2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

An Unsupervised Convolution Neural Network for Deformable Registration of Mono/Multi-Modality Medical Images

Xianyu Wang, Guochen Ning, Ne Yang, Xinran Zhang, Hui Zhang, Hongen Liao*, Senior Member, IEEE

Abstract— **Image registration is a fundamental and crucial step in medical image analysis. However, due to the differences between mono-mode and multi-mode registration tasks and the complexity of the corresponding relationship between multi-mode image intensity, the existing unsupervised methods based on deep learning can hardly achieve the two registration tasks simultaneously. In this paper, we proposed a novel approach to register both mono- and multi-mode images in a same framework. By approximately calculating the mutual information in a differentiable form and combining it with CNN, the deformation field can be predicted quickly and accurately without any prior information about the image intensity relationship. The registration process is implemented in an unsupervised manner, avoiding the need for the ground truth of the deformation field. We utilize two public datasets to evaluate the performance of the algorithm for mono-mode and multi-mode image registration, which confirms the effectiveness and feasibility of our method. In addition, the experiments on patient data also demonstrate the practicability and robustness of the proposed method.**

I. Introduction

Accurate integration of complementary information from different medical images is essential for assisting doctors in disease diagnosis and treatment, and the premise is that images are spatially aligned. Therefore, image registration is clinically significant [1]. However, the image registration problem is usually ill-posed, and medical images are susceptible to noise and artifacts. Besides, the non-linear correspondence between intensities of different mode images increases the difficulty of image registration. Therefore, medical image registration remains a challenging problem [2, 3].

In order to restore various deformations in medical images, researchers have made numerous efforts in traditional registration methods, and proposed many non-rigid registration algorithms, such as B-spline [4] and Large diffeomorphic distance metric mapping (LDDMM) [5]. However, such traditional registration methods require complex optimization and are time-consuming. Recently, to improve the computational speed of traditional registration methods, estimating the deformation field by neural networks has received much attention. Balakrishnan *et al*. [6] proposed an unsupervised CNN-based deformable registration algorithm (Voxelmorph). The registration of T1 MRI of different individuals was achieved by optimizing the cross-correlation between fixed and moving images. Later, other researchers suggested some improvement strategies, which enhanced the ability to deal with large deformation [7], or avoided the spatial folding of moving images [8]. Zhu *et al.* [9] integrated non-rigid registration network and affine alignment network to achieve actual end-to-end registration. But these methods are all for the mono-mode image registration. Fan *et al.* [10] developed an adversarial similarity network for multi-modality image registration, which introduced a discriminator network to determine whether the images were well aligned, and the registration network was trained under the guidance of the feedback of the discriminator. Although this method can realize multi-mode registration, the discriminator should be trained with aligned images, limiting its practical application.

Generally, the registration of mono-mode images collected at different times is helpful for disease follow-up, and the multi-mode image registration is beneficial for integrating multi-scale information for disease diagnosis. Although the DL-based image registration algorithms have been widely studied, as far as we know, and there is still a lack of method that can achieve mono-modality and multi-modality image registration at the same time. To address the aforementioned issues, we proposed an unsupervised convolutional neural network to complete fast and accurate deformable registration of mono-/multi-mode images in the same framework. The effectiveness and feasibility of the proposed method were verified on two public datasets, and the practicability and robustness were confirmed by experiments on patient data.

II. Methods

A. Problem Formulation

The proposed deformable registration framework that can achieve two types of registration tasks is shown in Fig.1. Let I_M, I_F be two sets of image data in 3D space, which are moving and fixed image in the registration task, respectively. The purpose of registration is to determine the deformation field Φ, so that the deformed moving image can be accurately aligned with the reference image. We assume that the fixed and the moving image have been affine registered during the preprocessing stage, so the sources of misalignment between them are only non-rigid deformations. Therefore, image registration can be expressed as such an optimization problem:

$$\widehat{\Phi} = arg \min_{\Phi} \mathcal{L}(I_F, I_M, \Phi) \tag{1}$$

$$\mathcal{L}(I_F, I_M, \Phi) = \mathcal{L}_{sim}(I_F, I_M \circ \Phi) + \lambda \mathcal{L}_{reg}(\Phi) \tag{2}$$

where, $\mathcal{L}_{sim}$ is the similarity loss, and $\mathcal{L}_{reg}$ is the regularizer term. In DL-based methods, the deformation field is usually expressed as $\Phi = f_\theta(I_F, I_M)$, and θ is the learnable parameters.

