DGMIR: Dual-Guided Multimodal Medical Image Registration based on Multi-view Augmentation and On-site Modality Removal

Gao Le[1,2], Yucheng Shu[1,2,3] ✉, Lihong Qiao[1,2,3], Lijian Yang[1,2], Bin Xiao[1,2], Weisheng Li[1,2], and Xinbo Gao[1,2]

[1] Chongqing University of Posts and Telecommunications, Chongqing 400065, China
[2] Chongqing Key Laboratory of Image Cognition, Chongqing 400065, China
[3] Chongqing Key Laboratory of Precision Diagnosis and Treatment for Kidney Diseases, Chongqing 400065, China
{shuyc, qiaolh, xiaobin, liws, gaoxb}@cqupt.edu.cn, d240201011@stu.cqupt.edu.cn, eeejyang@gmail.com

Abstract. Multi-modal medical image registration integrates complementary information from various modalities to deliver comprehensive visual insights for disease diagnosis, treatment planning, surgical navigation, etc. However, current methods often suffer from artifacts, computational overhead, or insufficient handling of modality-specific interference. Moreover, they still rely on specialized modules, such as generative trans-modal units, additional encoders, or handcrafted modality-invariant operators, without fully exploiting the inherent potential of registration features. To address these drawbacks in multimodal medical image registration, we propose a novel registration framework. First, a plug-and-play architecture is proposed to directly process multi-scale heterogeneous features, with active guidance only during deformation field generation stage. Second, we introduce a multi-view feature reorganization module that dynamically optimizes feature distributions via adaptive relation computation and global calibration. Finally, an in-network modality removal module is introduced to leverage multi-scale adaptive convolutions to explicitly eliminate modality-specific interference. Extensive experiments on the BraTS2018 and Learn2Reg2021 datasets confirm that our proposed method achieves state-of-the-art performance on multiple multimodal medical image registration metrics. (https://github.com/St-Antonio/DGMIR)

Keywords: multimodal medical image registration · feature disentanglement · feature removal.

1 Introduction

Deformable image registration (DIR) is a fundamental technique in medical image processing. It aims to establish spatial alignment between two or more images, facilitating a wide range of clinical downstream workflows, such as disease

diagnosis, treatment planning, surgical navigation, etc. Recent advances in deep learning have driven significant progress in DIR, leading to notable improvements in both computational speed and registration accuracy [2–4].

However, advancements in imaging technology have led to widespread use of multimodal medical images like CT, MRI, etc. Their complementary diagnostic value makes multimodal image registration a critical clinical requirement. Nevertheless, due to variations in acquisition protocols, the same or similar anatomical structures often exhibit significant visual differences across imaging modalities. While existing registration frameworks are still limited by this semantic-appearance mismatch phenomenon, recent research has increasingly focused on multimodal medical image registration methodologies.

Among these researches, image translation-based strategies are widely employed [5–8]. By mapping images from different modalities to a unified modality via generative network, multimodal registration can be simplified to a monomodal task, notably reducing the influence of visual heterogeneity. However, these methods often introduce artifacts and require additional computational resources [1]. Additionally, some studies [9, 10] have introduced cross-modal-interaction networks to enhance feature correlation between multimodal images. However, such methods merely integrate generic modules, such as attention mechanisms, into the registration task, failing to address the fundamental challenge of modality-specific confounding factors in cross-modal feature alignment.

Recently, some methods attempt to distill task-specific representations essential for multimodal registration [13–15]. For instance, Qin et al. [13] introduced a feature space decomposition approach, separating the shared shape space from the modality-specific appearance space. Wang et al. [14] developed distinct encoders to separately extract general and structural features, further optimizing structural representation via a self-similarity module. Mok et al. [15] proposed to leverage neighborhood self-similarity and anatomy-based contrastive learning to achieve highly discriminative, contrast-invariant representations.

