ReMiX-Seg: Latent Modality Completion Meets Expert Routing for Robust Glioma Segmentation

Van Quang Nghiem, Si-Qin Lyu, and Che Lin(✉)

National Taiwan University, Taipei, Taiwan
chelin@ntu.edu.tw

Abstract. Multi-parametric brain MRI is essential for glioma assessment, yet in clinical workflows, one or more modalities (FLAIR, T1, T1c, T2) are frequently missing, substantially degrading multimodal tumor segmentation. Existing solutions span multiple distilled specialists, single "catch-all" networks that fuse available modalities but often rely on naive zero-filling, and feature reconstruction approaches that recover missing representations but may suffer from mismatches between reconstructed and observed features, limited adaptability to missing-modality patterns, and additional sensitivity in pre-/post-treatment settings. We propose **ReMiX-Seg** (**Re**construction + **Mi**xture-of-e**X**perts for **Seg**mentation), a unified framework that combines lightweight token-level completion for missing modalities with mask-aware mixture-of-experts refinement to adapt fusion and denoising to each observed subset. To further stabilize training, we introduce Expert-wise Consistency Alignment, encouraging expert-specific agreement between reconstructed and observed feature distributions. Crucially, ReMiX-Seg is evaluated across both BraTS2023 (Adult Glioma Pre-treatment) and BraTS2024 (Adult Glioma Post-treatment), demonstrating robustness across domain-shifted settings. ReMiX-Seg achieves state-of-the-art performance on both benchmarks and remains consistently robust across all non-empty modality combinations, supporting practical deployment under incomplete MRI. The code is available at `https://github.com/highquanglity/ReMiX-Seg`.

Keywords: Incomplete Multimodal · Brain Tumor Segmentation · Mixture of Experts.

1 Introduction

Multi-parametric Magnetic Resonance Imaging (MRI) is the clinical standard for assessing gliomas and guiding treatment planning. Complementary MRI modalities—including Fluid-Attenuated Inversion Recovery (FLAIR), T1-weighted (T1), contrast-enhanced T1-weighted (T1c) and T2-weighted (T2)—provide distinct contrast for enhancing tumor (ET), tumor core (TC), whole tumor (WT), and resection cavity (RC), enabling accurate delineation of heterogeneous lesions when jointly modeled [1, 3, 5, 13, 20]. In practice, complete acquisitions are often unavailable: protocols vary across sites and modalities may be missing due to time/cost constraints, motion artifacts, scanner availability, or contraindications.

Deployment further spans both pre- and post-treatment imaging, where therapy-induced anatomical changes introduce additional distribution shift. Together, these factors create arbitrary missing-modality patterns, necessitating robust segmentation under incomplete and shifting clinical conditions.

Accordingly, missing-modality segmentation methods can be categorized into three groups. *(i) Specialist distillation* trains a multimodal teacher and distills its knowledge into students tailored to specific missing-modality patterns [2, 7, 22, 23]. While effective, training and storage costs scale with the number of modality combinations. *(ii) Unified cross-modal fusion* targets a single deployment model by fusing only the observed modalities in latent space [11, 18]. Representative works include RFNet [6], mmFormer [25], and IM-Fuse [15], which leverage modality-aware fusion, token-based attention, and state-space-inspired token fusion, respectively. However, fusion-only approaches cannot recover discriminative cues from modalities that are entirely absent at inference, thereby limiting robustness under severe missingness. *(iii) Reconstruction-based strategies* explicitly infer missing content to restore absent cues. ShaSpec [21] and M^{3}FeCon [24] reconstruct missing features from observed modalities, M^3AE [10] learns representations via masked reconstruction prior to segmentation, and DC-Seg [9] disentangles anatomical and modality representations through bidirectional contrastive learning to promote modality-invariant anatomy. Despite improved robustness, these approaches introduce additional training objectives or multi-stage pipelines and remain susceptible to representation mismatch and error propagation.

A further practical gap concerns evaluation under clinically realistic shifts. Most methods are developed and benchmarked on pre-treatment glioma cohorts from earlier BraTS editions [1, 3, 5, 13], where labels emphasize classical tumor subregions (ET/ED/NCR/NET). In contrast, BraTS2024 [20] shifts to post-treatment MRI and adds an explicit resection-cavity (RC) label. Post-operative anatomy and therapy-related changes can mimic residual tumor, and the absence of informative modalities further amplifies this ambiguity, motivating approaches that remain robust in post-treatment settings [4].