B. Deformable Registration Network

Research supported by National Natural Science Foundation of China (82027807, 81771940), and Beijing Municipal Natural Science Foundation (7212202).

X. Wang, G. Ning, N. Yang, X. Zhang, H. Zhang and H. Liao are with the department of Biomedical Engineering School of Medicine, Tsinghua University, Beijing 100084, China. (email: liao@tsinghua.edu.cn)

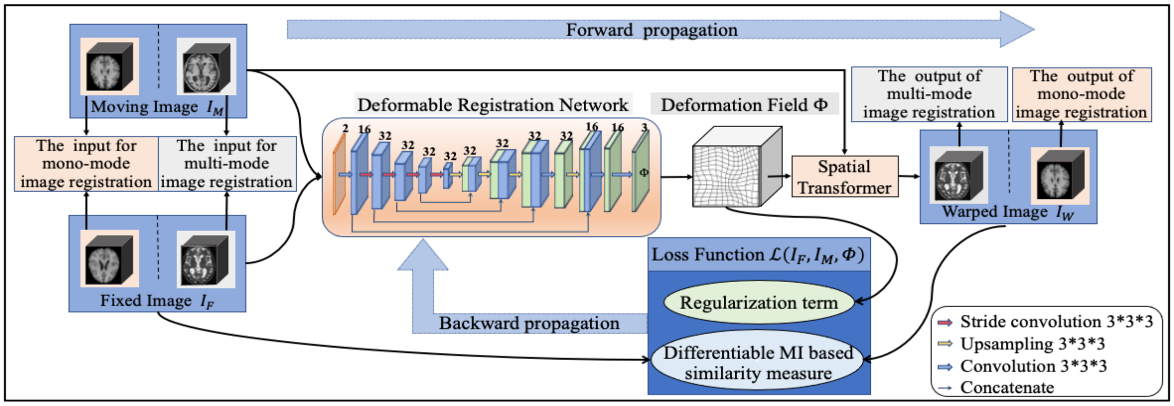
Forward propagation
Moving Image I_M
The input for mono-mode image registration
The input for multi-mode image registration
Fixed Image I_F
Deformable Registration Network
2
16
32
32
32
32
32
32
32
32
16
16
3
Φ
Deformation Field Φ
Spatial Transformer
The output of multi-mode image registration
The output of mono-mode image registration
Warped Image I_W
Loss Function $\mathcal{L}(I_F, I_M, \Phi)$
Regularization term
Differentiable MI based similarity measure
Backward propagation
Stride convolution 3*3*3
Upsampling 3*3*3
Convolution 3*3*3
Concatenate

Figure 1. Illustration of the proposed deformable registration framework. The input fixed and moving image can be in the same mode or different mode.

The deformable registration network is used to correct the local misalignment. As shown in Fig.1, the encoder down-samples the input images by stride convolution, and extracts the features at multiple resolutions. The decoder performs up-sampling and convolution to restore the resolution of the feature map. We performed Leaky ReLU activation after each convolution layer, and the last convolution layer converts the results into a flow field. The kernel size is $3 \times 3 \times 3$. The features of different receptive fields are transmitted to the decoder through the skip-connection, which can achieve multi-scale feature fusion and improve the robustness of the network.

C. Differentiable MI-based loss function

The proposed framework for both types of registration tasks obtains the optimal parameters by minimizing the loss function. To calculate the loss value in real-time during the training stage, we utilize the spatial transformer (STN) to obtain the deformed moving image $I_M(\Phi)$ [11].

