SAR-Net: Sample-Agnostic Retrieval Network for Robust Brain Tumor Segmentation with Missing Modalities

Shenhai Zheng[1,2], Weifu Zhou[1], Xue Tang[1], Laquan Li[2,3](✉), and Weisheng Li[1,2]

[1] School of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China
[2] Chongqing College of Artificial Intelligence, Chongqing 401339, China
[3] School of Mathematics and Statistics, Chongqing University of Posts and Telecommunications, Chongqing 400065, China
lilq@cqupt.edu.cn

Abstract. Incomplete multi-modal MRI data severely degrades clinical brain tumor segmentation. Existing solutions utilizing generative synthesis or implicit disentanglement often suffer from hallucinated artifacts or weak feature discriminability. To address these limitations, we propose a novel Sample-Agnostic Retrieval Network (SAR-Net), shifting the paradigm from "fabricating" missing features to "retrieving" latent style prototypes from a global distribution. We explicitly disentangle images into modality-agnostic "content" and sample-agnostic "styles", utilizing a Dynamic Style Memory Bank optimized via Momentum Contrastive Learning to ensure high discriminability. To impute missing modalities, a Global Consensus Voting mechanism retrieves compatible style prototypes, which are then fused with structural content via adaptive modulation. Extensive evaluations on BraTS 2020, together with ablation studies on BraTS 2018, demonstrate that our framework outperforms state-of-the-art methods under diverse missing-modality settings. The code is available at https://github.com/CoCodeRepo/FS2R.

Keywords: Brain tumor segmentation · Missing modality · Retrieval-based framework · Feature disentanglement · Contrastive learning.

1 Introduction

Gliomas, the most common primary malignant brain tumors, exhibit strong invasiveness and heterogeneity, posing a severe public-health threat [1, 15]. Accurate multi-modal MRI segmentation is therefore crucial for diagnosis, treatment planning, and post-operative monitoring. Standard protocols usually include T1-weighted (T1), T1-contrast enhanced (T1ce), T2-weighted (T2), and Fluid-Attenuated Inversion Recovery (FLAIR) scans [11,22], which provide complementary cues for edema, necrotic/non-enhancing tumor core, and enhancing tumor regions [26]. However, complete modality acquisition is often unavailable

in practice, and such missingness can severely degrade standard segmentation models.

To address this challenge, existing research primarily follows three routes. The first, *Modality Synthesis* (e.g., GANs or Diffusion Models), attempts to explicitly generate missing data [5,8,21,23,26]. Despite visual progress, these models are computationally heavy and prone to hallucinating spurious anatomical details under severe missingness, introducing misleading noise that corrupts segmentation [3]. The second route, *Robust Representation Learning*, shifts from external synthesis to internal feature alignment to build a modality-invariant latent space [2,6,10,16,20]. However, lacking an explicit global reference, these implicit methods often suffer from feature entanglement, struggling to establish clear boundaries for reliable cross-modal adaptation [25]. The third route, *Knowledge Distillation*, uses full-modality teacher networks to guide missing-modality students [12, 14, 17–19]. Yet, mainstream pixel-level probability alignment often neglects geometric consistency, leading to fuzzy structural boundaries [19].

Motivated by these limitations, we propose to "retrieve" missing style prototypes from the global data distribution rather than "fabricate" missing features from scratch. The resulting SAR-Net (Sample-Agnostic Retrieval Framework) follows a "Disentangle-Contrast-Retrieve" design logic to resolve the aforementioned issues: 1) Retrieval over Synthesis: To circumvent hallucinations, we utilize a memory bank to retrieve reliable style prototypes instead of generating them from noise; 2) Discriminative Optimization: To resolve feature entanglement, we introduce a Momentum Contrastive Learning mechanism to ensure that style prototypes are clustered into highly discriminative groups; 3) Adaptive Modulation: To preserve structural fidelity, we fuse the retrieved styles with structural content via adaptive feature modulation, maintaining sharp geometric boundaries. The primary contributions are summarized as follows: (1) We propose SAR-Net, a retrieval-based missing-modality framework that uses a Dynamic Style Memory Bank and Global Consensus Voting for feature compensation. (2) We introduce momentum contrastive learning to organize compact and discriminative modality-specific prototypes for reliable retrieval and feature modulation.

