Deep Learning-Based Medical Image Registration: A Review

Emily J. Carter[1], Daniel K. Thompson[2], and Sophia M. Reynolds[3]

[1] Department of Artificial Intelligence, Westbridge Institute of Technology, Seattle, WA 98101, USA
[2] Center for Biomedical Imaging and Analytics, North Valley University, Denver, CO 80202, USA
[3] School of Computer Science and Medical Engineering, Eastland University, Boston, MA 02115, USA
Corresponding author: Emily J. Carter (e-mail: e.carter@westbridge-tech.edu)

ABSTRACT Image registration refers to finding the spatial transformation relationship between two or more images and mapping one image to another, so that points corresponding to the same spatial position across images are aligned. It is a key step in image fusion. In the field of medical image analysis, medical image registration is widely applied in disease diagnosis, radiotherapy, surgical navigation, and other scenarios. Therefore, achieving efficient and accurate image registration has become a research hotspot. Traditional registration methods are limited by slow speed, high computational complexity, and poor applicability to multi-modal image registration, making it difficult to meet the demands of modern medical image analysis for efficiency, precision, and robustness. In recent years, with the rapid development of deep learning technology, deep learning networks represented by convolutional neural networks have received widespread attention due to their end-to-end and transferable advantages. This paper summarizes four mainstream deep learning technologies adopted in medical image registration research and discusses their future development trends.

INDEX TERMS Medical Image Registration, Deep Learning, Convolutional Neural Networks, Multi-modal Medical Imaging, Clinical Applications; Image Fusion.

I. INTRODUCTION

Modern medical imaging technologies have revolutionized clinical practice by enabling the widespread acquisition of high-volume, multi-modal patient data, forming a rich information foundation for disease diagnosis and treatment planning. Modalities such as computed tomography (CT) and magnetic resonance imaging (MRI) are routinely used, and their complementary information is often integrated via medical image registration to support comprehensive clinical decision-making.

Despite its critical role, medical image registration faces substantial hurdles stemming from the inherent complexity of clinical data. Cross-modality variations in image quality, anatomical deformation patterns, and non-linear intensity distributions make multi-modal registration a persistent challenge. Compounding these issues are the limited availability of annotated medical images and the absence of standardized gold standards for performance evaluation, which hinder the robust validation and clinical deployment of registration algorithms.

Follow an optimization-based paradigm: a similarity metric is defined, and iterative algorithms are used to search for parameters that maximize this measure. This approach suffers from inherent drawbacks: it is computationally intensive, requires extensive parameter tuning, and is prone to converging to local optima, which can severely degrade registration accuracy.

Deep learning has emerged as a transformative solution to these limitations, offering end-to-end frameworks that automatically learn the complex, non-linear spatial correspondences between images. These methods have demonstrated superior performance in terms of both accuracy and efficiency, particularly when scaling to large, high-resolution datasets. By minimizing manual intervention, they also advance the automation and clinical translatability of registration workflows, with far-reaching implications for precision medicine. This review provides a structured overview of key deep learning approaches to medical image registration, discusses existing challenges, and outlines promising avenues for future research.

II. Deep Learning-Based Medical Image Registration Techniques

Based on a review of recent literature on deep learning-driven medical image registration, existing deep learning-based registration models can be broadly categorized into four main paradigms: convolutional neural network

(CNN)-based registration, generative adversarial network (GAN)-based registration, deep reinforcement learning (DRL)-based registration, and Transformer-based registration. The following sections will elaborate on each of these four technical approaches in detail.

A. Convolutional Neural Network-Based Medical Image Registration

Since the concept of Convolutional Neural Networks (CNNs) was first proposed, the field has advanced rapidly, fueled by continuous improvements in computing power. In medical image registration, early applications of CNNs primarily focused on leveraging CNN architectures to automatically learn high-dimensional features extracted during the registration process, enabling the selection of optimal registration features. By learning adaptive features, these methods replaced handcrafted feature design to achieve automated registration[1].

