EDARFusion: Entropy-Difference-Aware Dynamic Registration for Unaligned Multimodal Medical Image Fusion

1st Shulan Zhang
College of Computer Science
Sichuan Normal University
Chengdu, China
20241393059@stu.sicnu.edu.cn

2nd Guihua Liao
College of Computer Science
Sichuan Normal University
Chengdu, China
136247048@qq.com

3rd Renbin Fang
College of Computer Science
Sichuan Normal University
Chengdu, China
20241302014@stu.sicnu.edu.cn

***Abstract*—Unaligned multimodal medical image fusion faces dual challenges of spatial misalignment and cross-modal information inconsistency. Existing collaborative registration–fusion frameworks often enforce global alignment, causing unreliable regions such as low-texture areas and noise to severely interfere with deformation estimation; they are also prone to inducing erroneous deformations and ghosting in regions with large modality discrepancies. To address these challenges, from the perspective of distinctly handling anatomical misalignment that should be aligned and modality-specific differences that should be preserved, we propose EDARFusion: Entropy-Difference-Aware Dynamic Registration for Unaligned Multimodal Medical Image Fusion. Specifically, the Feature-Reliability Guided Pre-Registration Adapter (FRG-PreReg) predicts a feature-level reliability prior map to suppress the interference of unreliable regions in deformation learning; the Entropy-Difference-Aware Flow Registration (EDA-FlowReg) generates a gating-weight map from local information entropy differences and imposes soft constraints on incremental deformation updates in multi-scale estimation, thereby enabling discrepancy-aware dynamic alignment and reducing ghosting and fusion artifacts caused by misregistration; in the fusion stage, an alignment-consistency constraint is introduced to dynamically integrate structural and modality-specific information on the basis of reliability-aware alignment. Extensive experiments and visualizations of the local information entropy-difference map, the gating-weight map, and the reliability map demonstrate that the proposed method achieves superior registration and fusion performance compared with multiple representative approaches, with interpretable decision-making. The source code is available at https://github.com/Adazsl/EDARFusion**

***Index Terms*—Medical Image Fusion, Multi-modal, Deformation Field, Entropy difference, Reliability prior**

I. Introduction

Due to the limitations of imaging mechanisms, a single medical imaging modality is often unable to simultaneously cover structural details and functional abnormalities. Multimodal Medical Image Fusion (MMIF) aims to integrate complementary information across modalities to generate fused results that are more interpretable for clinical practice [1], [2].

In recent years, deep learning has greatly advanced MMIF, yet most methods assume that the source images are pixel-wise aligned [3]. However, in real scenarios, this assumption frequently fails due to patient motion or device differences, and directly fusing misaligned images may introduce ghosting and pseudo-structures. To mitigate misalignment, traditional pipelines typically adopt a two-stage “registration-then-fusion” strategy, but once registration errors occur, they can be propagated and amplified in the fusion stage [4]. To alleviate error accumulation, recent studies have explored end-to-end registration–fusion collaborative frameworks, e.g., Zhong et al. (2024) proposed PAMRFuse [5] promotes generation and registration to adapt to misalignment; Dong et al. (2025) proposed PFRFusion [6] performs progressive feature alignment for end-to-end collaboration; Li et al. (2025) proposed BSAFusion [7] employs bidirectional step-wise alignment to improve adaptability to large misalignments; and Su et al. (2025) UniFuse [8] proposed incorporates alignment and restoration into a unified framework to enhance robustness against degradations and misalignments.

Nevertheless, these methods still face critical challenges under unaligned settings. First, globally enforced strong-consistency alignment makes unreliable regions significantly interfere with deformation estimation. Second, models often treat “anatomical misalignments that should be aligned” and “modality discrepancies that should be preserved” indiscriminately, which may cause incorrect deformations and ghosting in strong-discrepancy regions and further induce fusion artifacts. Therefore, introducing a controllable and interpretable mechanism into collaborative frameworks to adaptively distinguish “differences to be aligned” from “differences to be preserved”, while suppressing interference from unreliable regions, is key to breaking the performance bottleneck of unaligned MMIF.

