A 3D medical image registration method based on multi-scale feature fusion

Jinshuo Zhang¹, Fangquan Liu¹, Xuedou Yu², Yingjun Ma³, and Xiuyang Zhao*¹

¹University of Jinan, Jinan,250022
²Dezhou University, Dezhou,253023
³Shandong University of Political Science and Law, Jinan,250014
E-mail: xiuyangzhao@hotmail.com

Abstract. Deformable medical image registration refers to finding a certain transformation so that the corresponding points of two medical images can be aligned in space. This has important clinical applications. In this article, we propose an unsupervised end-to-end medical image registration method. In this method, the fixed image and the moving image are concatenated in series and input into the convolution neural network to obtain the feature images of different scales. In order to improve the ability of neural network to capture global and local information, we fuse feature maps of different scales. The spatial transformation network uses the deformation field to deform the moving image, so as to realize the registration of the two pairs. We validate our method in the ABIDE data set and compare it with some classic state-of-the-art methods. The experimental results show that our method improves the registration accuracy of image pairs.

1. Introduction
Deformable image registration is the process of establishing dense correspondence between two or more images [1]. It is a basic task in medical image processing and has important clinical applications, so it has been paid attention to by many scholars in recent years. Medical imaging equipment is diverse and each has its own advantages. In clinical applications, doctors often need to combine image information collected by multiple imaging equipment to make a diagnosis of the patient's condition. If medical images of different modalities can be registered and fused together, it will effectively assist doctors in diagnosing the condition. In addition, accurate medical image registration is of great significance for specifying tumor radiotherapy plans and monitoring the development of lesions.

The traditional image registration method generally optimizes the similarity measure of the gray information or feature information of the input image [2][3] This type of method usually has high registration accuracy, but often requires a lot of time and computing resources to optimize, which is very inconvenient in clinical diagnosis and therefore loses its competitive advantage in practical applications. In recent years, with the development of deep learning, some methods [4][5][6] based on deep learning have been applied to image registration, and many remarkable results have been achieved.

In this article, we follow the basic framework of learning-based medical image registration, and propose an image registration method fusing multi-scale deformation fields. This method superimposes fixed images and moving images and inputs them into a convolution neural network. In the process of convolution, different scale samples are taken, and different scale feature maps are
obtained by optimizing the parameter function. The fusion of feature images of different scales is conducive to the accurate deformation of the image and better image registration effect.

2. Related work

Traditional image registration algorithms have been paid attention to by many scholars in the past decades. Rangarajan et al. [7] first extracted the key shape feature points in the two images and then used the mutual information method to register the two image pairs. Rohr et al. [8] improved on the thin-plate spline interpolation method and realized the elastic registration of two images. Although these methods can achieve good registration accuracy, the amount of calculation is complicated and requires a lot of time. Therefore, in recent years, more scholars have begun to explore the use of deep learning methods for image registration. This makes the registration method based on learning a research hotspot. Such methods are usually divided into two types: supervised learning and unsupervised learning. Image registration based on supervised learning generally requires the use of anatomical information annotated by experts or traditional methods to generate simulated deformation fields, and use this information as ground truth to guide image registration. Cao et al. [9] introduced a balanced active point guided sampling strategy on the limited image data set to facilitate the accurate learning of CNN model. In recent years, many image registration methods based on unsupervised learning have emerged because these methods need additional auxiliary information and it is not easy to obtain the information. Voxelmorph [10] [11] is a typical representative of unsupervised learning based registration. It proposes a solution based on unsupervised learning, which does not need to obtain information such as ground truth correspondence or anatomical landmarks during training. Voxermorph uses convolutional neural network to extract information and optimizes the parametric function to achieve image registration.

3. method

Deformable medical image registration is described as the optimization problem of parametric function, because it has the effective ability to establish dense correspondence between images. Let $F$ and $M$ represent fixed image and moving image respectively, and deformable image registration can be described as:

$$\hat{\phi} = \arg\min( L(F, M \circ \phi) + \lambda L_p(\phi))$$

Where $M \circ \phi$ is the image of the moving image deformed by the deformation field $\phi$, $L_R(\phi)$ is the regularization term, which can make the deformation field smoother, and $\lambda$ is the parameter to measure the regularization degree.

