DOI: 10.53469/jtpes.2024.04(02).01

Deep Learning-Based Medical Image Registration Algorithm: Enhancing Accuracy with Dense Connections and Channel Attention Mechanisms

Yulu Gong[1,*], Houze Liu[2], Lianwei Li[3], Jingxiao Tian[4], Hanzhe Li[5]

[1]Computer & Information Technology, Northern Arizona University, Flagstaff, US
[2]Computer science, New York University, New York, US
[3]Computer science, University of Texas at Arlington, Arlington, US
[4]Electrical and Computer Engineering, San Diego State University, San Diego, US
[5]Computer Engineering, New York University, New York, US
**Correspondence Author, rowan199408@gmail.com*

Abstract: ***In critical clinical medical image analysis applications, such as surgical navigation and tumor monitoring, image registration is crucial. Recognizing the potential for enhanced accuracy in existing unsupervised image registration techniques for single-modal imagery, this research introduces an innovative deep learning-based image registration algorithm. Its novelty resides in integrating short and long connections to create a densely connected structure, markedly refining the feature map interconnectivity within the U-Net architecture. This advancement addresses the significant semantic gap issues arising from disparities in feature map sampling depths. Moreover, the algorithm incorporates a channel attention mechanism within the U-shaped network's decoder, significantly mitigating image noise and facilitating the generation of smoother deformation fields. This enhancement not only boosts the model's detail sensitivity but also markedly increases image registration precision, particularly evident when processing single-modal brain MRI datasets, thereby proving the algorithm's efficacy and utility. Extensive clinical application-based training and testing have underscored this algorithm's substantial contributions to medical image registration accuracy enhancement. Overall, by leveraging deep learning technologies and innovative algorithmic structures, this study addresses pivotal challenges in medical image registration, offering more precise and dependable support for clinical applications like surgical navigation and tumor surveillance.***

Keywords: Unsupervised Deep Learning, Medical Image Registration, Deep Learning, Convolutional Neural Network.

1. INTRODUCATION

The employment of computer vision for medical image processing plays a pivotal role in clinical medicine, significantly enhancing the efficiency of disease diagnosis by physicians and mitigating the risk of misdiagnosis due to visual fatigue. For instance, medical images captured by devices are utilized for surgical navigation and observing the growth of tumors through images taken at different times, among other clinical applications [1][2]. Without exception, the processing of these images involves registration alignment, that is, the fusion of two images to maintain consistency in spatial coordinates.

In recent years, deep learning methods, capable of enhancing iterative improvement and intensity-based registration performance, have received extensive application in the field of medical image registration [3]. Registration based on deep learning is categorized into supervised learning [4] methods and unsupervised learning methods [5]. Both approaches employ neural networks to estimate transformation parameters, offering higher generalizability compared to traditional registration algorithms. Supervised learning refers to the need for labeled data samples during training, i.e., the true deformation field, utilizing neural networks to execute the registration process. Dai et al. [6] introduced a transformation model based on direct image appearance, predicting each module sequentially to achieve medical image registration. This approach inputs image pairs composed of fixed and moving images, obtaining the deformation field through the initial momentum decoded by a U-Net, which is used to resample the moving image as a reference fixed image. Xu et al. [7] proposed a method to directly acquire the deformation field using multi-scale convolutional neural networks (CNN), surpassing the registration accuracy of traditional B-spline methods. Unsupervised learning methods, which can train using original images and overcome the dependency on labeled data, were explored by Bob et al. [4] through the Deep Learning Image Registration (DLIR) framework for unsupervised affine and deformable image registration. This framework trains

CNN using the image similarity between image pairs without the need for label data, learning to predict transformation parameters to form the deformation field through image pair analysis. Balakrishnan et al. [8] employed a CNN similar to U-Net [9] to obtain the deformation field, naming the algorithm VoxelMorph. It demonstrated significant improvements in registration speed and accuracy, gaining widespread recognition in the medical image registration field. However, VoxelMorph still employs long connections similar to U-Net at the encoding-decoding structure, where the connected feature maps may have a large semantic gap, affecting subsequent registration accuracy.

