RDMR: Recursive Inference and Representation Disentanglement for Multimodal Large Deformation Registration

Yibo, Hu[1], Ziqi Zhao[1], Qi Zhang[1], Lisa X.Xu[1,2], and Jianqi Sun[1,2,*]

1.School of Biomedical Engineering, Shanghai JiaoTong University, Shanghai, China.
2.National Engineering Research Center of Advanced Magnetic Resonance Technologies for Diagnosis and Therapy (NERC-AMRT), Shanghai
milesun@sjtu.edu.cn

Abstract. Multimodal large deformation image registration is a challenging task in medical imaging, primarily due to significant modality differences and large tissue deformations. Current methods typically employ dual-branch multiscale pyramid registration networks. However, the dual-branch structure fails to explicitly enforce that the model learns modality-invariant image registration features. Furthermore, in the multiscale registration process, only the deformation field is propagated, which restricts the model's capacity to accommodate more complex deformations. To enhance the model's ability to learn features from different modalities, we propose a modality representation disentanglement method, incorporating Multi-layer Contrastive Loss(MCL) to enforce the learning of modality-invariant features. To address the challenge of complex large deformations, we introduce a Multi-Scale Feature fusion Registration module(MSFR), which integrates features and deformation fields from different scales during the registration process. To explore the registration potential of the trained model, we propose a Recursive Inference enhancement strategy that further improves registration performance. This model is referred to as RDMR. Based on experimental results from both private and public datasets, the RDMR model outperforms other SOTA models. Compared to the baseline registration model (Voxel Morph), the RDMR model achieved improvements of 1.4 and 4.5 percentage points in the DSC metric, respectively. Our code is publicly available at:https://github.com/ybby2020/RDMR

Keywords: Multimodal deformable registration · Modality-invariant contrastive · Recursive inference.

1 Introduction

Aligning (registration) anatomical structures across multimodal medical images is a fundamental task in medical image analysis. This task has significant implications for clinical applications, including preoperative multimodal diagnosis, intraoperative image-guided surgical planning, and postoperative efficacy evaluation [1,2]. However, multimodal medical image registration faces significant challenges due to the inherent differences in imaging mechanisms across modalities. For example, MRI provides high-contrast images of soft tissues, whereas CT offers high-resolution images of bone structures. These modality-specific characteristics lead to significant discrepancies in intensity distribution profiles across imaging modalities. Furthermore, additional physiological factors[3] such as posture changes, respiratory-induced motion introduce complex spatial variations that frequently result in nonlinear, large-magnitude tissue deformations.

In recent years, deep learning-based deformable registration methods have achieved remarkable progress[4,5,6,7,8,9,10]. However, current studies predominantly focus on unimodal registration[12], with insufficient exploration of multimodal registration under large deformations. Current unimodal registration research addressing large deformations predominantly adopts dual-stream pyramid registration methods[13][14][15]. However, existing approaches exhibit some limitations that hinder their performance in multimodal large deformation registration tasks: (1) during the multiscale registration process, only the deformation field was transmitted, lacking interaction of image features, which leads to a decrease in registration accuracy in complex scenarios; (2) the current dual-branch design cannot ensure that the model learns the intrinsic, modality-invariant features of the image.

To address the challenges of large deformations and intensity discrepancies in multimodal deformable registration. We propose a novel Disentanglement-based Multimodal Registration model (DMR) and introduce an innovative recursive inference enhancement strategy to further improve model performance (RDMR). To the best of our knowledge, this is the first work to propose using recursive inference to enhance performance in multiscale multimodal registration. Extensive experiments on multi-center clinical datasets demonstrate that our model outperforms state-of-the-art(SOTA) approaches in both qualitative and quantitative evaluations. Specifically, our contributions can be summarized as follows:

(i)*Multi-scale Feature Fusion Registration:* We design a multiscale feature interaction module that fuses multiscale features(MSFR), breaking the limitations of conventional multiscale deformation field propagation and significantly enhancing the model's capability to perceive complex deformations.

(ii)*Disentanglement based on contrastive learning:* We propose a novel multilayer modality-invariant contrastive loss(MCLoss) to enforce feature distribution consistency across modalities at different encoder depths, thereby improving registration performance and model generalization.

(iii)*Low Computation Load Enhancement Mechanism:* We introduce a recursive inference strategy that boosts performance without retraining the network. By recursively adjusting the multi-scale modules of the trained model during inference, we can optimize its final structure with minimal computational overhead, further enhancing performance.

