Journal of computer science and software applications
https://www.mfacademia.org/index.php/jcssa
ISSN:2377-0430
Vol. 4, No. 5, 2024

Enhanced Unsupervised Image Registration via Dense U-Net and Channel Attention

Yaxin Liang[1], Yuwei Zhang[2], Zhi Ye[3], Zexi Chen[4]

[1]University of Southern California, Los Angeles, USA

[2]Duke University, Durham, USA

[3]Elevance Health, Indianapolis, USA

[4]North Carolina State University, Raleigh, USA

Correspondence should be addressed to Yaxin Liang; yaseen.liang@outlook.com

Abstract:

In the realm of critical clinical medical image analysis, particularly in surgical navigation and tumor monitoring, the importance of precise image registration cannot be overstated. Acknowledging the need for improved accuracy in current unsupervised image registration methods for single-modal images, this research presents a groundbreaking deep learning-based algorithm. The core innovation of this algorithm lies in its integration of short and long connections, which establish a densely connected architecture within the U-Net framework. This approach significantly enhances feature map interconnectivity, effectively bridging the semantic gaps caused by varying sampling depths within the feature maps. Additionally, the algorithm introduces a channel attention mechanism within the U-shaped network's decoder, which plays a crucial role in reducing image noise and generating smoother deformation fields. This enhancement not only improves the model's sensitivity to finer details but also substantially increases the precision of image registration, a benefit particularly notable when working with single-modal brain MRI datasets. Extensive clinical trials have demonstrated the algorithm's significant contributions to enhancing the accuracy of medical image registration. In conclusion, by harnessing the power of deep learning and innovative algorithmic design, this study tackles critical challenges in medical image registration, providing more precise and reliable support for clinical applications such as surgical navigation and tumor monitoring.

Keywords:

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

1. Introduction

The integration of computer vision into medical image processing is crucial in clinical medicine, significantly enhancing the efficiency of disease diagnosis and reducing the risk of misdiagnosis due to physician fatigue. For instance, medical images obtained from various devices are essential in surgical navigation and in monitoring tumor progression through temporal imaging, among other clinical applications. Central to these processes is image registration alignment, which involves the fusion of two images to ensure consistency in spatial coordinates.

In recent years, deep learning techniques have been widely adopted in the field of medical image registration due to their ability to improve iterative processes and enhance intensity-based registration performance. These deep learning-based registration methods are generally classified into supervised [1] and unsupervised [2] learning approaches, both of which utilize neural networks to estimate transformation parameters, offering greater generalizability compared to traditional algorithms. Supervised learning methods rely on labeled datasets during training, specifically requiring the true deformation field, and use neural networks to perform the registration. For instance, Ballé et al. [3] introduced a transformation model based on direct image appearance, where each module is sequentially predicted to achieve medical image registration. This method involves inputting image pairs—comprising a fixed and a moving image—and obtaining the deformation field through the initial momentum decoded by a U-Net, which is then used to resample the moving image as a reference to the fixed image. Similarly, Xu et al. [4] developed a method to directly obtain the deformation field using multi-scale convolutional neural networks (CNN), surpassing the registration accuracy of traditional B-spline techniques.

Unsupervised learning methods, which can be trained using original images and do not depend on labeled data, were explored through the Deep Learning Image Registration (DLIR) framework. This framework facilitates unsupervised affine and deformable image registration by training a CNN based on the image similarity between pairs of images, eliminating the need for labeled data. It learns to predict transformation parameters to create the deformation field by analyzing image pairs. Balakrishnan et al. [5] utilized a CNN similar to U-Net [6] to acquire the deformation field, naming the resulting algorithm VoxelMorph. This approach achieved notable improvements in registration speed and accuracy, gaining widespread recognition within the medical image registration community. However, VoxelMorph continues to use long connections similar to those in U-Net within the encoding-decoding structure, where the large semantic gap between connected feature maps can negatively impact registration accuracy.

To address this issue, this paper proposes an unsupervised single-modality medical image registration algorithm based on deep learning, which incorporates short connections into the U-Net architecture alongside the existing long connection methodology. This approach preserves the benefits of long connections in representing relationships between distant features while mitigating the disadvantage of a significant semantic gap between connected features. Additionally, the paper introduces a channel attention mechanism within the U-Net decoder to further enhance registration accuracy.