Based on this, this paper proposes an unsupervised single-modality medical image registration algorithm based on deep learning, which implements short connections on top of the U-Net foundation, combined with the original long connection method. This approach retains the advantage of long connections in representing relationships between features at a distance while overcoming the drawback of a large semantic gap between connected features. The paper also designs a channel attention mechanism on the decoder of the U-Net to further enhance registration accuracy.

2. 2 RELATED WORK

The objective of medical image registration is to find the optimal spatial correspondence between a fixed image I_F and a moving image I_M , and iteratively update this correspondence in reverse through a loss function derived from the energy function of traditional registration methods:

$$\hat{\varphi} = \arg\min_{\varphi} E(I_M, I_F, \varphi) \tag{1}$$

In this context, φ denotes a spatial transformation, while φ represents the optimal transformation. The goal of registration is to minimize the loss function to optimize the spatial correspondence between the fixed image I_F and the floating image I_M, until the optimal registered image is obtained.

In the field of image registration, neural networks are commonly used to parameterize the spatial mapping relationships. The advantage of using neural networks lies in their ability to autonomously learn and optimize parameters by minimizing the loss function, thereby identifying the matching model between images. For registration tasks, convolutional neural networks, particularly U-shaped networks, are typically employed to generate deformation fields. These networks downsample the input images to capture the spatial correspondence between image pairs, followed by upsampling for image reconstruction. The aim is to identify key points within the images and eliminate noise and other irrelevant features.

Regarding the spatial transformation (i.e., the deformation field), it refers to the vector displacement field by which the floating image transforms towards the fixed image. This field reflects the displacement needed for the floating image to register with the fixed image. Subsequently, using the deformation field to perform warping and interpolation on the floating image yields the registered image.

3. METHODOLOGY

3.1 Overall Network Framework

This paper introduces an unsupervised, single-modality medical image registration algorithm based on deep learning, whose overall framework is depicted in Figure 1. It is important to note that the registration algorithm proposed in this paper is independent of image dimensionality. For demonstration purposes, this paper uses two-dimensional brain magnetic resonance images (MR) as an example, although the algorithm is equally applicable to three-dimensional images. Initially, the fixed image I_F and the floating image IM are input as a dual-channel image pair into the network, where the convolutional neural network extracts features from the images and generates an estimated vector displacement field, namely the deformation field φ . Then, the Spatial Transformer Networks (STN) [10] applies the deformation field to the input floating image

IM and performs interpolation, using bilinear interpolation for two-dimensional images, to produce the registered image $\widehat{I}_F$. The similarity measure L_{sim} between the fixed image I_F and the registered image $\widehat{I}_F$, along with the

smoothness of the deformation field L_{smooth}, serve as the objective functions to iteratively train the model and update the network parameters.

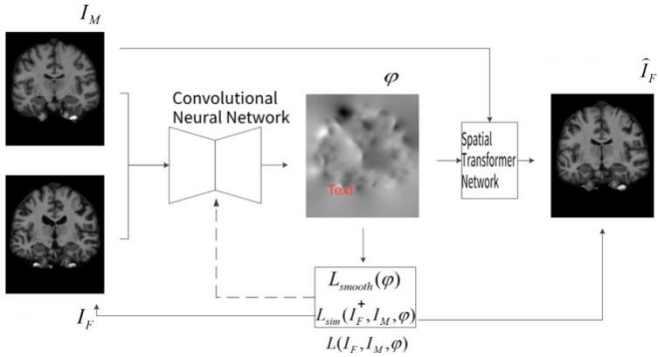

Figure 1: Overall framework

3.2 Network Architecture

The registration network proposed employs a convolutional neural network, specifically a U-shaped network akin to U-Net, comprising encoding and decoding phases. It extracts features from and transforms the input image pair to generate a deformation field. Additionally, a channel attention mechanism is deployed in the decoder to further enhance the realism of the transformed images produced by warping the moving image with the deformation field generated by the convolutional neural network. The structure of the registration network is illustrated in Figure 2.