In sum, while existing methods demonstrate competent multimodal registration performance, they predominantly rely on additional network components, such as generative trans-modal units, specialized encoding branches, or handcrafted modality-invariant feature engineering strategies, without fully exploiting the potential inherent in the registration features themselves. Therefore, in this paper, we argue that the scaling-law should not be taken for granted: before stacking various fancy network modules, it is imperative to first focus on generalized feature learning frameworks and conduct an in-depth analysis to explore their upper limits in multimodal registration tasks. Therefore, in this study, we propose a novel multimodal DIR model, namely DGMIR, to address the aforementioned issues. Our main contributions are as follows:

- We propose a flexible multimodal registration framework that directly integrates multi-scale heterogeneous features in a plug-and-play manner. By implementing active guidance only during the deformation field decoding and generation phases, our approach iteratively achieves accurate multimodal registration.

- We introduce a multi-view feature reorganization guidance module that dynamically modulates feature distributions through adaptive relation computation and global calibration factor, thereby enhancing the discriminative power of multimodal registration features.
- We propose a modality on-site removal guidance module, which leverages adaptive mean convolutions across multiple scales within the registration backbone, to explicitly learn and eliminate modality information for the continual optimization of the deformation field.

Extensive experiments on two public multimodal datasets, BraTS2018[22] and Learn2Reg2021[23], demonstrate our method's superiority over state-of-the-art approaches, achieving significantly improvements in multimodal registration.

2 Methodology

2.1 Overall Structure

The overall structure of the network is shown in Fig.1. Given a pair of fixed and moving image, I_F and I_M, the model outputs a deformation field ϕ: $U_\theta(I_F, I_M) = \phi$, where U_θ denotes the registration network.

The encoder adopts a four-stage architecture, where each stage contains just one convolutional layer followed by a ReLU activation function. The number of output channels in each stage is $\{16, 32, 32, 64\}$, with the first stage preserving the original resolution and the subsequent three stages successively reducing the feature map size by half. The encoder produces two sets of feature maps: $\{F_1, F_2, F_3, F_4\}$ and $\{M_1, M_2, M_3, M_4\}$. The decoder adopts a coarse-to-fine manner across four stages, each comprising a sequential combination of a multi-view feature reorganization guidance module and a modality on-site removal guidance module.

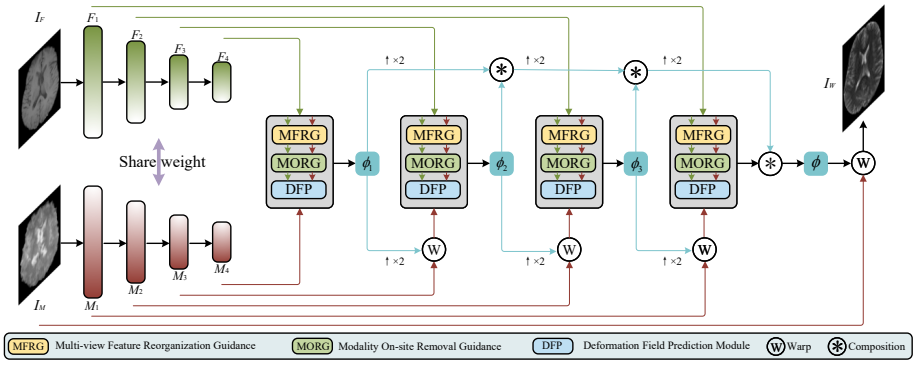

Fig. 1. The Overall structure of the proposed network.

2.2 Multi-view Feature Reorganization Guidance (MFRG)

Through preliminary experiments, we discovered that there are notable disparities between multi-modal image features due to different imaging protocols, and these disparities are predominantly manifested as gaps in the statistical distribution within the channel dimension[24]. In particular, the mean and maximum are two key statistics within each channel, which can offer different views to guide feature modulation. Therefore, we propose a multi-view guided feature enhancement mechanism, which integrates aforementioned statistics indicators to evaluate the importance of channel with different distribution. By combining these two distribution indicators, we achieve a more comprehensive characterization of feature distribution. Moreover, we incorporate a learnable global calibration factor to guide holistic feature expression by adaptively fusing these complementary indicators. Finally, we quantify inter-modal feature relationships and selectively strengthen the feature representation. The specific structure is shown in Fig.2.