Motivated by the above, we propose EDARFusion: Entropy-Difference-Aware Dynamic Registration for Unaligned Multimodal Medical Image Fusion. The main contributions are summarized as follows:

- We design a Feature-Reliability Guided Pre-Registration Adapter (FRG-PreReg), which estimates a feature-level reliability prior from cross-modal feature similarity and uses it to gate correspondence-enhancing context in-

jection, thereby suppressing the interference from low-texture/noisy regions and stabilizing deformation learning.
- We propose an Entropy-Difference-Aware Flow Registration module (EDA-FlowReg), which derives an entropy-difference-based gating weight and applies soft-gated incremental updates in a coarse-to-fine flow estimation, thereby reducing over-registration in strong-discrepancy regions and alleviating ghosting/pseudo-structures caused by misregistration.
- We propose EDARFusion, which performs flow-guided fusion under a unified optimization and targets both spatial misalignment and cross-modal inconsistency.
- We introduce an alignment-consistency regularization for fusion to enforce structural consistency under reliability- and discrepancy-aware alignment, improving robustness while preserving modality-specific patterns. Experiments and visualizations of entropy difference, gating weights, and reliability priors validate its effectiveness and interpretability.

II. Related Work

A. Aligned Multimodal Medical Image Fusion

Early deep MMIF methods mainly relied on CNN-based multi-scale aggregation to preserve structures and complementary information, e.g., DMC-Fusion [9] proposed by Zuo et al. (2021), which enhances detail preservation but still assumes pixel-wise alignment and may introduce ghosting and pseudo structures under misalignment. To compensate for limited long-range dependency modeling in CNNs, Tang et al. (2022) proposed MATR [10], which uses a multi-scale adaptive Transformer to capture cross-modal long-range relations and improve global consistency in aligned settings. Li et al. (2023) proposed DFENet [11], strengthening representations by combining local texture learning (CNN branch) and global semantic modeling (Transformer branch). To mitigate the entanglement of shared information, Zhao et al. (2023) proposed CDDFuse [12], performing correlation-driven feature decomposition to separate modality-shared and modality-specific components. With the rise of generative paradigms, Zhao et al. (2023) proposed DDFM (Denoising Diffusion Model for Multi-Modality Image Fusion) [13], formulating fusion as conditional generation to obtain more natural textures and salient functional patterns. These methods generally assume aligned inputs; misalignment can lead to ghosting and pseudo structures, motivating explicit modeling of alignment.

B. Cross-modal Medical Image Registration and Reliability Modeling

Traditional cross-modal registration often depends on information-theoretic similarity measures, where mutual information (MI) has been widely adopted, with classic representative works by Wells et al. (1996) and Maes et al. (1997) [14], [15]. However, such approaches typically rely on iterative optimization and may suffer from unstable gradients and local minima in low-texture or noisy regions. Deep learning enables end-to-end prediction of deformation fields: VoxelMorph [16] by Balakrishnan et al. (2019) predicts deformable fields in an unsupervised manner, enabling fast registration. Nevertheless, strong cross-modal discrepancies (e.g., lesions or functional hotspots) may still cause mismatch in similarity measures. To improve robustness, reliability mechanisms have been introduced to suppress irrelevant regions, e.g., attention gating [17] by Schlemper et al. (2019). Moreover, to enhance trustworthiness and safety in medical scenarios, studies have emphasized registration error/confidence and uncertainty modeling, e.g.Bierbrier et al. (2022) provided a taxonomy and scoping review on registration error and confidence estimation [18], and Zou et al. (2023) reviewed uncertainty estimation paradigms and applications in medical imaging [19]. These works inspire reducing registration drive in unreliable regions, yet task-coupled modeling for misaligned fusion still requires explicit treatment of cross-modal discrepancies.

C. Joint Registration–Fusion Frameworks for Misaligned MMIF

To reduce error accumulation in the "registration-then-fusion" paradigm, end-to-end joint frameworks have been explored. Xu et al. (2022) proposed RFNet [20] to couple registration and fusion in a unified network, while strong discrepancy regions may still be prone to over-registration. Tang et al. (2022) proposed SuperFusion [4], further introducing structural constraints to enhance synergy, but with increased complexity and training cost. To strengthen end-to-end collaboration, Xu et al. (2023) proposed MURF [21] with a mutually reinforcing mechanism for joint optimization, where deformation stability in strong discrepancy regions remains challenging. Zhong et al. (2024) proposed PAMRFuse [5], using cross-modal generation–registration mutual promotion for specific misaligned data, but its multi-stage design increases overhead.Tang et al. (2025) proposed C2RF [22], leveraging commonality mining and contrastive learning to disentangle common and specific features, alleviating conflicts between registration and fusion objectives. Li et al. (2025) proposed BSAFusion [7], using bidirectional stepwise feature alignment to handle large misalignments, while it still tends to enforce global consistency and may cause over-registration and artifact amplification in strong discrepancy regions.Overall, existing joint frameworks often lack explicit and interpretable mechanisms for suppressing unreliable regions and handling strong discrepancy regions, motivating our reliability-prior and discrepancy-aware improvements.