2 Method

The architecture of the proposed multimodal deformable registration model is shown in Fig.1. During the training phase, multimodal data are fed into the dual-branch feature extraction module to obtain image features at different scales. Registration starts at the smallest scale, where features from different modalities at the same scale are input into the MSFR (Multi-Scale Feature Registration) module, and the process continues until the final deformation field is obtained. During the inference phase, further improvement of the registration results is achieved by recursively applying the MSFR module at each scale. The number of recursions is associated with parameters such as a, b, c, and d in Fig.1.

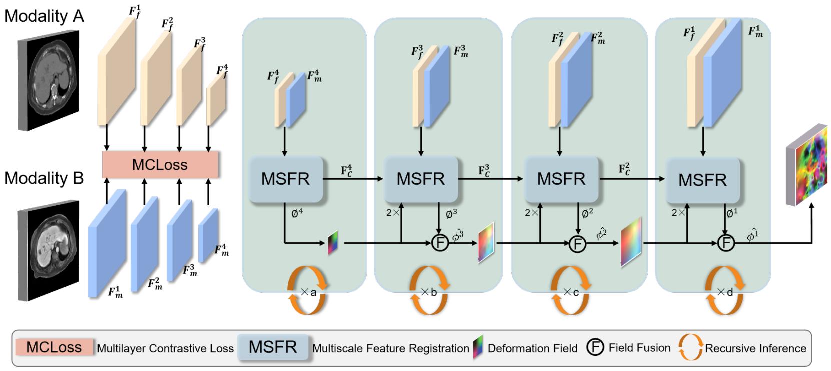

Fig. 1. Overview of the proposed RDMR architecture.

2.1 Dual-stream Encoding Branch

This module is used for image feature extraction, with the goal of extracting modality-invariant features to facilitate subsequent multiscale registration. We use convolutional neural network as the backbone to perform image feature extraction. Since the module is designed to encode images from different modalities, the weight parameters of the two branches are independent, with each branch encoding features for fixed and moving, respectively. The structure of both branch encoders is identical, and we describe one branch as an example. The encoder consists of four encoding blocks. The first block primarily increases the feature channels without changing the image feature dimensions. The following three encoding blocks have the same structure, as shown in Fig.2(a). These blocks include convolutional downsampling and convolutional processes (instance normalization and activation functions applied after each convolution). After each encoding block, the number of channels in the features doubles, while the spatial dimensions are reduced by half. The output features at each layer are denoted as F_m^i, F_f^i for $i = 1, \ldots, 4$.

2.2 Multi-Scale Feature Registration

The multi-scale deformable registration begins at the smallest scale, and the registration features and deformation field are propagated to the next scale for further refinement. The index i denotes the current scale, while i+1 refers to the previous (coarser) scale. Current multiscale pyramid registration models typically only propagate the deformation field backward. In this paper, we simultaneously propagate the image registration features and the deformation field. Specifically, we propose MSFR module to handle the input and output of both registration features and deformation fields.

The MSFR module(Fig.2(b)) receives four inputs: the previous deformation field ϕ^{i+1} and registration features F_c^{i+1}, the current scale's moving image features F_m^i and fixed image features F_f^i. The registration features F_c^{i+1} are first processed by a transposed convolution to adjust dimensions to $\hat{F}_c^{i+1}$, while the moving image features F_m^i are warped by the deformation field to obtain $\hat{F}_m^i$. Then $\hat{F}_c^{i+1}$,F_f^i and $\hat{F}_m^i$ are concatenated and fed into a two-stage image feature extraction module, which consists of convolutional layers, normalization, and activation functions, ultimately outputting the registration features F_c^i. These registration features can be directly passed to the next scale, while also being input into the registration head to predict the deformation field(ϕ^i) at the current scale. The output deformation field at each scale is combined with the previously obtained deformation field. The registration head mainly consists of convolutional modules and a diffeomorphism layer[27]. The processing flow of MSFR can be represented by Equation (1).

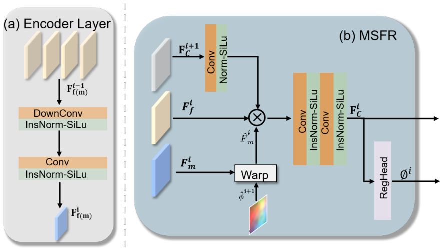

Fig. 2. (a)Details of the encoder block architecture.(b)Details of the MSFR architecture.

$$\text{MSFR}\begin{cases}\hat{F}_c^{i+1} = ConvUP(F_c^{i+1}),\\ \hat{F}_m^i = warp(F_m^i, \hat{\phi}^{i+1}),\\ F = concat(F_f^i, \hat{F}_m^i, \hat{F}_c^{i+1})\\ F_c^i = CConv(F),\\ \phi^i = RegHead(F_c^i),\\ i = 3, 2, 1\end{cases} \tag{1}$$

where "CConv" denotes a two-layer convolution operation.

