HGMAMBA: A HETEROGENEOUS GRAPH MAMBA FRAMEWORK FOR INTERPRETABLE MISSING MODALITY BRAIN TUMOR SEGMENTATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Multimodal MRI provides complementary information for brain tumor segmentation. In clinical practice, however, one or more modalities are often unavailable, which severely degrades models trained with complete modalities. Existing methods address this problem in limited ways. Some substitute missing modalities with zero-filled inputs, leading to biased feature representations. Others adopt multi-stage optimization with additional pretraining, increasing computational cost. Moreover, most methods overlook modality-specific characteristics and treat all modalities as equally informative for different tumor regions. To address these limitations, we propose HGMamba, an interpretable heterogeneous graph framework with Mamba-based interaction for missing-modality brain tumor segmentation. HGMamba models imaging modalities and hierarchical tumor regions, including the whole tumor, tumor core, and enhancing tumor, as heterogeneous graph nodes to capture modality-region dependencies. A data-driven affinity prior is introduced to quantify modality-region discriminability without additional learnable parameters and guide modality-to-region aggregation. A lightweight Mamba-based operator further enables efficient graph interaction, while a reliability gate and nested region interaction mechanism reduce the interference from missing modalities. Extensive experiments on BraTS 2018 and BraTS 2020 demonstrate that HGMamba achieves comparable Dice performance to the strongest competitor while substantially improving HD95. Furthermore, it achieves these improvements with an order of magnitude fewer parameters and over $2\times$ faster inference.

1 INTRODUCTION

Multimodal magnetic resonance imaging (MRI) is the clinical standard for diagnosing and assessing brain tumors (Abidin et al., 2024). Four modalities are routinely acquired: T1-weighted (T1), contrast-enhanced T1-weighted (T1ce), T2-weighted (T2), and Fluid-Attenuated Inversion Recovery (FLAIR). Each highlights different tissue properties, and their complementarity supports accurate delineation of the hierarchical tumor sub-regions, namely the whole tumor (WT), tumor core (TC), and enhancing tumor (ET). Deep-learning models trained on the full modality set have reached strong segmentation accuracy (Wang et al., 2023c). In practice, however, a complete modality set is rarely available. Modalities are often missing due to image corruption, varied acquisition protocols, contrast-agent allergies, or scan artifacts (Zhou et al., 2023), and full-modality models then degrade sharply. This makes robustness to missing modalities a central obstacle to clinical use (Zeng et al., 2024).

Existing approaches fall into several families. One line trains a dedicated network for every modality subset (Havaei et al., 2016; Chen et al., 2019; Shi et al., 2024), whose cost grows exponentially with the number of modalities. Another line synthesizes the absent modalities (Zhang et al., 2025a; Jia et al., 2026), which is limited by generation quality and often adds noise. A single model that fills missing inputs with zero maps (Feng et al., 2023) avoids retraining but biases the features, and recent multi-stage pipelines distill from a full-modality teacher or rely on generative pretrain-

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

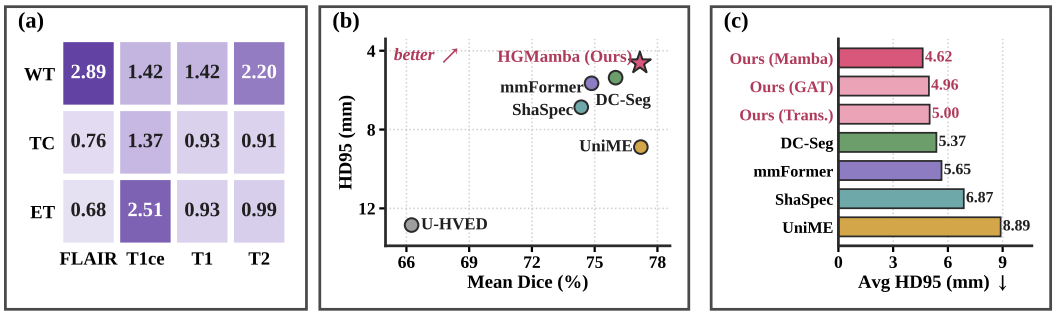

Figure 1: (a) Data-driven affinity prior (Cohen's d): each modality separates each tumor region to a different degree. (b) Mean Dice and HD95 on BraTS 2018 (HD95 axis inverted). (c) Average HD95 on BraTS 2018 for the baselines and for HGMamba with its NodeMamba, GAT, and Transformer operators; all three instantiations have a lower HD95 than every baseline.

ing (Shi et al., 2024; Zhu et al., 2026; Pipoli et al., 2025), which is complex and heavy. Despite their differences, these methods share one blind spot: they treat every modality as equally informative for every region (Abidin et al., 2024). This contradicts radiological knowledge that each modality is most discriminative for specific sub-regions; T1ce sharply reveals the enhancing tumor (Lerner et al., 2024), while FLAIR best outlines the whole tumor (Figure 1a). Graph- and attention-based fusion (Peng et al., 2024; Wang et al., 2024; D'Souza et al., 2024; Karimijafarbigloo et al., 2024; Qiu et al., 2025; Kang et al., 2026) does not close this gap either. These methods relate modalities to one another but never model how each modality relates to each tumor region, and their interactions become ill-defined when a modality is absent.

We therefore make modality-region relations explicit. HGMamba represents MRI modalities and the nested tumor regions as two node types, so a missing modality simply becomes an inactive node and the model runs on any subset without zero-filling or pretraining. Rather than asking attention to learn from scratch which modality serves which region, a data-driven affinity prior, estimated once from training data without learnable parameters, biases the modality-to-region routing, and a reliability gate adapts this fixed preference to the modalities actually present. The region nodes then exchange information along WT$\supset$TC$\supset$ET. Node interaction uses a lightweight state-space operator, NodeMamba, though graph attention or Transformer operators also work. On BraTS 2018 and 2020, HGMamba matches the strongest existing method in Dice while producing clearly more accurate boundaries in HD95 (Figure 1b,c), with an order of magnitude fewer parameters and over $2\times$ faster inference.

Our main contributions are summarized as follows:

- We propose a heterogeneous modality-region fusion framework for missing-modality brain tumor segmentation, which represents MRI modalities and nested tumor regions as two node types and routes information between them under any modality subset.
- We introduce a data-driven affinity prior that estimates the discriminative relevance of each modality to each region without learnable parameters, guiding the routing and making the fusion interpretable.
- We instantiate node interaction with NodeMamba; capacity-matched GAT and Transformer variants show that the framework is not tied to a specific operator.
- Experiments on BraTS 2018 and 2020 over all 15 modality combinations, plus zero-shot transfer to unseen scanners and to brain metastases, show Dice on par with the strongest competitor, clearly better HD95, and an order of magnitude fewer parameters.

2 Related Work

2.1 Brain Tumor Segmentation with Missing Modalities

Multimodal MRI segmentation performs well with all modalities (Isensee et al., 2024), but clinical scans are often incomplete (Chen et al., 2022) due to corruption, contraindications, or cost. Existing methods fall into three groups. Synthesis-based methods generate the missing modalities before segmentation (Zhang et al., 2025a; Jia et al., 2026), which is unreliable when the missing modality is diagnostically critical. Knowledge-distillation methods train a student under a full-modality teacher (Kang et al., 2024; Zhu et al., 2025; Wang et al., 2023b), but still need complete data and generalize poorly when key modalities are absent. Shared-latent-space methods learn modality-invariant representations, including U-HVED (Dorent et al., 2019), mmFormer (Zhang et al., 2022), ShaSpec (Wang et al., 2023a), and M3FeCon (Zeng et al., 2024), yet tend to suppress the modality-specific cues that matter most when a key modality is missing. Closest to us, DC-Seg (Li et al., 2026) uses region-aware fusion for the differing sensitivity of modalities to tumor regions, and UniME (Song et al., 2026) decouples representation learning from segmentation with two-stage masked-image-modeling pretraining.

2.2 Attention and Graph Models for Medical Image Segmentation

Attention and graph models are two common ways to capture the relations among features in medical image segmentation (Mienye & Viriri, 2025). Graph neural networks build graphs over image regions and pass messages between nodes (Singh et al., 2026). Early methods construct graphs over supervoxels or superpixels and classify each node. Later works apply graph-attention fusion or modality-adaptive graphs to relate different modalities (Bae et al., 2024). Transformers take another route. They use self-attention to relate spatial or modality tokens, and are often placed between a convolutional encoder and decoder to model long-range context. Both families improve multimodal fusion. Yet they are usually attached to a conventional backbone, and they rarely model the hierarchy among tumor sub-regions or the region-dependent contribution of each modality.

2.3 State Space Models and Mamba

State space models have recently become efficient alternatives to Transformers for long-sequence modeling, and Mamba (Gu & Dao, 2023) is a leading example. Through a selective scan, it captures long-range dependencies with linear complexity. This is attractive for volumetric imaging, where self-attention is costly. Several works bring Mamba into missing-modality brain tumor segmentation. LS3M (Zhang et al., 2025b) combines learnable modality sorting with state-space modeling to reorder modality sequences. M2SegMamba (Zhang et al., 2026) adds semantic-level masking to a multi-encoder Mamba architecture. These methods use Mamba mainly as a sequence backbone over spatial or modality tokens. They do not use it to model the structured relations among modalities and tumor regions. In contrast, we use Mamba as the interaction operator inside a heterogeneous graph over modality and region nodes.

3 Method

3.1 Overview

Let $\mathcal{M} = \{\mathrm{FLAIR}, \mathrm{T1ce}, \mathrm{T1}, \mathrm{T2}\}$ with $M{=}4$, and let $\boldsymbol{a} \in \{0,1\}^M$ indicate which modalities are available. We segment three nested regions, whole tumor (WT), tumor core (TC), and enhancing tumor (ET), where $\mathrm{ET} \subset \mathrm{TC} \subset \mathrm{WT}$. Figure 2(a) shows the pipeline. Each available modality is encoded by its own 3D U-Net into a bottleneck feature $\boldsymbol{f}_m \in \mathbb{R}^{C\times d\times h\times w}$ and skip features. We pool the bottlenecks into modality nodes and fuse them with learnable region nodes in our Heterogeneous Graph Fusion (HGF) module, guided by a data-driven affinity prior. The fused feature and the aggregated skips are then decoded into the final segmentation. Skip features at each level are combined by a masked mean over the available modalities,

$$\bar{\boldsymbol{s}}^{(l)} = \frac{\sum_m a_m \, \boldsymbol{s}_m^{(l)}}{\sum_m a_m}. \tag{1}$$

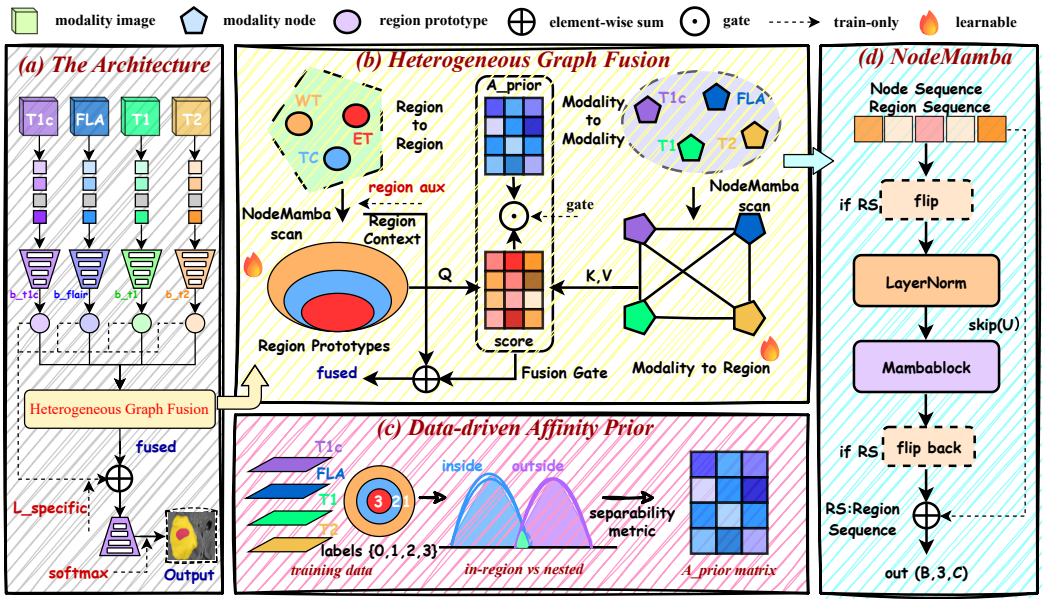

Figure 2: Overview of HGMamba. Imaging modalities and the nested WT/TC/ET regions are modeled as two node types and fused by a shared NodeMamba operator under a training-free, Cohen's d–based affinity prior; a collapse-safe fusion gate then re-weights the bottleneck features for decoding.