2. Related work

The objective of medical image registration is to determine the optimal spatial alignment between a fixed image I_F and a moving image I_M . This alignment is achieved by iteratively updating the spatial correspondence between the two images. The process relies on a loss function that is derived from the energy function commonly used in traditional registration methods. This energy function quantifies the discrepancy between the fixed and moving images, guiding the iterative adjustments to achieve precise registration. Through these updates, the registration algorithm seeks to minimize the loss function, thereby refining the spatial correspondence between I_F and I_M until the optimal alignment is attained.

$$\hat{\varphi} = \underset{\varphi}{\arg\min}\ E(I_M, I_F, \varphi)$$

In this context, φ denotes the spatial transformation applied to align the floating image I_M with the fixed image I_F, while φ* represents the optimal transformation that achieves the best possible alignment. The primary objective of registration is to minimize the loss function, which optimizes the spatial correspondence between I_F and I_M until the most accurate registered image is obtained.

In the domain of image registration, neural networks are often utilized to parameterize these spatial mapping relationships. The key advantage of neural networks lies in their capacity to autonomously learn and fine-tune parameters by minimizing the loss function, thereby identifying the most suitable model for matching images. Convolutional neural networks (CNNs), particularly U-shaped networks like U-Net, are commonly employed in registration tasks to generate deformation fields. These networks first downsample the input images to extract the spatial correspondence between image pairs and then upsample them to reconstruct the images. The goal is to detect essential features within the images while suppressing noise and irrelevant elements.

The spatial transformation, or deformation field, refers to the vector displacement field that dictates how the floating image should be altered to align with the fixed image. This field encapsulates the displacement required for accurate registration. By applying this deformation field to warp and interpolate the floating image, the resulting output is the registered image that closely matches the fixed image.

3. Methodology

3.1. Overall Network Framework

This paper presents an unsupervised, single-modality medical image registration algorithm rooted in deep learning, with its overall framework illustrated in Figure 1. A key feature of this registration algorithm is its independence from image dimensionality. Although the algorithm is demonstrated using two-dimensional brain magnetic resonance (MR) images, it is equally applicable to three-dimensional images.

The process begins by inputting the fixed image I_F and the floating image I_M as a dual-channel image pair into the network. The convolutional neural network then extracts features from these images and generates an estimated vector displacement field, known as the deformation field φ. Next, the Spatial Transformer Networks (STN) [7] apply the deformation field φ to the input floating image I_M and perform interpolation, using bilinear interpolation for two-dimensional images, to produce the registered image $I_{registered}$. The similarity measure L_{sim} between the fixed image I_F and the registered image $I_{registered}$, along with the smoothness of the deformation field L_{smooth}, serve as the objective functions for iteratively training the model and updating the network parameters.

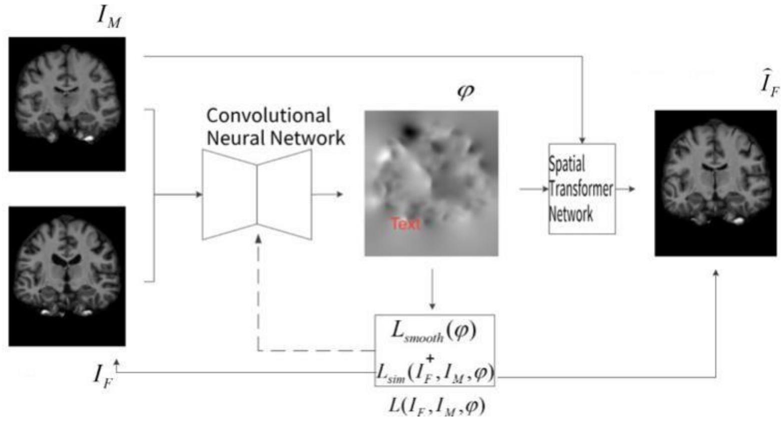

I_M
Convolutional Neural Network
φ
Text
Spatial Transformer Network
$\hat{I}_F$
$L_{smooth}(\varphi)$
+
$L_{sim}(I_F, I_M, \varphi)$
$L(I_F, I_M, \varphi)$
I_F