Given the automated and efficient nature of CNN training, many researchers later applied CNNs to regression tasks for estimating model parameters in image registration. For instance, Miao et al.[2] proposed a CNN-based regression approach to address two key limitations of conventional intensity-based 2D/3D registration methods: slow computational speed and small capture range.

For non-rigid registration, the Voxelmorph model proposed by Balakrishnan et al.[3] stands as a prominent example, with its core framework illustrated in Figure 1. Built entirely on CNN architectures, this model learns the mapping relationships between images in an end-to-end fashion. Its lightweight design offers two major advantages over comparable tools: it achieves higher computational efficiency and imposes significantly lower memory requirements. The successful application of the Voxelmorph framework in medical image registration, particularly for 3D volumes, has made it a widely adopted baseline in subsequent research[4]. This has inspired numerous improved variants of the Voxelmorph model, all aimed at enhancing registration performance.

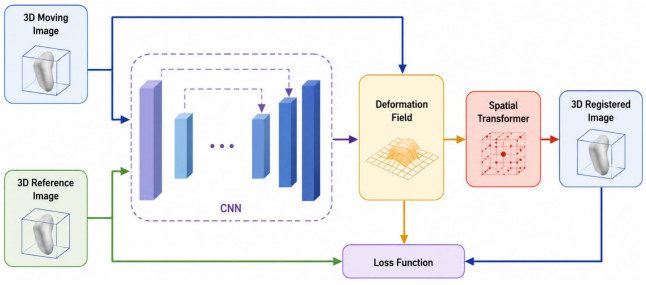

FIGURE 1. **Deep Learning Framework for 3D Medical Image Registration**

B. Medical Image Registration Using Generative Adversarial Networks

Early medical image registration methods were predominantly built on supervised learning paradigms. However, the growing volume of medical imaging data has exposed two key limitations of supervised approaches: the difficulty of acquiring high-quality annotations and the scarcity of labeled datasets. These challenges have driven researchers to explore unsupervised registration frameworks.

Generative Adversarial Networks (GANs) offer a distinct advantage in this context: they require neither ground-truth labels nor hand-designed similarity metrics during training. Qiao et al.[5] addressed multi-modal registration tasks by proposing a GAN-based model that uses a single generator-discriminator architecture to handle intensity inconsistencies across modalities. Wang et al.[6] leveraged CycleGAN to achieve robust registration and fusion of multi-sequence brain MR images, as well as multi-modal abdominal CT and MR datasets. Beyond aligning fixed and moving images, this method also preserves high visual fidelity in the registered outputs. For MRI-CT multi-modal registration, Yang et al.[7] developed the DTR-GAN network, whose superior performance was validated through extensive experiments.

While GAN-based models can generate more detailed and realistic deformation fields without relying on annotated data, they also present notable drawbacks. The inherent adversarial training process often leads to prolonged training times and challenges with model convergence[8].

C.Deep Reinforcement Learning-Based Medical Image Registration

Deep Reinforcement Learning (DRL) is a machine learning paradigm that integrates deep learning (DL) and reinforcement learning (RL). It leverages deep neural networks to approximate the policy or value functions in RL, enabling the solution of complex decision-making problems[9]. Compared to conventional registration methods, DRL offers notable advantages such as fewer parameters and superior inference performance, making it particularly effective for multi-modal image registration tasks.

Traditional image registration workflows are highly susceptible to human bias, especially in the selection of feature representations and similarity metrics, which can introduce significant errors into registration outcomes. To address this limitation, researchers have proposed an agent-based model composed of a policy network and a value network. This framework guides the moving image toward alignment with the reference image, achieving precise registration[10]. The policy network, a neural network module, learns to predict the probability distribution of possible actions under specific states, thereby guiding the agent ' s decision-making process. Meanwhile, the value network provides a mechanism for evaluating the potential utility of different states and actions, helping the agent identify optimal decision paths.