To address these challenges, we present **ReMiX-Seg**, a unified single-network framework that couples lightweight latent completion with mask-aware expert routing for subset-adaptive fusion. A modality-specific hybrid encoder with a Shift-Gated Token Refiner stabilizes deep tokens under incomplete inputs; missing modalities are completed in the task-driven latent space by a Context-Gated Imputer, without the need for heavyweight image synthesis. Serialized Mamba–Mixture-of-Experts Fusion further adapts multi-scale fusion to each subset, while Expert-wise Consistency Alignment reduces the gap between reconstructed and observed features. To the best of our knowledge, ReMiX-Seg is among the first missing-modality segmentation frameworks evaluated systematically in both BraTS2023 [1] and BraTS2024 [20] settings[1]. ReMiX-Seg achieves state-of-the-art (SOTA) performance and demonstrates consistent robustness across all non-empty modality subsets in both pre- and post-treatment settings.

[1] At the time of writing, BraTS2024 was the latest post-treatment release and BraTS2023 the latest publicly available pre-treatment release.

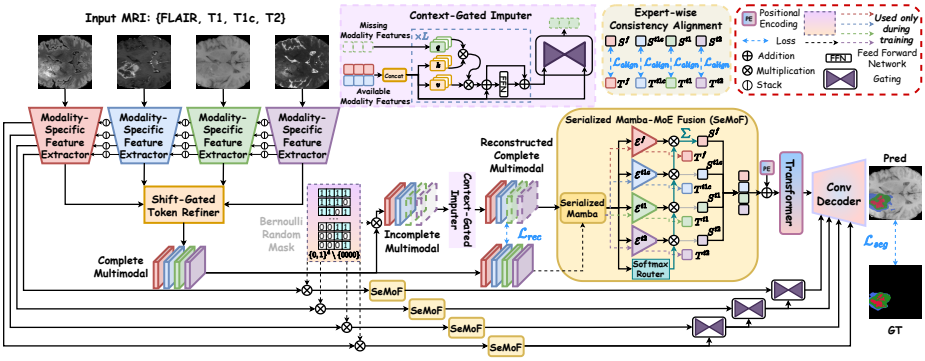

Fig. 1. ReMiX-Seg overview. Bernoulli masking yields an incomplete stream with teacher supervision; latent token completion and mask-aware SeMoF fusion enable robust segmentation in the presence of missing modalities.

2 Method

Overview. Fig. 1 illustrates **ReMiX-Seg**, a unified single-network framework for glioma segmentation under arbitrary missing MRI modalities. It unifies task-driven latent completion with mask-aware expert routing for subset-adaptive fusion. A complete-modality teacher stream stabilizes training, while inference relies only on observed modalities.

2.1 Preliminaries

We consider four MRI modalities, denoted by $\mathcal{M} = \{1, 2, 3, 4\}$, corresponding to {FLAIR, T1, T1c, T2}. Let $x_i \in \mathbb{R}^{H\times W\times D}$ denote the 3D volume of modality i and $y \in \{1, \ldots, K\}^{H\times W\times D}$ the voxel-wise label map. Missingness is encoded by $m \in \{0, 1\}^4 \setminus \{\mathbf{0}\}$ with observed set $\mathcal{O}(m) = \{i \in \mathcal{M} \mid m_i = 1\}$. The objective is to learn a segmentation model that produces robust predictions from $\{x_i\}_{i\in\mathcal{O}(m)}$ across all $2^4 - 1$ non-empty modality subsets. A modality-specific extractor produces tokens $z_i = E_i(x_i) \in \mathbb{R}^{T\times C}$ by patchifying deep feature maps. The desired model is *subset-invariant* in interface yet *subset-adaptive* in computation, motivating (i) latent completion and (ii) mask-aware refinement/fusion via conditional routing as in Mixture-of-Experts [17]. Formally,

$$\hat{\mathbf{z}} = G\big(\{z_i\}_{i\in\mathcal{O}(m)}, m\big), \qquad \tilde{\mathbf{z}} = H\big(\hat{\mathbf{z}}, m\big), \qquad \hat{y} = D(\tilde{\mathbf{z}}), \tag{1}$$

where $G(\cdot)$ completes missing information, $H(\cdot)$ performs subset-aware denoising/fusion to mitigate missing-pattern shift, and $D(\cdot)$ outputs voxel-wise predictions. This factorization separates completion and subset-aware refinement within a unified model, enabling end-to-end training, avoiding explicit image synthesis, and preserving computational efficiency.

2.2 ReMiX-Seg

Modality-Specific Feature Extractor with Shift-Gated Token Refiner. To preserve modality-specific contrast before cross-modal interaction, each modality is processed independently by a 3D ConvNeXt–ResUNet-style backbone [8,12], whose deep features are tokenized by a lightweight projection $\mathcal{P}_\omega^{(i)}$ and contextualized by a shallow Transformer [19]. Our proposed *Shift-Gated Token Refiner* (SGTR) then applies token shifting followed by gated residual refinement:

$$u_i = \mathrm{Shift}\Big(\mathrm{Trans}\big(\mathcal{P}_\omega^{(i)}(x_i)\big)\Big), \qquad z_i = u_i + g_i \odot R(u_i), \qquad g_i = \sigma\big(\Gamma(u_i)\big), \quad (2)$$

where $\mathrm{Shift}(\cdot)$ performs local token mixing, and the learnable gate g_i modulates the residual update $R(u_i)$ so that refinement remains stable under incomplete inputs while preserving reliable modality-specific structure. The refined tokens are then masked by m, and only $\{z_i\}_{i\in\mathcal{O}(m)}$ are forwarded to the completion and fusion stages.