2 Method

2.1 Overview

The proposed architecture is illustrated in Fig. 1. SAR-Net first uses parallel dual-branch encoders to decouple multi-modal MRI inputs into modality-agnostic structural content and sample-agnostic modality style features. A Multi-Layer Dynamic Style Memory Bank, optimized with momentum contrastive learning, then improves the discriminability and representativeness of style prototypes. For missing modalities, the available modality styles serve as collaborative queries to retrieve the corresponding missing style prototypes, which are fused with structural content through Adaptive Feature Modulation for feature reconstruction and robust segmentation.

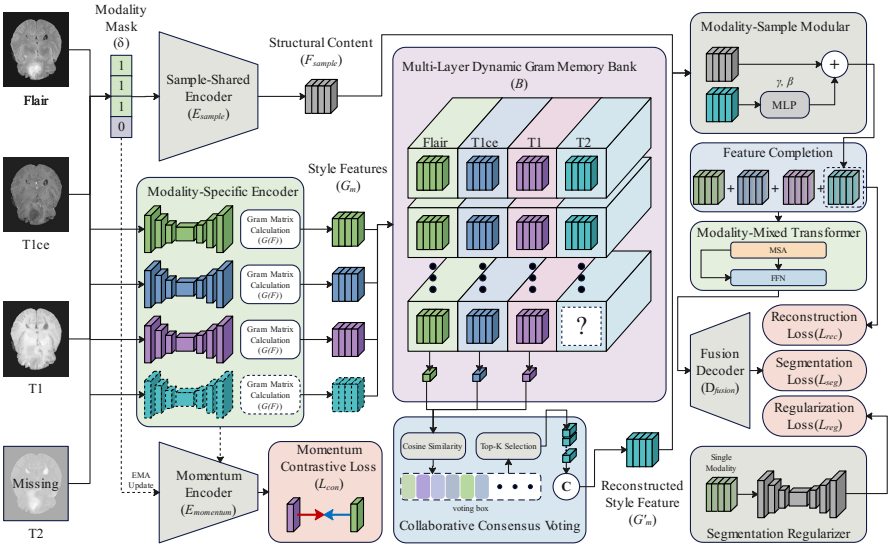

Fig. 1. Overview of SAR-Net. (a) A dual-branch encoder disentangles structural content and style codes, with styles optimized via Momentum Contrastive Learning. (b) For missing modalities, available modalities vote to retrieve the optimal style prototype. (c) The retrieved style modulates structural content via adaptive modulation for accurate segmentation.

2.2 Disentangled Representation Learning

To mitigate feature entanglement, we design a parallel dual-branch architecture. Given an input $\mathbf{X}$ and availability mask δ, a sample-shared encoder E_{sample} extracts modality-agnostic structural content:

$$\mathcal{F}_{sample} = E_{sample}(\mathbf{X} \odot \delta). \tag{1}$$

By sharing weights and processing randomized modality combinations, E_{sample} is constrained to capture robust, modality-invariant geometric information.

Concurrently, for each available modality m, a modality-specific encoder $E^m_{modality}$ extracts the corresponding modality feature map from the isolated input $\mathbf{X}^m$:

$$\mathcal{F}^m_{modality} = E^m_{modality}(\mathbf{X}^m). \tag{2}$$

To explicitly decouple the latent style from the spatial structural content, we compute the Gram matrix for these extracted features. For a feature map F at any layer (flattened to $F_{flat} \in \mathbb{R}^{C\times\Omega}$, where C is channels and Ω is spatial resolution), its Gram matrix is computed to capture global texture and intensity distributions:

$$G(F) = \frac{1}{\Omega} F_{flat} F^{\mathsf{T}}_{flat}. \tag{3}$$

This operation effectively eliminates spatial dimensions, yielding a compact, sample-agnostic style feature independent of individual anatomical variations.

