learning, large deformation feature enhancement, large deformation feature guidance*

1. Introduction

Image registration is an important technique in medical image processing. Its purpose is to align medical images acquired at different locations or at different times, so that doctors can more accurately diagnose and treat diseases. Image registration technology is widely used in the field of medical images, including surgical navigation, disease diagnosis, treatment plan formulation, and drawing of organ structure maps[1].

The core of the medical image registration problem lies in how to determine the similarity between images and how to spatially map the images. Traditional grayscale-based registration methods[2-3] usually use the grayscale information of the image itself to calculate the similarity between images. However, when the similarity measure function has multiple extreme values, it is easy to fall into local optimality. In addition, traditional registration methods require iterative calculations, usually with high calculation costs and slow registration speed, which are difficult to meet the registration efficiency requirements of clinical applications.

With the development of deep learning, some scholars have introduced deep learning[4-5] into medical image registration and have achieved many research results. According to whether label information is needed, medical image registration based on deep learning can be divided into supervised medical image registration methods and unsupervised medical image registration methods.

In the supervised medical image registration method, the input image pair and the gold standard registration field are input into the registration network for training, and the trained registration network is used to directly predict the test sample to obtain the registration field. The gold standard registration field is generally obtained by two methods: random generation and traditional registration method. For example, Hessam et al.[6] proposed nonrigid image registration using multi-scale 3D convolutional neural networks. This network uses randomly generated registration fields in the training phase, which improves the efficiency of label generation for training samples. RegNet performance Outperforms single-resolution B-spline registration algorithms. However, the randomly generated registration field can only synthesize limited spatial transformations, and the synthesized registration field cannot effectively simulate the real movement of human tissues and organs, so the generalization effect in clinical practice is poor. Some scholars use the registration field obtained by traditional

ISSN 2616-5775 Vol. 6, Issue 9: 102-109, DOI: 10.25236/AJCIS.2023.060915

methods as the gold standard field to avoid the problem of large differences with the real deformation. However, this method is difficult to circumvent the problems of traditional methods, and the final registration accuracy is difficult to surpass the used traditional method.

The unsupervised medical image registration method does not require a gold standard registration field. It only needs to input fixed images and moving images into the registration network to directly predict the registration field. The unsupervised registration network VoxelMorph proposed by Balakrishnan et al.[7] which realizes three-dimensional brain image registration. During the training process, the network uses the spatial transformation network to calculate the image similarity loss, penalizes the appearance difference between the moved image and the fixed image, and finally achieves registration accuracy comparable to SyN. However, the input image of VoxelMorph requires additional affine registration for preprocessing, which increases the registration time. To spve this question, the VTN network proposed by Zhao et al.[8] cascades the affine network into the registration network to achieve true end-to-end registration. However, this method has poor registration effect in large deformation areas of the image. In order to solve this problem, some scholars are working on stacking multiple networks to improve the registration accuracy. For example, Bob et al.[9] proposed a recursive cascade registration network, which gradually distorts the image by recursively cascading multiple VoxelMorphs, thus improving the registration accuracy. However, as the number of cascade layers increases, it is difficult to maintain the smoothness of the registration field, and the cascade leads to a substantial increase in the number of network model parameters, resulting in longer training time. In order to reduce the number of model parameters, CycleMorph proposed by Kim et al.[10] uses a cascade of two registration sub-networks to obtain input by switching the order of fixed images and moving images through cycle consistency, which can be expanded to large volumes to be registered. Multi-scale registration is performed on the images, so that the model captures different levels of transformation relations.