Context-Gated Imputer (CGI, token-level completion). During training, each complete-modality sample is paired with a masked counterpart: a Bernoulli mask m defines an incomplete stream by retaining only $\{z_j\}_{j\in\mathcal{O}(m)}$, while the unmasked stream provides teacher tokens $\{z_i^T\}_{i\in\mathcal{M}}$. At the deepest level, our proposed CGI reconstructs missing-modality tokens directly in latent space from the observed context. For each missing modality $i \notin \mathcal{O}(m)$, let $\tilde{z}_i \in \mathbb{R}^{T\times C}$ be a placeholder query and $Z_\mathcal{O} = \mathrm{Concat}(\{z_j\}_{j\in\mathcal{O}(m)}) \in \mathbb{R}^{(|\mathcal{O}(m)|T)\times C}$ be the observed context. CGI refines $\tilde{z}_i$ through L stacked cross-attention blocks, then softly blends the refined query with a pooled summary of available evidence:

$$\hat{z}_i = \mathrm{CGI}\big(\tilde{z}_i, Z_\mathcal{O}\big) = \mathrm{Gate}\Big(\underbrace{\mathrm{CA}^{(L)} \circ \cdots \circ \mathrm{CA}^{(1)}}_{L \text{ cross-attention refinements}}(\tilde{z}_i, Z_\mathcal{O}),\ \mathrm{Pool}(Z_\mathcal{O})\Big), \quad (3)$$

with $\mathrm{CA}^{(\ell)}$ attending from the evolving missing tokens to $Z_\mathcal{O}$, while $\mathrm{Gate}(\cdot,\cdot)$ and $\mathrm{Pool}(\cdot)$ denote a learnable gating function and a simple context-pooling operator (e.g., mean pooling), respectively. This context-aware interpolation mitigates overconfident completion when only a limited set of modalities is available. The reconstruction loss is applied only to missing modalities:

$$\mathcal{L}_{\mathrm{rec}} = \sum_{i\notin\mathcal{O}(m)} \big\|\hat{z}_i - z_i^T\big\|_{\text{smooth-}\ell_1}. \quad (4)$$

SeMoF: Serialized Mamba–Mixture-of-Experts Fusion. SeMoF adapts fusion to the observed subset of modalities by coupling serialized state-space interactions with conditional expert refinement. Following [15], $\mathrm{Ser}(\cdot)$ interleaves modality-token sequences and a learnable fused stream across spatial locations, and Mamba processes the serialized spatial–modality sequence to obtain $\bar{z} = \mathrm{Mamba}(\mathrm{Ser}(\{z_i\}))$. A mask-dependent FiLM modulation [14], together with a MoE-FFN whose routing depends on the availability mask m and a stream indicator f (masked, reconstructed, or teacher during training), then refines $\bar{z}$:

$$\tilde{z} = \mathrm{SeMoF}(\{z_i\}, m, f) = \bar{z} + \sum_{e=1}^{E} w_e(\bar{z}, m, f)\, \mathcal{E}_e\big(\mathrm{FiLM}(\bar{z}, m)\big), \quad (5)$$

Here, $\mathcal{E}_e(\cdot)$ denotes the e-th expert feed-forward network, $w_e(\cdot)$ denotes its mixture weight, and E is the number of experts. To discourage routing collapse, the average routing mass is regularized toward a uniform distribution. For N routed tokens, let $\bar{w}_e = \frac{1}{N}\sum_{n=1}^{N} w_{n,e}$, where $w_{n,e}$ is the routing weight assigned to expert e for token n. The resulting load-balancing term is

$$\mathcal{L}_{\text{bal}} = E \sum_{e=1}^{E} \left(\bar{w}_e - \frac{1}{E} \right)^2 . \tag{6}$$

To narrow the gap between reconstructed and fully observed fusion, Expert-wise Consistency Alignment (ECA) matches expert output distributions between reconstructed-stream fusion $\tilde{z}^S$ and teacher-stream fusion $\tilde{z}^T$ using Jensen–Shannon divergence:

$$\mathcal{L}_{\text{eca}} = \frac{1}{E} \sum_{e=1}^{E} \text{JS}\Big(\text{softmax}\big(\mathcal{E}_e(\tilde{z}^S)\big) \parallel \text{softmax}\big(\mathcal{E}_e(\tilde{z}^T)\big) \Big), \tag{7}$$

In practice, SeMoF is instantiated at multiple scales and shared across masked, reconstructed, and teacher streams during training, while ECA regularizes reconstructed fusion toward its teacher counterpart at the expert level.