![Figure 1](image.png)

As shown in Figure 1, we connect the fixed image and the moving image as input. The registration network outputs the registration deformation field, and then uses the spatial transformation layer [12]...
to deform the moving image to the registered image. The loss function is calculated between the deformed image and the fixed image to optimize the network parameters.

3.1. Network Architecture

As shown in the figure the fixed image and the moving image are concatenated and input into the neural network, we use 3×3×3 convolutional layers (stride is 2) followed by a ReLU activation function to implement downsampling. The relative value of the size of the image feature map and the number of channels are marked in the figure. For example, 1/2,16 means that the size of the feature map at this time is half of the original size and the current number of channels is 16. In order to fuse the feature map information of multiple scales, we sample the feature maps of different sizes into the same size and use the summation method to fuse the feature maps. Feature maps with small resolutions can often capture more global information, which is helpful to improve the registration of large deformed regions in human tissues. Feature maps with large resolutions often focus on the contextual information of local details and can more fully express tiny anatomical structures. Feature maps of different scales can be combined to obtain multi-scale information, which is useful for improving the information extraction ability of neural networks.

3.2. Loss Function

The loss function is composed of image similarity items and regular items. We use the cross-correlation function (CC) commonly used in the field of medical image registration to evaluate the similarity between image pairs. The higher the CC value, the higher the similarity between the two images. The image phase loss between the registered image and the fixed image can be described as follows:

\[ L(F,M;\phi) = -CC(F,M;\phi) \]  

where:

\[ L(F,M;\phi) \]
In order to make the deformation field smoother and prevent overfitting to a certain extent, we use a regular term:

\[ L_\lambda (\varphi) = \| \nabla \varphi \|^2 \]  

(4)

The total loss function is as follows:

\[ L = -CC(F, M \circ \varphi) + \lambda L_\lambda(\varphi) \]  

(5)

4. Experiments

4.1. Data set

ABIDE (Autism Brain Imaging Data Exchange) [13] is a large-scale evaluation data set of the internal brain structure of autism patients. The data includes more than 1,046 collected from patients with autism spectrum disorders and typical controls aged 7 to 64 years. Samples. We need to perform some preprocessing on the data: first use FreeSurfer [14] to strip the human brain image, and then the size of the image is uniformly cropped to 160*192*224, and the resolution is resampled to the same resolution. We use ANTs [15] to affine the image pairs. During the experiment, we used 1569 images for training and 100 images for testing.

4.2. Results

We use Dice Similarity coefficient in the ABIDE dataset to evaluate the proposed method and compare it with some classic image registration methods such as affine, Symmetric Normalization (SyN) [16] and voxelmorph. Table 1 shows the average DSC of our proposed method, affine alignment, and SyN.

![Figure3. One sample results of the proposed method.](image)

Table 1. Average DSC of affine alignment, SyN, Voxel morph and our method.

|                      | Affine alignment | SyN    | Voxel morph | Ours   |
|----------------------|------------------|--------|-------------|--------|
| DSC(%)               | 58.81            | 58.64  | 71.69       | 71.79  |
Figure 3 is a sample we randomly selected and shows the visual image registration results from three angles of sagittal plane, coronal plane, and transverse plane.

In order to further measure the effect of the proposed method, we selected 15 important anatomical structures in the human brain image, and calculated the dice coefficients of the method on these anatomical structures. The experimental results show that our method is compared with the classic affine alignment and SyN methods. Excellent registration effects have been achieved on these cross-sectional structures.

| Name | Name | Name |
|------|------|------|
| 1. Left-Cerebral-White-Matter | 6. Left-Thalamus-Proper | 11. 4th-Ventricle |
| 2. Left-Cerebral-Cortex | 7. Left-Caudate | 12. Brain-Stem |
| 3. Left-Lateral-Ventricle | 8. Left-Putamen | 13. Left-Hippocampus |
| 4. Left-Cerebellum-White-Matter | 9. Left-Pallidum | 14. Left-Amygdala |
| 5. Left-Lateral-Ventricle | 10. 3rd-Ventricle | 15. CSF |

Figure 4. The average DSC evaluated on 15 ROIs.