Although the above methods use the network cascade method to improve the registration accuracy, it leads to the accumulation of interpolation artifacts and affects the quality of the registration field. To solve this problem, some scholars try to estimate the registration field based on a multi-resolution strategy. Mok et al.[11] simulated the traditional multi-resolution strategy and proposed the L-level Laplacian pyramid framework (LapIRN), creating an image pyramid on the input image pair, use the network to calculate the registration field at different resolutions, distort the upper-level image, train and expand the network on higher-resolution data, and finally calculate the multi-resolution registration field from the pyramid structure to estimate the final registration field. Kang et al.[12] proposed a dual-stream pyramid network Dual-PRNet, which calculates two feature pyramids respectively to directly estimate the registration field in the feature space. The refinement of the registration field and convolution features is carried out layer by layer, sequentially and from coarse to fine, and the fixed image and the moving image are more accurately aligned in the feature space. Compared with LapIRN,improved registration in large deformation areas. However, when the above network performs feature extraction, it continuously uses downsampling to obtain multi-scale features, and the feature fusion method at skip connections is relatively simple, and detailed information is lost, resulting in poor registration results in large deformation areas.

In order to solve this problem, we propose a large deformation features guided network for medical image registration, referred to as LDGNet. The main contributions of this paper include:

(1) We design a large deformation feature enhancement module, which helps the network to supplement the detailed features containing large deformation information lost during the encoding and decoding process.

(2) We design a large deformation feature guidance module, which enhance the encoding feature map and large deformation feature fusion, so as to fully pay attention to the large deformation features when performing feature fusion.

2. Methodology

2.1 The whole frame

The overall registration framework of the LDGNet network proposed in this paper is shown in Figure 1. The registration process can be roughly divided into three steps: (1) The fixed image and the moving image are input into the registration network LDGNet for feature extraction to obtain the registration field; (2) The spatial transformation network uses the registration field to distort the

ISSN 2616-5775 Vol. 6, Issue 9: 102-109, DOI: 10.25236/AJCIS.2023.060915

moving image to obtain the registration image; (3) Use the similarity measure of the moved image and the fixed image as a loss function to update the network parameters.

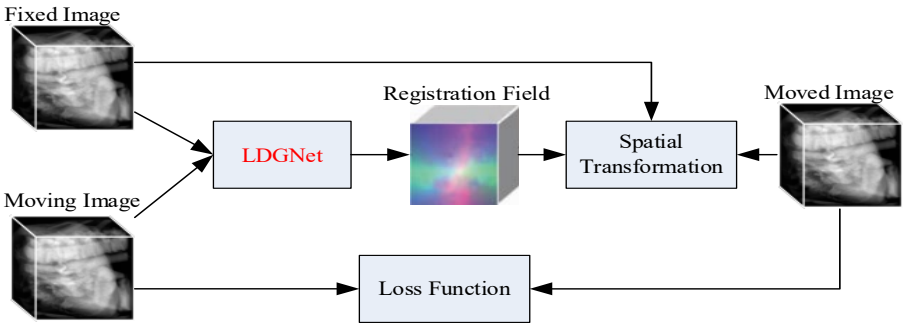

Figure 1: LDGNet registration framework

The LDGNet registration network is shown in Figure 2. First, the fixed and moving images are concatenated by channel as input, and then the input image pair is encoded and decoded to extract image features for estimating the final registration field ϕ. In the encoding part, the network continuously uses four three-dimensional convolutions with a size of $3\times3\times3$ and a step size of 2 for feature extraction. The spatial dimension of each convolutional feature map is halved, and the size of the last layer of feature maps is 1/16 of the input image, although the network retains the information of the image to the greatest extent, it still inevitably loses a lot of detail information containing large deformation information, so a large deformation feature enhancement module is designed behind the minimum feature map to help the network enhance the extraction of large deformation features . In the decoding part, the network upsamples the smallest feature map, and then stitches the upsampled feature map with the feature map of the corresponding layer of the encoding part through skip connections to supplement the detailed information. However, it is relatively simple to use splicing for feature fusion. The detailed information from the encoded feature map cannot be well fused, so a large deformation feature guidance module is designed at the skip connection to help the network better fuse the large deformation feature information.