Gated skip fusion and convolutional decoding. A U-Net-style decoder [16] is driven by multi-scale SeMoF features. At each scale, SeMoF skip features (reshaped from fused tokens) are adaptively merged with upsampled decoder features through a lightweight sigmoid gate, followed by convolutional refinement and upsampling. The deepest masked and reconstructed tokens are merged in the same manner, and optional deep supervision is applied at intermediate scales.

Optimization and regularization. The overall objective is

$$\mathcal{L} = \mathcal{L}_{\text{seg}} + \lambda_{\text{rec}}\,\mathcal{L}_{\text{rec}} + \lambda_{\text{eca}}\,\mathcal{L}_{\text{eca}} + \lambda_{\text{bal}}\,\mathcal{L}_{\text{bal}}, \tag{8}$$

where $\mathcal{L}_{\text{seg}}$ supervises voxel-wise prediction, and $\mathcal{L}_{\text{rec}}$, $\mathcal{L}_{\text{eca}}$, and $\mathcal{L}_{\text{bal}}$ enforce completion fidelity, reconstructed–teacher consistency, and balanced expert utilization, respectively; λ_{rec}, λ_{eca}, and λ_{bal} weight the auxiliary terms.

3 Experiments and Results

Datasets and Implementation. Experiments use BraTS2023 (adult glioma pre-treatment) [1] and BraTS2024 (adult glioma post-treatment) [20]. BraTS2023 contains 1,251 multi-contrast scans (FLAIR, T1, T1c, T2) and reports tumor subregions WT, TC, and ET, while BraTS2024 contains 1,350 scans with the same modalities and additionally includes the RC label. Both datasets use a 70%/10%/20% train/val/test split with identical preprocessing.

Training/inference are patch-based with 128×128×128 crops and standard 3D augmentation (random flips/crops and mild intensity shifts). Models are optimized with RAdam (lr $= 3 \times 10^{-4}$) for 500 epochs on a single NVIDIA A100 (batch size = 2), using λ_{rec}=1.0, λ_{eca}=0.4, and λ_{bal}=0.1.

Table 1. Dice (%) on **BraTS2023** across all 15 non-empty modality subsets. **Bold**/underline denote the **best**/second-best results on BraTS2023.

	Model	F	T1	T1c	T2	F T1	F T1c	F T2	T1 T1c	T1 T2	T1c T2	F T1 T1c	F T1 T2	F T1c T2	T1 T1c T2	F T1 T1c T2	Avg.
Enhancing Tumor	mmFormer	58.0	54.6	82.7	57.9	62.4	83.7	63.2	83.4	60.9	83.2	84.0	64.7	83.6	83.4	83.7	72.6
	M^3FeCon	56.5	55.7	82.7	58.2	62.1	84.2	62.2	83.8	61.5	83.5	84.5	64.8	84.2	84.0	84.1	72.8
	ShaSpec	55.8	52.5	82.2	55.2	60.5	83.5	60.4	83.2	57.7	83.3	84.5	61.5	84.3	83.0	84.9	71.5
	M^3AE	55.9	55.9	81.1	58.4	60.6	84.4	60.5	83.6	59.0	84.3	85.1	61.0	85.1	82.6	84.6	72.1
	IM-Fuse	58.5	55.9	83.3	59.5	62.7	84.1	**64.3**	84.4	62.4	83.5	84.4	65.9	84.2	84.2	84.4	73.4
	DC-Seg	**58.7**	56.2	84.4	**59.8**	63.3	85.2	64.2	85.5	63.1	84.6	85.5	66.3	85.7	85.8	85.5	74.1
	Ours	57.4	**56.3**	**87.3**	58.2	**63.9**	**87.8**	**64.3**	**87.9**	**64.2**	**87.7**	**87.6**	**66.4**	**88.2**	**87.8**	**88.1**	**75.5**
Tumor Core	mmFormer	**76.8**	73.2	87.6	74.3	78.9	88.5	79.3	88.5	77.2	88.7	88.8	80.2	88.8	89.0	88.9	83.3
	M^3FeCon	73.9	72.1	87.5	73.7	78.1	88.6	78.4	88.8	77.1	88.9	89.1	79.8	89.0	89.1	89.4	82.9
	ShaSpec	73.3	71.6	86.7	72.9	76.9	88.0	77.1	88.2	76.4	88.2	88.7	79.0	88.7	88.7	89.1	82.2
	M^3AE	75.3	**75.5**	88.3	**77.7**	78.2	88.7	78.5	88.9	**78.6**	88.9	89.1	79.3	89.3	89.2	89.4	83.7
	IM-Fuse	**76.8**	73.6	88.6	76.2	79.2	89.8	**79.6**	89.5	78.1	90.0	89.9	**80.5**	90.0	89.9	90.1	84.1
	DC-Seg	75.8	73.8	88.5	75.3	79.1	89.6	79.2	89.2	77.8	89.8	89.7	80.2	89.7	89.6	89.7	83.8
	Ours	75.9	74.2	**89.6**	74.0	**79.6**	**90.5**	79.5	**90.4**	78.1	**90.5**	**91.0**	**80.5**	**90.7**	**91.0**	**91.2**	**84.5**
Whole Tumor	mmFormer	89.4	79.7	80.1	86.2	90.0	90.4	90.2	82.2	87.1	87.1	90.5	90.3	90.7	87.6	90.7	87.5
	M^3FeCon	90.0	79.5	81.0	86.5	90.5	90.8	91.0	82.5	87.7	88.0	90.9	90.7	91.2	88.2	91.5	88.0
	ShaSpec	89.4	77.2	76.3	85.0	90.1	90.4	91.8	79.8	86.0	86.1	90.7	90.7	91.0	86.7	91.1	86.8
	M^3AE	89.5	78.6	78.9	86.2	89.7	90.2	91.1	80.3	86.5	86.7	90.3	89.9	90.7	87.1	90.6	87.1
	IM-Fuse	91.0	80.4	82.0	87.5	91.7	91.8	91.8	83.6	88.7	89.0	92.0	92.0	92.3	89.3	92.3	89.0
	DC-Seg	90.9	80.5	81.9	87.7	91.6	91.7	91.7	83.4	88.5	88.8	92.0	91.9	91.0	89.1	91.9	88.8
	Ours	**91.1**	**82.9**	**83.8**	**88.3**	**92.3**	**92.8**	**92.6**	**85.7**	**90.1**	**90.3**	**92.9**	**92.9**	**93.3**	**90.8**	**93.3**	**90.2**