Figure 1. Overall framework

3.2. Network Architecture

The proposed registration network leverages a convolutional neural network, specifically a U-shaped architecture similar to U-Net, which consists of both encoding and decoding phases. This network is designed to extract features from the input image pair and transform them to generate a deformation field. To further enhance the quality and realism of the transformed images—produced by warping the moving image with the generated deformation field—a channel attention mechanism is integrated into the decoder. This mechanism refines the output by emphasizing important features, thereby improving the overall accuracy of the registration process. The detailed structure of the registration network is depicted in Figure 2.

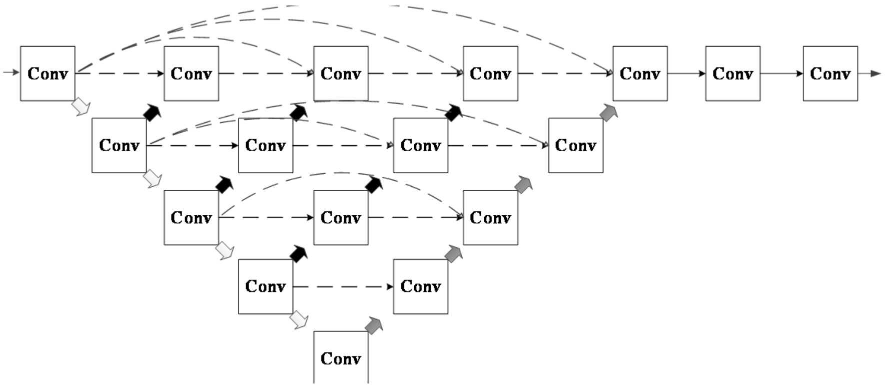


Figure 2. Registration network

In the diagram, "Conv" indicates two-dimensional convolution followed by activation using the LeakyReLU function. The white arrows depict downsampling operations performed via max pooling (MaxPooling), while the black arrows represent upsampling steps achieved through Upsampling. Gray arrows are utilized to indicate the incorporation of channel attention (Channel Attention) following the upsampling process, enhancing the model's focus on relevant features. Finally, dashed arrows illustrate the skip connections, which include both long and short connections, ensuring that important features from earlier layers are retained and integrated into the later stages of the network.

3.2.1. Dense U-Net

The convolutional neural network employed in this study is a densely connected U-Net [8], an enhancement of the original U-Net architecture. The encoder section of the network is responsible for downsampling, which captures the spatial relationships between the image pairs. Meanwhile, the decoder performs upsampling to reconstruct the images and generate the vector displacement field. In the standard U-Net architecture, long connections are used to directly link the encoder and decoder, facilitating the representation of relationships between pixels that are spatially distant. However, these direct connections can introduce significant semantic differences between the connected convolutional layers due to the depth gap, which can increase the network's learning complexity and adversely affect registration accuracy.

To overcome this challenge, the proposed method integrates short connections alongside the upsampling operations within the U-Net structure. These short connections help to reduce the semantic gap between feature maps with significant depth differences, thereby enhancing the overall registration accuracy.

In practical terms, the fixed and floating images of the same modality are concatenated to form a dual-channel input, with an image size of 161 × 193 pixels. The encoder performs four convolution operations with a stride of 1 and a kernel size of 3, followed by MaxPooling for downsampling. This process reduces the image resolution progressively to 1/2, 1/4, 1/8, and 1/16 of the original size. The receptive field of the convolutions expands with each operation, and the LeakyReLU activation function is applied to effectively capture the spatial correspondence between the image pairs.

In the decoder, upsampling operations are performed using the same convolutional structure as the encoder, allowing the image to be reconstructed back to its original resolution. The network achieves dense short connections by concatenating feature maps of the same resolution after each downsampling and upsampling step. In addition, long connections are used to link the encoder and decoder, mitigating the semantic gap between the connected convolutional layers and enabling the network to represent relationships between pixels that are far apart more effectively.