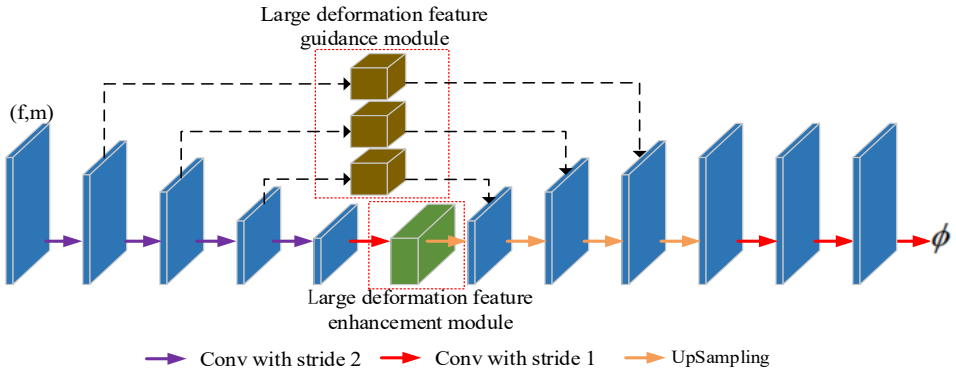

Figure 2: The network structure of LDGNet

2.2 Large deformation feature enhancement module

In order to enhance the network's extraction of large deformation features, this paper designs a large deformation feature enhancement module at the codec connection, as shown in Figure 3.

(1) Extract features at multiple scales. Specifically, for the input feature map, three dilated convolutions with a size of $3\times3\times3$, and dilation rates of 6, 12, and 18 are used. Using dilated convolutions to extract multi-scale features can not only expand the receptive field of the convolution kernel, thereby improving the range of features extracted by the convolution kernel, but also does not lose the resolution of the image, reducing the loss of Large deformation feature information.

(2) Multi-scale feature fusion. For the extracted features of multiple scales, the concat operation is used to splicing to obtain the spliced feature map, and then the convolution with a size of $1\times1\times1$ is used for feature fusion, and reduce the output feature map to the same dimension as the input feature map.

ISSN 2616-5775 Vol. 6, Issue 9: 102-109, DOI: 10.25236/AJCIS.2023.060915

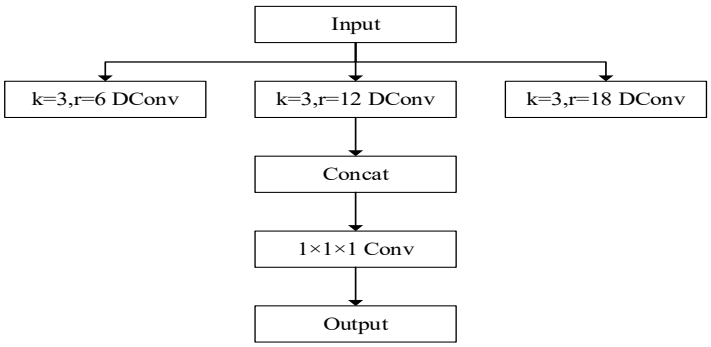

Figure 3: Large deformation feature enhancement module

2.3 Large deformation feature guidance module

The feature map from the encoder contains more detailed feature information containing large deformations, while the feature map obtained by upsampling from the decoder has more global information.In order to fully fuse the feature maps of the encoder and decoder and guide the registration of large deformation, a large deformation feature guidance module is designed. The specific structure can be divided into two steps as shown in Figure 4: encoding feature map enhancement and large deformation feature fusion.