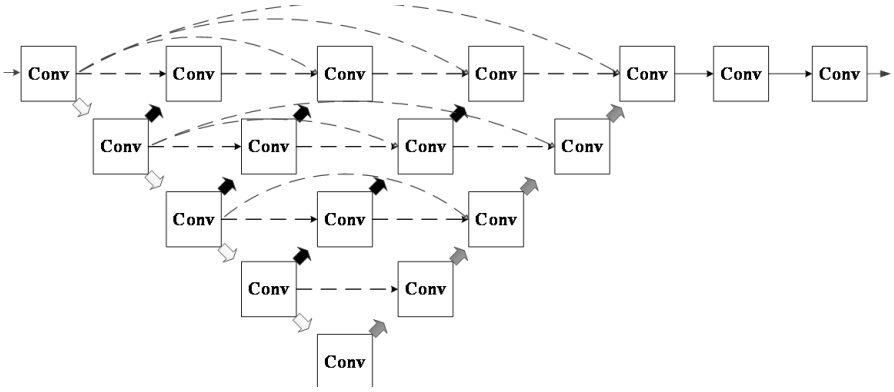

Figure 2: Registration network

In the diagram, "Conv" represents two-dimensional convolution and activation by the LeakyReLU function; white arrows indicate downsampling via max pooling (MaxPooling); black arrows denote upsampling through Upsampling; gray arrows are used to signify the addition of channel attention (Channel Attention) after upsampling; dashed arrows represent skip connections (long and short connections).

3.2.1 Dense U-Net

The convolutional neural network utilized in this study is a densely connected U-Net, an improvement based on the original U-Net architecture. The encoder performs downsampling to capture the spatial correspondence between image pairs, while the decoder performs upsampling for image reconstruction and obtaining the vector displacement field [11]. Typically, U-Net directly employs a simple long connection between the encoder and decoder, which can represent the relationship between two pixels that are far apart. However, direct connections may cause a significant semantic difference between two connected convolutional layers due to the large depth gap, increasing the network's learning difficulty and subsequently affecting registration accuracy [12]. To address this, the proposed method incorporates short connections with upsampling operations in the U-Net structure, reducing the semantic gap between feature maps with large depth differences.

Specifically, image pairs composed of fixed and floating images of the same modality are concatenated into a dual-channel and inputted into the convolutional neural network, with an input image size of 160 × 192. The encoder uses four convolution operations with a stride of 1 and a kernel size of 3, followed by MaxPooling for

downsampling, reducing the image resolution to 1/2, 1/4, 1/8, and 1/16 of the original image, respectively. The receptive field of convolution expands progressively, and then the LeakyReLU activation function is applied to capture the spatial correspondence between image pairs. The decoder performs upsampling following the same convolution operations as the encoder, allowing the image to be reconstructed to its original resolution. The network forms dense short connections by concatenating feature maps of the same resolution after each downsampling and upsampling operation. Additionally, long connections from the encoder to the decoder are deployed, reducing the semantic gap between two connected convolutional layers and representing the relationship between two pixels that are far apart.

3.2.2 Attention Module

The attention mechanism has been proven to enhance registration performance in registration algorithms [13]. Inspired by ECA-Net [14], this paper introduces a channel attention ECA module into the convolutional neural network. It emphasizes effective features and suppresses noise on a global scale, introducing only a minimal number of parameters to the network model while improving subsequent registration accuracy.

The ECA module is deployed in the decoder of the U-shaped network. First, the feature map W × H × C obtained by upsampling in the decoder undergoes Global Average Pooling (GAP) for channel compression, resulting in a dimension of 1 × 1 × C, where W, H represent the dimensions of the feature map, and C represents the number of channels. These dimensions are determined by the image size and channel number after each upsampling in the decoder. Then, the weight information for each channel is generated through a fast one-dimensional convolution with a kernel size of K, followed by a Sigmoid activation function, where K represents the coverage rate of local cross-channel interaction and should be adjusted based on the number of channels C. The formula for determining K is:

$$K = \psi(C) = \left| \frac{\log_2(C)}{r} + \frac{b}{r'} \right|_{\text{odd}} \tag{2}$$

In this context, \(r\) and \(b\) represent coefficients. In this study, the values are set as: \(r = 2\), \(b = 1\).

4. EXPERIMENTAL RESULTS

This study utilizes the OASIS (Open Access Series of Imaging Studies) dataset [15], a commonly employed dataset in the field of medical image registration, selecting brain cross-sectional MR data from 416 individuals spanning young, middle-aged, non-demented, and demented elderly demographics.