2.3 Contrastive-Enhanced Momentum Memory Bank

To ensure style features evolve into highly discriminative prototypes, we maintain a Multi-Layer Dynamic Gram Memory Bank $\mathcal{B}$, storing the most recent Gram matrices for each modality. To prevent distributional shifts caused by network parameter fluctuations, we employ a momentum encoder $E_{momentum}$ whose parameters θ_k update via an exponential moving average (EMA):

$$\theta_k^{(t)} \leftarrow \mu\theta_k^{(t-1)} + (1-\mu)\theta_{main}^{(t)}. \tag{4}$$

Furthermore, we formulate a Modality Contrastive Loss ($\mathcal{L}_{con}$) based on InfoNCE to cluster same-modality styles and repel different-modality styles. To quantify the stylistic affinity between two Gram matrices $\mathbf{A}$ and $\mathbf{B}$, we utilize the normalized Frobenius inner product as the similarity metric:

$$\text{sim}(\mathbf{A}, \mathbf{B}) = \frac{\text{Tr}(\mathbf{A}^\intercal \mathbf{B})}{\|\mathbf{A}\|_F \cdot \|\mathbf{B}\|_F + \epsilon}. \tag{5}$$

Given a query style G_q, positive samples $\mathcal{S}^+$, and negative samples $\mathcal{S}^-$ sampled from the memory bank, the contrastive objective is defined as:

$$\mathcal{L}_{con} = -\mathbb{E}\left[\log \frac{\sum_{G^+ \in \mathcal{S}^+} \exp(\text{sim}(G_q, G^+)/\tau)}{\sum_{G \in \mathcal{S}^+ \cup \mathcal{S}^-} \exp(\text{sim}(G_q, G)/\tau)}\right]. \tag{6}$$

This explicitly enforces inter-class separability, establishing a robust and discriminative search space for subsequent retrieval.

2.4 Collaborative Retrieval and Modulation

For missing modalities $m \in \mathcal{M}_{miss}$, we propose a Global Consensus Voting strategy that uses available modalities $\mathcal{M}_{avail}$ as collaborative anchors to infer the latent style. Each available modality n queries its queue $\mathcal{B}_n$ using Gram-feature similarity, producing top-k local indices and Softmax-normalized weights. These weighted votes are accumulated in a shared index space to form the global voting score S_{global}. The missing style prototype $\hat{G}^m$ is then reconstructed via Softmax-weighted aggregation of the top-K global consensus samples in the target memory bank:

$$\hat{G}^m = \sum_{j \in \text{Top-K}(S_{global})} \frac{\exp(S_{global}[j])}{\sum_{p \in \text{Top-K}(S_{global})} \exp(S_{global}[p])} \cdot \mathcal{B}_m[j]. \tag{7}$$

To transform the retrieved style prototype $\hat{G}^m$ into a spatially resolved feature map, we introduce an Adaptive Feature Modulation module. Inspired by

AdaIN [7], this modulator uses a Multi-Layer Perceptron to generate affine parameters (γ, β) from $\hat{G}^m$, performing channel-wise scaling and shifting on the normalized structural content F_{sample}:

$$\hat{F}^m_{modality} = \gamma(\hat{G}^m) \cdot \frac{F_{sample} - \mu(F_{sample})}{\sigma(F_{sample}) + \epsilon} + \beta(\hat{G}^m). \tag{8}$$

This effectively injects the target statistical distribution into the generic anatomical structure. We explicitly constrain this generative mapping using a multi-level reconstruction loss ($\mathcal{L}_{rec}$) comprising L1 and MSE metrics.