III. PROPOSED METHOD

In this section, we provide a detailed introduction to EDARFusion, as shown in Fig. 1. EDARFusion comprises three modules: feature preparation, reliability- and entropy-difference-aware flow registration, and flow-guided fusion. The key designs are FRG-PreReg and EDA-FlowReg, which respectively suppress unreliable regions and prevent over-registration in strong-discrepancy areas.

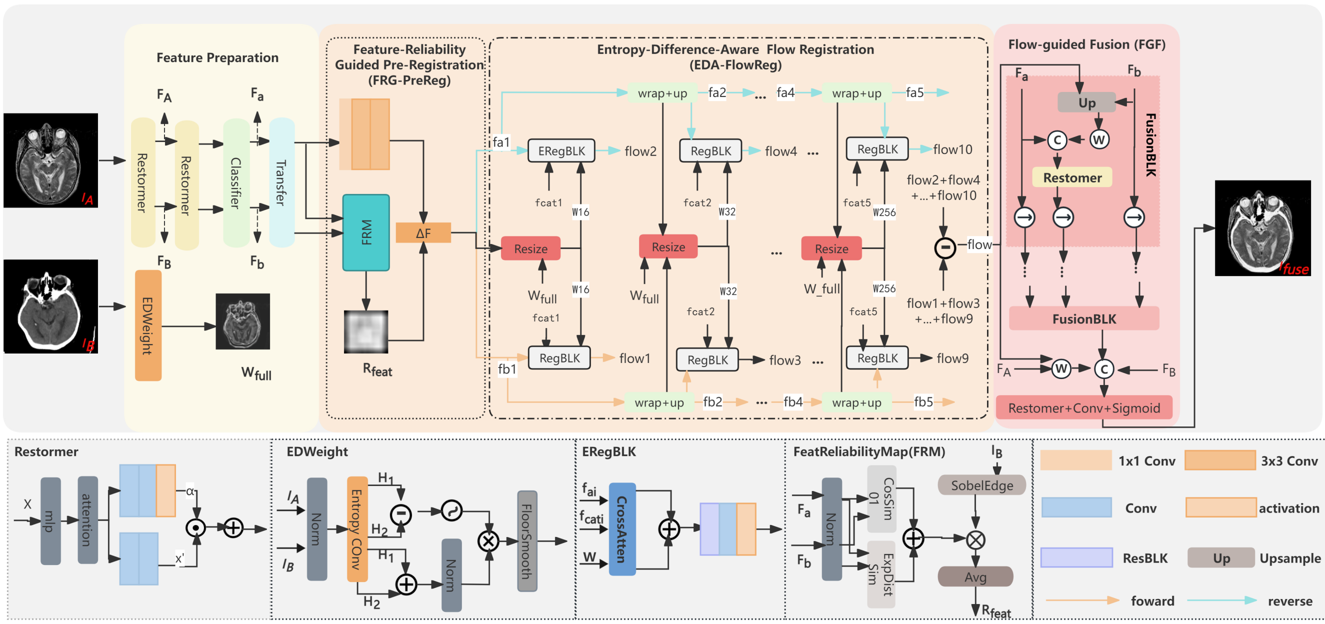

Fig. 1. Model structure of the EDARFusion.

A. *Feature Preparation*

Given an unaligned multimodal input pair $\{I_A, I_B\} \in \mathbb{R}^{1\times H\times W}$, we first adopt a Restormer encoder to extract modality-specific feature maps, and then apply a Transfer module to perform cross-modal interaction to mitigate modality discrepancy. The final feature maps $\{F_a, F_b\}$ are obtained for multimodal alignment, while the base features $\{F_A, F_B\}$ remain available for subsequent fusion.

To adaptively distinguish alignable anatomical misalignment from modality-specific discrepancy, we construct an entropy-difference weight generator that outputs the local entropy maps $\{H_A, H_B\}$, the entropy-difference map ΔH, and a gating weight W_{full} at full resolution. We first normalize each modality to $[0, 1]$, estimate a local histogram-based distribution in a neighborhood $\Omega(x)$, and compute Shannon entropy:

$$H(x) = -\sum_{b=1}^{B} p_b(x) \log_2 (p_b(x)) \tag{1}$$

where B is the number of bins and $p_b(x)$ is the estimated local probability of the bin b at location x. The local entropy difference is defined as:

$$\Delta H(x) = |H_A(x) - H_B(x)| \tag{2}$$

Intuitively, a large ΔH usually indicates a strong cross-modal discrepancy (e.g., functional hotspots or modality-induced appearance changes), which should not be over-registered.