2.3 Multi-layer Modality-invariant Contrastive Representation

The major challenge in multimodal registration is the intensity differences between images. In multiscale deformable registration research, even with dual-branch encoders, the model is not guaranteed to learn modality-invariant image features. To address this issue, we propose a multi-layer modality-invariant contrastive loss. We first introduce the design philosophy. *For the same patient, even with images acquired at different times and from different modalities, there tends to be a higher structural similarity.* In contrast, images from different patients or the same modality across patients exhibit noticeable structural differences. The fundamental idea behind contrastive learning[16][17] is to establish correlations between two signals, the ‘query’ and its ‘positive’ example, while contrasting them with other examples in the dataset, referred to as ‘negatives.’ Therefore, we propose treating images from the same patient across different modalities as positive pairs and images from different patients as negative pairs. This approach helps constrain the model to learn more essential or modality-independent feature information. Due to the presence of multiple output layers in the encoder branch, we propose computing the contrastive loss for the feature output at each layer. In summary, we define the contrastive loss[18] as follows:

$$\mathcal{L} = -\log\left[\frac{\exp(\boldsymbol{v}\cdot\boldsymbol{v}^{+}/\tau)}{\exp(\boldsymbol{v}\cdot\boldsymbol{v}^{+}/\tau)+\sum_{n=1}^{N}\exp(\boldsymbol{v}\cdot\boldsymbol{v}_n^{-}/\tau)}\right], \mathcal{L}_{\text{MCL}}(F_f,F_m)=\sum_{l=1}^{L} w_l\mathcal{L}^l \tag{2}$$

Here, τ represents a temperature parameter used to scale the distance between the query and other examples, with a default value of 0.5. $(\boldsymbol{v},\boldsymbol{v}^{+})$ represents positive pairs from the same patient across different modalities. $(\boldsymbol{v},\boldsymbol{v}_n^{-})$ represents the different patients across different modalities are referred to as negative pairs. l represents the layer index, and w_l represents the weight associated with a specific layer.

2.4 Recursive Inference Optimization

The multi-scale registration model can be viewed as an independent registration process at each individual scale. By recursively applying the model at each scale, it mimics traditional iterative optimization methods in registration[19]. In this paper, we propose that recursive-based architectural refinements can effectively enhance the performance of pre-trained multi-scale image registration models, achieving significant improvements with minimal additional computational overhead. Equation 3 illustrates the computation process.

$$\phi = f_\phi^1(f_\phi^2(f_\phi^3(f_\phi^4(F_m^4,F_f^4,\phi^4)\times a, F_m^3,F_f^3)\times b, F_m^2,F_f^2)\times c, F_m^1,F_f^1)\times d \tag{3}$$

Here, f_ϕ^i represents the prediction function for the deformation and registration features at different scales, while F_m^i, F_f^i denote the input features at each scale, and a, b, c, d represent the number of recursions at the corresponding scales. The values of these four hyperparameters should be determined through inference testing.

Specially, the four hyperparameters for recursive inference determine the number of recursions at each scale. We begin with a single recursion per scale to identify which scale yields the greatest improvement in registration quality.

We then increment the recursion that scales until no further gains. Unlike tuning learning rates or loss weights, which typically requires retraining, this strategy adjusts recursions post-training, greatly enhancing usability. Dataset-specific tuning is necessary but easy to perform.

2.5 Loss Function

This paper focuses on unsupervised multi-modal medical image deformable registration. To measure the similarity between multimodal images, we adopt the MIND metric[20]. Additionally, we apply a regularization function that imposes constraints on the gradients of the deformation fields. Finally, we utilize the proposed MCLoss. Consequently, the overall loss function consists of three components.

$$L(I_f, I_m, \phi) = \lambda_1 L_{MIND}(I_f, I_m \circ \phi) + \lambda_2 L_{\mathrm{MCL}}(F_f, F_m) + \sum_{p \in \Omega} \|\nabla \phi(p)\|^2 \quad (4)$$

where λ_1, λ_2 are the hyperparameters used to balance the contribution of loss functions. P represents the image voxel.