2.5 Transformer Integration and Optimization

We dynamically substitute missing modality channels with the reconstructed features $\hat{F}^m_{modality}$ to form a complete feature set. To capture global contextual dependencies and exploit inter-modal complementary information, the deepest semantic features are processed through a Modality-Mixed Transformer employing standard Multi-Head Self-Attention (MSA) and Feed-Forward Networks.

Certain modalities exhibit high sensitivity to specific tumor regions (e.g., FLAIR, T1ce). To mitigate "Shortcut Learning"—where the model over-relies on these dominant modalities—we introduce a Segmentation Regularizer ($\mathcal{L}_{reg}$) that forces weight-sharing auxiliary decoders to perform segmentation on single-modal features. The total optimization objective of the proposed framework is defined as:

$$\mathcal{L} = \mathcal{L}_{con} + \mathcal{L}_{rec} + \mathcal{L}_{reg} + \mathcal{L}_{ds} + \mathcal{L}_{seg} \tag{9}$$

where $\mathcal{L}_{seg}$ and $\mathcal{L}_{ds}$ represent the primary and deep-supervised segmentation losses (Weighted Cross-Entropy and Dice loss), respectively.

3 Experiments

Experimental Setup To comprehensively evaluate the robustness and generalizability of our proposed framework, we conducted experiments on two benchmarks from the Multimodal Brain Tumor Segmentation Challenge (BraTS) [11]: BraTS 2018 and BraTS 2020. All datasets comprise multi-contrast MRI scans with four co-registered modalities: s: Flair, T1c, T1, and T2. For the standard benchmarks BraTS 2018 (285 cases) and BraTS 2020 (369 cases), we strictly adhere to the evaluation protocol defining three nested tumor sub-regions: Whole Tumor (WT), Tumor Core (TC), and Enhancing Tumor (ET). To ensure a fair comparison with existing unified models, we adopt specific data splits: 199/29/57 subjects for training/validation/testing on BraTS 2018, and 219/50/100 subjects on BraTS 2020.

Also, we employ the Dice Similarity Coefficient (DSC) and the 95th percentile of Hausdorff Distance (HD95) for quantitative evaluation. Ablations are conducted on BraTS 2018 under the same protocol. Implemented in PyTorch, SAR-Net is trained with Adam (1000 epochs, batch size 1, learning rate 2×10^{-4},

weight decay 1×10^{-4}, polynomial decay), 128^3 crops, standard augmentation, a 250-entry modality-layer memory bank, momentum 0.995, and equal loss weights.

Table 1. Results of the proposed method and state-of-the-art unified models on BraTS 2020 dataset. Dice similarity coefficient (%) is employed for evaluation with every combination of modality settings. • and ∘ denote available and missing modalities, respectively. Method abbreviations denote RF = RFNet [4], MMF = mmFormer [24], M3AE = M^{3}AE [9], and M2FT = M^{2}FTrans [13].