B. *Reliability-and Entropy-Difference-Aware Flow Registration*

Our registration module contains two complementary components: FRG-PreReg and EDA-FlowReg.

1) Feature-Reliability Guided Pre-Registration Adapter:

In unaligned MMIF, globally enforcing alignment often allows unreliable regions to dominate deformation learning, leading to unstable flow updates and artifact amplification. We therefore introduce a reliability prior R_{feat} to indicate where cross-modal feature correspondence is trustworthy and use it to pre-condition the alignment features.Given alignment-oriented features $\{F_a, F_b\}$, FRG-PreReg predicts a reliability map $R_{\text{feat}} \in [0, 1]^{1\times H\times W}$ by combining cosine similarity and distance-based confidence:

$$\begin{aligned} R_{\text{feat}}(x) = \frac{1}{2}\Bigg(& \frac{1 + \cos\big(F_a(x), F_b(x)\big)}{2} \\ & + \exp\Big(- \frac{\|F_a(x) - F_b(x)\|_1}{\tau_f} \Big)\Bigg) \end{aligned} \tag{3}$$

where τ_f controls the sensitivity. In practice, apply lightweight smoothing and a small floor value to avoid vanishing gradients in large low-confidence regions; an optional edge-suppression term can be used to reduce the domination of extremely strong boundaries.

Instead of directly predicting deformation, FRG-PreReg injects registration-friendly correlation context into features. We first build a correlation descriptor:

$$\begin{aligned} Z &= \text{Conv}_{1\times1}\Big([F_a,\ F_b,\ |F_a - F_b|,\ F_a \odot F_b]\Big) \\ \Delta &= \text{Conv}_{3\times3}\big(\sigma(\text{Conv}_{3\times3}(Z))\big) \end{aligned} \tag{4}$$

where $\sigma(\cdot)$ is a non-linear activation (e.g., ReLU). Then we update the alignment features with reliability gating:

$$\begin{aligned} f_a &= F_a + \alpha\, R_{\text{feat}} \odot \Delta \\ f_b &= F_b + \alpha\, R_{\text{feat}} \odot \Delta \end{aligned} \tag{5}$$

where α controls injection strength and $\odot$ denotes element-wise multiplication. To avoid over-sharing in low-reliability regions and preserve modality-specific characteristics, we further add a lightweight fallback term using raw features $\bar{F}_a, \bar{F}_b$:

$$\begin{aligned} f_a &\leftarrow f_a + \beta(1 - R_{\text{feat}}) \odot \bar{F}_a \\ f_b &\leftarrow f_b + \beta(1 - R_{\text{feat}}) \odot \bar{F}_b \end{aligned} \tag{6}$$

where β is a small weight. Eqs. (5)–(6) ensure that reliable regions receive stronger correspondence enhancement, while unreliable regions are prevented from dictating subsequent deformation learning. The refined features $\{f_a, f_b\}$ are fed into EDA-FlowReg.

2) Entropy-Difference-Aware Flow Registration: From the entropy difference in Eq. (2), we construct a soft gating weight map $W_{\text{full}} \in [0,1]^{1\times H\times W}$:

$$\begin{aligned} w(x) &= \exp\Big(-\frac{\Delta H(x)}{\tau + \epsilon}\Big) \\ g(x) &= \text{Norm}\Big(\frac{H_A(x) + H_B(x)}{2}\Big) \end{aligned} \tag{7}$$

$$W_{\text{full}}(x) = \text{clip}\Big(w(x)^{\gamma} \cdot g(x),\ w_{\min},\ 1\Big) \tag{8}$$

where τ controls discrepancy tolerance, γ adjusts the softness of the gate, $\text{Norm}(\cdot)$ denotes min–max normalization, $w_{\min}$ is a floor value to prevent gate collapse and ϵ is a small constant. We resize W_{full} to each pyramid scale s via bilinear interpolation, denoted as W_s.

EDA-FlowReg adopts a coarse-to-fine estimation with K stages. At stage i, the network predicts a deformation increment $\Delta\text{flow}_i^{\text{pred}}$ from the current-stage features (optionally with entropy-aware cross-attention), and we apply a *soft* gate to avoid flow collapse in strong-discrepancy regions:

$$\Delta\text{flow}_i = \Delta\text{flow}_i^{\text{pred}} \odot \Big(\alpha_u + (1-\alpha_u)W_i\Big), \quad \alpha_u > 0 \tag{9}$$

The stage-wise deformation is then updated in an incremental manner:

$$\text{flow}_i = \text{Up}(\text{flow}_{i-1}) + \Delta\text{flow}_i, \qquad \text{flow}_0 = \mathbf{0} \tag{10}$$

where $\text{Up}(\cdot)$ upsamples the flow (with appropriate scale compensation) to match the current resolution.