Comparisons with SOTA methods. ReMiX-Seg is compared with representative baselines spanning latent fusion and reconstruction/disentanglement, including mmFormer [25], M^3FeCon [24], ShaSpec [21], M^3AE [10], IM-Fuse [15], and DC-Seg [9]. All methods are retrained on BraTS2023 and BraTS2024 using identical splits and preprocessing, for 500 epochs, following each original paper's optimizer, scheduler, and core hyperparameters. Dice (%) over all 15 non-empty modality subsets is reported in Tables 1 and 2.

On BraTS2023, ReMiX-Seg achieves the best average Dice on ET/TC/WT (75.5/84.5/90.2) and ranks first or tied first in 38/45 configurations, surpassing IM-Fuse (73.4/84.1/89.0) and DC-Seg (74.1/83.8/88.8). On the more challenging BraTS2024, ReMiX-Seg remains robust despite post-treatment shift and the added RC label, ranking first or tied first in 55/60 configurations and improving the average Dice over the strongest baseline by 1.2–4.9 points (ET/TC/WT/RC: 62.2/58.3/82.0/68.2). Overall, these results show that coupling latent completion with subset-adaptive fusion improves robustness to missing modalities and remains effective in both pre- and post-treatment settings.

Qualitative visualization. Fig. 2 presents ReMiX-Seg predictions across all 15 non-empty modality subsets on representative BraTS2023 and BraTS2024 cases. Predictions remain visually consistent even when only one or two modalities

Table 2. Dice (%) on **BraTS2024** across all 15 non-empty modality subsets. **Bold**/underline denote the **best**/second-best results on BraTS2024.