3.2 Heterogeneous Graph Fusion

Nodes We cast fusion as a graph problem. A missing modality then becomes a simple change of which nodes are active, not a change of the network structure. Each bottleneck is pooled by global average pooling into a *modality node*, giving a node sequence $\boldsymbol{Z} = [\boldsymbol{z}_1; \ldots; \boldsymbol{z}_M] \in \mathbb{R}^{M\times C}$. The three regions are represented by learnable *region nodes* $\boldsymbol{P} \in \mathbb{R}^{R\times C}$ with R=3. These nodes act as region-level queries and carry region knowledge shared across cases. Both node types are updated by a shared operator, NodeMamba. Crucially, the region nodes are free parameters optimized end-to-end, not embeddings of any ground-truth mask. The forward pass consumes only the images and the availability mask $\boldsymbol{a}$, so no annotation is used at inference; region labels enter solely through the auxiliary training losses of Section 3.4, exactly as deep supervision does in standard segmentation networks.

NodeMamba NodeMamba treats a short node sequence $\boldsymbol{U}$ as a 1D sequence. It updates the sequence with a residual selective state-space block:

$$\mathrm{NodeMamba}(\boldsymbol{U}) = \boldsymbol{U} + \mathrm{Mamba}(\mathrm{LN}(\boldsymbol{U})). \tag{2}$$

The number of nodes is small, so this scan is cheap. It still lets every node exchange information with the others and needs no fixed connectivity; other operators can be used in its place (Section 4.4).

Modality self-interaction The modality nodes first exchange information to become context-aware,

$$\tilde{\boldsymbol{Z}} = \mathrm{NodeMamba}(\boldsymbol{Z}) \in \mathbb{R}^{M\times C}. \tag{3}$$

Affinity-modulated routing This is the core step. Each region node attends to the modality nodes and draws mainly from the ones that best describe it. The region nodes act as queries, and the modality nodes act as keys and values,

$$\boldsymbol{Q} = \boldsymbol{P}\boldsymbol{W}_Q, \quad \boldsymbol{K} = \tilde{\boldsymbol{Z}}\boldsymbol{W}_K, \quad \boldsymbol{V} = \tilde{\boldsymbol{Z}}\boldsymbol{W}_V. \tag{4}$$

The learned region knowledge therefore enters through the queries. We then add the affinity prior $\boldsymbol{A}_{\mathrm{prior}} \in \mathbb{R}^{R\times M}$ to the attention score. Columns of absent modalities are masked, and a row-wise

softmax gives the routing weights,

$$\mathbf{\Pi} = \mathrm{softmax}\Big(\frac{\boldsymbol{Q}\boldsymbol{K}^\top}{\sqrt{C}} + \gamma\, \boldsymbol{A}_{\mathrm{prior}}\Big) \in \mathbb{R}^{R\times M}. \tag{5}$$

A reliability gate then rescales each modality by a scalar. It is a single linear projection followed by a sigmoid,

$$\boldsymbol{g} = \sigma(\boldsymbol{V}\boldsymbol{w}_g) \in \mathbb{R}^M, \qquad \boldsymbol{R}_1 = (\mathbf{\Pi} \odot \boldsymbol{g}^\top)\, \boldsymbol{V} \in \mathbb{R}^{R\times C}. \tag{6}$$

The rows are not renormalized, so the gate acts as an amplitude modulation. The prior gives a stable preference for each region, and the gate adjusts it to the current input.

Nested region interaction So far the region nodes are treated separately, but the three regions are not independent. They follow a strict containment $\mathrm{ET} \subset \mathrm{TC} \subset \mathrm{WT}$. A confident enhancing-tumor response is therefore good evidence that the same location also belongs to the tumor core and the whole tumor. We use this fact by letting the region nodes exchange information along the hierarchy. The region sequence is flipped to the order ET, TC, WT, processed by NodeMamba, and flipped back, so the finest region informs the coarser ones in one pass:

$$\boldsymbol{R}_2 = \mathrm{Flip}^{-1}\big(\mathrm{NodeMamba}(\mathrm{Flip}(\boldsymbol{R}_1))\big) \in \mathbb{R}^{R\times C}. \tag{7}$$

This encourages the three predictions to stay anatomically consistent. It also helps the smaller regions, which are usually the hardest to segment. Because the regions are nested, we always proceed from the smallest to the largest, so the reverse order (WT to ET) has no anatomical justification and is not used.

Collapse-safe fusion A graph built from a few pooled nodes can sometimes learn a weak pattern, which would hurt the final prediction. We avoid this by not rebuilding the segmentation feature from the nodes. Instead, the main feature is a re-weighting of the original bottlenecks. The final prediction thus always has a strong convolutional feature to rely on, and the graph only adds to it. A fusion gate gives a weight per modality from the context-aware modality nodes, with absent modalities masked before the softmax,

$$\boldsymbol{w} = \mathrm{softmax}(\mathrm{MLP}(\tilde{\boldsymbol{Z}})) \in \mathbb{R}^M. \tag{8}$$

The fused bottleneck is the weighted sum of the original modality features,

$$\boldsymbol{F} = \sum_{m=1}^{M} w_m\, \boldsymbol{f}_m. \tag{9}$$

Finally, the refined region nodes are projected into a context term and added to $\boldsymbol{F}$. This injects region-level semantics into the fused feature. The result is decoded together with the aggregated skips into the final segmentation. In this way the graph only reweights and enriches the bottleneck features. It can help the segmentation, but it never replaces the main path.

3.3 Data-Driven Affinity Prior

The affinity prior $\boldsymbol{A}_{\mathrm{prior}}$ scores how discriminative each modality is for each region. It is estimated once from the training set, using only present-modality cases when the training set is itself incomplete, has no learnable parameters, and so adds prior knowledge at almost no cost and cannot overfit (Fig. 2(c)). For region r and modality m we contrast the intensity of m inside r with its nested complement, ET against the rest of TC, TC against the rest of WT, and WT against the rest of the brain, measured by Cohen's d:

$$d_{r,m} = \frac{|\mu^+_{r,m} - \mu^-_{r,m}|}{\sqrt{((\sigma^+_{r,m})^2 + (\sigma^-_{r,m})^2)/2}}, \tag{10}$$

averaged over training cases, where $\mu^\pm, \sigma^\pm$ are the mean and standard deviation of the in-region and complement voxels. A larger $d_{r,m}$ means cleaner separation, and the resulting matrix matches clinical knowledge: T1ce scores highest for ET and FLAIR for WT. We inject this matrix directly into the attention score of equation 5. To avoid a common misreading, we distinguish three things: the prior $\boldsymbol{A}_{\mathrm{prior}}$ is estimated from data by equation 10 and is training-free and fixed; the routing

Table 1: Dice (%) on BraTS 2018 across all 15 modality combinations. ● and ○ denote available and missing modalities. Red: best; blue: second best; our method is highlighted in green.

Type	Flair	●	○	○	○	●	●	●	○	○	○	●	●	●	○	●	Avg
	T1	○	○	●	○	○	●	○	●	○	●	●	○	●	●	●	
	T1ce	○	●	○	○	●	○	○	●	●	○	●	●	○	●	●	
	T2	○	○	○	●	○	○	●	○	●	●	○	●	●	●	●	
WT	U-HVED	83.98	65.37	57.92	69.47	83.88	84.73	83.27	70.98	78.18	77.38	83.55	85.21	83.77	82.04	85.18	78.33
	mmFormer	88.17	72.34	74.27	83.28	89.02	89.13	89.20	76.94	85.77	85.27	89.07	89.77	89.61	85.84	89.79	85.16
	ShaSpec	**88.78**	73.70	73.14	82.28	89.74	89.01	89.88	77.87	86.79	86.69	89.49	90.65	89.87	87.53	90.49	85.73
	DC-Seg	88.48	79.34	77.70	85.00	89.78	89.21	89.79	82.43	87.02	86.92	89.80	90.33	90.02	87.47	90.32	86.91
	UniME	88.75	77.07	78.00	84.75	90.02	**90.02**	90.06	81.81	87.39	**87.49**	90.37	90.71	**90.34**	**88.44**	**90.93**	87.08
	Ours	88.63	**80.31**	**79.74**	**85.86**	**90.62**	89.75	**90.21**	**83.56**	**88.39**	87.34	**90.49**	**91.14**	90.19	88.29	90.81	**87.69**
TC	U-HVED	58.46	77.32	46.42	31.14	81.15	66.51	58.52	78.78	76.84	55.83	81.42	77.87	66.26	78.50	79.03	67.60
	mmFormer	71.95	82.80	69.33	70.57	84.07	76.07	74.74	85.03	83.72	74.07	85.11	84.12	76.88	84.82	84.89	79.21
	ShaSpec	70.24	81.59	58.52	66.02	84.21	76.43	73.03	84.02	83.32	73.91	85.15	84.28	77.67	84.56	85.03	77.87
	DC-Seg	72.12	83.03	**73.51**	73.66	83.86	77.76	75.07	**85.72**	**85.98**	76.95	85.90	85.66	77.44	**86.39**	86.39	80.63
	UniME	**75.53**	83.09	73.08	73.76	**85.79**	**79.68**	77.72	85.27	85.22	**77.73**	**86.70**	**86.00**	**79.81**	86.24	**86.90**	**81.50**
	Ours	75.05	**83.85**	70.71	**74.95**	85.73	77.93	**78.90**	85.02	85.46	77.09	85.55	85.77	79.39	85.74	85.45	81.11
ET	U-HVED	32.84	68.70	25.18	20.72	71.90	39.82	31.06	70.00	71.45	33.44	72.31	72.05	38.44	72.00	72.27	52.81
	mmFormer	41.55	73.01	40.22	42.08	73.65	46.31	47.16	73.76	73.34	47.02	74.08	73.72	49.50	73.47	73.84	60.18
	ShaSpec	38.45	72.88	30.66	39.85	74.18	44.96	47.46	73.86	74.17	45.94	74.93	74.70	49.90	74.78	75.27	59.47
	DC-Seg	38.76	74.79	38.38	44.35	75.19	42.47	47.15	75.38	75.55	46.00	75.41	75.34	47.01	75.69	75.36	60.46
	UniME	43.46	75.15	**41.39**	**47.60**	**75.73**	**48.90**	**51.50**	75.68	75.84	**50.55**	75.74	**75.85**	**56.39**	75.95	75.86	**63.04**
	Ours	**44.94**	**75.35**	34.89	45.68	75.50	45.28	49.42	**78.35**	**78.89**	50.04	**78.29**	75.31	52.05	**78.24**	**78.01**	62.68

attention over modalities is learned; and the clinical assignment (FLAIR for WT, T1ce for ET) is neither hand-set nor a separate learned map, but emerges as the empirical $\arg\max$ of the estimated prior. Because volumes are intensity-normalized per case, the prior depends on normalized rather than raw intensities and stays stable under the shifts across scanners and protocols.

Appendix B justifies both choices: under a Gaussian model $d_{r,m}$ is a strictly monotone, parameter-free proxy for the mutual information $I(x_m; y_r)$, so $\boldsymbol{A}_{\text{prior}}$ ranks modalities by their information about a region (Prop. 1); and adding it to the logits gives the product-of-experts posterior $\Pi_r(m) \propto p_\theta(m \mid \cdot)\, \pi(m \mid r)$, the Bayes-consistent combination that multiplicative injection lacks (Prop. 2).