Moreover, the maps $\{H_A, H_B, \Delta H, W_{\text{full}}, R_{\text{feat}}\}$ provide explicit evidence for "where to align" and "where to preserve".

C. Flow-guided Fusion

The fusion module $\mathcal{G}(\cdot)$ takes the reference features and the aligned moving features to generate a fused image I_F. Given flow_K, we warp modality-B image (and optionally features) toward modality A:

$$\begin{aligned} I_{B\to A} &= \mathcal{W}(I_B, \text{flow}_K) \\ \tilde{F}_B &= \mathcal{W}(F_B, \text{flow}_K) \end{aligned} \tag{11}$$

Then the fused result is obtained by:

$$I_F = \mathcal{G}\big(F_A, \tilde{F}_B, I_A, I_{B\to A}\big) \tag{12}$$

Although the fusion backbone is lightweight and can be implemented by multi-scale feature aggregation, the key benefit of our method lies in that the proposed FRG-PreReg and EDA-FlowReg produce reliability- and discrepancy-aware alignment, yielding higher-quality $\tilde{F}_B$ and $I_{B\to A}$ and thus substantially reducing ghosting/pseudo-structures caused by misregistration. In particular, by suppressing unreliable regions and avoiding over-registration in high-ΔH areas, EDARFusion better preserves modality-specific patterns (e.g., hotspots) while maintaining anatomical consistency in alignable regions.

IV. Experiments

A. Dataset

We evaluate EDARFusion on CT–MRI, PET–MRI, and SPECT–MRI image pairs collected from the Harvard Whole Brain Atlas [23]. Following the common unaligned-fusion setting, we treat MRI as the reference (fixed) modality and synthetically generate misaligned image pairs by applying a mixture of rigid (e.g., rotation and translation) and non-rigid deformations to the other modalities (CT/PET/SPECT). To increase sample diversity, we randomly resample the mixed deformations at each epoch and further apply random rotations and flips as data augmentation.

B. Experimental Settings and Metrics

All experiments are implemented in PyTorch (Python 3.9) and conducted on a single NVIDIA RTX 4060 GPU. We train each sub-dataset for 3,000 epochs with batch size 32 using Adam, with an initial learning rate of 5×10^{-5}. A cosine annealing schedule is adopted with a minimum learning rate of 5×10^{-7}.

To quantitatively evaluate fusion quality, we employ five metrics: $Q_{AB/F}$ [24], Q_{CV} [25], Q_{SSIM} [26], Q_{VIF} [27], and Q_S [28]. Here, $Q_{AB/F}$ evaluates edge information transfer, Q_{CV} measures fusion artifacts, Q_{SSIM} evaluates structural similarity, Q_{VIF} measures visual information fidelity, and Q_S reflects structure-preserving fusion quality. Higher values indicate better performance for all metrics except Q_{CV}, where lower values are better. Following the standard implementation of [25], Q_{CV} is computed on grayscale intensity images rather than normalized ratio scores; therefore, its numerical scale is larger than that of the other metrics.

C. Comparison Experiments

Unaligned MMIF requires collaborative registration–fusion modeling to compensate for spatial misalignment, reduce ghosting artifacts, and preserve modality-specific discrepancies without over-registration. Accordingly, we compare EDARFusion with representative baselines from joint registration–fusion, misalignment-robust fusion, and generative fusion paradigms, including SuperFusion [4], MURF [21], C2RF [22], BSAFusion [7], and DDFM [13]. Whenever possible, official implementations are used; otherwise, faithful reproductions are adopted under the same dataset protocol.

Fig. 2 shows qualitative comparisons. EDARFusion produces sharper boundaries, fewer double contours, and better