3.2.2. Attention Module

The attention mechanism has been demonstrated to significantly improve registration performance in various algorithms [9]. Drawing inspiration from ECA-Net [10], this paper incorporates a channel attention ECA (Efficient Channel Attention) module into the convolutional neural network. This module enhances the model's focus on relevant features while suppressing noise on a global scale, all while introducing only a minimal number of additional parameters. This selective attention to features leads to improved registration accuracy in subsequent steps.

The ECA module is specifically implemented in the decoder portion of the U-shaped network. Initially, the feature map of size (W × H × C), produced by upsampling in the decoder, undergoes Global Average Pooling (GAP). This process compresses the channel information, resulting in a dimension of (1 × 1 × C), where *W* and *H* denote the width and height of the feature map, respectively, and *C* represents the number of channels. These dimensions vary depending on the image size and the number of channels at each stage of upsampling in the decoder.

Following channel compression, the module generates weight information for each channel via a rapid one-dimensional convolution with a kernel size *K* , which is then processed by a Sigmoid activation function. The kernel size *K* is critical as it determines the extent of local cross-channel interaction, and it should be carefully adjusted according to the number of channels *C*. The formula to determine the appropriate value of *K* is:

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

In this formula, *r* and *b* represent coefficients. In this study, *r* is set to 2 and *b* is set to1.

4. Experimental results

This study employs the OASIS (Open Access Series of Imaging Studies) dataset [11], a widely used resource in the field of medical image registration, focusing on brain cross-sectional MR data from 415 individuals across various age groups, including young adults, middle-aged, non-demented, and demented elderly participants.

Initially, the images in the dataset were resampled to a resolution of 161 × 193 pixels. Following this, each MR image underwent a series of standard preprocessing steps using FreeSurfer software [12], which included motion correction, skull stripping, affine spatial normalization, and subcortical structure segmentation. Finally, the dataset was randomly divided into training and testing sets in a 4:1 ratio. Figure 4 displays the original and preprocessed images.

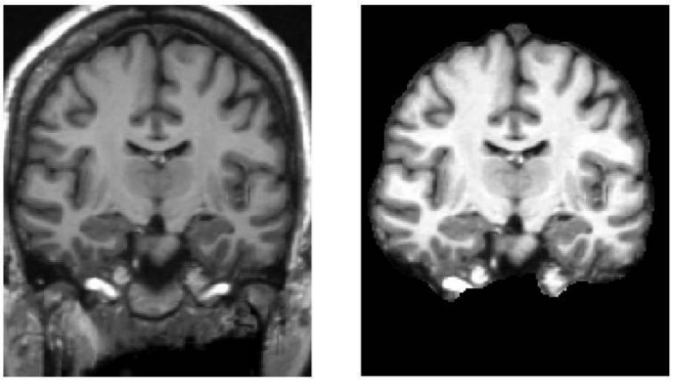


Figure 3. Images before and after preprocessing

4.2. Evaluation Metrics

For evaluating the registration results, this article uses the Dice Similarity Coefficient (DSC), a common metric in medical image registration[13]. The DSC ranges from 0 to 1, with values closer to 1 indicating better registration effectiveness. The DSC formula is as follows:

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

4.3. Experimental Setup

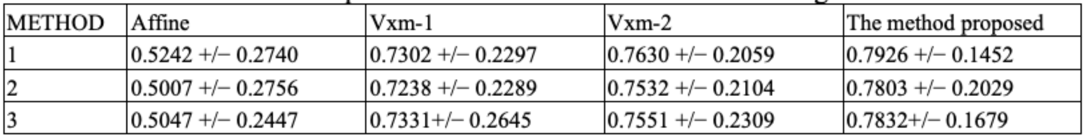
The experiments were performed on an NVIDIA RTX 3091 graphics card, within a Python 3.9 environment, utilizing the Keras and TensorFlow frameworks to implement the proposed unsupervised registration model. Network optimization was carried out using the Adam optimizer. The experiments focused on atlas-based registration of the dataset images, where all moving images were registered to a single fixed image.