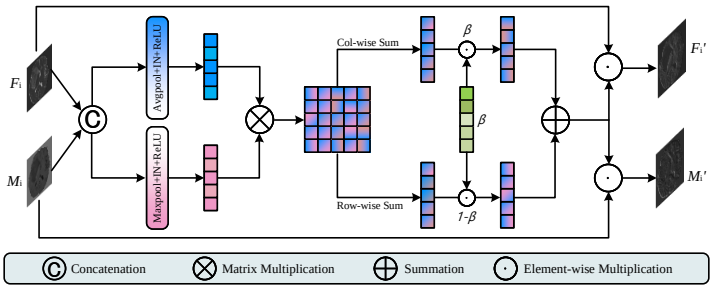

Fig. 2. Multi-view Feature Reorganization Guidance module.

For fixed and moving image features $F_i, M_i \in \mathbb{R}^{c\times h\times w\times d}$, where i indicates the encoder stage, $h \times w \times d$ specifies the spatial dimensions of the feature map, and c denotes channel number. We begin by concatenating F_i and M_i, then independently applying max pooling and average pooling layers to the combined features. Subsequently, we use two $1\times1\times1$ convolution layers to both shrink and recover the feature descriptors based on the pooled outputs, thereby eliminating redundant information while retaining salient features. After deriving the feature global expressions, we calculate inter-modal cross-correlation and intra-modal self-correlation through matrix multiplication. Inspired by [25, 26], we utilize row-wise and column-wise summations to capture correlations between the channels and quantify feature distribution's importance relative to all others, then integrate these importance scores with a calibration factor β. Ultimately, feature expressions are refined through element-wise multiplication of the adaptive enhancement factor and the original features. By introducing β, the network not only adaptively modulates feature expressiveness but also suppresses extraneous representations.

2.3 Modality On-site Removal Guidance (MORG)

In our multimodal registration experiments, we found the modality feature exhibits smaller gradient variations and a relatively uniform distribution. We also decomposed the image into high and low-frequency components and reconstructing it only with the high-frequency part, we found that the modality-independent features can be effectively preserved. Based on these observations, we assume that uniformly distributed modality features may interfere the inter-modal correspondence to accurately predicting the deformation field. However, current methods do not explicitly exclude interference from these modality-specific features. Therefore, we propose a guided coding method that explicitly decodes modality information and progressively eliminates modality-specific features at multiple scales during the decoding process. To the best of our knowledge, this marks the first attempt to explicitly perform modality removal within the feature decoding and flow generation process.

In digital image processing, mean filtering attenuates high-frequency components while preserving low-frequency information. It utilizes a fixed-size sliding window, computing the arithmetic mean of the enclosed pixels to replace the center pixel. Therefore, we incorporate mean filtering properties as prior knowledge into the convolution operation. This approach not only guides the convolution to be more sensitivity to uniformly distributed features but also offers the flexibility to fine-tune the convolution weights for different modalities. Guided by these considerations, we designed a modality feature removal module.

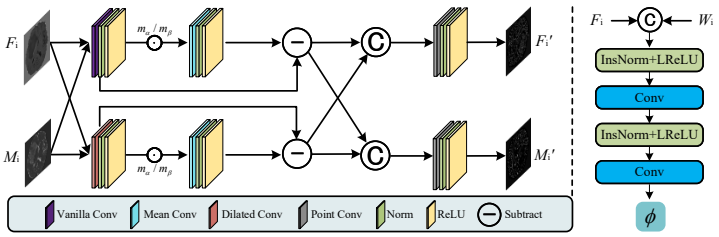

Fig. 3. Left: Modality On-site Removal Guidance module. Right: Deformation Field Prediction module.

Next, we provide the detailed implementation of our method as illustrated in Fig.3. Initially, we apply both a 3×3×3 vanilla convolution and a dilated convolution to the input features, respectively. These two distinct convolution techniques facilitate feature extraction across varying receptive field ranges. Next, the vanilla convolution outputs undergo a 3×3×3 mean convolution, whereas the dilated convolution outputs undergo a 5×5×5 mean convolution. Specifically, the weights of these mean convolutions are initialized using a uniform distribution. Subsequently, we subtract the mean convolution output from the vanilla convolution output, producing de-modality features at the current scale.

Finally, the subtracted outputs are concatenated and fed into a pointwise convolution layer, yielding the module's final output. Where m_α and m_β are learnable parameters act as modality-specific parameters to dynamically adjust feature representations for different modalities by Hadamard product with features.