It is well known that multi-modality images have more complicated intensity correspondences than mono-modality, which causes the failure to promote mono-mode registration methods to multi-mode registration tasks. To solve this problem, we harnessed mutual information (MI) as the similarity measure, which is the key to ensure the success of the experiments. The definition of MI is as follows:

$$MI(I_M, I_F) = \sum_{m,f} p(m,f) log\,\frac{p(m,f)}{p(m)p(f)} \qquad (3)$$

m and f are the gray values in the moving and the fixed image, respectively. $p(m)$ and $p(f)$ are the marginal distributions, and $p(m,f)$ is their joint distribution. The better the alignment, the greater the MI. But the discrete MI cannot be used in the optimization, so we approximate it in a differentiable form.

We divided the image into several gray levels. Each voxel contributes not only to the marginal distribution of the gray level it falls into, but also a series of gray levels. Parzen density estimation was employed to describe this process. For a given series of sample B, the contribution of each sample b to $p(m)$ is a function of its distance from m:

$$p_B(m) = \frac{1}{N}\sum_{b\in B} W(m-b) \qquad (4)$$

and Gaussian function was adopted as the weight function:

$$\mathrm{W(m-b)} = \left(1/\sigma\sqrt{2\pi}\right)\mathrm{e}^{-\frac{(\mathrm{m-b})^2}{2\sigma^2}} \qquad (5)$$

Similarly, the joint probability distribution can be obtained:

$$p_{B,C}(m,f) = \frac{1}{N}\sum_{b\in B,c\in C} W(m-b)W(f-c) \qquad (6)$$

By substituting (4)-(6) into (3), the differentiable MI can be obtained. Therefore, the similarity loss is:

$$\mathcal{L}_{sim}(I_F, I_M \circ \Phi) = -MI(I_M(\Phi), I_F) \qquad (7)$$

Besides, we exploited the spatial gradient of the deformation field to regularize its smoothness and continuity:

$$\mathcal{L}_{reg}(\Phi) = \sum_{p\in\Omega} \|\nabla\Phi(p)\|^2 \qquad (8)$$

Thus, the entire loss function is shown in (2), and λ is the weight of the regularization. The network was trained by minimizing the loss function with standard back-propagations.

III. Experiments and Results

A. Dataset and implementation details

We first evaluated the proposed algorithm on two public datasets, LPBA40 [12] and IXI [13]. LPBA40 includes 40 T1w MRI volumes, and each volume has a segmentation mask with 56 anatomical labels. 992 pairs of data composed of 32 volumes were used as the training set, and 56 pairs of data consisting of 8 volumes were used as the test set. Multi-modality registration used 483 groups of T1w and T2w MRI images in the IXI dataset, and the ratio of training to test data is 4:1. Firstly, we performed a standardized preprocess on all IXI data, including spatial resample, crop and affine alignment.

For mono-mode registration, we compared the proposed method with the classic Voxelmorph. For multi-mode image registration, our method was compared with two of the state-of-the-art algorithms: SyN [14] and B-spline [4]. The two methods were implemented by ANTs and Elastix, respectively, and MI was used as the similarity measure. Notably, when reproducing the algorithms, we tried various parameters and finally selected the one with the best registration effect.

B. Results on public dataset

In the experimental stage, the proposed deformable registration framework for both mono- and multi-mode images was evaluated qualitatively and quantitatively. The speed and accuracy were compared with the baseline methods.

The qualitative results of mono-mode and multi-mode image registration are shown in Fig. 2 and Fig. 3, respectively. It can be seen intuitively that the results gained by our method are fully aligned with the fixed images in key structures, and the registration effect is better than that of baseline methods.

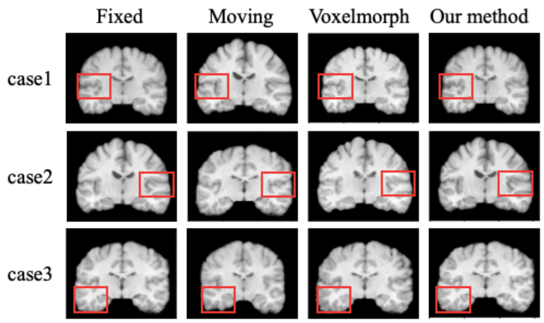

Figure 2. Comparison of mono-mode registration results obtained different methods. At the position indicated by the red boxes, the proposed method obtained better effect than the baseline method.