3.4 Learning Objective

The main output is supervised by a combined Dice and cross-entropy loss, $\mathcal{L}_{\text{seg}}(\cdot) = \mathcal{L}_{\text{Dice}}(\cdot) + \mathcal{L}_{\text{CE}}(\cdot)$. We add two auxiliary branches, used only in training and dropped at inference. The first keeps every encoder well trained. When a modality is down-weighted or dropped in fusion, its encoder gets little gradient. To avoid this, each bottleneck $\boldsymbol{f}_m$ is decoded on its own by a weight-shared decoder into a prediction $\hat{\boldsymbol{y}}_m$, supervised by the segmentation loss. The second branch supervises the region nodes. We correlate the refined region nodes with the fused feature at every location,

$$\boldsymbol{A}_{\text{aux}} = \text{einsum}(\boldsymbol{R}_2, \boldsymbol{F}') \in \mathbb{R}^{R \times d \times h \times w}, \tag{11}$$

and supervise it with region labels at the bottleneck resolution. The two auxiliary losses are

$$\mathcal{L}_{\text{sp}} = \sum_{m:a_m=1} \mathcal{L}_{\text{seg}}(\hat{\boldsymbol{y}}_m, \boldsymbol{y}), \quad \mathcal{L}_{\text{rg}} = \mathcal{L}_{\text{seg}}(\boldsymbol{A}_{\text{aux}}, \boldsymbol{y}_{\text{region}}), \tag{12}$$

where $\boldsymbol{y}_{\text{region}}$ are the WT/TC/ET labels at the bottleneck resolution. The overall objective is

$$\mathcal{L} = \mathcal{L}_{\text{seg}}(\hat{\boldsymbol{y}}, \boldsymbol{y}) + \lambda_{\text{sp}}\mathcal{L}_{\text{sp}} + \lambda_{\text{rg}}\mathcal{L}_{\text{rg}}, \tag{13}$$

where λ_{sp} and λ_{rg} balance the auxiliary terms against the main loss.

4 Experiments

4.1 Datasets and Evaluation

We evaluate HGMamba on two widely used brain tumor segmentation benchmarks, BraTS 2018 and BraTS 2020 (Menze et al., 2015), comprising 285 and 369 multimodal cases, respectively.

Table 2: HD95 (mm, lower is better) on BraTS 2018 across all 15 modality combinations. ● and ○ denote available and missing modalities. Red: best; blue: second best; our method is highlighted in green.

Type	Flair	●	○	○	○	●	●	●	○	○	○	●	●	●	○	●	Avg
	T1	○	○	●	○	○	●	○	●	○	●	●	○	●	●	●	
	T1ce	○	●	○	○	●	○	○	●	●	○	●	●	○	●	●	
	T2	○	○	○	●	○	○	●	○	●	●	○	●	●	●	●	
WT	U-HVED	24.19	20.15	41.55	11.91	11.96	16.09	8.86	23.65	9.72	20.44	10.59	10.92	11.07	7.61	7.44	15.74
	mmFormer	5.00	9.38	9.65	8.34	3.43	3.20	3.04	7.29	4.30	8.41	3.34	3.00	**2.93**	4.24	3.06	5.24
	ShaSpec	5.52	9.72	10.21	14.32	3.15	3.72	5.11	8.26	9.52	7.98	3.45	3.08	3.92	7.62	2.93	6.57
	DC-Seg	**4.07**	7.98	9.35	8.03	3.67	3.72	3.16	5.62	7.23	8.06	3.57	2.91	2.94	4.82	2.90	5.20
	UniME	8.37	20.76	19.52	16.62	5.07	**2.93**	4.33	10.14	9.97	11.94	2.86	5.74	5.80	9.55	4.07	9.18
	Ours	4.67	**6.84**	**6.43**	**7.76**	**2.72**	3.18	**2.97**	**5.45**	**4.22**	**3.89**	**2.67**	**2.53**	3.01	**3.57**	**2.62**	**4.17**
TC	U-HVED	32.13	13.23	23.00	14.64	12.19	15.96	12.68	9.49	8.78	13.73	8.23	6.50	14.24	8.28	6.58	13.31
	mmFormer	7.71	6.04	8.95	10.17	5.12	6.05	7.30	4.36	4.82	9.07	**4.54**	4.98	5.90	4.56	4.60	6.28
	ShaSpec	10.49	7.33	13.16	14.59	5.16	6.56	10.23	5.14	9.00	11.14	4.88	7.29	8.67	4.48	4.54	8.18
	DC-Seg	11.28	**4.84**	10.01	7.48	5.15	6.22	6.94	3.96	**4.19**	**6.46**	4.74	4.53	5.88	4.19	**4.31**	6.01
	UniME	19.48	19.20	18.95	14.93	5.43	**5.74**	9.55	9.26	11.68	10.49	4.90	8.74	9.40	9.06	4.77	10.77
	Ours	**7.20**	5.24	**8.14**	**6.70**	**4.89**	5.97	**5.72**	**3.48**	4.28	6.50	4.93	**4.37**	**5.56**	**4.14**	4.48	**5.44**
ET	U-HVED	30.67	7.31	20.33	11.20	6.71	14.40	9.90	3.51	3.08	11.15	3.23	3.12	11.25	3.32	3.24	9.49
	mmFormer	8.32	2.80	10.09	10.59	2.82	6.87	7.75	2.78	3.00	8.42	2.80	2.96	6.34	3.03	2.97	5.44
	ShaSpec	7.39	3.07	13.52	14.74	2.95	7.51	6.93	2.84	2.91	8.64	2.74	2.88	6.10	2.79	2.68	5.85
	DC-Seg	7.83	2.14	10.79	7.43	2.76	8.08	6.80	2.23	2.65	6.46	2.76	2.76	**5.33**	2.65	2.72	4.89
	UniME	7.33	8.78	13.35	11.43	2.64	7.08	6.82	6.06	6.11	10.62	2.62	2.58	6.52	6.28	2.52	6.72
	Ours	**6.69**	**1.74**	**8.36**	**7.13**	**2.03**	**6.67**	**6.53**	**1.94**	**2.07**	**6.31**	**1.92**	**2.09**	6.36	**1.92**	**1.94**	**4.25**

Each case provides four co-registered, skull-stripped MRI sequences (FLAIR, T1ce, T1, and T2) at a resolution of $155 \times 240 \times 240$, together with expert annotations. Following common practice, the annotations are merged into three nested regions for evaluation: whole tumor (WT), tumor core (TC), and enhancing tumor (ET). We randomly split each dataset into training, validation, and testing subsets with a ratio of approximately 8:1:1, yielding 228/28/29 cases for BraTS 2018 and 295/36/38 for BraTS 2020. We adopt two complementary metrics, the Dice Similarity Coefficient (DSC) that measures regional overlap and the 95% Hausdorff Distance (HD95) that measures boundary accuracy, evaluating every model under all 15 possible modality combinations and reporting the per-region average. Cases where a region is empty in the prediction or the ground truth are excluded from the HD95 average, identically for all methods.

4.2 Implementation Details

We use a 3D U-Net with four modality-specific encoders as the backbone. Input volumes are randomly cropped to 128^3, with random rotation, intensity shifting, and flipping. To simulate missing modalities, we randomly drop 0 to 3 modalities per case while keeping at least one, and deactivate the absent ones as inactive graph nodes instead of zero-filling. Each NodeMamba block uses a Mamba layer with state dimension 16, convolution width 2, and expansion factor 2. The reliability gate is a single linear layer with a sigmoid, and the affinity prior is injected with a fixed strength γ=1.0. We train with RAdam at a learning rate of 2×10^{-4} for 300 epochs and batch size 1 on a single NVIDIA A800 80 GB GPU, running Ubuntu with PyTorch 2.5.1 and CUDA 12.4, with λ_{sp}=1.0 and λ_{rg}=0.5. Each model is trained with a fixed random seed. At inference, we evaluate all 15 combinations with sliding-window prediction.

All baselines are run with their official implementations and original network architectures. To ensure a fair comparison, we retrain every method from scratch on the same data split, with the same preprocessing, training schedule, and 15-combination evaluation protocol as HGMamba. We keep each method's own architecture unchanged, so the comparison reflects the fusion design of each method under an identical data and evaluation setting.

4.3 Comparison with State-of-the-Art Methods

We compare HGMamba with five representative missing-modality methods that span the main fusion families: a hetero-modal variational model (U-HVED (Dorent et al., 2019)), a trans-

Table 3: Average Dice (%, ↑) over all 15 modality combinations on BraTS 2020.

Method	WT	TC	ET	Avg
U-HVED	77.43	65.81	50.72	64.65
mmFormer	86.98	79.49	64.09	76.85
ShaSpec	84.98	75.25	62.28	74.17
DC-Seg	86.11	79.26	**64.84**	76.74
UniME	87.40	**80.33**	64.76	**77.50**
Ours	**87.79**	79.84	64.80	77.48

Table 4: Average HD95 (mm, ↓) over all 15 modality combinations on BraTS 2020.

Method	WT	TC	ET	Avg
U-HVED	21.12	15.81	13.91	16.95
mmFormer	6.35	6.73	6.71	6.60
ShaSpec	8.76	8.36	7.92	8.35
DC-Seg	8.24	7.53	7.96	7.91
UniME	9.97	8.74	7.72	8.81
Ours	**5.60**	**6.32**	**5.95**	**5.96**

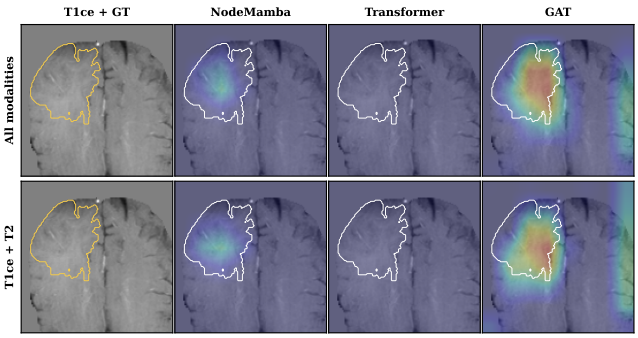

Figure 3: Grad-CAM of the fused features for the three operators (whole-tumor target, per-model normalization). The case shown is the BraTS2018 test volume where NodeMamba's in-tumor activation most exceeds the others' without a lower WT Dice.

former fusion model (mmFormer (Zhang et al., 2022)), a shared-specific representation model (ShaSpec (Wang et al., 2023a)), a disentangled region-aware model (DC-Seg (Li et al., 2026)), and a pretraining-based unified model (UniME (Song et al., 2026)). For fairness, all methods are re-run by us with their original architectures on the same data split and evaluation protocol, so the differences reflect the fusion design. Every score is averaged over all 15 modality combinations.

On BraTS 2018 (Tables 1 and 2), HGMamba matches the strongest competitor UniME in Dice (within 0.1%, paired Wilcoxon $p > 0.05$) while using an order of magnitude fewer parameters, and attains the best HD95 on all three regions, cutting the average to 4.62 mm from 5.37 mm (DC-Seg) and 8.89 mm (UniME), the gain over UniME being significant ($p < 0.01$). BraTS 2020 shows the same pattern (Tables 3 and 4; per combination in Tables 7 and 8): comparable Dice and the best average HD95 (5.96 mm), so the boundary advantage does not depend on one dataset.

Qualitative comparisons are given in Appendix A.1; HGMamba yields coherent masks where the baselines produce false positives, consistent with its lower HD95.

4.4 Ablation Study

We conduct ablations on BraTS 2018, retraining every variant under the same settings and evaluating it over all 15 modality combinations. Table 5 removes each key component. Replacing the data-driven affinity prior with a uniform prior causes the largest drop, reducing the mean Dice by 1.53%; removing the reliability gate costs 0.87%, and removing the nested region interaction costs 0.55%. The enhancing tumor degrades most in all three cases, indicating that it is the region most sensitive to fusion. Table 6 then replaces NodeMamba with a Graph Attention Network (GAT) and a Transformer under matched parameter counts, keeping all other components unchanged. The framework remains effective with every operator: the GAT and Transformer variants reach 76.46% and 76.79%, already above all baselines except UniME, and NodeMamba gives the highest score (77.16%). Changing the operator costs at most 0.70%, less than half the drop from removing the prior, and all three instantiations reach a lower average HD95 than every baseline (Figure 1c; Table 11), with NodeMamba the lowest. With T1ce available, NodeMamba also leads GAT and the Transformer on ET Dice by 2.8 and 2.3 points (Table 10). Grad-CAM (Figure 3) shows NodeMamba's activation staying inside the tumor, whereas GAT spreads beyond the boundary and the Transformer places little activation on the lesion slice. The gain thus comes mainly from the modality-region formulation and the affinity prior, with NodeMamba as the best-performing instantiation.