Eventually, we adopt a coarse-to-fine strategy for predicting the deformation field. In the initial decoding stage, the network inputs are the original encoder features F_1 and M_1. In later stages, the features become F_i and $M_i \circ \phi_{i-1}(W_i)$, where ϕ_{i-1} is obtained through $2\times$ upsampling of the previous stage's sub-deformation field, and the operator $\circ$ is applied through the Spatial Transformation Network[16]. Lastly, the deformation field at each stage is derived by merging the output of the deformation field prediction module (DFP) with the sub-deformation field from the preceding stage. Implementation specifics of the DFP appear in Fig.3, and its key formula is presented below:

$$\phi_i = \begin{cases} \phi_i^{'},\ \phi_i^{'} = \mathrm{DFP}(F_i,\ M_i)\ i = 1 \\ (\phi_{i-1} \circ \phi_i^{'}) + \phi_i^{'},\ \phi_i^{'} = \mathrm{DFP}(F_i,\ M_i \circ \phi_{i-1})\ i = 2, 3, 4 \end{cases} \tag{1}$$

3 Experiments

Datasets and Experimental details: In this study, we employ two public datasets: BraTS2018 and Learn2Reg2021. BraTS2018 comes from the MICCAI Brain Tumor Segmentation (BraTS) 2018 Challenge, consisting of 285 imaging volumes across four MRI modalities (T1, T1ce, T2, FLAIR), each accompanied by expert segmentation labels. All scans measure $240\times240\times155$ at 1 mm isotropic resolution. We validate our proposed model using T1 and T2 modalities. Since the original dataset is pre-aligned, we introduce random misalignment by applying elastic transformations and Gaussian smoothing to T2 images, followed by cropping them to $160\times192\times128$. Finally, we randomly split the dataset into training, validation, and test sets, with 185, 35, and 65 examples, respectively. Learn2Reg2021 (L2R2021) is an abdominal multimodal dataset comprising CT and MRI scans from the same patients, each accompanied by expert segmentation labels. All scans measure $192\times160\times192$ at 1 mm isotropic resolution and have been rigidly pre-aligned. The dataset is then divided into training, validation, and test subsets, comprising 5, 1, and 2 scans respectively.

We implement our network in PyTorch and train it with the Adam optimizer with a learning rate of 1e-4, using a batch size of 1 for 500 epochs. The result for each method is reported on the same PC with 3.2GHz CPU and RTX 4090 GPU. For evaluation, we use Dice score (DCS), 95% Hausdorff distance (HD95), and the percentage of negative values of the Jacobian determinant ($\%|J_\phi| \leq 0$) to assess registration performance.

In this study, we use $\mathcal{L}_{\mathrm{MIND}} = \frac{1}{N}\sum_{x\in\Omega}\sum_{n\in\mathcal{N}}\exp(-||D_F(x,n) - D_W(x,n)||^2)$ [21] as similarity loss function, where $D(x,n) = \left(\frac{|I(x)-I(x+n)|^2}{V(x)}\right)$, $I(x)$ represents the intensity value of the image at position x, $(x+n)$ denotes a point within the neighborhood, $V(x)$ is the local variance at position x. When segmentation

labels are available, we also introduce a weakly supervised loss function $\mathcal{L}_{\text{Dice}} = -\frac{1}{C}\sum_{i=1}^{C}\frac{2|S_{Fi}\cap S_{Wi}|}{|S_{Fi}|+|S_{Wi}|}$, where S_F and S_W represent the segmentation labels of the fixed and warped images, respectively, and C denotes the number of categories in the segmentation labels. In addition to similarity-based loss, we incorporate a regularization term $\mathcal{L}_{\text{reg}} = \sum \|\nabla\varphi\|^2$ to ensure the continuity and smoothness of the deformation field. The final loss function is $\mathcal{L} = \mathcal{L}_{\text{MIND}} + \lambda_1\mathcal{L}_{\text{Dice}} + \lambda_2\mathcal{L}_{\text{reg}}$, where λ_1 and λ_2 are the trade-off parameters between different terms, and in this experiment, they are set to 1 and 0.5, respectively.

We compare our model with a series of state-of-the-art deep learning methods, including VoxelMorph[3], TransMorph[2], GroupMorph[18], TransMatch[19], ModeT[17], and CorrMLP[20]. All comparison methods use the same loss function aforementioned, and all hyperparameter settings are kept unchanged with the original code.