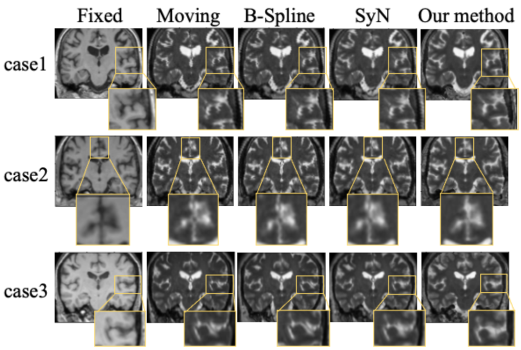

Figure 3. Comparison results of various methods for multi-mode image registration. The areas marked by yellow boxes show that the proposed algorithm has a significant improvement over the baseline methods.

In addition, we use the Dice similarity coefficient (DSC) of the warped and the reference segmentation mask to quantitatively compare the performance of mono-mode registration, and employ MI to evaluate the results of multi-mode registration. Besides, the structural similarity (SSIM) and peak signal-to-noise ratio (PSNR) are also used to quantify the registration effect of mono- and multi-mode. As Table 1 and Table 2 show, deformable transformation can significantly improve the registration effect, and the proposed method can obtain better quantitative indicators than the existing methods.

Registration speed is another crucial factor in evaluating the practicality of registration algorithms. As shown in Table 1, we counted the running time of the mono-mode registration algorithms on NVIDIA GeForce GTX 1080Ti GPU. Both methods can complete the entire registration process within 0.4s. As shown in Table 2, since ANTs and Elastix software do not have GPU versions, to make a more intuitive comparison, we measured the running time of our method on Intel Core i7-7800X CPU, and the average speed is more than 18 times faster than B-Spline and 300 times faster than SyN.

C. *Results on patient data*

We also validated in practical clinical applications. Due to the limited amount of patient data, we did not retrain the network, but directly tested the trained model on patient data.

A typical clinical application of mono-mode image registration is disease follow-up. In the chemotherapy of brainstem glioma, the tumor volume change is vital for assisting doctors in evaluating efficacy. However, to obtain the tumor volume, doctors need to manually mark the tumor boundary on each MRI scan, which is time-consuming and laborious. Therefore, we hope to realize the labeling of all data based on the tumor label of the baseline image.

Approved by the ethics committee of Beijing Tiantan Hospital Affiliated to Capital Medical University and the informed consent of patient, we collected the pre and post-chemotherapy MRI image of a brainstem glioma patient, which were used as the moving and the fixed image respectively, and the predicted deformation field was applied to the baseline tumor labels. The qualitative results were shown in Fig. 4, the difference between the fixed and moving image is significantly reduced, which shows that the deformation field accurately describes the misalignment between the two images, which ensure the accuracy of the warped tumor label, and the DSC between the predicted results and the ground truth can reach 88.70%.

Multi-mode registration is usually used to integrate the information of different modality images to assist doctors in disease diagnosis. Taking Alzheimer's disease (AD) as an example, intracranial vasoconstriction is an early symptom of AD. In the diagnostic images, T1w images can reflect the atrophy pattern and degree, and T2w images can reveal the degree of vascular changes. Hence, comprehensive analysis of the changes in brain tissue and vascular are contributed to the early diagnosis of AD. But, during image acquisition, many factors such as respiratory and head movement may cause non-rigid deformations, which will lead to errors in image fusion.

To verify the practicability of our method in solving this problem, three groups of patient data were randomly selected from the OASIS-3 [15] dataset, and the obtained results are shown in Fig.5. The qualitative results indicated that our method could align image details, and the average MI value was also improved from 0.5464 to 0.7692. It can also be seen that the structure in the post-registration difference map is clearer and closer to the reference image, which demonstrates that the structure of the warped image corresponds to that of fixed image, and only the difference of image intensity exists.