4.5 Efficiency Analysis

Figure 4 compares efficiency on BraTS 2018 (details in Appendix A.9). HGMamba matches UniME's accuracy with $18\times$ fewer parameters (8.74M vs. 158.12M) and runs over $2\times$ faster than

Table 5: Ablation study on BraTS2018. Mean Dice (%) over all modality combinations.

Method	WT	TC	ET	Avg
All	**87.69**	**81.11**	**62.68**	**77.16**
w/o A_prior	86.29	79.94	60.66	75.63
w/o gate	86.79	80.47	61.61	76.29
w/o nested	86.95	80.82	62.05	76.61

Table 6: Interaction operator comparison on BraTS2018 under matched parameter counts.

Method	WT	TC	ET	Avg
GAT	86.47	80.62	62.28	76.46
Transformer	87.33	80.95	62.09	76.79
NodeMamba (Ours)	**87.69**	**81.11**	**62.68**	**77.16**

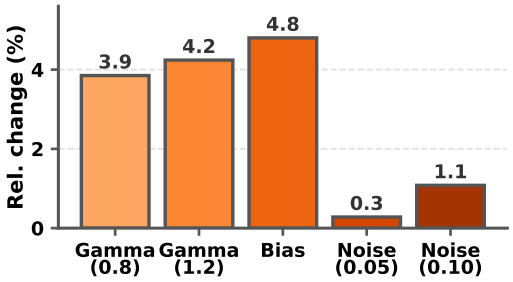

Figure 5: Stability of the affinity prior under intensity perturbations. The relative change stays below 5%.

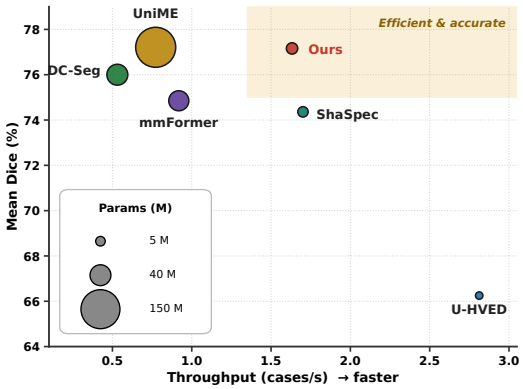

Figure 4: Efficiency vs. accuracy on BraTS 2018 (marker area $\propto$ parameters). HGMamba sits in the top-right region.

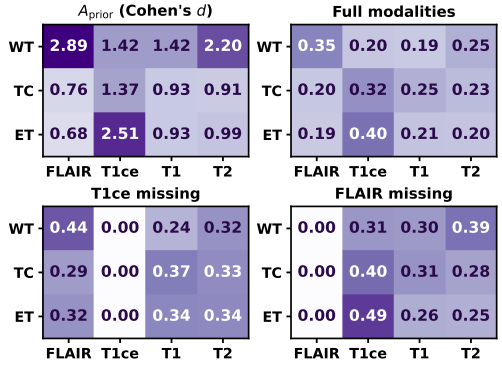

Figure 6: Affinity prior $\boldsymbol{A}_{\text{prior}}$ and routing weights under full, T1ce-missing, and FLAIR-missing inputs.

UniME and 3$\times$ faster than DC-Seg. The saving comes mostly from the lightweight encoders; the graph fusion adds only 0.31M parameters (3.5%).

4.6 Interpretability and Stability of the Prior

HGMamba lets its fusion be inspected (Fig. 6). The prior assigns FLAIR to the whole tumor and T1ce to the enhancing tumor, matching clinical knowledge, and the routing weights follow it under full modalities. When a modality is absent, its column drops to zero and the weight shifts sensibly, for example to T2 when FLAIR is missing. Under intensity perturbations, the re-estimated prior changes by less than 5% and the top modality per region never flips (Fig. 5).

5 Conclusion

We propose HGMamba, a heterogeneous modality-region fusion framework for missing-modality brain tumor segmentation that treats modalities and nested tumor regions as two node types. A parameter-free affinity prior guides the routing between them and makes the fusion interpretable, and the framework stays effective with NodeMamba, graph attention, or Transformer operators. On BraTS 2018 and 2020, HGMamba matches the strongest competitor in Dice with better HD95 and far fewer parameters, and the boundary advantage carries over to unseen scanners and brain metastases. Two limitations remain: full-modality Dice is on par with, not ahead of, the strongest pretrained baseline, and the fusion is validated only on a lightweight backbone.

AI USE STATEMENT

In this work, we used generative AI tools solely to polish the language and improve the readability of the manuscript. We did not use generative AI tools for designing experiments, analyzing data, generating research ideas, or contributing to the scientific findings of this work. All AI-assisted text was reviewed by the authors. We take responsibility for the final content of this work, including all text, claims, and artifacts.

ETHICS STATEMENT

Our study does not involve human subjects, sensitive data, or personally identifiable information. The datasets we use are publicly available and commonly adopted in the community. We are not aware of potential ethical risks related to privacy, fairness, or misuse, and we have read and adhere to the ICLR Code of Ethics.

REPRODUCIBILITY STATEMENT

We are committed to ensuring the reproducibility of our results. Implementation details, including the model architecture and training setup, are provided in Section 3 and the Implementation Details of the main paper. The datasets and their processing steps are described in the Experiments section, and we further include the source code and scripts in the supplementary materials to facilitate replication.

REFERENCES

Zain Ul Abidin, Rizwan Ali Naqvi, Amir Haider, Hyung Seok Kim, Daesik Jeong, and Seung Won Lee. Recent deep learning-based brain tumor segmentation models using multi-modality magnetic resonance imaging: a prospective survey. *Frontiers in Bioengineering and Biotechnology*, 12, 2024. URL https://api.semanticscholar.org/CorpusID:271344292.

Joseph Bae, Saarthak Kapse, Lei Zhou, Kartik Mani, and Prateek Prasanna. HoG-Net: Hierarchical Multi-Organ Graph Network for Head and Neck Cancer Recurrence Prediction from CT Images . In *proceedings of Medical Image Computing and Computer Assisted Intervention – MICCAI 2024*, volume LNCS 15005. Springer Nature Switzerland, October 2024.

Cheng Chen, Qi Dou, Yueming Jin, Hao Chen, Jing Qin, and Pheng-Ann Heng. 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.

Cheng Chen, Qi Dou, Yueming Jin, Quande Liu, and Pheng Ann Heng. Learning with privileged multimodal knowledge for unimodal segmentation. *IEEE Transactions on Medical Imaging*, 41 (3):621–632, 2022. doi: 10.1109/TMI.2021.3119385.

Reuben Dorent, Samuel Joutard, Marc Modat, Sébastien Ourselin, and Tom Vercauteren. Hetero-modal variational encoder-decoder for joint modality completion and segmentation. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part II*, pp. 74–82, Berlin, Heidelberg, 2019. Springer-Verlag. ISBN 978-3-030-32244-1. doi: 10.1007/978-3-030-32245-8_9. URL https://doi.org/10.1007/978-3-030-32245-8_9.

Niharika S. D'Souza, Hongzhi Wang, Andrea Giovannini, Antonio Foncubierta-Rodriguez, Kristen L. Beck, Orest Boyko, and Tanveer F. Syeda-Mahmood. Fusing modalities by multiplexed graph neural networks for outcome prediction from medical data and beyond. *Medical Image Analysis*, 93:103064, 2024. ISSN 1361-8415. doi: https://doi.org/10.1016/j.media.2023.103064. URL https://www.sciencedirect.com/science/article/pii/S1361841523003249.

Xue Feng, Kanchan Ghimire, Daniel D. Kim, Rajat S. Chandra, Helen Zhang, Jian Peng, Binghong Han, Gaofeng Huang, Quan Chen, Sohil Patel, Chetan Bettegowda, Haris I. Sair, Craig Jones, Zhicheng Jiao, Li Yang, and Harrison Bai. Brain tumor segmentation for multi-modal mri

with missing information. *Journal of Digital Imaging*, 36(5):2075–2087, 2023. doi: 10.1007/s10278-023-00860-7.

Zhenyu Gong, Tao Xu, Nan Peng, Xing Cheng, Chen Niu, Benedikt Wiestler, Fan Hong, and Hongwei Bran Li. A multi-center, multi-parametric MRI dataset of primary and secondary brain tumors. *Scientific Data*, 11(1):789, 2024. doi: 10.1038/s41597-024-03634-0.

Albert Gu and Tri Dao. Mamba: Linear-time sequence modeling with selective state spaces. *ArXiv*, abs/2312.00752, 2023. URL https://api.semanticscholar.org/CorpusID:265551773.

Mohammad Havaei, Nicolas Guizard, Nicolas Chapados, and Yoshua Bengio. Hemis: Hetero-modal image segmentation. In Sebastien Ourselin, Leo Joskowicz, Mert R. Sabuncu, Gozde Unal, and William Wells (eds.), *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*, pp. 469–477, Cham, 2016. Springer International Publishing. ISBN 978-3-319-46723-8.

Fabian Isensee, Tassilo Wald, Constantin Ulrich, Michael Baumgartner, Saikat Roy, Klaus Maier-Hein, and Paul F. Jäger. nnu-net revisited: A call for rigorous validation in 3d medical image segmentation. In Marius George Linguraru, Qi Dou, Aasa Feragen, Stamatia Giannarou, Ben Glocker, Karim Lekadir, and Julia A. Schnabel (eds.), *Medical Image Computing and Computer Assisted Intervention – MICCAI 2024*, pp. 488–498, Cham, 2024. Springer Nature Switzerland. ISBN 978-3-031-72114-4.

Xuefeng Jia, Biyuan Li, Jinying Ma, Chunjie Lv, Xiao Tian, Lianhao Huo, and Mengyao Lun. Multi-modal medical image synthesis via dual-branch wavelet encoding and deformable feature interaction. *Artificial Intelligence in Medicine*, 171:103287, 2026. ISSN 0933-3657. doi: https://doi.org/10.1016/j.artmed.2025.103287. URL https://www.sciencedirect.com/science/article/pii/S0933365725002222.

Ming Kang, Fung Fung Ting, Shier Nee Saw, Rapharl Phan, Zongyuan Ge, and Chee-Ming Ting. A multimodal feature distillation with mamba-transformer network for brain tumor segmentation with incomplete modalities. *IEEE Transactions on Artificial Intelligence*, 2024. URL https://api.semanticscholar.org/CorpusID:269293244.

Ming Kang, Fung Fung Ting, Shier Nee Saw, Raphaël C. W. Phan, Zongyuan Ge, and Chee-Ming Ting. A multimodal feature distillation with mamba-transformer network for brain tumor segmentation with incomplete modalities, 2026. URL https://arxiv.org/abs/2404.14019.

Sanaz Karimijafarbigloo, Reza Azad, Amirhossein Kazerouni, Saeed Ebadollahi, and Dorit Merhof. Mmcformer: Missing modality compensation transformer for brain tumor segmentation. In Ipek Oguz, Jack Noble, Xiaoxiao Li, Martin Styner, Christian Baumgartner, Mirabela Rusu, Tobias Heinmann, Despina Kontos, Bennett Landman, and Benoit Dawant (eds.), *Medical Imaging with Deep Learning*, volume 227 of *Proceedings of Machine Learning Research*, pp. 1144–1162. PMLR, 10–12 Jul 2024. URL https://proceedings.mlr.press/v227/karimijafarbigloo24b.html.