Quantitative and Qualitative Analysis: Table 1 presents the quantitative comparison results of various methods on two datasets. In both datasets, the proposed DGMIR demonstrates a substantial advantage across all primary evaluation metrics. In terms of the DCS metric, our method surpasses the second-best approach by 2.2% and 1.6% on BraTS2018 and L2R2021, respectively. Regarding the HD95 metric, our method outperforms all competitors on BraTS2018 but trails the top-performing method by 0.98 on L2R2021. We also visualized the registration results of all methods on BraTS2018, as shown on the left panel of Fig.4. To confirm the statistical significance of these results, we conducted two-sided Wilcoxon signed-rank tests on the DSC and HD95 metrics, comparing our method against the second-best approach in each dataset. The obtained p-values ($p \ll 0.05$) consistently reject the null hypothesis, confirming the statistical significance of our method's superior performance. In addition, the same order of magnitude or lower $\%|J_\phi| \leq 0$ results in the table show that DGMIR maintains a good anatomical topology while maintaining excellent registration accuracy.

To assess the effectiveness of the proposed modules, we conducted systematic ablation experiments on the BraTS2018 dataset, as presented in Table 2. The baseline model includes only the deformation field prediction module in the decoder. Experimental findings reveal that adding MFRG and MORG individually boosts DSC by 8.8% and 8.9%, respectively, confirming each component's independent effectiveness. Furthermore, we examine the efficacy of mean convolution in MORG. In Table 2, where *no_ grad* signifies that the mean convolution weights are not learned via backpropagation, *ord_ conv* denotes using the standard convolution, *lncc* denotes LNCC loss function. We also visualize the MORG output on the right panel of Fig.4, where the features exhibit a high degree of uniformity after modality removal yet preserve the underlying structural information. By integrating the tabular data with the visualization results, one can see that mean convolution effectively isolates and extracts modality-specific features. Notably, employing the image intensity-based LNCC loss function yields results comparable to MIND. This is chiefly because, after the modality removal

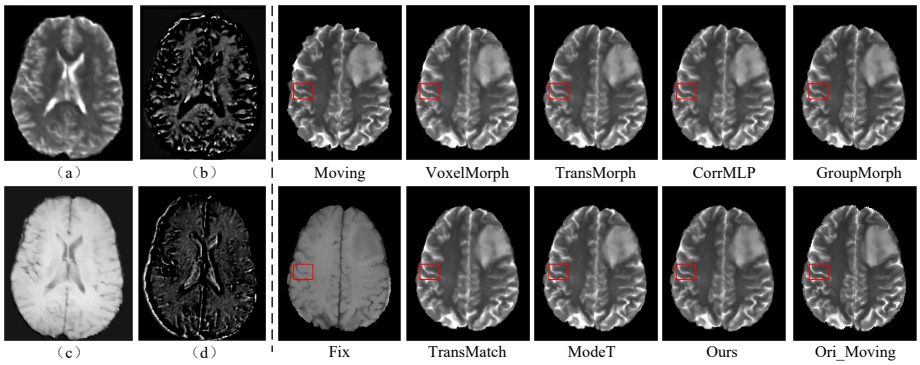

Fig. 4. Left: Visualization of the output of the MORG, where (a) and (c) denote moving and fixed feature from difference modalities (T1 and T2), and (b) and (d) denote the feature after modality removal. Right: Visualization of the registration results, where Ori_Moving denotes the original volume without random elastic transformation.

Table 1. The quantitative results of different registration methods on two datasets.

	BraTS2018			Learn2Reg2021		
	DSC (%)	HD95	$\%\|J_\phi\| \leq 0$	DSC (%)	HD95	$\%\|J_\phi\| \leq 0$
VoxelMorph[3]	75.5	2.546	0.07e-4	59.9	24.77	7.31e-3
TransMorph[2]	77.1	2.334	0.84e-4	60.9	24.42	1.07e-2
CorrMLP[20]	77.2	2.379	0.07e-4	63.6	25.37	2.19e-4
GroupMorph[18]	76.8	2.267	0.85e-4	60.0	24.09	4.56e-3
TransMatch[19]	76.8	2.256	0.06e-4	61.8	**22.69**	6.13e-3
ModeT[17]	74.4	2.602	0.66e-4	63.9	27.27	3.77e-4
Ours	**79.4**	**1.931**	1.52e-4	**65.5**	23.67	9.08e-4

step, the feature space predominantly contains structural features resembling a unimodal distribution, allowing unimodal loss to perform comparably well.