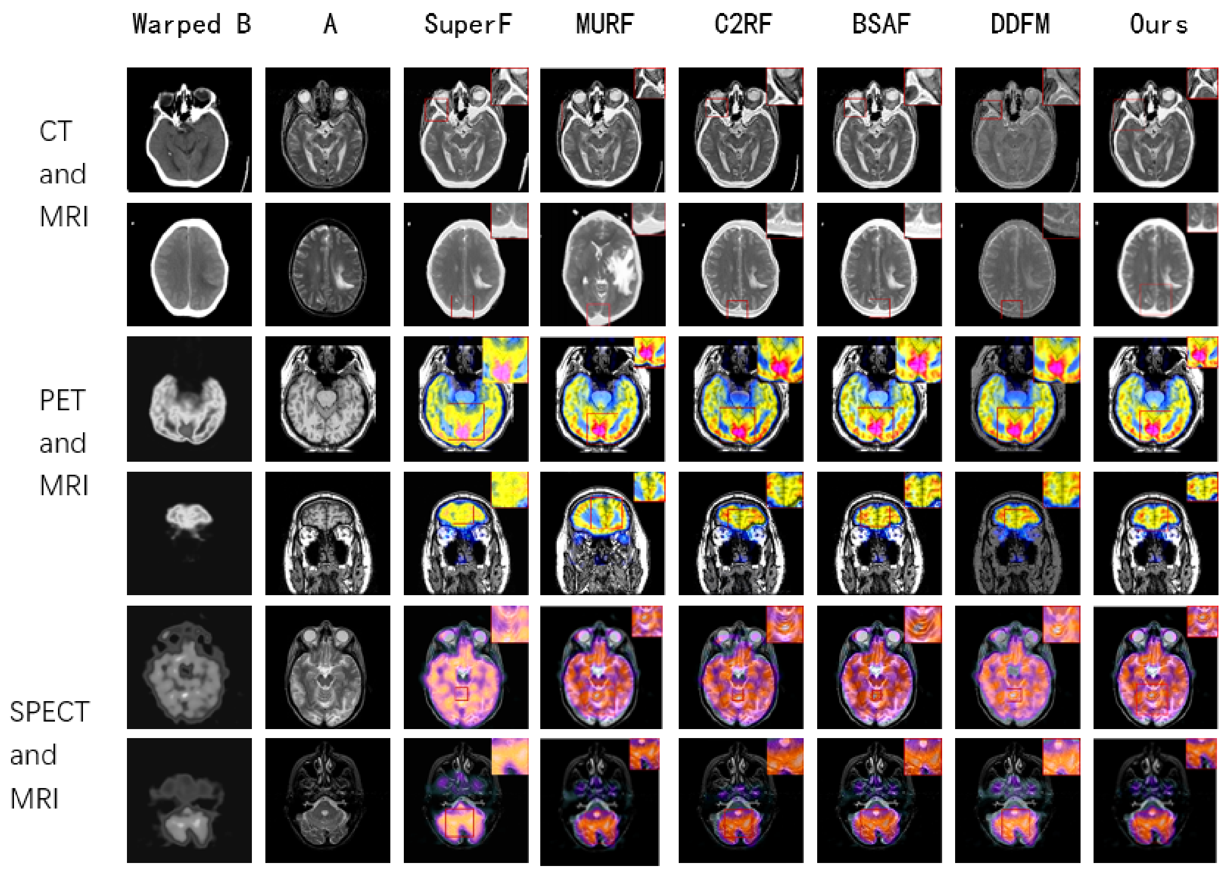

Fig. 2. Visual comparison of fusion results. The first column shows the deformed image to be fused, the second column displays the corresponding label, and the third column presents the MRI image to be fused. Columns 4 to 9 show the results obtained by different fusion methods.

TABLE I
QUANTITATIVE COMPARISON ON CT-MRI, PET-MRI AND SPECT-MRI DATASETS.

Model	CT-MRI					PET-MRI					SPECT-MRI				
	$Q_{AB/F}$	Q_{CV}	Q_{SSIM}	Q_{VIF}	Q_S	$Q_{AB/F}$	Q_{CV}	Q_{SSIM}	Q_{VIF}	Q_S	$Q_{AB/F}$	Q_{CV}	Q_{SSIM}	Q_{VIF}	Q_S
SuperFusion [4]	0.28	4137	0.94	0.30	0.24	0.6	1326	0.95	0.29	0.64	0.45	867	1.03	0.48	0.61
MURF [21]	0.31	5560	1.25	0.21	0.67	0.53	576	0.86	0.26	0.63	0.48	456	0.82	0.27	0.57
C2RF [22]	0.36	3862	1.34	0.32	0.71	0.68	705	0.92	0.44	0.59	0.63	524	0.99	0.60	0.53
BSAFusion [7]	0.38	4728	1.41	0.28	0.72	0.78	434	**1.03**	0.46	**0.76**	0.77	**326**	1.02	0.49	0.76
DDFM [13]	0.32	**2150**	1.46	0.29	0.78	0.52	1081	0.94	0.48	0.63	0.51	446	0.95	**0.61**	0.62
Ours	**0.45**	2828	**1.52**	**0.38**	**0.79**	**0.80**	**323**	1.01	**0.58**	0.74	**0.79**	374	**1.06**	0.54	**0.78**

preservation of modality-specific patterns. Table I summarizes the quantitative results on three datasets. EDARFusion achieves the best overall performance across most metrics, with higher $Q_{AB/F}$ [24] and Q_S [28], competitive Q_{VIF} [27] and Q_{SSIM} [26], and lower Q_{CV} [25], indicating fewer misregistration-induced artifacts.