Targeting the challenge of aligning heterogeneous features in multi-modal registration, Hu et al.[11] reformulated the registration process as a sequential decision problem, solved by an agent trained via asynchronous reinforcement learning. Their architecture incorporates convolutional LSTM layers after stacked convolutions, allowing it to extract spatio-temporal image features and implicitly learn a similarity metric. Evaluated on a dataset of nasopharyngeal carcinoma CT-MR images, the model demonstrated highly competitive performance in medical image registration tasks.

D. Transformer-Based Medical Image Registration

The Transformer architecture excels at capturing global contextual information through its attention mechanism, enabling it to model long-range dependencies and extract highly discriminative image features.

Chen et al.[12] proposed ViTVNet, a hybrid framework combining Vision Transformer (ViT) with the VNet architecture. This design aims to capture both global contextual cues and multi-scale image features, addressing key limitations of purely CNN-based methods. The network effectively leverages long-range dependencies while maintaining multi-scale feature extraction capabilities.

Building on the complementary strengths of CNNs and Transformers, Song et al.[13] first employed CNNs to generate feature maps, then used Transformer encoders to capture global context. In experiments on brain MRI datasets, this approach achieved a 1% performance improvement over competing models.

Further advancing the ViTVNet framework, Chen et al.[14] introduced TransMorph, which replaces the vanilla ViT module with Swin Transformer. However, the registration performance of this method is sensitive to window size selection, and repeated sliding window operations introduce significant computational overhead.

To mitigate these challenges — reducing computational complexity and model parameters — Ma et al.[15] proposed the SymTrans model. By incorporating an efficient multi-head self-attention mechanism, SymTrans effectively cuts down both parameter count and computational cost.

III. Future Development Trends

Several quantitative metrics are commonly used to evaluate the performance of medical image registration algorithms.

Dice Similarity Coefficient (DSC):

$$DSC = \frac{2|X \cap Y|}{|X| + |Y|} \quad (1)$$

Mean Squared Error (MSE):

$$MSE = \frac{1}{N}\sum (x_i - y_i)^2 \quad (2)$$

Root Mean Squared Error (RMSE):

$$RMSE = \sqrt{\left(\frac{1}{N}\sum (x_i - y_i)^2\right)} \quad (3)$$

Sum of Squared Differences (SSD):

$$SSD(I, J) = \sum (I(x, y) - J(x, y))^2 \quad (4)$$

Target Registration Error (TRE) measures the distance between corresponding landmarks after registration.

Hausdorff Distance evaluates the maximum geometric discrepancy between two point sets.

IV. Future Development Trends

While deep learning-based methods have been widely studied for medical image registration in recent years, their registration accuracy has not yet shown significant superiority over traditional algorithms. These methods offer two key advantages: first, parallel GPU acceleration significantly improves computational efficiency; second, end-to-end network architectures enhance model automation and transferability.

Supervised image registration relies heavily on precise data annotations, but high-quality labeled samples are extremely scarce in real-world clinical applications. This scarcity often leads to overfitting during model training, limiting generalization performance. To mitigate this issue, various data augmentation strategies are typically employed to expand the available training data.

Unsupervised registration, on the other hand, eliminates the need for labeled data, effectively alleviating the problem of insufficient training datasets. However, it still faces substantial challenges in multi-modal registration tasks, as quantifying similarity across different image modalities remains difficult. As a result, unsupervised methods are currently most commonly applied to single-modal registration problems, while semi-supervised approaches are predominantly used for multi-modal registration. Given the unique characteristics of medical imaging data, unsupervised image registration remains a major focus of ongoing research. Notably, GAN-based models excel at unsupervised learning, while Transformer architectures demonstrate strong capabilities in capturing global image features. Consequently, deep learning techniques based on GANs and Transformers will continue to be key research directions in medical image registration.