Initially, the images within the dataset were resampled to a size of 160 × 192. Subsequently, each MR image in the dataset underwent a series of standard preprocessing steps using FreeSurfer software [16], including motion correction, skull stripping, affine spatial normalization, and subcortical structure segmentation. Ultimately, the dataset was randomly divided into training and testing sets in a 4:1 ratio. Figure 4 showcases the original and preprocessed images.

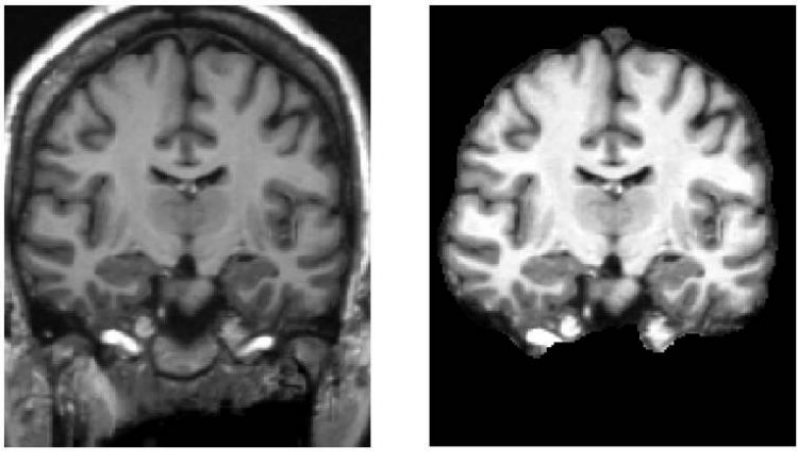

Figure 3: Images before and after preprocessing

4.2 Evaluation Metrics

For the assessment of registration results, this article employs the Dice Similarity Coefficient (DSC), a widely used evaluation metric in medical image registration. The DSC coefficient is utilized to calculate and assess the alignment level between two images, with a value range of [0,1]. A DSC value closer to 1 indicates better registration effectiveness [17]. The expression for the DSC coefficient is as follows:

$$DSC\left(s_F^k, s_M^k, \phi\right) = \frac{2|s_F^k \cap (s_M^k \circ \phi)|}{|s_F^k| + |s_M^k \circ \phi|} \tag{3}$$

Here, S_F and S_M denote the segmentation labels for the fixed image I_F and the moving image I_M, and k represents the k^{th} label; $S^k{}_M \circ \phi$ signifies the segmentation labels for the image $\widehat{I}_F$ after registration. In contrast to the interpolation method used during training for the moving image, the resampling for $S^k{}_M \circ \phi$ employs nearest neighbor interpolation to achieve the desired effect.

4.3 Experimental Setup

The experiments were conducted on an NVIDIA RTX 3090 graphics card, within a software environment of Python 3.8, using the Keras and TensorFlow frameworks to implement the proposed unsupervised registration model, with network optimization driven by the Adam optimizer. The experiments involved atlas-based registration of the dataset images, where all moving images are registered to a single fixed image. The initial learning rate was set to 10^{-4}, with 2000 epochs planned for the experiment. The weight hyperparameter σ in the loss function was set to 0.01, with a batch size of 8, meaning that in each iteration, 8 moving images and 8 reference images were randomly selected from the training set to form 8 pairs for training.

4.4 Experimental Results

To evaluate the performance of the proposed unsupervised registration algorithm, the selected comparison algorithms were Affine [18] and VoxelMorph. Notably, the convolutional neural network of the VoxelMorph algorithm also utilized a U-Net-like U-shaped network[19][20] to generate the deformation field. The experiments were conducted on two frameworks proposed by the algorithm, Vxm-1 and Vxm-2[21], to assess the registration effects of each algorithm through the DSC coefficients of 24 segmentation labels per image in the test set and through visualization of the results.

Table 1 presents the registration results of randomly selected 3 moving images to the same fixed image in the test set by different algorithms. It is observed that the method proposed in this article achieves higher DSC coefficients across all image groups compared to the comparison algorithms, indicating a significant advantage in registration accuracy of the proposed method over the comparative algorithms.