3 Experiments

3.1 Datasets and Comparison Methods

The first dataset is a proprietary dataset comprising 3D multimodal MR-CT liver-paired data(The data has undergone an ethical review). The dataset comprised 122 cases: hospital A(HA) included 40 cases, hospital B(HB) included 68 cases, and hospital C(HC) included 14 cases. For model training, we utilized all data from HA and 60 cases from HB, others were reserved for testing. These data were resampled to a voxel spacing of $1.75 \times 1.75 \times 2$, then were cropped to $160 \times 160 \times 64$. The second dataset is the publicly multi-modal dataset AbdomenCT-1K(AMOS)[21]. We used 30 unpaired but annotated MR and CT, with 24 MR/CT volumes allocated for training and 6 for testing. Considering the combination of MR and CT, the actual training data consists of 576 pairs, with 36 pairs for testing. Additionally, we performed resampling and cropping on this dataset to ensure consistency in data dimensions.

To evaluate the performance of our proposed model, we compared it with several SOTA deformable registration algorithms, including VoxelMorph(VM)[11], TransMorph(TM)[22], VIT-V-Net[23], LKU-Net[25], RDP[26], and PIVIT[24]. We employed the Dice Similarity Coefficient (DSC) as an evaluation metric. Furthermore, to assess the plausibility of the deformation fields, we computed the percentage of voxels with a negative Jacobian determinant(($\%|J_\phi| \leq 0$), which indicates local folding within the deformation field. In addition, we also compared the model's parameters and inference time.

3.2 Training Settings

All models are implemented using the pytorch. The model training was performed using an NVIDIA GeForce RTX 3090 GPU and an Intel i9-12900K CPU. We implement the model using Adam optimizer with a learning rate of $1e^{-4}$. The batch size is set as 4 and the networks are trained for 200 epochs. For the paired private datasets, λ_1 and λ_2 were set to 10 and 2, respectively. For the unpaired AMOS, we relax the weighting of the MCL, λ_1 and λ_2 were set to 10 and 0, respectively.

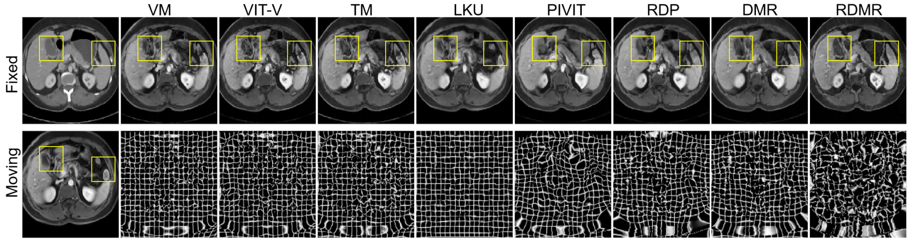

Fig. 3. The visualization of the results and the corresponding deformation fields

4 Results

Fig.3 shows the visualization results of different models. The content within the yellow boxes highlights that our proposed model exhibits a structure more similar to the fixed image than other models. The visualization of the deformation field (second row) shows that our proposed model is capable of producing a wider range of deformations, particularly in areas corresponding to complex anatomical structures. Table.1 summarizes the comparative results of different models across three multimodal liver datasets. Our proposed DMR model demonstrates superior performance on all evaluation metrics compared to baseline methods. Notably, the RDMR with recursive inference achieves further improvements over the DMR. In terms of the DSC metric, RDMR achieves the best performance across all three datasets. Specifically, on the HB dataset, the DSC score improves by 1.4% relative to the baseline model VM, and on the AMOS dataset, the improvement is 4.5%. The proportion of negative Jacobian determinants about most models remain below 0.1%, indicating most models handle the image folding problem better.

The models tested on the HC dataset were trained on the HB dataset. From the metrics in Table.1, it is evident that each model performs worse on the HC than on the HB, as changes in data distribution can lead to a decline in the model's generalization ability. However, our proposed model still achieves the best performance on the HC dataset, demonstrating that the modality-invariant

Table 1. Comparison of results from different methods across three datasets.