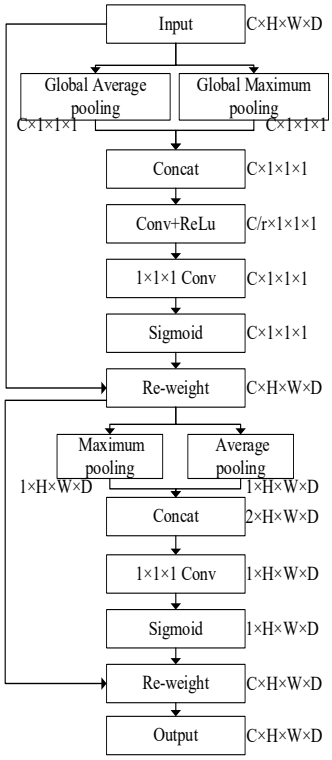

Figure 4: Large deformation feature guidance module

(1) In encoding feature map enhancement, since the feature map from the encoder part contains more detailed information containing large deformations, more weight is given to the encoding feature map. Specifically: first, the feature map from the encoder part and the feature map upsampled by the decoder are spliced by channel to obtain the spliced feature map, and then a convolution of size $1\times1\times1$ is used to perform feature fusion on the spliced feature map, and then average pooling and maximum pooling are used to perform feature fusion on the input features The feature information aggregation operation is followed by compressing the spatial dimension of the input feature map, summing and merging element by element, and finally generating the encoding feature enhanced attention map after passing through the Sigmoid activation function. At this time, the attention map places more weight in the channel encoding the feature map.

(2) In large deformation feature fusion, the enhanced feature map is given more weight to the fusion

ISSN 2616-5775 Vol. 6, Issue 9: 102-109, DOI: 10.25236/AJCIS.2023.060915

of detailed feature information. Specifically, average pooling and maximum pooling are first used to compress the enhanced feature map in height and width. , then perform the concat operation, and then use the convolution with a size of $1\times1\times1$ operation to reduce the dimension of the feature map to 1. Finally, the feature fusion attention map is generated after the Sigmoid activation function. At this time, the attention map puts more weight on the fusion of large deformation features.

2.4 Loss function

In this paper, the optimal registration field ϕ is calculated by maximizing the loss function, which is shown in formula (1):

$$\phi = Max(Sim(I_f, I_m \circ \phi) + Reg(\phi)) \tag{1}$$

$Sim(I_f, I_m \circ \phi)$ is used to calculate the similarity between the fixed image I_f and the registered image $I_m \circ \phi$, and $\mathrm{Re}g(\phi)$ is used to calculate the regularization of the registration field ϕ.

We achieve $Sim(I_f, I_m \circ \phi)$ by calculating NCC, as shown in Equation (2):

$$Sim(I_f, I_m \circ \phi) = NCC(I_f, I_m \circ \phi) \tag{2}$$

Maximizing $Sim(I_f, I_m \circ \phi)$ will make $m \circ \phi$ closer and closer to f, but may cause the deformation field to fold. Therefore, this paper uses the spatial gradient of the deformation field to smooth the deformation field, as shown in Equation (3):

$$Reg(\phi) = \sum_{p\in\Omega} \left\| \nabla u(p) \right\|^2 \tag{3}$$

$\nabla u(p)$ is the spatial gradient, and is calculated as shown in Equation (4):

$$\nabla u(p) = \left(\frac{\partial u(p)}{\partial x}, \frac{\partial u(p)}{\partial y}, \frac{\partial u(p)}{\partial z} \right) \tag{4}$$

We use the difference between adjacent voxels in the deformation field to be approximately equal to the spatial gradient, and the calculation of $\frac{\partial u(p)}{\partial x}$ is shown in Equation (5):

$$\frac{\partial u(p)}{\partial x} \approx u((p_x + 1, p_y, p_z)) - u((p_x, p_y, p_z)) \tag{5}$$

The calculations of $\frac{\partial u(p)}{\partial y}$ and $\frac{\partial u(p)}{\partial z}$ are similar to $\frac{\partial u(p)}{\partial x}$.

3. Results and discussion

3.1 Dataset

We use the IXI[13] dataset for registration experiments. The IXI dataset contains 576 sets of images,the training set, verification set, and test set are 403, 58, and 115 (7:1:2) respectively. The evaluation metrics are calculated using the segmentation values of 29 anatomical structures that come with the image.

3.2 Experimental setup

The hardware environment configuration of this experiment CPU is Intel(R) Xeon(R) Gold 6326, memory 50GB, GPU is Nvidia Tesla A30, video memory 24GB. The experiment in this paper was conducted on Ubuntu 20.04. We use PyTorch to implement the deep learning framework.