Anna Lerner, Kieran Palmer, Tom Campion, Thomas O Millner, Emily Scott, Cressida Lorimer, Dimitrios Paraskevopoulos, Grainne McKenna, Silvia Marino, Rachel Lewis, and Nick Plowman. Gliomas in adults: Guidance on investigations, diagnosis, treatment and surveillance. *Clinical Medicine*, 24(5):100240, 2024. ISSN 1470-2118. doi: https://doi.org/10.1016/j.clinme.2024.100240. URL https://www.sciencedirect.com/science/article/pii/S1470211824054253.

Haitao Li, Ziyu Li, Yiheng Mao, Zhengyao Ding, and Zhengxing Huang. Dc-seg: Disentangled contrastive learning for brain tumor segmentation with missing modalities. In James C. Gee, Daniel C. Alexander, Jaesung Hong, Juan Eugenio Iglesias, Carole H. Sudre, Archana Venkataraman, Polina Golland, Jong Hyo Kim, and Jinah Park (eds.), *Medical Image Computing and Computer Assisted Intervention – MICCAI 2025*, pp. 138–148, Cham, 2026. Springer Nature Switzerland. ISBN 978-3-032-04984-1.

Bjoern H. Menze, Andras Jakab, Stefan Bauer, et al. The multimodal brain tumor image segmentation benchmark (brats). *IEEE Transactions on Medical Imaging*, 34(10):1993–2024, 2015.

Ibomoiye Domor Mienye and Serestina Viriri. Graph neural networks in medical imaging: Methods, applications and future directions. *Information*, 16(12), 2025. ISSN 2078-2489. doi: 10.3390/info16121051. URL https://www.mdpi.com/2078-2489/16/12/1051.

Liang Peng, Songyue Cai, Zongqian Wu, Huifang Shang, Xiaofeng Zhu, and Xiaoxiao Li. Mmgpl: Multimodal medical data analysis with graph prompt learning. *Medical Image Analysis*, 97:103225, 2024. ISSN 1361-8415. doi: https://doi.org/10.1016/j.media.2024.103225. URL https://www.sciencedirect.com/science/article/pii/S1361841524001506.

Vittorio Pipoli, Alessia Saporita, Kevin Marchesini, Costantino Grana, Elisa Ficarra, and Federico Bolelli. IM-Fuse: A Mamba-based Fusion Block for Brain Tumor Segmentation with Incomplete Modalities . In *proceedings of Medical Image Computing and Computer Assisted Intervention – MICCAI 2025*, volume LNCS 15967. Springer Nature Switzerland, September 2025.

Yansheng Qiu, Kui Jiang, Hongdou Yao, Zheng Wang, and Shin'ichi Satoh. Does adding a modality really make positive impacts in incomplete multi-modal brain tumor segmentation? *IEEE Transactions on Medical Imaging*, 44(5):2194–2205, 2025. doi: 10.1109/TMI.2025.3526818.

Junjie Shi, Caozhi Shang, Zhaobin Sun, Li Yu, Xin Yang, and Zengqiang Yan. Passion: Towards effective incomplete multi-modal medical image segmentation with imbalanced missing rates. In *Proceedings of the 32nd ACM International Conference on Multimedia*, pp. 456–465, 2024.

Yashbir Singh, Yassine Himeur, Colleen M. Farrelly, Peguy Kem-Meka Tiotsop Kadzue, Jennifer Z. Rozenblit, Amina Kunovac, Isabelle C. Pappas, Ashok Choudhary, Sara Salehi, Shadi Atalla, and Quincy A. Hathaway. Graph neural networks for medical imaging analysis and biological data: Integrating topology, geometry, radiomics, and generative ai. *Bioengineering*, 13(6), 2026. ISSN 2306-5354. doi: 10.3390/bioengineering13060638. URL https://www.mdpi.com/2306-5354/13/6/638.

Peibo Song, Xiaotian Xue, Jinshuo Zhang, Zihao Wang, Jinhua Liu, Shujun Fu, Fangxun Bao, and Si Yong Yeo. Uni-encoder meets multi-encoders: Representation before fusion for brain tumor segmentation with missing modalities. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 15627–15638, June 2026.

Hu Wang, Yuanhong Chen, Congbo Ma, Jodie Avery, Louise Hull, and G. Carneiro. Multi-modal learning with missing modality via shared-specific feature modelling. *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 15878–15887, 2023a. URL https://api.semanticscholar.org/CorpusID:260164982.

Hu Wang, Congbo Ma, Jianpeng Zhang, Yuan Zhang, Jodie Avery, Louise Hull, and Gustavo Carneiro. Learnable cross-modal knowledge distillation fornbsp;multi-modal learning withnbsp;missing modality. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2023: 26th International Conference, Vancouver, BC, Canada, October 8–12, 2023, Proceedings, Part IV*, pp. 216–226, Berlin, Heidelberg, 2023b. Springer-Verlag. ISBN 978-3-031-43900-1. doi: 10.1007/978-3-031-43901-8_21. URL https://doi.org/10.1007/978-3-031-43901-8_21.

Pengyu Wang, Qiushi Yang, Zhibin He, and Yixuan Yuan. Vision transformers in multi-modal brain tumor mri segmentation: A review. *Meta-Radiology*, 1(1):100004, 2023c. ISSN 2950-1628. doi: https://doi.org/10.1016/j.metrad.2023.100004. URL https://www.sciencedirect.com/science/article/pii/S2950162823000048.

Ruomei Wang, Wei Guo, Yongjie Wang, Xin Zhou, Jonathan Cyril Leung, Shuo Yan, and Lizhen Cui. Hybrid multimodal fusion for graph learning in disease prediction. *Methods*, 229:41–48, 2024. ISSN 1046-2023. doi: https://doi.org/10.1016/j.ymeth.2024.06.003. URL https://www.sciencedirect.com/science/article/pii/S1046202324001531.

Zhilin Zeng, Zelin Peng, Xiaokang Yang, and Wei Shen. Missing as Masking: Arbitrary Cross-modal Feature Reconstruction for Incomplete Multimodal Brain Tumor Segmentation . In *proceedings of Medical Image Computing and Computer Assisted Intervention – MICCAI 2024*, volume LNCS 15008. Springer Nature Switzerland, October 2024.

Xinyue Zhang, Ali Bahri, Christian Desrosiers, Hui Liu, and Fangxun Bao. M^2segmamba: Mamba-based incomplete multimodal learning for brain tumor segmentation with few samples. *IEEE Journal of Biomedical and Health Informatics*, 30(3):2455–2465, 2026. doi: 10.1109/JBHI.2025.3600652.

Yao Zhang, Nanjun He, Jiawei Yang, Yuexiang Li, Dong Wei, Yawen Huang, Yang Zhang, Zhiqiang He, and Yefeng Zheng. mmformer: Multimodal medical transformer for incomplete multimodal learning of brain tumor segmentation. In Linwei Wang, Qi Dou, P. Thomas Fletcher, Stefanie Speidel, and Shuo Li (eds.), *Medical Image Computing and Computer Assisted Intervention – MICCAI 2022*, pp. 107–117, Cham, 2022. Springer Nature Switzerland. ISBN 978-3-031-16443-9.

Yue Zhang, Chengtao Peng, Qiuli Wang, Dan Song, Kaiyan Li, and S. Kevin Zhou. Unified multi-modal image synthesis for missing modality imputation. *IEEE Transactions on Medical Imaging*, 44(1):4–18, 2025a. doi: 10.1109/TMI.2024.3424785.

Zheyu Zhang, Yayuan Lu, Feipeng Ma, Yueyi Zhang, Huanjing Yue, and Xiaoyan Sun. Incomplete multi-modal brain tumor segmentation via learnable sorting state space model. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 25982–25992, June 2025b.

Tongxue Zhou, Su Ruan, and Haigen Hu. A literature survey of mr-based brain tumor segmentation with missing modalities. *Computerized Medical Imaging and Graphics*, 104:102167, 2023. ISSN 0895-6111. doi: https://doi.org/10.1016/j.compmedimag.2022.102167. URL https://www.sciencedirect.com/science/article/pii/S0895611122001379.

Shenghao Zhu, Yifei Chen, Weihong Chen, Yuanhan Wang, Chang Liu, Shuo Jiang, Feiwei Qin, and Changmiao Wang. Bridging the Gap in Missing Modalities: Leveraging Knowledge Distillation and Style Matching for Brain Tumor Segmentation . In *proceedings of Medical Image Computing and Computer Assisted Intervention – MICCAI 2025*, volume LNCS 15967. Springer Nature Switzerland, September 2025.

Shenghao Zhu, Yifei Chen, Weihong Chen, Shuo Jiang, Guanyu Zhou, Yuanhan Wang, Feiwei Qin, Changmiao Wang, and Qiyuan Tian. No modality left behind: Adapting to missing modalities via knowledge distillation for brain tumor segmentation. *Medical Image Analysis*, 112:104108, 2026. ISSN 1361-8415. doi: https://doi.org/10.1016/j.media.2026.104108. URL https://www.sciencedirect.com/science/article/pii/S1361841526001775.

A Additional Results

A.1 Qualitative Results

Figure 7 shows qualitative results under four missing-modality settings. Removing an informative modality hurts every method, and the effect is strongest when T1ce is absent. The differences stand out in the harder cases. When FLAIR is missing, UniME predicts false positives in regions where no lesion can exist. ShaSpec fails in a similar way when only T2 is available. HGMamba stays spatially consistent and gives more coherent masks, which agrees with its lower HD95.

A.2 Per-Combination Results on BraTS 2020

In the main text we reported the averaged Dice and HD95 on BraTS 2020. Here we give the full per-combination results. Table 7 reports Dice and Table 8 reports HD95 across all 15 modality combinations. The trends match BraTS 2018: HGMamba leads on the whole tumor in Dice, stays first or second in most combinations, and obtains the best average HD95 on every region, confirming that its advantage in boundary accuracy transfers across datasets rather than being tuned to one benchmark.

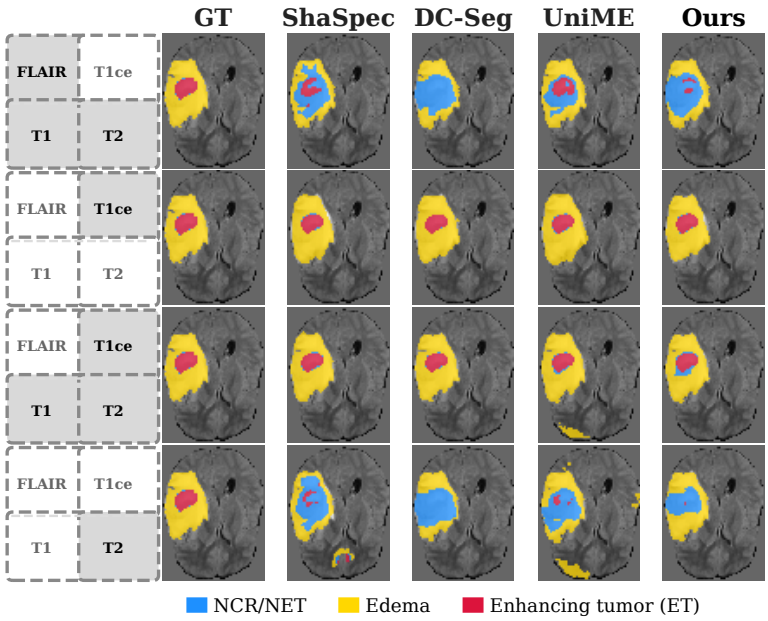

Figure 7: Qualitative results on BraTS2018 under different input modality combinations. Blue, gold, and red indicate the NCR/NET, peritumoral edema, and enhancing tumor (ET).

Table 7: Dice (%) on BraTS 2020 across all 15 modality combinations. ● and ○ denote available and missing modalities. Red: best; blue: second best; our method is highlighted in green.