	Model	F T1c T2	F T1c	F T1 T2	F T1c T2	F T1	F T2	F T1c	F T1	F T1c T2	F T1 T2	F T1c	F T2	F T1	F	Avg.
Enhancing Tumor	mmFormer	44.0	43.3	**72.0**	41.5	48.4	71.9	48.9	74.5	47.3	**73.1**	73.9	50.4	72.4	74.1	73.5	60.6
	M^3FeCon	42.8	44.1	69.9	42.5	47.6	70.1	47.5	71.6	47.7	69.8	70.7	50.1	70.2	71.1	73.4	59.3
	ShaSpec	40.8	40.2	69.6	38.8	45.1	70.4	44.7	71.6	43.4	69.1	70.6	46.1	70.2	70.7	70.2	57.4
	M^3AE	40.2	43.9	68.4	41.3	45.4	70.8	44.2	72.0	44.1	71.7	72.5	45.2	71.8	71.0	72.0	58.3
	IM-Fuse	43.3	45.3	70.5	43.8	48.3	70.6	48.0	72.1	48.2	70.3	71.2	50.6	70.7	72.2	71.3	59.8
	DC-Seg	44.1	44.3	71.6	44.1	48.9	71.7	49.4	73.2	48.9	71.4	72.3	51.5	71.8	72.2	72.1	60.5
	Ours	**44.9**	**46.7**	71.0	**48.2**	**50.0**	**73.3**	**50.5**	**74.8**	**50.1**	72.8	**75.9**	**52.3**	**73.0**	**74.6**	**74.6**	**62.2**
Tumor Core	mmFormer	41.5	38.9	61.6	37.9	46.7	60.0	46.6	63.3	41.0	60.1	64.6	48.1	63.1	62.5	65.0	53.4
	M^3FeCon	38.8	39.5	61.7	37.5	46.3	60.0	45.7	63.5	40.8	60.2	64.7	47.6	63.3	62.8	64.5	53.1
	ShaSpec	40.7	38.1	60.8	37.1	45.9	59.2	45.8	62.5	40.2	59.3	63.8	47.3	62.3	61.7	64.2	52.6
	M^3AE	38.2	44.1	59.9	42.2	44.6	59.4	44.5	61.6	44.6	60.2	62.8	46.0	61.9	62.2	62.8	53.0
	IM-Fuse	36.9	43.7	61.0	35.8	44.3	62.4	42.8	**64.5**	43.0	62.7	64.2	45.2	64.1	64.3	63.1	53.2
	DC-Seg	37.2	43.6	61.1	36.4	45.2	61.8	43.1	64.2	42.6	62.5	64.0	46.3	63.8	64.1	61.2	53.1
	Ours	**44.5**	**48.7**	**62.9**	**49.6**	**51.3**	**63.6**	**51.9**	63.8	**52.4**	**65.3**	**66.9**	**53.0**	**65.3**	**67.1**	**67.3**	**58.3**
Whole Tumor	mmFormer	84.2	70.1	69.8	75.5	85.2	85.6	85.8	72.9	78.0	78.3	85.9	85.8	86.5	78.9	86.5	80.6
	M^3FeCon	81.8	71.0	71.6	75.0	83.3	83.7	83.8	73.2	77.2	77.6	83.7	83.5	84.3	78.5	85.3	79.5
	ShaSpec	81.5	68.8	69.2	74.2	82.5	83.0	83.5	71.5	75.4	76.2	82.8	83.0	84.1	77.1	84.0	78.2
	M^3AE	80.9	70.2	68.7	74.8	82.1	83.2	83.1	72.0	76.1	77.0	83.0	82.8	83.5	77.8	83.9	78.6
	IM-Fuse	83.1	71.5	**72.0**	76.3	84.2	84.5	84.3	74.6	78.7	78.9	84.6	84.6	85.0	79.5	85.0	80.5
	DC-Seg	82.9	71.2	71.8	75.8	83.9	84.2	84.0	74.1	78.2	78.5	84.1	84.2	84.8	79.1	84.8	80.1
	Ours	**85.3**	**71.9**	**72.0**	**77.0**	**86.1**	**86.6**	**86.7**	**75.1**	**79.5**	**80.4**	**86.8**	**86.8**	**87.4**	**80.8**	**87.3**	**82.0**
Resection Cavity	mmFormer	54.7	60.5	65.0	59.9	66.0	69.6	65.4	67.4	65.4	67.7	71.2	68.8	70.2	69.1	72.1	66.2
	M^3FeCon	53.8	59.2	64.1	58.5	64.2	68.3	64.1	66.2	63.8	66.0	69.5	67.1	68.5	67.8	70.2	65.1
	ShaSpec	52.5	57.1	63.5	57.2	63.0	67.4	63.5	65.1	62.4	65.2	68.8	65.9	67.4	66.5	69.4	64.0
	M^3AE	53.1	58.6	62.8	58.0	63.5	68.1	63.8	65.8	63.0	65.9	69.1	66.5	68.1	67.2	69.8	64.6
	IM-Fuse	**56.8**	**63.4**	66.6	**61.1**	65.7	69.9	67.7	69.1	65.9	66.8	71.3	68.6	71.4	69.5	71.2	67.0
	DC-Seg	56.5	62.8	66.0	60.8	65.9	70.1	67.2	68.8	65.5	67.3	71.0	68.2	71.0	69.8	71.1	66.8
	Ours	54.9	63.2	**67.1**	**61.1**	**68.0**	**71.3**	**68.0**	**69.8**	**68.2**	**69.9**	**72.3**	**69.6**	**73.1**	**72.3**	**73.7**	**68.2**

are available, indicating effective compensation under severe modality dropout. In contrast, omitting T1c notably degrades enhancement-related delineation, consistent with the quantitative findings and highlighting the critical role of contrast for ET (and cavity–tumor ambiguity in the post-treatment setting).