Table 1: Comparison of DSC coefficients of different algorithms

METHOD	Affine	Vxm-1	Vxm-2	The method proposed
1	0.5243 +/− 0.2741	0.7303 +/− 0.2298	0.7631 +/− 0.2060	0.7927 +/− 0.1453
2	0.5008 +/− 0.2757	0.7239 +/− 0.2290	0.7533 +/− 0.2105	0.7804 +/− 0.2030
3	0.5048 +/− 0.2448	0.7332+/− 0.2646	0.7552 +/− 0.2310	0.7833+/− 0.1680

5. CONCLUSION

This paper introduces an unsupervised registration method for brain unimodal magnetic resonance imaging (MRI) based on deep learning. By utilizing the differences between the deformed moving image and the fixed image, the method iteratively optimizes the parameters within the convolutional neural network (CNN) in reverse, eliminating the time and financial costs associated with manual annotation. The proposed method initially enhances the connectivity within the CNN by designing a densely connected U-Net, a U-shaped network similar to U-Net, which employs short connections between the encoder and decoder. This approach addresses the issue of significant semantic gaps[22][23] caused by large sampling depth differences between two connected convolutional layers while also preserving the long connection advantage of representing the relationship between two distantly located pixels. Furthermore, the method incorporates a channel attention mechanism during the upsampling phase in the decoder[24], which through feature recalibration, effectively emphasizes useful features

and suppresses noise during image reconstruction, thereby improving the registration effect in the subsequent generation of registered images. Experimental results demonstrate that the proposed unsupervised image registration algorithm outperforms comparative methods such as Affine and Voxelmorph in terms of the Dice Similarity Coefficient (DSC).

REFERENCES

[1] Liu, B., Yu, L., Che, C., Lin, Q., Hu, H., & Zhao, X. (2023). Integration and Performance Analysis of Artificial Intelligence and Computer Vision Based on Deep Learning Algorithms. arXiv preprint arXiv:2312.12872.

[2] Yan, X., Xiao, M., Wang, W., Li, Y., & Zhang, F. (2024). A Self-Guided Deep Learning Technique for MRI Image Noise Reduction. Journal of Theory and Practice of Engineering Science, 4(01), 109–117. https://doi.org/10.53469/jtpes.2024.04(01).15

[3] Weimin WANG, Yufeng LI, Xu YAN, Mingxuan XIAO, & Min GAO. (2024). Enhancing Liver Segmentation: A Deep Learning Approach with EAS Feature Extraction and Multi-Scale Fusion. International Journal of Innovative Research in Computer Science & Technology, 12(1), 26–34. Retrieved from https://ijircst.irpublications.org/index.php/ijircst/article/view/21

[4] De Vos, B.D., Berendsen, F.F., Viergever, M.A., et al. (2019) A Deep Learning Framework for Unsupervised Affine and Deformable Image Registration. Medical Image Analysis, 52, 128-143. https://doi.org/10.1016/j.media.2018.11.010

[5] Li, H. and Fan, Y. (2018) Non-Rigid Image Registration Using Self-Supervised Fully Convolutional Networks without Training Data. Proceedings of the 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018),Washington DC, 4-7 April 2018, 1075-1078. https://doi.org/10.1109/ISBI.2018.8363757

[6] Dai, W., Tao, J., Yan, X., Feng, Z., & Chen, J. (2023, November). Addressing Unintended Bias in Toxicity Detection: An LSTM and Attention-Based Approach. In 2023 5th International Conference on Artificial Intelligence and Computer Applications (ICAICA) (pp. 375-379). IEEE.

[7] Xu, H., & Colmenares, J. A. (2023). Admission Control with Response Time Objectives for Low-latency Online Data Systems. arXiv preprint arXiv:2312.15123.

[8] Balakrishnan, G., Zhao, A., Sabuncu, M.R., et al. (2019) VoxelMorph: A Learning Framework for Deformable Medi- cal Image Registration. IEEE Transactions on Medical Imaging, 38, 1788-1800. https://doi.org/10.1109/TMI.2019.2897538

[9] Ronneberger, O., Fischer, P. and Brox, T. (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Mu- nich, 5-9 October 2015, 234-241. https://doi.org/10.1007/978-3-319-24574-4_28

[10] Tianbo, S., Weijun, H., Jiangfeng, C., Weijia, L., Quan, Y., & Kun, H. (2023, January). Bio-inspired Swarm Intelligence: a Flocking Project With Group Object Recognition. In 2023 3rd International Conference on Consumer Electronics and Computer Engineering (ICCECE) (pp. 834-837). IEEE.