D. Ablation Experiments

We conduct three ablation variants on PET–MRI to validate the contributions of FRG-PreReg and EDA-FlowReg. Table II reports the quantitative results.

1) Experimental Method 1: Only FRG-PreReg: We keep FRG-PreReg to predict R_{feat} and pre-condition alignment features, while removing entropy-difference gating (setting $W_i \equiv 1$). This variant improves robustness in low-texture/noisy regions, but may still over-align strong-discrepancy areas, leaving residual ghosting.

2) Experimental Method 2: Only EDA-FlowReg: We remove FRG-PreReg (using $R_{\text{feat}} \equiv 1$) while retaining entropy-difference-based gating $\{W_i\}$. It reduces over-registration in high-ΔH regions and improves perceptual quality (e.g., lower Q_{CV}), but remains sensitive to unreliable regions, limiting structural gains.

3) Experimental Method 3: EDARFusion: Combining FRG-PreReg and EDA-FlowReg yields the best overall performance. FRG-PreReg stabilizes deformation learning by sup-

TABLE II
ABLATION ON PET-MRI.

Setting	$Q_{AB/F} \uparrow$	$Q_{CV} \downarrow$	$Q_{\text{SSIM}} \uparrow$	$Q_{\text{VIF}} \uparrow$	$Q_S \uparrow$
Only FRG-PreReg	0.78	360	**1.06**	0.55	0.76
Only EDA-FlowReg	0.79	353	1.05	0.54	**0.77**
FRG + EDA (Ours)	**0.80**	**323**	1.02	**0.58**	0.74

pressing unreliable interference, while EDA-FlowReg prevents erroneous deformation in strong-discrepancy regions via soft gating.

As a result, EDARFusion achieves more accurate alignment and higher-quality fusion with fewer artifacts, as validated by Table II.

V. CONCLUSION

This paper presents EDARFusion for unaligned multimodal medical image fusion, targeting spatial misalignment and cross-modal inconsistency. We introduce FRG-PreReg to suppress unreliable regions via a feature-level reliability prior, and EDA-FlowReg to softly gate multi-scale flow updates using local entropy differences, reducing over-registration in high-discrepancy areas. Experiments on CT–MRI, PET–MRI, and SPECT–MRI demonstrate improved fusion quality over representative baselines, and ablations verify the complementary contributions of the two components. Future work will extend to real clinical misalignment and 3D volumes.

REFERENCES

[1] M. A. Azam, K. B. Khan, S. Salahuddin, E. Rehman, S. A. Khan, M. A. Khan, S. Kadry, and A. H. Gandomi, "A review on multimodal medical image fusion: Compendious analysis of medical modalities, multimodal databases, fusion techniques and quality metrics," *Computers in Biology and Medicine*, vol. 144, p. 105253, 2022.

[2] S. Karim, G. Tong, J. Li, A. Qadir, U. Farooq, and Y. Yu, "Current advances and future perspectives of image fusion: A comprehensive review," *Information Fusion*, vol. 90, pp. 185–217, 2023.

[3] T. Zhou, Q. Cheng, H. Lu, Q. Li, X. Zhang, and S. Qiu, "Deep learning methods for medical image fusion: A review," *Computers in Biology and Medicine*, vol. 160, p. 106959, 2023.

[4] L. Tang, Y. Deng, Y. Ma, J. Huang, and J. Ma, "SuperFusion: A versatile image registration and fusion network with semantic awareness," *IEEE/CAA Journal of Automatica Sinica*, vol. 9, no. 12, pp. 2121–2137, 2022.

[5] Y. Zhong, S. Zhang, Z. Liu, X. Zhang, Z. Mo, Y. Zhang, H. Hu, W. Chen, and L. Qi, "Unsupervised fusion of misaligned PAT and MRI images via mutually reinforcing cross-modality image generation and registration," *IEEE Transactions on Medical Imaging*, vol. 43, no. 5, pp. 1702–1714, 2024.

[6] A. Dong, J. Xu, and L. Wang, "A fusion network for multimodality medical image registration with progressive feature alignment," *Knowledge-Based Systems*, vol. 317, p. 113427, 2025.

[7] H. Li, D. Su, Q. Cai, and Y. Zhang, "BSAFusion: A bidirectional stepwise feature alignment network for unaligned medical image fusion," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 39, no. 5, 2025, pp. 4725–4733.