The Adam optimizer is used during training, the maximum number of training iterations is set to

ISSN 2616-5775 Vol. 6, Issue 9: 102-109, DOI: 10.25236/AJCIS.2023.060915

100, the batch size is 1, and the initial learning rate is 0.0001.

3.3 Evaluation metrics

In this paper, visual analysis and quantitative analysis are used to evaluate the registration results. Visual analysis mainly uses registration result visualization and difference result visualization to measure the quality of the registration results; quantitative analysis uses Dice and Jacobian to evaluate the performance of the registration algorithm. The Dice represents the overlapping degree of the moving image and the fixed image in the corresponding anatomical area, and the calculation of the Dice is shown in formula (6):

$$\text{Dice(A, B)} = 2 \times \frac{|A \cap B|}{|A| + |B|} \tag{6}$$

Where A and B represent the anatomical areas corresponding to the moving image and the fixed image respectively. When the Dice value is closer to 1, it means that the similarity between the two images in the corresponding area is higher, that is, the registration accuracy is higher; otherwise, it is lower.

The Jacobian is a commonly used metric to evaluate the topological properties of dense displacement fields. Taking three-dimensional data as an example, the Jacobian determinant at each point in the dense displacement field is calculated as shown in Equation (7):

$$\det(J(i,j,k)) = \begin{vmatrix} \frac{\partial i}{\partial x} & \frac{\partial j}{\partial x} & \frac{\partial k}{\partial x} \\ \frac{\partial i}{\partial y} & \frac{\partial j}{\partial y} & \frac{\partial k}{\partial y} \\ \frac{\partial i}{\partial z} & \frac{\partial j}{\partial z} & \frac{\partial k}{\partial z} \end{vmatrix} \tag{7}$$

When $\det(J(i,j,k)) = 1$, it means that there is no spatial change at this point; $\det(J(i,j,k)) > 1$ means that expansion occurs; when $0 < \det(J(i,j,k)) < 1$ means that shrinkage occurs; when $\det(J(i,j,k)) <= 0$ means that the image collapses at this point. In the following description, use $|J_\phi| \leqslant 0$ to represent the percentage of the voxel number of $\det(J(i,j,k)) <= 0$ in the overall prime number.

3.4 Comparative experiment

In order to verify the effectiveness of LDGNet, the experiment in this section compares LDGNet with mainstream medical image registration methods such as SyN[14], LDDMM[15], VoxelMorph[9], CycleMorph[12], VTN[10], and NICE-Net[16].

The quantification results on the IXI dataset are shown in Table 1. On the IXI dataset, LDGNet achieved the best Dice score, which was 0.1223, 0.0853, 0.0118, 0.0100, 0.0130, and 0.0025 higher than SyN, LDDMM, VoxelMorph, CycleMorph, VTN, and NICE-Net, respectively.

Table 1: Quantitative evaluations on IXI dataset

Dataset	IXI	
Evaluation metrics	Dice (mean±std)	$\|J_\phi\| \leqslant 0$ (mean±std)
SyN[14]	0.6420±0.040	<.0.00001
LDDMM[15]	0.6790±0.044	<.0.00001
VoxelMorph[9]	0.7525±0.017	0.860±0.072
CycleMorph[12]	0.7543±0.015	0.711±0.061
VTN[10]	0.7513±0.012	0.881±0.054
NICE-Net[16]	0.7618±0.019	0.672±0.087
LDGNet	0.7643±0.021	0.790±0.055