5. conclusion

In this article, we propose an unsupervised and deformable medical image registration method. This method facilitates the neural network to capture more global and local information by fusing feature maps of multiple scales, thereby improving the matching of medical images. The experimental results show that our method has better results than some state-of-the-art methods.

acknowledgments

This research was supported by Project of Jinan Scientific Research Leader’s Laboratory (No. 2018GXRC023), Natural Science Foundation of Shandong province (Nos. ZR2019MF 013, ZR2019BF026)

References

[1] Zhang Z, Sejdić E 2019 Radiological images and machine learning: trends, perspectives, and prospects. Computers in biology and medicine 108: pp 354–370

[2] Khader M, Schiavi E, Hamza AB 2017 A multicomponent approach to nonrigid registration of diffusion tensor images. Applied Intelligence 46(2): pp 241–253
[3] Maes F, Collignon A, Vandermeulen D, Marchal G, Suetens P 1997 Multimodality image registration by maximization of mutual information. IEEE transactions on Medical Imaging 16(2): pp 187–198

[4] Fan J, Cao X, Yap PT, Shen D 2019 Birnet: Brain image registration using dual-supervised fully convolutional networks. Medical image analysis 54: pp 193–206

[5] de Vos BD, Berendsen FF, Viergever MA, Staring M, Işgum I 2017 End-to-end unsupervised deformable image registration with a convolutional neural network. Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support pp 204–212

[6] Liao, R., Miao, S., Tourneire, P., Grbic, S., Kamen, A., Mansi, T., and Comaniciu, D. 2016. An artificial agent for robust image registration. In national conference on artificial intelligence (2016), pp 4168-4175.

[7] Rangarajan, A., Chui, H., Duncan, J. S. 2000. Rigid point feature registration using mutual information. Medical Image Analysis,3(4), pp 425-440.

[8] Rohr, K., Stiehl, H. S., Sprengel, R., Buzug, T. M., Weese, J., & Kuhn, M. H.. 2002. Landmark-based elastic registration using approximating thin-plate splines. IEEE Transactions on Medical Imaging,20(6), pp 526-534.

[9] Cao, X., Yang, J., Zhang, J., Nie, D., Kim, M., Wang, Q., and Shen, D. 2017. Deformable image registration based on similarity-steered cnn regression. In International Conference on Medical Image Computing and Computer-Assisted Intervention, pp 300–308.

[10] Balakrishnan G, Zhao A, Sabuncu MR, Guttag J, Dalca AV 2019 Voxelmorph: a learning framework for deformable medical image registration. IEEE transactions on medical imaging 38(8): pp 1788–1800

[11] Balakrishnan, G., Zhao, A., Sabuncu, M.R., Guttag, J., and Dalca, A.V. 2018. An unsupervised learning model for deformable medical image registration. In Proceedings of the IEEE conference on computer vision and pattern recognition (2018), pp 9252-9260.

[12] Jaderberg M, Simonyan K, Zisserman A, et al. 2015 Spatial transformer networks. Advances in neural information processing systems pp 2017–2025

[13] Di Martino A, Yan CG, Li Q, Denio E, Castellanos FX, Alaerts K, Anderson JS, Assaf M, Bookheimer SY, Dapretto M, et al. 2014 The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. Molecular psychiatry 19(6): pp 659–667

[14] Fischl B 2012 Freesurfer. Neuroimage 62(2): pp 774–781

[15] Avants BB, Tustison NJ, Song G, Cook PA, Klein A, Gee JC 2011 A reproducible evaluation of ants similarity metric performance in brain image registration. Neuroimage 54(3): pp 2033–2044

[16] Avants BB, Epstein CL, Grossman M, Gee JC 2008 Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. Medical image analysis 12(1): pp 26–41