Type	Flair / T1ce / T1 / T2																Avg
	Flair	●	○	○	○	●	●	●	○	○	○	●	●	●	○	●	
	T1ce	○	●	○	○	●	○	○	●	●	○	●	●	○	●	●	
	T1	○	○	●	○	○	●	○	●	○	●	●	○	●	●	●	
	T2	○	○	○	●	○	○	●	○	●	●	○	●	●	●	●	
WT	U-HVED	84.21	59.30	50.57	71.43	86.13	85.29	86.23	60.17	78.22	74.78	86.71	86.76	86.01	78.60	87.00	77.43
	mmFormer	**88.71**	78.96	78.91	84.54	89.66	89.70	89.54	81.47	87.33	87.21	90.10	90.15	89.90	88.21	90.33	86.98
	ShaSpec	85.59	76.68	70.86	81.97	89.32	88.31	88.28	77.69	86.48	84.12	89.65	89.97	89.09	86.64	90.09	84.98
	DC-Seg	87.69	80.85	77.21	82.44	89.69	89.07	87.48	83.27	85.90	84.99	89.68	89.08	88.46	86.73	89.10	86.11
	UniME	86.78	79.62	79.23	**85.67**	90.19	89.45	89.35	83.38	**88.73**	**88.03**	90.39	90.61	90.06	**88.99**	90.67	87.41
	Ours	88.30	**81.67**	**80.03**	84.97	**90.59**	**90.18**	**90.03**	**84.04**	88.34	86.81	**90.77**	**91.01**	**90.62**	88.50	**91.02**	**87.79**
TC	U-HVED	58.90	74.84	31.08	41.26	79.37	62.64	59.23	75.37	78.10	47.29	80.03	78.91	63.11	77.89	79.10	65.81
	mmFormer	72.59	82.13	71.28	69.82	83.85	75.92	75.97	83.71	84.47	76.33	84.38	84.46	77.37	**85.32**	84.72	79.49
	ShaSpec	66.35	80.35	57.89	64.16	82.13	71.81	71.64	80.27	81.64	68.85	82.55	82.38	73.91	82.08	82.72	75.25
	DC-Seg	71.36	84.45	69.34	69.73	84.85	75.94	74.46	84.52	84.57	75.40	**84.79**	84.53	76.32	84.42	84.20	79.26
	UniME	72.24	84.92	71.64	**72.46**	**84.95**	77.20	**76.33**	85.43	**85.17**	**77.03**	84.70	**84.99**	**77.93**	85.18	**84.75**	**80.33**
	Ours	**73.24**	**85.75**	**71.72**	71.16	84.24	**77.40**	75.87	**85.50**	84.59	75.70	84.07	83.19	77.90	83.98	83.28	79.84
ET	U-HVED	24.40	66.01	7.13	20.16	71.04	35.87	33.21	72.13	72.78	26.35	72.57	71.98	40.86	74.04	72.29	50.72
	mmFormer	43.80	78.56	39.43	45.76	78.86	47.95	51.00	78.98	78.92	49.88	79.34	79.05	51.65	79.11	79.05	64.09
	ShaSpec	43.08	74.90	30.69	43.16	78.33	45.82	48.69	77.52	78.40	45.25	**79.42**	**79.51**	50.16	79.38	**79.83**	62.28
	DC-Seg	45.93	**78.83**	42.67	45.11	**79.25**	49.54	51.50	**80.24**	**79.60**	49.17	79.39	79.34	52.73	**79.98**	79.38	**64.84**
	UniME	**47.25**	77.57	40.99	**49.04**	77.21	**51.32**	**53.96**	77.62	77.69	**52.87**	77.55	77.29	**55.72**	77.89	77.44	64.76
	Ours	46.55	78.80	**43.47**	46.65	77.84	50.77	51.91	78.78	79.59	50.06	78.09	78.60	53.33	79.08	78.50	64.80

A.3 Full Ablation and Operator Results on BraTS 2018

Tables 9 and 10 expand the averaged ablation and operator studies of the main text to all 15 modality combinations. Removing the affinity prior causes the largest drop across almost every combination, and the full model is best or second on most columns. Across interaction operators, the three variants stay close on every combination, confirming that the framework does not depend on a specific operator; NodeMamba attains the highest average on all three regions. Table 11 reports HD95 for the same variants. NodeMamba attains the lowest average HD95 on all three regions, while GAT

Table 8: HD95 (mm, lower is better) on BraTS 2020 across all 15 modality combinations. ● and ○ denote available and missing modalities. Red: best; blue: second best; our method is highlighted in green.

Type	Flair	●	○	○	○	●	●	●	○	○	○	●	●	●	○	●	Avg
	T1ce	○	●	○	○	●	○	○	●	●	○	●	●	○	●	●	
	T1	○	○	●	○	○	●	○	●	○	●	●	○	●	●	●	
	T2	○	○	○	●	○	○	●	○	●	●	○	●	●	●	●	
WT	U-HVED	16.57	32.15	41.54	19.50	18.26	15.74	16.52	33.22	18.52	20.45	16.32	17.35	14.00	20.23	16.45	21.12
	mmFormer	6.52	8.33	9.95	**8.58**	5.72	**4.79**	6.36	7.77	**5.67**	6.78	4.58	5.60	4.82	**4.59**	5.19	6.35
	ShaSpec	8.31	14.33	18.47	8.95	6.77	7.09	7.64	11.07	6.01	8.78	6.52	6.65	7.28	7.15	6.43	8.76
	DC-Seg	7.83	8.02	13.22	12.41	6.31	6.85	8.57	6.77	9.54	10.17	6.47	6.66	7.31	6.53	6.89	8.24
	UniME	13.60	17.58	18.14	9.49	5.95	11.07	10.30	9.30	5.69	9.83	8.29	5.85	8.58	7.82	8.11	9.97
	Ours	**5.81**	**7.04**	**8.59**	8.63	**4.40**	5.25	**4.67**	**6.11**	5.92	**5.27**	**4.22**	**4.32**	**4.31**	5.11	**4.32**	**5.60**
TC	U-HVED	15.27	18.80	40.50	20.67	9.90	18.62	12.79	16.69	10.62	22.51	10.56	8.38	15.80	9.95	6.12	15.81
	mmFormer	**8.27**	6.29	11.23	12.08	4.75	**6.73**	**6.81**	6.27	5.31	9.62	4.53	4.32	**6.32**	4.28	**4.14**	6.73
	ShaSpec	12.33	7.23	17.97	11.62	4.62	10.03	7.97	7.33	5.16	11.14	4.78	5.21	8.83	6.46	4.69	8.36
	DC-Seg	11.29	**4.50**	13.78	14.59	**4.45**	9.98	10.66	**4.32**	**4.39**	7.68	**4.40**	**4.30**	10.10	**4.24**	4.22	7.53
	UniME	16.34	16.73	15.84	10.93	4.56	10.47	10.35	8.08	4.68	**6.54**	4.59	4.51	8.59	4.46	4.49	8.74
	Ours	10.68	5.65	**7.94**	**8.71**	4.71	6.75	7.55	4.71	4.86	6.94	4.87	4.93	6.61	4.88	5.00	**6.32**
ET	U-HVED	16.73	17.50	33.96	20.46	9.19	21.55	14.99	10.07	8.24	19.06	5.52	6.04	15.60	4.40	5.40	13.91
	mmFormer	**9.52**	4.45	10.54	11.84	4.35	8.68	**7.76**	3.91	4.37	10.74	**4.03**	4.26	7.91	4.18	**4.15**	6.71
	ShaSpec	11.37	3.93	17.23	10.64	4.88	11.48	9.40	4.95	4.85	12.09	5.07	4.94	9.51	3.48	4.99	7.92
	DC-Seg	12.20	4.74	15.12	15.57	4.32	11.90	9.79	4.06	4.19	8.89	4.56	4.43	10.95	4.14	4.58	7.96
	UniME	11.94	8.07	15.70	12.05	4.12	10.61	11.38	**3.62**	3.84	7.87	4.04	**4.11**	10.21	4.17	4.15	7.72
	Ours	11.64	**3.85**	**9.44**	**7.70**	**4.06**	**8.23**	7.83	3.97	**2.58**	**7.58**	4.04	4.33	**7.02**	**2.62**	4.39	**5.95**

Table 9: Component ablation of HGMamba on BraTS 2018, reporting Dice (%) across all 15 modality combinations. ● and ○ denote available and missing modalities. Red: best; blue: second best; the full model is highlighted in green.

Type	Flair	●	○	○	○	●	●	●	○	○	○	●	●	●	○	●	Avg
	T1ce	○	●	○	○	●	○	○	●	●	○	●	●	○	●	●	
	T1	○	○	●	○	○	●	○	●	○	●	●	○	●	●	●	
	T2	○	○	○	●	○	○	●	○	●	●	○	●	●	●	●	
WT	w/o gate	88.50	76.77	76.65	85.10	89.84	89.73	**90.41**	80.93	87.79	85.87	90.09	91.00	**90.57**	87.82	90.83	86.79
	w/o nested	88.11	76.50	78.21	85.78	89.81	89.22	89.84	81.86	87.74	87.05	90.25	90.82	90.12	88.07	**90.91**	86.95
	w/o A_prior	88.28	77.78	73.15	**86.07**	89.70	88.37	90.38	79.58	87.58	86.33	89.29	90.69	89.70	87.21	90.28	86.29
	Full	**88.63**	**80.31**	**79.74**	85.86	**90.62**	**89.75**	90.21	**83.56**	**88.39**	**87.34**	**90.49**	**91.14**	90.19	**88.29**	90.81	**87.69**
TC	w/o gate	74.15	82.73	**71.62**	73.57	84.66	78.05	76.65	84.95	85.24	75.27	85.45	85.28	78.44	85.64	85.34	80.47
	w/o nested	74.94	83.45	71.34	73.15	85.25	**78.38**	77.78	84.87	85.16	76.80	85.44	85.43	79.12	85.61	**85.56**	80.82
	w/o A_prior	74.17	**84.56**	69.62	72.01	84.71	77.26	76.99	84.64	84.87	74.84	84.49	84.27	78.20	84.37	84.18	79.95
	Full	**75.05**	83.85	70.71	**74.95**	**85.73**	77.93	**78.90**	**85.02**	**85.46**	**77.09**	**85.55**	**85.77**	**79.39**	**85.74**	85.45	**81.11**
ET	w/o gate	44.51	74.97	39.19	**48.42**	74.80	46.48	49.21	74.98	75.71	48.32	74.09	74.60	50.02	75.10	73.78	61.61
	w/o nested	41.70	74.98	**43.81**	47.77	74.96	**47.19**	**49.79**	74.98	75.37	**50.18**	74.29	74.83	51.71	75.05	74.18	62.05
	w/o A_prior	43.84	73.72	37.24	45.37	73.91	44.66	49.02	74.00	74.84	47.33	73.73	74.14	50.01	74.30	73.74	60.66
	Full	**44.94**	**75.35**	34.89	45.68	**75.50**	45.28	49.42	**78.35**	**78.89**	50.04	**78.29**	**75.31**	**52.05**	**78.24**	**78.01**	**62.68**

is often best on individual TC combinations. On ET, NodeMamba gives the lowest HD95 in every combination that includes T1ce but not in most combinations without it, mirroring the Dice results.

A.4 Stability of the Affinity Prior

Table 12 lists the full affinity prior (Cohen's d) re-estimated under each intensity perturbation, together with the mean relative change with respect to the clean prior. The most discriminative modality per region (red) is preserved in every setting, and the relative change never exceeds 5%, confirming that the prior is robust to the intensity shifts that arise across scanners and protocols.

A.5 Per-Case Distribution of Dice and HD95

To look beyond averaged scores, Figure 8 shows the per-case distribution of Dice and HD95 across all 15 modality combinations on BraTS 2018, comparing HGMamba with the two strongest baselines (DC-Seg and UniME). Each ridge is a kernel density estimate over all case–combination sam-

Table 10: Interaction operator comparison on BraTS 2018, reporting Dice (%) across all 15 modality combinations. All three variants are matched in parameter count. ● and ○ denote available and missing modalities. Red: best; blue: second best; NodeMamba (ours) is highlighted in green.