ISSN 2616-5775 Vol. 6, Issue 9: 102-109, DOI: 10.25236/AJCIS.2023.060915

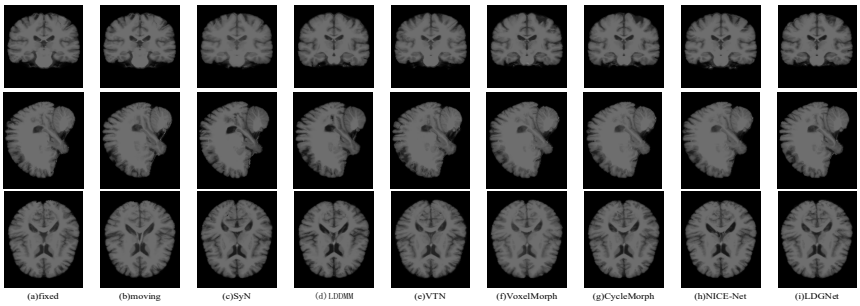

Figure 5: Visualization of registration results of seven methods on dataset IXI

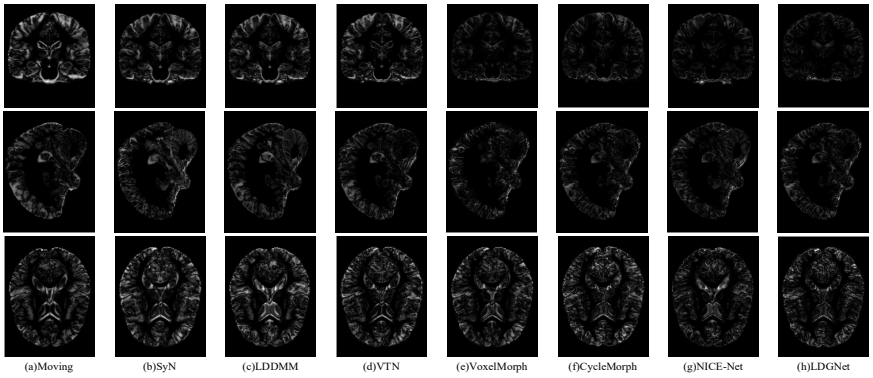

Figure 6: Visualization of difference results of seven methods on dataset IXI

The visualization results on the IXI dataset are shown in Figure 5 and Figure 6. Figure 5 is the visual registration results of seven methods on the brain dataset IXI, where Figure 5(a) and Figure 5(b) are fixed images and moving images, Figure 5(c), 5(d), 5(e), 5(f), 5(g), and 5(h) are images registered using SyN, LDDMM, VTN, VoxelMorph, CycleMorph, and NICE-Net respectively, and Figure 5(g) is registered using the method in this paper image after. Figure 6 shows the differential results of the seven methods on the brain dataset LPBA40. Figure 6(a) is the difference result of fixed image and floating image, Figure 6(b), 6(c), 6(d), 6(e), 6(f), and 6(g) are obtained using SyN, The differential results of LDDMM, VTN, VoxelMorph, CycleMorph, and NICE-Net moved images and fixed images. Figure 6(h) is the differential result of moved images and fixed images using this method. The difference result can be used to judge the quality of the moved result. When the difference result is closer to black, the registration result is better, otherwise, the registration result is worse. As can be seen from Figures 5 and 6, compared with other methods, this method achieves better registration results in brain fold areas containing large deformations.

3.5 Ablation study

In order to verify the effectiveness of the large deformation feature enhancement and large deformation feature guidance modules, this paper uses VoxelMorph as the benchmark method to conduct ablation experiments. Model A means that the large deformation feature enhancement and large deformation feature guidance modules are not used, that is, only the baseline model is used for training; Model B means that only the large deformation feature guidance module is added to the baseline model; Model C means that only the large deformation feature enhancement module is added to the baseline model; Model D means adding large deformation feature enhancement and large deformation feature guidance modules at the same time, which is the LDGNet network in this paper.

The results of ablation experiments on the IXI dataset are shown in Table 2. As can be seen from Table 2, on the IXI dataset, the Dice score obtained without using the large deformation feature enhancement and large deformation feature guidance modules (VoxelMorph) is 0.7525; while using the large deformation feature enhancement and large deformation feature guidance modules respectively, the Dice score is increased to 0.7601 and 0.7585; the Dice score obtained by using bothlarge deformation feature enhancement and large deformation feature guidance modules finally increased to 0.7643. To sum up, on the IXI dataset, the large deformation feature enhancement and large deformation feature guidance modules respectively contributed to improving the registration accuracy, and combines use them further improve the registration accuracy.