4 Conclusion

We introduce a flexible multimodal medical image registration network that achieves better registration results by actively guiding the prediction of deformation fields at the decoding stage. A multi-view feature reorganization guidance module integrates the global representation of feature distributions under different views and selectively enhances their representation. In addition, we employ a modality on-site removal module, which can explicitly capture modality-specific features and progressively eliminate them during decoding. The experimental results confirm the superior performance of the proposed method.

Table 2. The results of ablation experiment.

Methods	DSC(%)	HD95	$\%\|J_\phi\| \leq 0$
Baseline	69.5	2.936	-
Baseline+MFRG	78.3	2.155	2.35e-4
Baseline+MORG	78.4	1.999	1.53e-4
Baseline+MORG(no_grad)	77.6	2.132	1.53e-4
Baseline+MORG(ord_conv)	77.6	2.171	1.18e-4
Baseline+MFRG($lncc$)	76.5	2.237	1.84e-3
Baseline+MFRG+MORG($lncc$)	78.9	2.130	1.47e-4
Baseline+MFRG+MORG(Ours)	79.4	1.931	1.52e-4

Acknowledgements. This work was supported by the National Natural Science Foundation of China (Nos. 62331008, 62276040), the Natural Science Foundation of Chongqing (Nos. CSTB2023NSCQ-LZX0047, CSTB2024TIAD-KPX0040), National Key Research and Development Program (2024YFB4710100), the Technology Innovation and Application Development Key Projects of Chongqing (CSTB2025TIAD-KPX0009).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Kong, L., Lian, C., Huang, D., Hu, Y., Zhou, Q.: Breaking the dilemma of medical image-to-image translation. Advances in Neural Information Processing Systems, **34**, 1964-1978. (2021)
2. Chen, J., Frey, E.C., He, Y., Segars, W.P., Li, Y., Du, Y.: Transmorph: Transformer for unsupervised medical image registration. Medical Image Analysis 82, 102615 (2022)
3. Balakrishnan, G., Zhao, A., Sabuncu, M.R., Guttag, J., Dalca, A.V.: VoxelMorph: A learning framework for deformable medical image registration. IEEE Transactions on Medical Imaging 38(8), 1788–1800 (2019)
4. Shu, Y., Wang, H., Xiao, B., Bi, X., Li, W.: Medical image registration based on uncoupled learning and accumulative enhancement. In International Conference on Medical Image Computing and Computer-Assisted Intervention pp. 3-13. Cham: Springer International Publishing. (2021)
5. Chen, Z. K., Wei, J., Li, R.: Unsupervised Multi-Modal Medical Image Registration via Discriminator-Free Image-to-Image Translation. In IJCAI. IJCAI Proceedings (2022)
6. Liu, Y., Wang, W., Li, Y., Lai, H., Huang, S., Yang, X.: Geometry-consistent adversarial registration model for unsupervised multi-modal medical image registration. IEEE Journal of Biomedical and Health Informatics, **27**(7), 3455-3466 (2023)
7. Wei, D., Ahmad, S., Huo, J., Huang, P., Yap, P. T., Xue, Z., Sun, J., Li, W., Shen, D., Wang, Q.: SLIR: Synthesis, localization, inpainting, and registration for