Modalities				**Complete (WT)**					**Core (TC)**					**Enhancing (ET)**				
F	T1	T1c	T2	RF	MMF	M3AE	M2FT	**Ours**	RF	MMF	M3AE	M2FT	**Ours**	RF	MMF	M3AE	M2FT	**Ours**
∘	∘	∘	•	86.1	85.5	86.1	87.2	**88.3**	71.0	63.3	70.3	72.3	**75.5**	46.3	49.0	46.0	**51.5**	51.4
∘	∘	•	∘	76.8	78.0	73.9	78.8	**79.6**	81.5	81.5	81.4	81.9	**83.3**	74.9	78.3	72.4	82.6	**83.2**
∘	•	∘	∘	77.2	76.2	76.7	79.2	**79.9**	66.0	63.2	66.0	66.8	**67.2**	37.3	37.6	39.9	40.9	**41.3**
•	∘	∘	∘	87.3	86.5	86.5	88.7	**89.2**	69.2	64.6	68.0	72.2	**73.5**	38.2	36.6	40.5	43.4	**45.3**
∘	∘	•	•	87.7	87.5	87.4	**88.7**	88.6	83.5	82.6	83.0	**84.6**	84.3	75.9	77.2	76.8	**82.4**	**82.4**
∘	•	•	∘	81.1	80.7	78.1	82.4	**82.8**	83.4	82.8	82.4	83.7	**85.8**	78.0	81.7	75.4	83.8	**86.2**
•	•	∘	∘	89.7	88.7	89.4	**90.3**	89.6	73.1	71.7	73.8	74.4	**74.5**	41.0	42.9	43.2	47.0	**48.4**
∘	•	∘	•	87.7	86.9	87.2	**88.3**	88.1	73.1	67.7	72.5	73.6	**74.9**	45.7	49.1	46.6	49.9	**50.1**
•	∘	∘	•	89.9	89.4	89.3	90.6	**91.3**	74.1	70.3	75.0	75.4	**77.0**	49.3	49.0	47.3	53.9	**55.2**
•	∘	•	∘	89.9	89.3	89.5	90.4	**90.8**	84.7	83.7	82.0	85.5	**86.7**	76.7	79.4	74.7	83.1	**84.0**
•	•	•	∘	90.7	89.7	90.0	**91.0**	90.6	85.1	84.4	82.4	85.8	**87.6**	76.8	80.6	75.9	84.1	**84.8**
•	•	∘	•	90.6	89.8	90.4	**90.9**	**90.9**	75.2	72.4	75.1	76.1	**77.6**	49.9	50.0	48.2	53.3	**53.4**
•	∘	•	•	90.7	90.4	90.2	91.2	**92.2**	85.0	83.9	83.1	85.3	**85.7**	77.1	78.7	77.1	81.2	**84.0**
∘	•	•	•	88.3	87.6	88.6	89.0	**89.3**	83.5	79.0	84.1	84.9	**86.0**	77.0	68.3	77.4	82.4	**84.6**
•	•	•	•	91.1	90.5	90.6	**91.4**	90.7	85.2	84.6	84.4	85.4	**87.6**	78.0	79.9	78.0	82.2	**83.9**
Average				87.0	86.4	86.3	87.9	**88.1**	78.2	75.7	77.6	79.2	**80.5**	61.5	62.6	61.3	66.8	**67.9**

Comparison with State-of-the-Arts To rigorously assess the robustness of our Sample-Agnostic Retrieval Network (SAR-Net) under complex clinical scenarios, we compare it against four representative methods: RFNet [4] (CNN fusion), mmFormer [24] (Transformer fusion), M^{3}AE [9] (Generative masked autoencoder), and M^{2}FTrans [13] (Modality-masked Transformer). To simulate real-world clinical constraints, we report the segmentation performance across all 15 possible missing modality combinations.

As detailed in Table 1, SAR-Net achieves comprehensive superiority across almost all modalities. Compared to the leading competitor M^{2}FTrans, our framework yields average DSC gains of +0.20%, +1.30%, and +1.10% for WT, TC, and ET, respectively. The pronounced improvements in the challenging TC and ET regions underscore the efficacy of our consensus retrieval mechanism.

Particularly noteworthy is the model's resilience under severe missingness. For instance, in the extreme "T2-only" setting, SAR-Net surpasses the runner-up by 3.20% (75.5% vs. 72.3%) for TC segmentation. Furthermore, when segmenting the Enhancing Tumor without T1c (e.g., "FLAIR + T2" row), our method outperforms the best competitor by 1.30% (55.2% vs. 53.9%). This demonstrates SAR-Net's unique capability to infer latent enhancing characteristics by leveraging correlated style prototypes, overcoming the collapse seen in standard generative methods.