ISSN 2616-5775 Vol. 6, Issue 9: 102-109, DOI: 10.25236/AJCIS.2023.060915

Table 2: Ablation studies on IXI dataset

Model	large deformation feature guidance	large deformation feature enhancement	IXI (Dice)
A	×	×	0.7525±0.017
B	√	×	0.7601±0.031
C	×	√	0.7585±0.022
D	√	√	0.7643±0.021

4. Conclusion

In order to improve the registration effect of large deformation areas, this paper proposes a large deformation features guided network(LDGNet) for medical image registration. First, a large deformation feature enhancement module is used at the codec connection to help the network enhance the extraction of large deformation features. Secondly, a large deformation feature guidance module is designed at the skip connection to fully integrate the large deformation information supplemented by the encoder part to promote the network to more accurately generate the registration field of large deformation areas. Finally, the experimental results on the public brain dataset IXI show that LDGNet has achieved better registration results.

References

[1] Haskins G, Kruger U, Yan P. Deep learning in medical image registration: a survey[J]. Machine Vision and Applications, 2020, 31: 1-18.
[2] Sun W, Niessen W J, Klein S. Randomly perturbed B-splines for nonrigid image registration[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(7): 1401-1413.
[3] Sharifi H, Zhang H, Bagher-Ebadian H, et al. Utilization of a hybrid finite-element based registration method to quantify heterogeneous tumor response for adaptive treatment for lung cancer patients[J]. Physics in Medicine and Biology, 2018, 63(6): 065017.
[4] Bob D, Berendsen F F, Viergever M A. A deep learning framework for unsupervised affine and deformable image registration[J]. Medical Image Analysis, 2019, 52: 128-143.
[5] Cao X H, Yang J H, Zhang J, et al. Deformable image registration using a cue-aware deep regression network[J]. IEEE Transactions on Biomedical Engineering, 2018, 4(9): 1900-1911.
[6] Sokooti H, Vos B D, Berendsen F, et al. Nonrigid image registration using multi-scale 3D convolutional neural networks[C]//Proceeding of International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI). 2017: 232-239.
[7] Krebs J, Delingette H, Mailhe B, et al. Learning a probabilistic model for diffeomorphic registration[J]. IEEE transactions on medical imaging, 2019, 38(9): 2165-2176.
[8] Zhao S, Lau T F, Luo J, et al. Unsupervised 3d end-to-end medical image registration with volume tweening network[J]. IEEE Journal of Biomedical and Health Informatics, 2020, 24(5): 1394-1404.
[9] Vos B D, Berendsen F F, Viergever M A, et al. A deep learning framework for unsupervised affine and deformable image registration[J]. Medical image analysis, 2019, 52:128-143.
[10] Kim B, Kim D H, Park S H, et al. CycleMorph: cycle consistent unsupervised deformable image registration[J]. Medical Image Analysis, 2021, 71: 102036.
[11] Mok T C W, Chung A C S. Large deformation diffeomorphic image registration with Laplacian pyramid networks[C]// Medical Image Computing and Computer Assisted Intervention, 2020: 211-221.
[12] Kang M, Hu X, Huang W, et al. Dual-stream pyramid registration network[J]. Medical Image Analysis, 2022, 78: 102379.
[13] The Information eXtraction from Images (IXI) dataset. https://brain-development.org/ixi-dataset/. Accessed 19 Jan 2022.
[14] Avants B B, Epstein C L, Grossman M, et al. Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain[J]. Medical Image Analysis, 2008, 12(1): 26-41.
[15] Beg, M F, Miller, M I, Trouv´E, A, Younes, L. Computing large deformation metric mappings via geodesic flows of diffeomorphisms[J]. International journal of computer vision ,2005, 61:139-157.
[16] Meng M, Bi L, Feng D, et al. Non-iterative Coarse-to-Fine Registration Based on Single-Pass Deep Cumulative Learning[C]//Proceeding of International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI). 2022: 88-97.