IV. Discussion and Conclusions

Due to the difference between various registration tasks and the complexity of the intensity relationship between multi-mode images, it is a challenging task to complete the two registration tasks simultaneously in an unsupervised manner. In this article, we proposed a novel registration algorithm. By introducing the differentiable MI as the similarity measure, the problem of intensity correspondence between different mode images was effectively solved. Meanwhile, the linear relationship between mono-mode image intensities was also taken into

Table 1. Quantitative comparison of the speed and accuracy of mono-mode image registration using the proposed method and the baseline method.

Method	**Quantitative evaluation index**				
	DSC	***SSIM***	***PSNR***	***CPU time(s)***	***GPU time(s)***
Affine	0.5207 (0.0022)	0.9828 (0.0003)	16.6749 (1.8787)	0	0
Voxelmorph	0.5766 (0.0015)	0.9927 (0.0001)	26.7577 (4.8810)	11.1556 (0.0390)	0.3683 (0.0010)
Our Method	**0.5821 (0.0016)**	**0.9928 (0.0001)**	**27.0761 (5.0311)**	**11.0296 (0.0932)**	**0.3533 (0.2158)**

Table 2. Quantitative comparison of the speed and accuracy of multi-mode image registration using the proposed method and the baseline methods.

Method	Quantitative evaluation index				
	MI	***SSIM***	***PSNR***	***CPU time(s)***	***GPU time(s)***
Affine	0.6626 (0.0099)	0.8072 (0.0100)	12.0913 (1.2901)	0	0
B-Spline	0.7181 (0.0113)	0.8441 (0.0076)	12.5352 (1.5859)	247.7010 (41.7766)	
SyN	0.8842 (0.0110)	0.8476 (0.0088)	12.6835 (1.5199)	4130.6250 (558.3230)	
Our Method	**1.0174 (0.0080)**	**0.8550 (0.0083)**	**12.8267 (1.9141)**	**13.2584 (1.0252)**	**0.4709 (0.0086)**

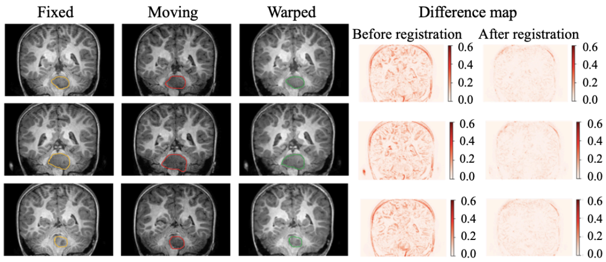

Figure 4. The image registration results of brainstem glioma patients and the tumor segmentation results based on the proposed strategy. Yellow: the manually delineated tumor boundary in the post-chemotherapy image. Red: the tumor boundary in the pre-chemotherapy image. Green: the tumor boundary obtained by the proposed segmentation strategy

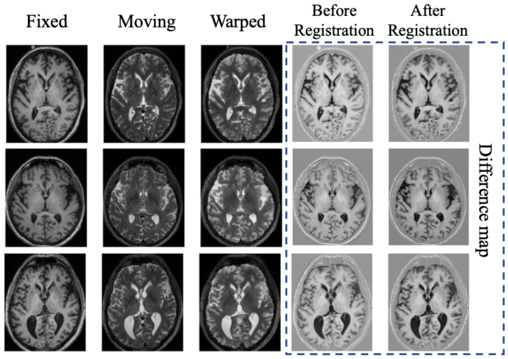

Figure 5. The registration results of T1w and T2w MRI images for Alzheimer's patients and the difference maps.

account, which reduced the development cost of designing various algorithms for different registration tasks.

In previous works, Zhu *et al.* [16] proposed a multi-mode image registration algorithm based on structural representation. Although this method has the potential for mono-mode registration, the accuracy relies heavily on the effect of feature extraction. Xu *et al*. [17] utilized the improved Cycle-GAN to generate images from one mode to another, and used the texture information to constrain the predicted deformation field. However, the accuracy of image generation is difficult to guarantee. Therefore, we predict the deformation field directly based on original images is more robust and universal.