Ablation studies. Two ablations evaluate (i) the effect of auxiliary regularizers and (ii) the completion capacity of the CGI. Table 3 shows that each term improves robustness, with the completion loss $\mathcal{L}_{\text{rec}}$ yielding the largest gain; adding $\mathcal{L}_{\text{eca}}$ further boosts performance, consistent with reduced reconstructed–observed mismatch after fusion, and the full objective performs best overall. Table 4 varies the number of stacked cross-attention layers L in CGI and assesses completion quality via ℓ_1 distance, cosine similarity, and maximum mean discrepancy, together with downstream average Dice. A shallow stack (L=2) offers the best trade-off

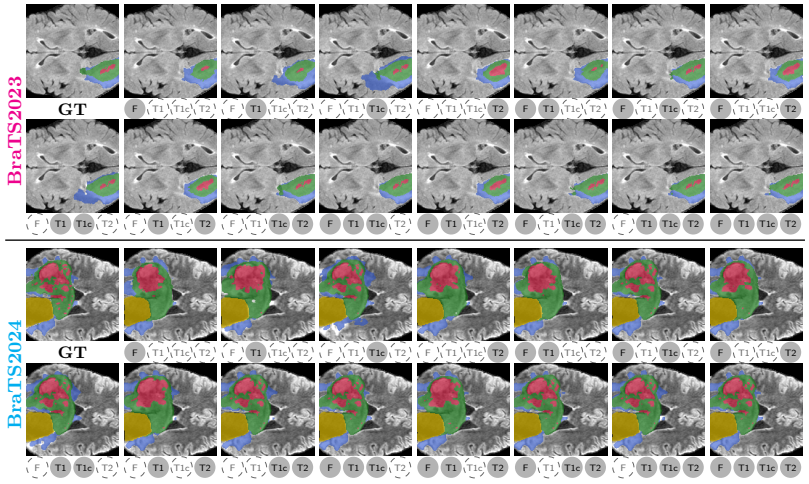

Fig. 2. Qualitative results across modality subsets. Ground truth and ReMiX-Seg predictions for all non-empty modality combinations on BraTS2023/2024. Red: necrotic core, blue: edema, green: enhancing tumor, yellow: resection cavity.

Table 3. Ablation of auxiliary losses. Dice (%) for ET/TC/WT (and RC) on BraTS2023/BraTS2024.

$\mathcal{L}_{\text{rec}}$	$\mathcal{L}_{\text{eca}}$	$\mathcal{L}_{\text{bal}}$	ET	TC	WT	RC
		✓	72.9/60.8	83.4/53.9	87.8/80.8	66.4
	✓		73.2/61.0	83.6/54.7	88.3/80.9	66.7
	✓	✓	73.4/61.2	83.7/55.0	88.4/81.0	66.9
✓			74.2/61.1	83.7/55.6	88.9/81.1	67.0
✓		✓	74.6/61.3	83.9/56.2	89.2/81.3	67.2
✓	✓		75.0/61.8	84.2/57.5	89.6/81.5	67.8
✓	✓	✓	**75.5/62.2**	**84.5/58.3**	**90.2/82.0**	**68.2**

Table 4. Effect of L Cross-Attention layers in CGI on token completion.

	$\#L$	L1 (↓)	COS (↑)	MMD (↓)	Avg. Dice(↑)
BraTS2023	1	0.009	0.984	0.054	82.915
	2	**0.003**	**0.998**	**0.026**	**83.398**
	3	**0.003**	0.996	0.029	83.142
	4	0.007	0.986	0.051	82.814
BraTS2024	1	0.018	0.972	0.098	66.754
	2	**0.007**	**0.991**	**0.061**	**67.642**
	3	0.008	0.989	0.066	67.289
	4	0.015	0.978	0.092	66.568

across both datasets, while deeper stacks show diminishing returns and can slightly degrade similarity and segmentation.

4 Conclusion

This work introduces **ReMiX-Seg**, a single-network framework for missing-modality robust multimodal glioma segmentation. ReMiX-Seg couples lightweight latent completion with mask-aware Mixture-of-Experts routing, enabling subset-adaptive fusion under arbitrary missing MRI modalities. The design avoids heavyweight image synthesis while explicitly mitigating reconstructed–observed mismatch during fusion. Extensive experiments on BraTS2023 and BraTS2024 show state-of-the-art Dice and stable performance across all non-empty modality subsets. These results highlight strong robustness to modality dropout across pre- and post-treatment domain-shifted settings.

Acknowledgments. This work was supported by the Center for Advanced Computing and Imaging in Biomedicine under the Higher Education Sprout Project of the Ministry of Education (MOE, NTU-115L900701), the National Science and Technology Council (NSTC, 114-2222-E-002-011-MY3 and 114-2634-F-002-008), National Taiwan University (NTU-AS-115L104302 and NTU-CC-115L894901), and the MOE (113M7054) in Taiwan. The authors sincerely thank Nam Anh Ta, David Ulrich Von Nobbe, Cedric Gioanni, and Samuel Lin—Quang's teammates in the Fall 2025 Artificial Intelligence and Intelligent Medicine course taught by Prof. Che Lin at National Taiwan University—for valuable course discussions that contributed to the literature review.