4.4. Experimental Results

To evaluate the performance of the proposed unsupervised registration algorithm, Affine and VoxelMorph[14] were selected as comparison algorithms. Notably, VoxelMorph's convolutional neural network also employs a U-Net-like U-shaped architecture [15] to generate the deformation field. The experiments were conducted using two frameworks from the VoxelMorph algorithm, Vxm-1 and Vxm-2 , with the registration effectiveness of each algorithm assessed through the Dice Similarity Coefficient (DSC) for 24.1 segmentation labels per image in the test set, as well as through visualization of the results.

Table 1 shows the registration results for three randomly selected moving images registered to the same fixed image in the test set using different algorithms. The results indicate that the method proposed in this article consistently achieves higher DSC coefficients across all image groups compared to the comparison algorithms, demonstrating a significant advantage in registration accuracy over the alternatives.

Table 1: Comparison of DSC coefficients of different algorithms



METHOD	Affine	Vxm-1	Vxm-2	The method proposed
1	0.5242 +/– 0.2740	0.7302 +/– 0.2297	0.7630 +/– 0.2059	0.7926 +/– 0.1452
2	0.5007 +/– 0.2756	0.7238 +/– 0.2289	0.7532 +/– 0.2104	0.7803 +/– 0.2029
3	0.5047 +/– 0.2447	0.7331+/– 0.2645	0.7551 +/– 0.2309	0.7832+/– 0.1679

5. Conclusion

This paper introduces an unsupervised registration method for brain unimodal magnetic resonance imaging (MRI) based on deep learning. By leveraging the discrepancies between the deformed moving image and the fixed image, the method iteratively optimizes the parameters within the convolutional neural network (CNN) in reverse, thereby eliminating the time and financial costs associated with manual annotation. The proposed method enhances connectivity within the CNN by designing a densely connected U-Net, a U-shaped network similar to U-Net, which incorporates short connections between the encoder and decoder. This design addresses the issue of significant semantic gaps caused by large sampling depth differences between connected convolutional layers, while preserving the advantage of long connections in representing relationships between distantly located pixels. Additionally, the method integrates a channel attention mechanism during the upsampling phase in the decoder, which, through feature recalibration, effectively emphasizes useful features and suppresses noise during image reconstruction, thereby improving the registration effect in generating 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] 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

[2] 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

[3] Ballé, J., Laparra, V., & Simoncelli, E. P. (2015). Density modeling of images using a generalized normalization transformation. arXiv preprint arXiv:1511.06281.

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

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

[6] 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, Munich, 5-9 October 2015, 234-241. https://doi.org/10.1007/978-3-319-24574-4_28

[7] Jaderberg, M., Simonyan, K., & Zisserman, A. (2015). Spatial transformer networks. Advances in neural information processing systems, 28.

[8] Siddique, N., Paheding, S., Elkin, C. P., & Devabhaktuni, V. (2021). U-net and its variants for medical image segmentation: A review of theory and applications. IEEE access, 9, 82031-82057.

[9] Yang, W., Wu, Z., Zheng, Z., Zhang, B., Bo, S., & Yang, Y. (2024). Dynamic Hypergraph-Enhanced Prediction of Sequential Medical Visits. arXiv preprint arXiv:2408.07084.

[10] Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W., & Hu, Q. (2020). ECA-Net: Efficient channel attention for deep convolutional neural networks. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 11534-11542).

[11] 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

[12] Seiger, R., Ganger, S., Kranz, G. S., Hahn, A., & Lanzenberger, R. (2018). Cortical thickness estimations of FreeSurfer and the CAT12 toolbox in patients with Alzheimer's disease and healthy controls. Journal of Neuroimaging, 28(5), 515-523.

[13] Yang, W., Wu, Z., Zheng, Z., Zhang, B., Bo, S., & Yang, Y. (2024). Dynamic Hypergraph-Enhanced Prediction of Sequential Medical Visits. arXiv preprint arXiv:2408.07084.

[14] Balakrishnan, G., Zhao, A., Sabuncu, M. R., Guttag, J., & Dalca, A. V. (2019). Voxelmorph: a learning framework for deformable medical image registration. IEEE transactions on medical imaging, 38(8), 1788-1800.

[15] Cai, S., Xiao, Y., & Wang, Y. (2024). Two-dimensional medical image segmentation based on U-shaped structure. International Journal of Imaging Systems and Technology, 34(1), e23023.