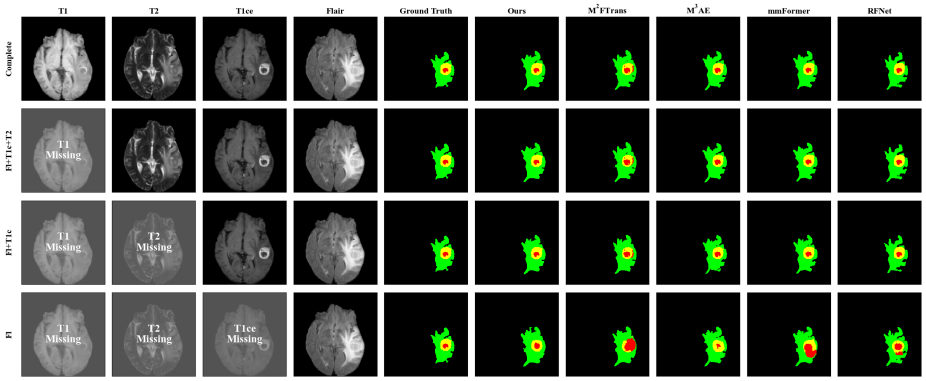

Fig. 2. Visual comparison of segmentation results on BraTS 2020 across diverse missing modality scenarios. Our method consistently preserves geometric fidelity and avoids hallucinated artifacts.

Table 2. Component-wise ablation study on the BraTS 2018 dataset.

Variant	Configuration					Metrics	
	Explicit Disentangle	Memory Bank	Contrastive Learning	Global Consensus	Imputation Source	DSC (%)	HD95 (mm)
A (Baseline)	-	-	-	-	Zero Input	74.40	7.42
B	✓	-	-	-	Content Only	75.15	7.10
C	✓	✓	-	-	Nearest Style	75.80	6.85
D	✓	✓	✓	-	Nearest Style	76.55	6.50
E (Ours)	✓	✓	✓	✓	Voting Style	**77.33**	**6.21**

Ablation Study To validate the "Disentangle-Contrast-Retrieve" paradigm, we conducted component-wise ablations on BraTS 2018 by progressively adding disentanglement (Dis.), memory bank (Mem.), contrastive learning (Con.), and consensus voting (Vote). As shown in Table 2, a baseline U-Net with zero-filling (Variant A) yields an average DSC of 74.40%. Introducing our explicit dual-branch encoder for structural content imputation (Variant B) improves DSC to 75.15%. Injecting nearest-neighbor style via the memory bank and AdaIN (Variant C) brings the DSC to 75.80%. Enforcing Momentum Contrastive Learning (Variant D) ensures style discriminability, jumping to 76.55%. Finally, our full model (Variant E) with Global Consensus Voting optimally filters out single-modality retrieval noise, peaking at 77.33% DSC and 6.21 mm HD95, confirming the synergy of each architectural design.

Loss Function Completeness We performed a "leave-one-out" study on the total objective function (Table 3). Removing the Contrastive Loss ($\mathcal{L}_{con}$) or Reconstruction Loss ($\mathcal{L}_{rec}$) leads to significant performance drops, consistent with our mechanism analysis. Specifically, the absence of $\mathcal{L}_{con}$ causes a severe degradation in geometric accuracy, increasing the HD95 by 0.61 mm, while removing

$\mathcal{L}_{rec}$ leads to the largest DSC drop of 1.10%. Notably, excluding the Regularization Loss ($\mathcal{L}_{reg}$) results in a DSC decline of 0.48% and a boundary error increase of 0.24 mm, confirming its role in preventing modality dominance and shortcut learning. Finally, Deep Supervision ($\mathcal{L}_{ds}$) further stabilizes training, contributing to a 0.38% gain in DSC and a 0.19 mm reduction in HD95.

Table 3. Ablation of core objective terms and auxiliary training stabilizers.