	HB		HC		AMOS			
Model	DSC↑	$\%\lvert J_\phi \rvert \leq 0$↓	DSC↑	$\%\lvert J_\phi \rvert \leq 0$↓	DSC↑	$\%\lvert J_\phi \rvert \leq 0$↓	T(S)	P(K)
VM	86.6±8.1	<0.03%	85.2±11.1	<0.15%	80.4±5.9	<0.15%	0.04	301.4
VIT-V	86.8±8.2	<0.02%	85.9±12.1	<0.08%	80.5±5.8	<0.16%	0.04	31519.8
TM	86.0±8.1	<0.05%	84.2±13.9	<0.10%	80.1±5.6	<0.16%	0.06	46771.2
LKU	84.3±9.1	<0.001%	78.3±17.4	<0.001%	79.0±5.8	<0.001%	0.03	2087.6
PIVIT	85.9±7.5	<0.001%	85.1±9.1	<0.01%	84.6±5.4	<0.08%	0.02	664.9
RDP	86.9±7.6	<0.001%	85.6±8.2	<0.001%	84.5±5.3	<0.03%	0.08	2240.8
DMR	87.7±8.2	<0.001%1	86.2±8.5	<0.001%	84.7±5.7	<0.05%	0.07	1904.2
RDMR	**88.0**±8.0	<0.001%	**87.0**±11.3	<0.001%	**84.9**±5.8	<0.04%	0.08	1904.2

constraints enable the model to focus on the essential information in the images. This also indicates that our method helps mitigate the model generalization issues caused by multi-center data shift. From the structural perspective, models that employ a multi-scale registration strategy, such as PIVIT, RDP, DMR, and RDMR tend to achieve better results when addressing large deformation registration problems. The trend is particularly pronounced in the test results of the AMOS dataset. The last two columns of Table.1 presents each model's inference time and parameters. It is evident that all models exhibit fast inference times. Regarding the model parameter, our proposed model falls into the mid-range category. Thus, considering both performance and model complexity, our model achieves a better balance.

Ablation Study on Recursive Inference: Table 2 presents the model's performance based on different recursion depths. The choice of recursion follows two rules: (1) select the setting yielding the highest DSC; (2) if DSC is equal, choose the configuration with lower computational cost. We first conducted a single recursive analysis across different scales, corresponding to columns 3-6 in Table 2. For the HB and HC datasets, the results improved when the configuration was set to (a,b,c,d) = 1121. We further tested the configuration 1131 and found that the performance declined. Therefore, the best model structure for the HB and HC datasets was 1121. Similarly, we analyzed the AMOS dataset and found that the performance improved with the configuration (a,b,c,d) = 1211, but declined when set to 1311. Thus, the most suitable model structure for the AMOS dataset was 1211. The experimental results indicate that, for the trained model, a simple inference test is sufficient to determine its final structure, further enhancing performance.

5 Conclusion

We propose a novel multimodal deformable registration model RDMR, which effectively addresses the challenges of multi-modal registration, large deformations,

Table 2. shows the recursive registration results at different scales. The four-digit number represents the number of recursions at each scale, corresponding to the model's structural parameters a, b, c, and d.

	1111	2111	1211	1121	1112	1131	1311
HB	87.7±8.2	87.1±9.4	87.2±8.8	**88.0**±8.0	87.0±8.3	87.9±7.9	-
HC	86.2±15.0	80.7±14.5	83.9±13.1	**87.0**±11.3	86.8±14.2	86.8±12.9	-
AMOS	84.7±5.7	82.7±6.4	**84.9**±5.8	84.6±5.8	84.8±6.1	-	84.8±6.0

and model generalizability. We propose a multiscale registration feature fusion mechanism and a multi-layer modality-invariant contrastive loss constraint to address the challenges of large deformations and multimodal registration, respectively. Additionally, we introduce a recursive inference strategy that further enhances the model's performance. This strategy is versatile and can be adopted by researchers in related fields. Our future research direction involves further optimizing the concept of recursive inference, allowing it to dynamically determine the appropriate number of recursions.

Acknowledgments. This work was supported in part by National Natural Science Foundation of China Grant 62471297, in part by the Ministry of Science and Technology of the People's Republic of China under Grant 2023YFC2411401, in part sponsored by the Fundamental Research Funds for the Central Universities(No.YG2025ZD03).

Disclosure of Interests. The authors have no competing interests to declare relevant to this article's content.

References

1. Zou J, Gao B, Song Y, et al. A review of deep learning-based deformable medical image registration[J]. Frontiers in Oncology, 2022, 12: 1047215.
2. Xiao H, Teng X, Liu C, et al. A review of deep learning-based three-dimensional medical image registration methods[J]. Quantitative Imaging in Medicine and Surgery, 2021, 11(12): 4895.
3. Mok T C W, Chung A C S. Large deformation diffeomorphic image registration with laplacian pyramid networks[C]//Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part III 23. Springer International Publishing, 2020: 211-221.
4. Zheng J Q, Wang Z, Huang B, et al. Residual Aligner-based Network (RAN): Motion-separable structure for coarse-to-fine discontinuous deformable registration[J]. Medical Image Analysis, 2024, 91: 103038.
5. Song X, Guo H, Xu X, et al. Cross-modal attention for MRI and ultrasound volume registration[C]//Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part IV 24. Springer International Publishing, 2021: 66-75.