Nevertheless, there are still areas that need improvement. First, we have only conducted experiments on a small clinical dataset, and directly used models trained on public data. Further experimental verification is needed on enough patient data. Second, registering the structural and functional images to get the anatomical position of functional information is a more challenging task, which is essential for diagnosing neurodegenerative diseases, so it should be further studied.

In conclusion, we demonstrated the performance of our method on the same and different mode brain images. The algorithm can register images without the ground truth of the deformation field. Moreover, the results showed that the method was compatible with two registration tasks, and can obtain comparable performance with the most representative methods. Meanwhile, the experiments on patient data also demonstrated the capability of the method in clinical problems. Besides, although the patient data were distinctly different from the data of normal volunteers, a stable result was still obtained, which verified the robustness of the method.

REFERENCES

[1] Maintz J B A, Viergever M A. "A survey of medical image registration." Medical image analysis, vol. 2, no. 1, pp. 1-36, 1998.

[2] Oliveira F P M, Tavares J M R S. "Medical image registration: a review." Computer methods in biomechanics and biomedical engineering, vol. 17, no. 2, pp. 73-93, 2014.

[3] El-Gamal F E Z A, Elmogy M, Atwan A. "Current trends in medical image registration and fusion." Egyptian Informatics Journal, vol. 17, no. 1, pp. 99-124, 2016.

[4] Xie Z, Farin G E. "Image registration using hierarchical B-splines." IEEE Transactions on visualization and computer graphics, vol. 10, no. 1, pp-85-94. 2004.

[5] Beg M F, Miller M I, Trouv´e A, et al. "Computing large deformation metric mappings via geodesic flows of diffeomorphisms." International Journal of Computer Vision, vol. 61, no. 3, pp. 139-157, 2005

[6] Balakrishnan G, Zhao A, Sabuncu M R, et al. "An unsupervised learning model for deformable medical image registration." Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 9252-9260, 2018.

[7] Zhao S, Dong Y, Chang E I, et al. "Recursive cascaded networks for unsupervised medical image registration." Proceedings of the IEEE International Conference on Computer Vision, pp. 10600-10610, 2019.

[8] Kuang D, Schmah T. "Faim–a convnet method for unsupervised 3d medical image registration." International Workshop on Machine Learning in Medical Imaging. Springer, Cham, pp. 646-654, 2019.

[9] Zhu Z, Cao Y, Qin C, et al. "Unsupervised 3D End-to-end Deformable Network for Brain MRI Registration." 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). IEEE, pp. 1355-1359, 2020.

[10] Fan J, Cao X, Wang Q, et al. "Adversarial learning for mono-or multi-modal registration." Medical image analysis, vol. 58: 101545, 2019.

[11] Jaderberg M, Simonyan K, Zisserman A, et al. "Spatial transformer networks." Advances in neural information processing systems. 2015.

[12] Shattuck D W, Mirza M, Adisetiyo V, et al. "Construction of a 3D probabilistic atlas of human cortical structures." Neuroimage, vol. 39, no. 3, pp. 1064-1080, 2008.

[13] IXI -Information eXtraction from images. www.brain-development.org

[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." Medical Image Analysis, vol. 12, no. 1, pp. 26-41, 2008.

[15] LaMontagne P J, Benzinger T L S, Morris J C, et al. "OASIS-3: longitudinal neuroimaging, clinical, and cognitive dataset for normal aging and Alzheimer disease." MedRxiv, 2019.

[16] Zhu, X., Ding, M., Huang, T., Jin, X., & Zhang, X. "PCANet-based structural representation for non-rigid multimodal medical image registration." Sensors, vol.18, no.5, pp. 1477, 2018

[17] Xu Z, Luo J, Yan J, et al. "Adversarial uni-and multi-modal stream networks for multimodal image registration." International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham, pp. 222-232, 2020.