Experiment	DSC (%)	HD95 (mm)	Δ DSC	Δ HD95
Full Objective	**77.33**	**6.21**	-	-
w/o $\mathcal{L}_{\text{con}}$	76.70	6.82	-0.63	+0.61
w/o $\mathcal{L}_{\text{rec}}$	76.23	6.77	-1.10	+0.56
w/o aux. decoder	76.85	6.45	-0.48	+0.24
w/o deep supervision	76.95	6.40	-0.38	+0.19

4 Conclusion

In this paper, we proposed the Sample-Agnostic Retrieval Network (SAR-Net) to robustly address the challenge of missing modalities in brain tumor MRI segmentation. Moving beyond traditional generative and disentanglement methods, we introduced a novel "Disentangle-Contrast-Retrieve" paradigm. By explicitly decoupling modality-agnostic structural content from sample-agnostic styles and leveraging a contrastive-enhanced Dynamic Style Memory Bank, our method transforms challenging modality imputation into a tractable feature retrieval task. The Global Consensus Voting mechanism further ensures high-fidelity style reconstruction, effectively circumventing the hallucination artifacts inherent in generative models. Extensive evaluations demonstrate that SAR-Net achieves state-of-the-art performance and exceptional geometric accuracy across diverse missing-modality scenarios. Future work will explore adaptive memory update strategies to handle extreme pathological heterogeneity and extend this retrieval paradigm to broader multi-modal clinical applications, such as PET-CT co-segmentation.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China(Nos. 62331008, 62221005, 61902046 and 61901074), the Science and Technology Research Program of Chongqing Municipal Education Commission (No. KJZD-M202620605) and the Natural Science Foundation of Chongqing (No. 2023NSCQ-LZX0045).

Disclosure of Interests

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

References

1. Bahadure, N.B., Ray, A.K., Thethi, H.P.: Image analysis for mri based brain tumor detection and feature extraction using biologically inspired bwt and svm. International journal of biomedical imaging **2017**(1), 9749108 (2017)
2. Chen, C., Dou, Q., Jin, Y., Chen, H., Qin, J., Heng, P.A.: Robust multimodal brain tumor segmentation via feature disentanglement and gated fusion. In: International conference on medical image computing and computer-assisted intervention. pp. 447–456. Springer (2019)
3. Cohen, J.P., Luck, M., Honari, S.: Distribution matching losses can hallucinate features in medical image translation. In: International conference on medical image computing and computer-assisted intervention. pp. 529–536. Springer (2018)
4. Ding, Y., Yu, X., Yang, Y.: Rfnet: Region-aware fusion network for incomplete multi-modal brain tumor segmentation. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 3975–3984 (2021)
5. Dorent, R., Joutard, S., Modat, M., Ourselin, S., Vercauteren, T.: Hetero-modal variational encoder-decoder for joint modality completion and segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 74–82. Springer (2019)
6. Havaei, M., Guizard, N., Chapados, N., Bengio, Y.: Hemis: Hetero-modal image segmentation. In: International conference on medical image computing and computer-assisted intervention. pp. 469–477. Springer (2016)
7. Huang, X., Belongie, S.: Arbitrary style transfer in real-time with adaptive instance normalization. In: Proceedings of the IEEE international conference on computer vision. pp. 1501–1510 (2017)
8. Huang, Z., Lin, L., Cheng, P., Peng, L., Tang, X.: Multi-modal brain tumor segmentation via missing modality synthesis and modality-level attention fusion. arXiv preprint arXiv:2203.04586 (2022)
9. Liu, H., Wei, D., Lu, D., Sun, J., Wang, L., Zheng, Y.: M3ae: Multimodal representation learning for brain tumor segmentation with missing modalities. In: Proceedings of the AAAI conference on artificial intelligence. vol. 37, pp. 1657–1665 (2023)
10. Liu, T., Jiang, H., Huang, K.: Kmd: Koopman multi-modality decomposition for generalized brain tumor segmentation under incomplete modalities. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 15663–15671 (2025)
11. Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., et al.: The multimodal brain tumor image segmentation benchmark (brats). IEEE transactions on medical imaging **34**(10), 1993–2024 (2014)
12. Qiu, Y., Chen, D., Yao, H., Xu, Y., Wang, Z.: Scratch each other's back: Incomplete multi-modal brain tumor segmentation via category aware group self-support learning. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 21317–21326 (2023)