6. Shi J, He Y, Kong Y, et al. Xmorpher: Full transformer for deformable medical image registration via cross attention[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer Nature Switzerland, 2022: 217-226.
7. Liu L, Huang Z, Liò P, et al. PC-SwinMorph: Patch representation for unsupervised medical image registration and segmentation[J]. arXiv preprint arXiv:2203.05684, 2022.
8. Guo T, Wang Y, Shu S, et al. Mambamorph: a mamba-based framework for medical mr-ct deformable registration[J]. arXiv preprint arXiv:2401.13934, 2024.
9. Ma T, Zhang S, Li J, et al. IIRP-Net: Iterative Inference Residual Pyramid Network for Enhanced Image Registration[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2024: 11546-11555.
10. Tan Z, Zhang L, Lv Y, et al. GroupMorph: Medical Image Registration via Grouping Network with Contextual Fusion[J]. IEEE Transactions on Medical Imaging, 2024.
11. 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, 38(8): 1788-1800.
12. Mok T C W, Li Z, Bai Y, et al. Modality-Agnostic Structural Image Representation Learning for Deformable Multi-Modality Medical Image Registration[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2024: 11215-11225.
13. Pham X L, Luu M H, van Walsum T, et al. CMAN: Cascaded Multi-scale Spatial Channel Attention-guided Network for large 3D deformable registration of liver CT images[J]. Medical Image Analysis, 2024, 96: 103212.
14. Kang M, Hu X, Huang W, et al. Dual-stream pyramid registration network[J]. Medical image analysis, 2022, 78: 102379.
15. Hu B, Zhou S, Xiong Z, et al. Recursive decomposition network for deformable image registration[J]. IEEE Journal of Biomedical and Health Informatics, 2022, 26(10): 5130-5141.
16. Park T, Efros A A, Zhang R, et al. Contrastive learning for unpaired image-to-image translation[C]//Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part IX 16. Springer International Publishing, 2020: 319-345.
17. Hu Y, Zhang S, Li W, et al. Unsupervised medical image synthesis based on multi-branch attention structure[J]. Biomedical Signal Processing and Control, 2025, 104: 107495.
18. Chen T, Kornblith S, Norouzi M, et al. A simple framework for contrastive learning of visual representations[C]//International conference on machine learning. PMLR, 2020: 1597-1607.
19. Zhao S, Lau T, Luo J, et al. Unsupervised 3D end-to-end medical image registration with volume tweening network[J]. IEEE journal of biomedical and health informatics, 2019, 24(5): 1394-1404.
20. Heinrich M P, Jenkinson M, Bhushan M, et al. MIND: Modality independent neighbourhood descriptor for multi-modal deformable registration[J]. Medical image analysis, 2012, 16(7): 1423-1435.
21. Ma J, Zhang Y, Gu S, et al. Abdomenct-1k: Is abdominal organ segmentation a solved problem?[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 44(10): 6695-6714.
22. Chen J, Frey E C, He Y, et al. Transmorph: Transformer for unsupervised medical image registration[J]. Medical image analysis, 2022, 82: 102615.

23. Chen J, He Y, Frey E C, et al. Vit-v-net: Vision transformer for unsupervised volumetric medical image registration[J]. arXiv preprint arXiv:2104.06468, 2021.
24. Ma T, Dai X, Zhang S, et al. PIViT: Large deformation image registration with pyramid-iterative vision transformer[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer Nature Switzerland, 2023: 602-612.
25. Jia X, Bartlett J, Zhang T, et al. U-net vs transformer: Is u-net outdated in medical image registration?[C]//International Workshop on Machine Learning in Medical Imaging. Cham: Springer Nature Switzerland, 2022: 151-160.
26. Wang H, Ni D, Wang Y. Recursive deformable pyramid network for unsupervised medical image registration[J]. IEEE Transactions on Medical Imaging, 2024.
27. Dalca A V, Balakrishnan G, Guttag J, et al. Unsupervised learning of probabilistic diffeomorphic registration for images and surfaces[J]. Medical image analysis, 2019, 57: 226-236.