Type	Flair	●	○	○	○	●	●	●	○	○	○	●	●	●	○	●	Avg
	T1ce	○	●	○	○	●	○	○	●	●	○	●	●	○	●	●	
	T1	○	○	●	○	○	●	○	●	○	●	●	○	●	●	●	
	T2	○	○	○	●	○	○	●	○	●	●	○	●	●	●	●	
WT	GAT	87.61	77.24	76.95	84.66	89.11	88.67	**90.23**	80.14	87.86	86.83	89.08	90.86	**90.21**	87.13	90.49	86.47
	Transformer	88.62	**80.47**	79.33	84.35	90.46	89.31	90.09	83.04	87.96	86.33	90.31	90.90	90.01	88.03	90.69	87.33
	NodeMamba (Ours)	**88.63**	80.31	**79.74**	**85.86**	**90.62**	**89.75**	90.21	**83.56**	**88.39**	**87.34**	**90.49**	**91.14**	90.19	**88.29**	**90.81**	**87.69**
TC	GAT	**75.05**	83.31	73.63	71.63	84.83	**78.80**	77.38	85.03	84.51	76.71	85.13	84.65	78.70	85.11	84.77	80.62
	Transformer	74.02	83.58	**74.53**	72.91	84.87	77.38	77.67	**85.53**	84.63	**78.10**	85.52	85.16	79.06	**85.80**	85.43	80.95
	NodeMamba (Ours)	75.05	**83.85**	70.71	**74.95**	**85.73**	77.93	**78.90**	85.02	**85.46**	77.09	**85.55**	**85.77**	**79.39**	85.74	**85.45**	**81.11**
ET	GAT	**46.17**	74.71	**43.75**	44.53	74.76	**51.78**	49.54	74.68	74.58	49.77	74.32	74.16	**53.31**	74.45	73.67	62.28
	Transformer	45.16	74.41	40.23	**47.77**	74.78	46.80	**50.64**	75.26	75.29	**50.18**	74.70	74.88	51.25	75.45	74.58	62.09
	NodeMamba (Ours)	44.94	**75.35**	34.89	45.68	**75.50**	45.28	49.42	**78.35**	**78.89**	50.04	**78.29**	**75.31**	52.05	**78.24**	**78.01**	**62.68**

Table 11: Interaction operator comparison on BraTS 2018, reporting HD95 (mm, lower is better) across all 15 modality combinations. ● and ○ denote available and missing modalities. Red: best; blue: second best; NodeMamba (ours) is highlighted in green.

Type	Flair	●	○	○	○	●	●	●	○	○	○	●	●	●	○	●	Avg
	T1ce	○	●	○	○	●	○	○	●	●	○	●	●	○	●	●	
	T1	○	○	●	○	○	●	○	●	○	●	●	○	●	●	●	
	T2	○	○	○	●	○	○	●	○	●	●	○	●	●	●	●	
WT	GAT	5.24	10.38	9.31	8.33	3.85	3.56	**2.92**	7.26	**4.21**	4.27	3.55	2.75	**2.84**	4.64	2.87	5.06
	Transformer	**3.65**	6.98	9.89	8.24	3.14	3.50	3.38	6.05	4.98	7.36	3.10	2.81	3.35	4.74	2.81	4.93
	NodeMamba (Ours)	4.67	**6.84**	**6.43**	**7.76**	**2.72**	**3.18**	2.97	**5.45**	4.22	**3.89**	**2.67**	**2.53**	3.01	**3.57**	**2.62**	**4.17**
TC	GAT	7.83	6.15	7.35	9.04	**4.44**	**5.33**	6.40	4.09	**4.23**	**5.85**	**4.22**	4.49	**5.26**	**3.86**	**4.30**	5.52
	Transformer	7.29	**5.11**	**7.04**	9.01	4.70	5.89	6.33	4.40	5.63	6.62	4.43	4.57	5.52	5.18	4.41	5.74
	NodeMamba (Ours)	**7.20**	5.24	8.14	**6.70**	4.89	5.97	**5.72**	**3.48**	4.28	6.50	4.93	**4.37**	5.56	4.14	4.48	**5.44**
ET	GAT	8.08	2.18	**7.17**	7.10	2.56	**5.82**	7.27	2.06	2.21	**5.48**	2.46	2.29	**5.29**	2.06	2.23	4.29
	Transformer	7.08	2.38	7.22	**7.03**	2.77	6.12	**6.32**	2.38	2.22	5.50	2.65	2.70	5.66	2.13	2.64	4.32
	NodeMamba (Ours)	**6.69**	**1.74**	8.36	7.13	**2.03**	6.67	6.53	**1.94**	**2.07**	6.31	**1.92**	**2.09**	6.36	**1.92**	**1.94**	**4.25**

ples, and the dashed line marks the mean. HGMamba attains a Dice comparable to the strongest competitor, while for HD95 it consistently shows both the lowest mean and the smallest spread on all three regions: its distribution is concentrated near small values with a short tail, whereas UniME exhibits a heavy tail of large boundary errors. This indicates that HGMamba is not only accurate on average but also stable across cases, rarely producing the large boundary failures seen in the baselines.

A.6 Robustness to the Number of Available Modalities

Figure 9 groups the per-case scores by the number of available modalities (1 to 4). As fewer modalities are available, every method degrades, but HGMamba degrades the least on the boundary metric: under a single available modality its HD95 stays clearly the lowest on all regions (e.g., on the whole tumor, 6.4 mm versus 7.4 mm for DC-Seg and 16.3 mm for UniME), and its boxes remain short while UniME's grow tall and long. This shows that the advantage of HGMamba is largest exactly where the missing-modality problem is hardest.

A.7 Feature-Space Visualization

To inspect what the fused representation encodes, we take the feature map just before the final classifier and run t-SNE on voxel features sampled per tissue class (background, NCR/NET, edema, enhancing tumor). Figure 10 shows the eight modality subsets with the clearest class separation, ranked by a silhouette score on the tumor classes. Two observations stand out. First, the four tissue classes form well-separated clusters, confirming that the fused feature is discriminative rather than collapsed. Second, the subsets that separate best all contain T1ce, which agrees with the learned

Table 12: Stability of the affinity prior under intensity perturbations. Relative change is the mean absolute change of $\boldsymbol{A}_{\text{prior}}$ against the clean estimate. The most discriminative modality per region (WT: FLAIR, TC/ET: T1ce) is preserved in every case.

Perturbation	Rel. change	Top modality preserved		
	(%)	WT	TC	ET
Clean	–	FLAIR	T1ce	T1ce
Gamma (0.8)	3.85	✓	✓	✓
Gamma (1.2)	4.24	✓	✓	✓
Bias field	4.80	✓	✓	✓
Noise (0.05)	0.28	✓	✓	✓
Noise (0.10)	1.08	✓	✓	✓

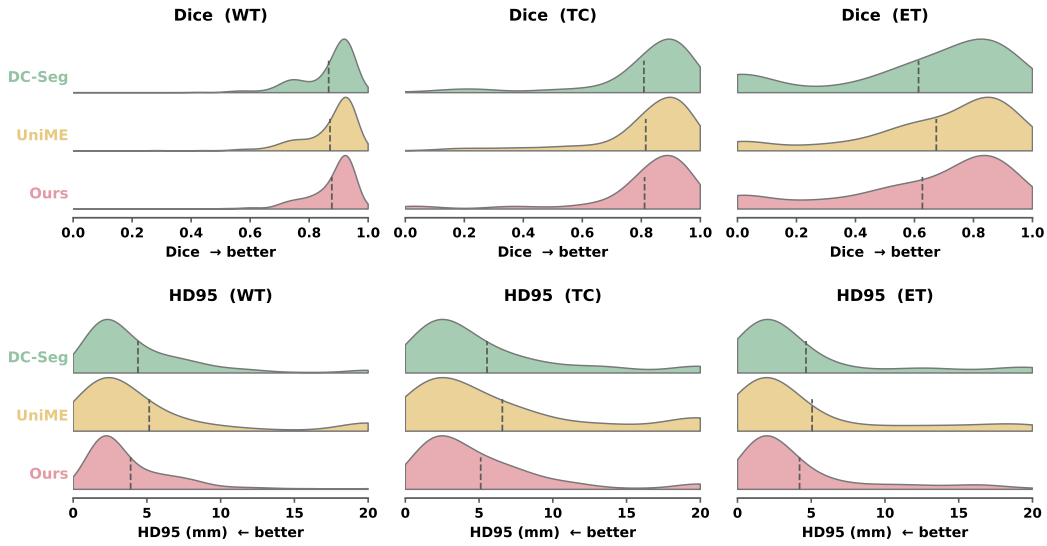

Figure 8: Per-case distribution (kernel density) of Dice (↑, top) and HD95 (↓, bottom) over all 15 modality combinations on BraTS 2018. Dashed lines denote the mean. HGMamba (rose) yields the lowest mean HD95 with the smallest variance on every region, while UniME (gold) shows a long tail of large boundary errors.

affinity prior that assigns T1ce the highest discriminative weight for the enhancing tumor. The feature geometry therefore mirrors the region-aware routing our model is designed to perform.

A.8 Cross-Domain Generalization

To probe robustness beyond the training distribution, we evaluate all BraTS2018-trained models on MOTUM (Gong et al., 2024), a multi-center collection acquired on different scanners and protocols, under two shifts: gliomas from unseen scanners (cross-center, n=29) and brain metastases (cross-disease, n=35), the latter differing sharply in morphology from the training gliomas. Each model is run zero-shot with no fine-tuning. MOTUM annotates two regions derived from FLAIR and contrast-enhanced T1, so we report whole tumor (WT) and enhancing tumor (ET); volumes are resampled to $1\,\mathrm{mm}$ isotropic and normalized as in training. Table 13 summarizes the results. On both shifts HGMamba attains the best WT Dice and the best HD95 on WT and ET, and is within $0.05\,\mathrm{mm}$ of the best ET HD95 on gliomas. Under the harder metastasis shift every method drops, yet HGMamba keeps the highest whole-tumor overlap and the lowest boundary error, so its boundary advantage transfers to an out-of-distribution tumor type. The one exception is ET Dice, which sits within the baseline range rather than leading it, while its ET HD95 stays best; this matches our overall finding that HGMamba draws sharper boundaries at comparable overlap. The gain indicates that

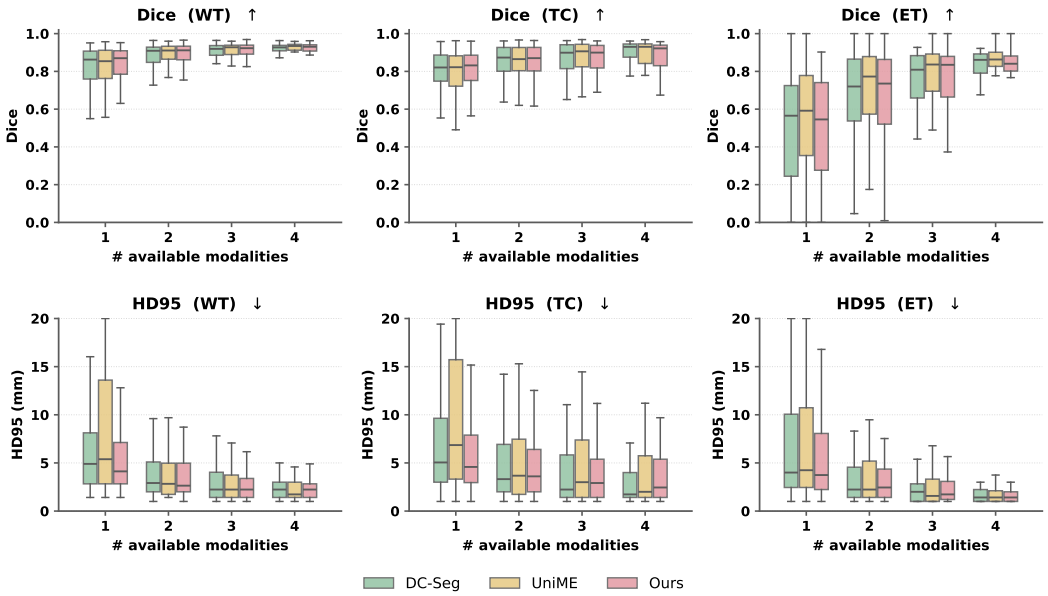

Figure 9: Dice (↑, top) and HD95 (↓, bottom) grouped by the number of available modalities on BraTS 2018. The fewer the available modalities, the larger the gap in HD95 in favour of HGMamba, whose boxes stay short and low while UniME's spread widens sharply.