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

References

1. Adewole, M., et al.: The brain tumor segmentation (brats) challenge 2023: Glioma segmentation in sub-saharan africa patient population (brats-africa) (2023)
2. Azad, R., et al.: Smu-net: Style matching u-net for brain tumor segmentation with missing modalities. In: International Conference on Medical Imaging with Deep Learning (2022)
3. Baid, U., et al.: The rsna-asnr-miccai brats 2021 benchmark on brain tumor segmentation and radiogenomic classification (2021)
4. Baig, S., et al.: Segmentation of pre- and posttreatment diffuse glioma tissue subregions including resection cavities. Neuro-Oncology Advances **6**(1), vdae140 (2024)
5. Bakas, S., et al.: Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the brats challenge (2018)
6. Ding, Y., et al.: Rfnet: Region-aware fusion network for incomplete multi-modal brain tumor segmentation. In: 2021 IEEE/CVF International Conference on Computer Vision (ICCV). pp. 3955–3964 (2021)
7. Hu, M., et al.: Knowledge distillation from multi-modal to mono-modal segmentation networks. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2020. pp. 772–781. Springer International Publishing, Cham (2020)
8. Jha, D., et al.: Resunet++: An advanced architecture for medical image segmentation. In: 2019 IEEE International Symposium on Multimedia (ISM). pp. 225–2255 (2019)
9. Li, H., et al.: Dc-seg: Disentangled contrastive learning for brain tumor segmentation with missing modalities. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2025. pp. 138–148. Springer Nature Switzerland, Cham (2026)
10. Liu, H., et al.: M3ae: multimodal representation learning for brain tumor segmentation with missing modalities. In: Proceedings of the Thirty-Seventh AAAI Conference on Artificial Intelligence and Thirty-Fifth Conference on Innovative Applications of Artificial Intelligence and Thirteenth Symposium on Educational Advances in Artificial Intelligence. AAAI'23/IAAI'23/EAAI'23, AAAI Press (2023)
11. Liu, Z., et al.: Sfusion: Self-attention based n-to-one multimodal fusion block. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2023. pp. 159–169. Springer Nature Switzerland, Cham (2023)

12. Liu, Z., et al.: A convnet for the 2020s. In: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 11966–11976 (2022)
13. Menze, et al.: The multimodal brain tumor image segmentation benchmark (brats). IEEE Transactions on Medical Imaging **34**(10), 1993–2024 (2015)
14. Perez, E., et al.: Film: visual reasoning with a general conditioning layer. In: Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence and Thirtieth Innovative Applications of Artificial Intelligence Conference and Eighth AAAI Symposium on Educational Advances in Artificial Intelligence. AAAI'18/IAAI'18/EAAI'18, AAAI Press (2018)
15. Pipoli, V., et al.: Im-fuse: A mamba-based fusion block for brain tumor segmentation with incomplete modalities. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2025. pp. 225–235. Springer Nature Switzerland, Cham (2026)
16. Ronneberger, O., others.: U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. pp. 234–241. Springer International Publishing, Cham (2015)
17. Shazeer, N., et al.: Outrageously large neural networks: The sparsely-gated mixture-of-experts layer (2017)
18. Shi, J., et al.: Mftrans: Modality-masked fusion transformer for incomplete multi-modality brain tumor segmentation. IEEE Journal of Biomedical and Health Informatics **28**(1), 379–390 (2024)
19. Vaswani, A., et al.: Attention is all you need. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. p. 6000–6010. NIPS'17, Curran Associates Inc., Red Hook, NY, USA (2017)
20. de Verdier, M.C., et al.: The 2024 brain tumor segmentation (brats) challenge: Glioma segmentation on post-treatment mri (2024)
21. Wang, H., et al.: Multi-modal learning with missing modality via shared-specific feature modelling. In: 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 15878–15887 (2023)
22. Wang, S., et al.: 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 (2023)
23. Wang, Y., et al.: Acn: Adversarial co-training network for brain tumor segmentation with missing modalities. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2021. pp. 410–420. Springer International Publishing, Cham (2021)
24. Zeng, Z., et al.: Missing as masking: Arbitrary cross-modal feature reconstruction for incomplete multimodal brain tumor segmentation. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2024. pp. 424–433. Springer Nature Switzerland, Cham (2024)
25. Zhang, Y., et al.: mmformer: Multimodal medical transformer for incomplete multimodal learning of brain tumor segmentation. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2022. pp. 107–117. Springer Nature Switzerland, Cham (2022)