image-guided thermal ablation of liver tumors. Medical image analysis, **65**, 101763 (2020)
8. Huang, Z., Chen, B.: Unsupervised Multi-Modal Medical Image Registration via query-selected attention and decoupled Contrastive Learning. In 2024 IEEE International Conference on Multimedia and Expo (ICME) pp. 1-6. IEEE (2024)
9. Song, X., Chao, H., Xu, X., Guo, H., Xu, S., Turkbey, B., Wood, B., Sanford, T., Wang, G., Yan, P.: Cross-modal attention for multi-modal image registration. Medical Image Analysis, **82**, 102612 (2022)
10. Zhang, J., Qing, C., Li, Y., Wang, Y.: BCSwinReg: A cross-modal attention network for CBCT-to-CT multimodal image registration. Computers in Biology and Medicine, **171**, 107990 (2024)
11. Zhong, Y., Zhang, S., Liu, Z., Zhang, X., Mo, Z., Zhang, Y., Hu, H., Chen, W., Qi, L.: Unsupervised Fusion of Misaligned PAT and MRI Images via Mutually Reinforcing Cross-Modality Image Generation and Registration. IEEE Transactions on Medical Imaging. 2023.
12. Zhou, B., Augenfeld, Z., Chapiro, J., Zhou, S. K., Liu, C., Duncan, J. S.: Anatomy-guided multimodal registration by learning segmentation without ground truth: Application to intraprocedural CBCT/MR liver segmentation and registration. Medical image analysis, **71**, 102041 (2021)
13. Qin, C., Shi, B., Liao, R., Mansi, T., Rueckert, D., Kamen, A.: Unsupervised deformable registration for multi-modal images via disentangled representations. In International Conference on Information Processing in Medical Imaging pp. 249-261. Cham: Springer International Publishing (2019)
14. Wang, Z., Wang, H., Ni, D., Xu, M., Wang, Y.: Encoding matching criteria for cross-domain deformable image registration. Medical Physics. (2024)
15. Mok, T. C., Li, Z., Bai, Y., Zhang, J., Liu, W., Zhou, Y. J., Yan, K., Jin, D., Shi, Y., Yin, X., Lu, L. Zhang, L.: Modality-Agnostic Structural Image Representation Learning for Deformable Multi-Modality Medical Image Registration. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition pp. 11215-11225. (2024)
16. Jaderberg, M., Simonyan, K., Zisserman, A., et al.: Spatial transformer networks. Advances in Neural Information Processing Systems pp. 2017–2025 (2015)
17. Wang, H., Ni, D., Wang, Y.: ModeT: Learning deformable image registration via motion decomposition transformer. In International Conference on Medical Image Computing and Computer-Assisted Intervention pp. 740-749. Cham: Springer Nature Switzerland. (2023)
18. Tan, Z., Zhang, L., Lv, Y., Ma, Y., Lu, H.: GroupMorph: Medical Image Registration via Grouping Network with Contextual Fusion. IEEE Transactions on Medical Imaging. (2024)
19. Chen, Z., Zheng, Y., Gee, J. C.: Transmatch: A transformer-based multilevel dual-stream feature matching network for unsupervised deformable image registration. IEEE transactions on medical imaging, **43**(1), 15-27. (2023)
20. Meng, M., Feng, D., Bi, L., Kim, J.: Correlation-aware Coarse-to-fine MLPs for Deformable Medical Image Registration. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition pp. 9645-9654. (2024)
21. Heinrich, M. P., Jenkinson, M., Bhushan, M., Matin, T., Gleeson, F. V., Brady, M., Schnabel, J. A.: MIND: Modality independent neighbourhood descriptor for multi-modal deformable registration. Medical image analysis, **16**(7), 1423-1435. (2012)

22. Gollub, R. L., Shoemaker, J. M., King, M. D., White, T., Ehrlich, S., Sponheim, S. R., ..., Andreasen, N. C.: The MCIC collection: a shared repository of multi-modal, multi-site brain image data from a clinical investigation of schizophrenia. Neuroinformatics, 11, 367-388. (2013)
23. Hering, A., Hansen, L., Mok, T. C., Chung, A. C., Siebert, H., Häger, S., ..., Heinrich, M. P.: Learn2Reg: comprehensive multi-task medical image registration challenge, dataset and evaluation in the era of deep learning. IEEE Transactions on Medical Imaging, **42**(3), 697-712. (2022)
24. Hu, S., Liao, Z., Xia, Y. Devil is in channels: Contrastive single domain generalization for medical image segmentation. In International Conference on Medical Image Computing and Computer-Assisted Intervention pp. 14-23 (2023, October).
25. Gao, Z., Xie, J., Wang, Q., Li, P. Global second-order pooling convolutional networks. In Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition pp. 3024-3033 (2019).
26. Sun, H., Wen, Y., Feng, H., Zheng, Y., Mei, Q., Ren, D., Yu, M. Unsupervised bidirectional contrastive reconstruction and adaptive fine-grained channel attention networks for image dehazing. Neural Networks, 176, 106314 (2024).