[11] Ni, F., Zang, H., & Qiao, Y. (2024, January). Smartfix: Leveraging machine learning for proactive equipment maintenance in industry 4.0. In The 2nd International scientific and practical conference “Innovations in education: prospects and challenges of today”(January 16-19, 2024) Sofia, Bulgaria. International Science Group. 2024. 389 p. (p. 313).

[12] Hao Xu, Qingsen Wang, Shuang Song, Lizy Kurian John, and Xu Liu. 2019. Can we trust profiling results? Understanding and fixing the inaccuracy in modern profilers. In Proceedings of the ACM International Conference on Supercomputing. 284–295.

[13] Sun, W., Wan, W., Pan, L., Xu, J., & Zeng, Q. (2024). The Integration of Large-Scale Language Models Into Intelligent Adjudication: Justification Rules and Implementation Pathways. Journal of Industrial Engineering and Applied Science, 2(1), 13–20. https://doi.org/10.5281/zenodo.10607564

[14] Pan, L., Sun, W., Wan, W., Zeng, Q., & Xu, J. (2023). Research Progress of Diabetic Disease Prediction Model in Deep Learning. Journal of Theory and Practice of Engineering Science, 3(12), 15-21.

[15] Marcus, D.S., Wang, T.H., Parker, J., et al. (2007) Open Access Series of Imaging Studies (OASIS): Cross-Sectional MRI Data in Young, Middle Aged, Nondemented, and Demented Older Adults. Journal of Cognitive Neuroscience, 19, 1498-1507. https://doi.org/10.1162/jocn.2007.19.9.1498

[16] Fischl, B. (2012) FreeSurfer. Neuroimage, 62, 774-781. https://doi.org/10.1016/j.neuroimage.2012.01.021

[17] Wan, W., Xu, J., Zeng, Q., Pan, L., & Sun, W. (2023). Development and Evaluation of Intelligent Medical Decision Support Systems. Academic Journal of Science and Technology, 8(2), 22-25.

[18] Hao Hu, Shulin Li, Jiaxin Huang, Bo Liu, and Chang Che. 2023. Casting Product Image Data for Quality Inspection with Xception and Data Augmentation. Journal of Theory and Practice of Engineering Science 3, 10 (Oct. 2023), 42–46. https: //doi.org/10.53469/jtpes.2023.03(10).06

[19] Liu, B. (2023). Based on intelligent advertising recommendation and abnormal advertising monitoring system in the field of machine learning. International Journal of Computer Science and Information Technology, 1(1), 17-23.. (2023). International Journal of Computer Science and Information Technology, 1(1), 17-23. https://doi.org/10.62051/ijcsit.v1n1.03

[20] Xu, J., Pan, L., Zeng, Q., Sun, W., & Wan, W. (2023). Based on TPUGRAPHS Predicting Model Runtimes Using Graph Neural Networks. Frontiers in Computing and Intelligent Systems, 6(1), 66-69..

[21] Ma, D., Dang, B., Li, S., Zang, H., & Dong, X. (2023). Implementation of computer vision technology based on artificial intelligence for medical image analysis. International Journal of Computer Science and Information Technology, 1(1), 69-76.

[22] Hengyi Zang. (2024). Precision Calibration of Industrial 3D Scanners: An AI-Enhanced Approach for Improved Measurement Accuracy. Global Academic Frontiers, 2(1), 27-37. https://gafj.org/journal/article/view/30

[23] Li, L., Yang, Y., Zhan, S., & Wu, B. (2021, May). Sentence dependent-aware network for aspect-category sentiment analysis. In International Conference on Web Engineering (pp. 166-174). Cham: Springer International Publishing.

[24] Hao Xu, Shuang Song, and Ze Mao. 2023. Characterizing the Performance of Emerging Deep Learning, Graph, and High Performance Computing Workloads Under Interference. arXiv:2303.15763