[8] D. Su, Y. Zhang, H. Li, J. Li, and Y. Liu, "UniFuse: A unified all-in-one framework for multi-modal medical image fusion under diverse degradations and misalignments," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2025, pp. 14 238–14 247.

[9] Q. Zuo, J. Zhang, and Y. Yang, "DMC-Fusion: Deep multi-cascade fusion with classifier-based feature synthesis for medical multi-modal images," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 9, pp. 3438–3449, 2021.

[10] W. Tang, F. He, Y. Liu, and Y. Duan, "MATR: Multimodal medical image fusion via multiscale adaptive transformer," *IEEE Transactions on Image Processing*, vol. 31, pp. 5134–5149, 2022.

[11] W. Li, Y. Zhang, G. Wang, Y. Huang, and R. Li, "DFENet: A dual-branch feature enhanced network integrating transformers and convolutional feature learning for multimodal medical image fusion," *Biomedical Signal Processing and Control*, vol. 80, p. 104402, 2023.

[12] Z. Zhao, H. Bai, J. Zhang, Y. Zhang, S. Xu, Z. Lin, R. Timofte, and L. Van Gool, "CDDFuse: Correlation-driven dual-branch feature decomposition for multi-modality image fusion," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 5906–5916.

[13] Z. Zhao, H. Bai, Y. Zhu, J. Zhang, S. Xu, Y. Zhang, K. Zhang, D. Meng, R. Timofte, and L. Van Gool, "DDFM: Denoising diffusion model for multi-modality image fusion," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 8082–8093.

[14] W. M. Wells, P. Viola, H. Atsumi, S. Nakajima, and R. Kikinis, "Multimodal volume registration by maximization of mutual information," *Medical Image Analysis*, vol. 1, no. 1, pp. 35–51, 1996.

[15] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, "Multimodality image registration by maximization of mutual information," *IEEE Transactions on Medical Imaging*, vol. 16, no. 2, pp. 187–198, 1997.

[16] G. Balakrishnan, A. Zhao, M. R. Sabuncu, J. Guttag, and A. V. Dalca, "VoxelMorph: A learning framework for deformable medical image registration," *IEEE Transactions on Medical Imaging*, vol. 38, no. 8, pp. 1788–1800, 2019.

[17] J. Schlemper, O. Oktay, M. Schaap, M. Heinrich, B. Kainz, B. Glocker, and D. Rueckert, "Attention gated networks: Learning to leverage salient regions in medical images," *Medical Image Analysis*, vol. 53, pp. 197–207, 2019.

[18] J. Bierbrier, H.-E. Gueziri, and D. L. Collins, "Estimating medical image registration error and confidence: A taxonomy and scoping review," *Medical Image Analysis*, vol. 81, p. 102531, 2022.

[19] K. Zou, Z. Chen, X. Yuan, X. Shen, M. Wang, and H. Fu, "A review of uncertainty estimation and its application in medical imaging," *Meta-Radiology*, vol. 1, no. 1, p. 100003, 2023.

[20] H. Xu, J. Ma, J. Yuan, Z. Le, and W. Liu, "RFNet: Unsupervised network for mutually reinforcing multi-modal image registration and fusion," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 19 679–19 688.

[21] H. Xu, J. Yuan, and J. Ma, "MURF: Mutually reinforcing multi-modal image registration and fusion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 10, pp. 12 148–12 166, 2023.

[22] L. Tang, Q. Yan, X. Xiang, L. Fang, and J. Ma, "C2RF: Bridging multi-modal image registration and fusion via commonality mining and contrastive learning," *International Journal of Computer Vision*, vol. 133, no. 8, pp. 5262–5280, 2025.

[23] K. A. Johnson and J. A. Becker, "The whole brain atlas," Harvard Medical School and Brigham and Women's Hospital, n.d., online resource. Available: http://www.med.harvard.edu/aanlib/. Accessed: 2026-03-30.

[24] C. S. Xydeas and V. Petrovic, "Objective image fusion performance measure," *Electronics Letters*, vol. 36, no. 4, pp. 308–309, 2000.

[25] H. Chen and P. K. Varshney, "A human perception inspired quality metric for image fusion based on regional information," *Information Fusion*, vol. 8, no. 2, pp. 193–207, 2007.

[26] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[27] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Transactions on Image Processing*, vol. 15, no. 2, pp. 430–444, 2006.

[28] G. Piella and H. Heijmans, "A new quality metric for image fusion," in *Proceedings of the 2003 International Conference on Image Processing*, vol. 3. IEEE, 2003, pp. III–173–III–176.