13. Shi, J., Yu, L., Cheng, Q., Yang, X., Cheng, K.T., Yan, Z.: Mftrans: Modality-masked fusion transformer for incomplete multi-modality brain tumor segmentation. IEEE journal of biomedical and health informatics **28**(1), 379–390 (2023)
14. Su, J., Luo, Z., Wang, C., Lian, S., Lin, X., Li, S.: Reconstruct incomplete relation for incomplete modality brain tumor segmentation. Neural Networks **180**, 106657 (2024)
15. Wadhwa, A., Bhardwaj, A., Verma, V.S.: A review on brain tumor segmentation of mri images. Magnetic resonance imaging **61**, 247–259 (2019)
16. Wang, H., Chen, Y., Ma, C., Avery, J., Hull, L., Carneiro, G.: Multi-modal learning with missing modality via shared-specific feature modelling. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 15878–15887 (2023)
17. Wang, H., Hassan, S., Liu, Y., Ma, C., Chen, Y., Li, Q., Geng, J., Wang, B., Tian, Y., Xie, Y., et al.: Meta-learned modality-weighted knowledge distillation for robust multi-modal learning with missing data. arXiv preprint arXiv:2405.07155 (2024)
18. Wang, S., Yan, Z., Zhang, D., Wei, H., Li, Z., Li, R.: Prototype knowledge distillation for medical segmentation with missing modality. In: ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 1–5. IEEE (2023)
19. Xie, D., Wu, Y., Ai, Z., Min, J., Jiang, Z., Geng, S., Wang, L.: Ccsd: Cross-modal compositional self-distillation for robust brain tumor segmentation with missing modalities. arXiv preprint arXiv:2511.14599 (2025)
20. Yang, Q., Guo, X., Chen, Z., Woo, P.Y., Yuan, Y.: D 2-net: Dual disentanglement network for brain tumor segmentation with missing modalities. IEEE Transactions on Medical Imaging **41**(10), 2953–2964 (2022)
21. Zeng, Z., Peng, Z., Yang, X., Shen, W.: Missing as masking: Arbitrary cross-modal feature reconstruction for incomplete multimodal brain tumor segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 424–433. Springer (2024)
22. Zhang, D., Huang, G., Zhang, Q., Han, J., Han, J., Yu, Y.: Cross-modality deep feature learning for brain tumor segmentation. Pattern Recognition **110**, 107562 (2021)
23. Zhang, X., Chen, Q., He, H., Zhu, L., Xie, Z., Lu, Y., Cheng, F.: Generative learning-based lightweight mri brain tumor segmentation with missing modalities. Expert Systems with Applications **261**, 125478 (2025)
24. Zhang, Y., He, N., Yang, J., Li, Y., Wei, D., Huang, Y., Zhang, Y., He, Z., Zheng, Y.: mmformer: Multimodal medical transformer for incomplete multimodal learning of brain tumor segmentation. In: International conference on medical image computing and computer-assisted intervention. pp. 107–117. Springer (2022)
25. Zhang, Z., Yang, G., Zhang, Y., Yue, H., Liu, A., Ou, Y., Gong, J., Sun, X.: Tmformer: Token merging transformer for brain tumor segmentation with missing modalities. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 38, pp. 7414–7422 (2024)
26. Zhao, B., Zhang, Y., Zhao, M., Xu, G., Wang, C.: Adaptive auxiliary diffusion for multi-modal brain tumor segmentation with random missing modalities. Biomedical Signal Processing and Control **109**, 108015 (2025)