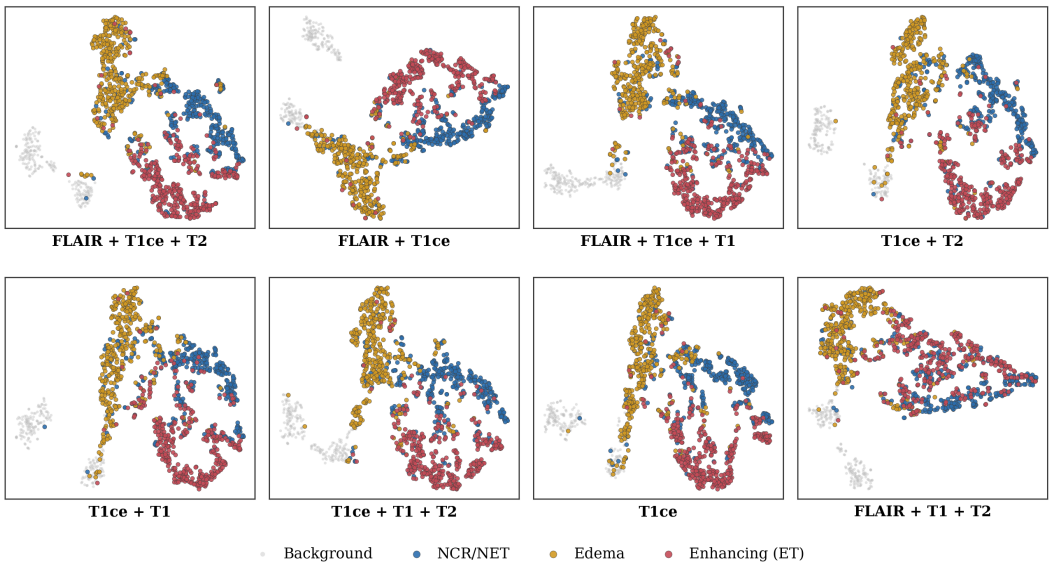

Figure 10: t-SNE of the fused feature (pre-classifier) on BraTS 2018, colored by tissue class. The eight modality subsets with the highest tumor-class separability are shown, ordered by silhouette score. Subsets containing T1ce give the cleanest separation, consistent with the affinity prior weighting T1ce most for the enhancing tumor.

the affinity-prior routing keeps modality-to-region assignment stable when intensities shift, which is where competing fusion schemes degrade most.

A.9 Efficiency and Complexity

Table 14 lists the full efficiency measurements behind Figure 4. All models are evaluated on the same GPU with the same input size under the same inference setting.

Table 13: Zero-shot cross-domain segmentation on MOTUM. All models are trained on BraTS2018 and evaluated without fine-tuning. Only WT and ET are annotated in MOTUM. Best in **red**, second in blue; our row is shaded.

Method	WT		ET	
	Dice↑	HD95↓	Dice↑	HD95↓
Glioma (cross-center, n=29)				
U-HVED	0.740	32.28	0.506	18.88
ShaSpec	0.806	20.37	0.556	15.66
mmFormer	0.797	14.06	0.556	16.62
DC-Seg	0.785	22.51	**0.577**	**15.22**
UniME	0.793	21.89	0.570	**15.22**
Ours	**0.842**	**7.69**	0.538	15.27
Metastasis (cross-disease, n=35)				
U-HVED	0.604	36.67	0.510	25.73
ShaSpec	0.731	21.58	0.497	23.71
mmFormer	0.709	20.71	**0.535**	17.18
DC-Seg	0.689	22.23	0.513	20.13
UniME	0.731	23.91	0.515	18.86
Ours	**0.766**	**16.63**	0.489	**15.74**

Table 14: Efficiency and complexity comparison on BraTS 2018: inference time, peak GPU memory, parameters, and FLOPs.

Method	Time (ms)	Memory (GB)	Params (M)	FLOPs (G)
U-HVED	355.6	1.31	1.25	61.7
ShaSpec	587.8	2.34	6.62	414.0
mmFormer	1088.9	1.80	36.39	123.8
DC-Seg	1883.3	5.90	40.04	280.7
UniME	1294.1	2.11	158.12	459.3
Ours	612.6	2.13	8.74	202.9

Parameters and memory. HGMamba uses 8.74M parameters, about 18× fewer than UniME (158.12M) and 4.6× fewer than DC-Seg (40.04M), while matching the Dice of UniME and achieving the lowest HD95 among all methods. Its peak memory (2.13 GB) is close to that of UniME and ShaSpec and less than half that of DC-Seg (5.90 GB). Most parameters sit in the four lightweight modality-specific encoders; the heterogeneous graph fusion adds only 0.31M parameters (3.5% of the model), because it operates on a handful of pooled nodes rather than on voxels, so NodeMamba scans the modality sequence and the region sequence separately, each with at most four nodes, and its cost is negligible.

Speed and FLOPs. HGMamba runs 2.1× faster than UniME and 3.1× faster than DC-Seg, and requires fewer FLOPs than ShaSpec, DC-Seg, and UniME. HGMamba is not the cheapest model on every axis: U-HVED and ShaSpec are faster, U-HVED and ShaSpec have fewer parameters, and U-HVED and mmFormer need fewer FLOPs. All three, however, trail HGMamba by more than 2% in mean Dice and 1 mm in HD95 (Tables 1 and 2).

Discussion. HGMamba therefore offers the best accuracy for its cost rather than the lowest cost overall. We present this as a favorable accuracy-cost trade-off and not as a capacity advantage: a lighter backbone lowers the expressive ceiling, and the fusion design is compatible with heavier encoders, which we leave to future work.

B THEORETICAL JUSTIFICATION OF THE AFFINITY PRIOR

We formalize the two claims used in Section 3.3: that $d_{r,m}$ measures region-discriminative information, and that adding it to the attention logits is the Bayes-consistent way to inject it.

Setup. Fix a region r and a modality m. Let $y_r \in \{\text{in}, \text{out}\}$ indicate membership of a voxel in r versus its nested complement, with equal class priors $\Pr(y_r = \text{in}) = \frac{1}{2}$, and let the intensity of modality m be

$$x_m \mid y_r{=}\text{in} \;\sim\; \mathcal{N}(\mu^+_{r,m}, \sigma^2_{r,m}), \qquad x_m \mid y_r{=}\text{out} \;\sim\; \mathcal{N}(\mu^-_{r,m}, \sigma^2_{r,m}). \tag{A1}$$

The equal-variance form is used for a clean statement. Under (A1) the pooled standard deviation in equation 10 reduces to $\sigma_{r,m}$, so Cohen's d coincides with the detectability index,

$$d_{r,m} = \frac{|\mu^+_{r,m} - \mu^-_{r,m}|}{\sqrt{\big((\sigma^+_{r,m})^2 + (\sigma^-_{r,m})^2\big)/2}} = \frac{|\mu^+_{r,m} - \mu^-_{r,m}|}{\sigma_{r,m}}. \tag{A2}$$

In the general unequal-variance case, $d_{r,m}$ keeps the pooled denominator of equation 10.

Proposition 1 (Cohen's d ranks discriminative information). *Under (A1), the Bayes error of the membership test from x_m is $\Phi(-d_{r,m}/2)$, and the mutual information $I(x_m; y_r)$ is a function of $d_{r,m}$ alone that is strictly increasing in it. Consequently $d_{r,m}$ is a strictly monotone, parameter-free proxy for $I(x_m; y_r)$, and*

$$\arg\max_m d_{r,m} \;=\; \arg\max_m I(x_m; y_r).$$

Proof. Write $\Delta\mu = \mu^+_{r,m} - \mu^-_{r,m}$, $\bar\mu = (\mu^+_{r,m} + \mu^-_{r,m})/2$ and $\sigma = \sigma_{r,m}$. Under (A1) with equal priors, the log-likelihood ratio

$$\log \frac{p(x_m \mid \text{in})}{p(x_m \mid \text{out})} = \frac{\Delta\mu}{\sigma^2}\,\big(x_m - \bar\mu\big)$$

is affine and monotone in x_m, so the Bayes rule thresholds x_m at $\bar\mu$ and its error is

$$\Phi\!\left(-\frac{|\Delta\mu|}{2\sigma}\right) = \Phi\!\left(-\frac{d_{r,m}}{2}\right),$$

which is strictly decreasing in $d_{r,m}$ by (A2). For the mutual information, let $\eta(x) = \Pr(y_r{=}\text{in} \mid x)$. Bayes' rule with equal priors gives

$$\eta(x) = \sigma_{\text{logit}}\!\left(\frac{\Delta\mu}{\sigma^2}\,(x - \bar\mu)\right), \qquad I(x_m; y_r) = \log 2 - \mathbb{E}_{x_m}\big[H_b(\eta(x_m))\big],$$

where H_b is the binary entropy. Substituting $u = (x - \bar\mu)/\sigma$ makes both the logistic slope and the sampling distribution of u depend on $(\mu^\pm, \sigma)$ only through $d_{r,m}$. Increasing $d_{r,m}$ sharpens η toward $\{0, 1\}$, which lowers $\mathbb{E}[H_b(\eta)]$ and raises I. Strict monotonicity gives the stated $\arg\max$ identity. □

Proposition 2 (Additive injection is Bayes-consistent routing). *Let $\boldsymbol{s}_r \in \mathbb{R}^M$ be the learned attention logits for region r, let $\boldsymbol{a}_r = \boldsymbol{A}_{\text{prior}}[r, \cdot]$, and define $\boldsymbol{\Pi}_r = \text{softmax}(\boldsymbol{s}_r + \gamma\,\boldsymbol{a}_r)$. If $\exp(\boldsymbol{s}_r) \propto p_\theta(m \mid \cdot)$ is read as learned evidence and $\exp(\gamma\boldsymbol{a}_r) \propto \pi_\gamma(m \mid r)$ as a prior, then*

$$\Pi_r(m) \;\propto\; p_\theta(m \mid \cdot)\,\pi_\gamma(m \mid r)$$

is the posterior of a product of experts. Additive injection is the log-linear combination that keeps $\boldsymbol{\Pi}_r$ a normalized categorical distribution for any $\boldsymbol{a}_r \in \mathbb{R}^M$ and leaves the logit gradient $\partial(\boldsymbol{s}_r + \gamma\boldsymbol{a}_r)/\partial\boldsymbol{s}_r = \boldsymbol{I}$ unmodulated by the prior. A multiplicative logit rule $\text{softmax}\big((\mathbf{1} + \gamma\boldsymbol{a}_r) \odot \boldsymbol{s}_r\big)$ rescales this gradient to $\text{diag}(\mathbf{1} + \gamma\boldsymbol{a}_r)$, and a probability-reweighting rule $\boldsymbol{\Pi}_r \propto \text{softmax}(\boldsymbol{s}_r) \odot \boldsymbol{a}_r$ requires $\boldsymbol{a}_r \geq \mathbf{0}$ and explicit renormalization and has no posterior interpretation.

Proof. For a categorical variable the softmax is the maximum-entropy distribution whose log-probabilities are linear in the input features. Adding the feature $\gamma\boldsymbol{a}_r$ to the logits multiplies the unnormalized distribution by $\exp(\gamma\boldsymbol{a}_r)$, and the normalizer $\sum_m \exp(s_{r,m} + \gamma a_{r,m})$ restores $\sum_m \Pi_r(m) = 1$, giving the product-of-experts posterior. The Jacobian claims follow by differentiating the two logit maps with respect to $\boldsymbol{s}_r$. The reweighting rule is normalized only after dividing by $\sum_m \text{softmax}(\boldsymbol{s}_r)_m\, a_{r,m}$, which is undefined when $\boldsymbol{a}_r$ has mixed signs and, being a function of $\boldsymbol{s}_r$, couples the prior to the learned logits. □

Remark. Proposition 1 assumes Gaussian intensities to obtain a closed-form link between $d_{r,m}$ and $I(x_m; y_r)$; it justifies using $d_{r,m}$ as a discriminability score rather than making a claim about final segmentation accuracy. Proposition 2 is distribution-free and explains why the additive form of equation 5 is preferable to multiplicative alternatives.