V. Conclusion

This paper presents a comprehensive review of deep learning techniques applied to medical image registration. It outlines the current state-of-the-art research progress of four mainstream deep learning approaches across various medical image registration tasks, discusses commonly used evaluation metrics and future development trends in the field, and highlights the significant research value of deep learning in this domain. Nevertheless, challenges such as the scarcity of annotated medical image data and the lack of standardized gold standards remain critical issues requiring sustained attention and further investigation by the research community.

ACKNOWLEDGMENT

The authors thank their colleagues at the Department of Computer Science, Redwood Institute of Technology, the

Medical Imaging Research Center at Lakeside University, the School of Biomedical Engineering at Atlantic State University, and the Department of Artificial Intelligence in Medicine at Horizon University for their valuable discussions and technical support throughout this work. The authors also appreciate the constructive suggestions provided by anonymous reviewers, which helped improve the quality and clarity of this manuscript.

REFERENCES

[1] ZHAO L Y, JIA K B. Deep adaptive log-demons: diffeomorphic image registration with very large deformations[J]. Computational and Mathematical Methods in Medicine, 2015, 2015(1): 836202.

[2] MIAO S, WANG Z J, LIAO R. A CNN regression approach for real-time 2D/3D registration[J]. IEEE Transactions on Medical Imaging, 2016, 35(5): 1352–1363.

[3] BALAKRISHNAN G, ZHAO A, SABUNCU M R, et al. VoxelMorph: a learning framework for deformable medical image registration[J]. IEEE Transactions on Medical Imaging, 2019.

[4] LI Y X, TANG H, WANG W, et al. Dual attention network for unsupervised medical image registration based on VoxelMorph[J]. Scientific Reports, 2022, 12: 16250.

[5] QIAO J, LAI Q, LI Y, et al. A GAN based multi-contrast modalities medical image registration approach[C]//2020 IEEE International Conference on Image Processing (ICIP). IEEE, 2020: 3000–3004.

[6] WANG C J, YANG G, PAPANASTASIOU G, et al. DiCyc: GAN-based deformation invariant cross-domain information fusion for medical image synthesis[J]. Information Fusion, 2021, 67: 147–160.

[7] YANG A L, YANG T J, ZHAO X, et al. DTR-GAN: an unsupervised bidirectional translation generative adversarial network for MRI-CT registration[J]. Applied Sciences, 2024, 14(1): 95.

[8] ZHOU T, LI Q, LU H L, et al. GAN review: Models and medical image fusion applications[J]. Information Fusion, 2023, 91: 134–148.

[9] ZHOU S K, LE H N, LUU K, et al. Deep reinforcement learning in medical imaging: a literature review[J]. Medical Image Analysis, 2021, 73: 102193.

[10] Yao Mingqing, Hu Jing. Multi-modal medical image registration based on deep reinforcement learning[J]. Journal of Computer-Aided Design & Computer Graphics, 2020, 32(8): 1236–1247.

[11] HU J, LUO Z W, WANG X, et al. End-to-end multimodal image registration via reinforcement learning[J]. Medical Image Analysis, 2021, 68: 101878.

[12] CHEN J Y, HE Y F, FREY E C, et al. ViT-V-net: vision transformer for unsupervised volumetric medical image registration [EB/OL]. 2021: 2104.06468. https://arxiv.org/abs/2104.06468v1.

[13] SONG L, LIU G X, MA M R. TD-Net: unsupervised medical image registration network based on Transformer and CNN[J]. Applied Intelligence, 2022, 52(15): 18201–18209.

[14] CHEN J Y, FREY E C, HE Y F, et al. TransMorph: Transformer for unsupervised medical image registration[J]. Medical Image Analysis, 2022, 82: 102615.

[15] MA M R, XU Y B, SONG L, et al. Symmetric transformer-based network for unsupervised image registration[J]. Knowledge-Based Systems, 2022(257): 109959.