Virtual Nodes Guided Dynamic Graph Neural Network for Brain Tumor Segmentation with Missing Modalities

Supplementary Material

1. Training Efficiency

Method	Memory Usage (MiB)	Time per Step (s)
MMFormer	11143	2.34
M3AE (pretrain)	9664	0.66
M3AE (train)	13208	2.16
Ours	10565	2.28

Table 1. Comparison of memory usage and time per step for different methods.

As shown in Tab 1, our method achieves lower memory consumption and comparable per-step training time compared to MMFormer. In contrast to M3AE, which involves a separate pretraining stage and higher peak memory during finetuning, our approach is more streamlined and efficient. Note that M3AE's two-stage training incurs additional overhead, while our single-stage pipeline simplifies the training process without sacrificing performance.

2. More Metrics

HD95↓	ET	TC	WT	Avg
mmF	7.71	7.74	6.99	7.48
M3AE	5.80	8.40	**6.90**	7.03
Ours	**4.79**	**6.17**	7.01	**5.99**
ASSD↓	**ET**	**TC**	**WT**	**Avg**
mmF	2.17	2.71	2.59	2.49
M3AE	1.76	2.96	**2.09**	2.27
Ours	**1.73**	**2.32**	2.25	**2.10**

Table 2. Comparison of HD95 and ASSD scores (lower is better) on the ET, TC, and WT regions. The reported results represent the average performance across all possible modality combinations.

As shown in Tab 2, our method performs well on ET and TC but is slightly less effective on WT. This is because the model, guided by our missing modality design (e.g., virtual nodes and dynamic connections), tends to prioritize the T1c modality, which is crucial for ET and TC. In contrast, WT relies more on the Flair modality, which receives less emphasis during fusion. Nevertheless, we believe this trade-off is reasonable and plan to explore better task balancing in future work.

Method	ET	TC	WT
single	63.2	79.8	86.0
channel	**66.0**	**81.4**	**86.7**

Table 3. Comparison of single and channel methods on ET, TC, and WT. The reported results represent the average performance across all possible modality combinations.

3. Node Construction Strategy

One concern raised is the rationale behind treating each feature channel as an individual node, instead of using the entire feature representation as a single node. In convolutional neural networks (CNNs), different channels are known to encode distinct semantic cues, such as edges, textures, or specific anatomical structures [5]. By treating each channel as a node, our graph explicitly preserves this semantic granularity, enabling more fine-grained and interpretable feature interaction across nodes. This design enhances the graph's capacity to model complex, high-order dependencies between semantic concepts, as shown in Tab 3. In contrast, collapsing the entire feature map into a single node would obscure these distinctions, effectively reducing the graph to a simple MLP-like structure and weakening its representational power.

4. More experiments

Method	ET	TC	WT	Avg
MFI	60.2	77.0	86.1	74.4
RfNet	61.5	78.2	87.0	75.6
Scratch	63.5	78.3	**87.2**	76.3
Hyper-GAE	63.9	79.3	86.7	76.6
Ours	**66.0**	**81.4**	86.7	**78.0**

Table 4. Comparative results with state-of-the-art methods. The reported results represent the average performance across all possible modality combinations.

We present comparative results with additional state-of-the-art methods, as shown in Tab 4. The compared methods include MFI [6], RfNet [1], Scratch [3], and Hyper-GAE [4]. Among them, MFI and Hyper-GAE are graph-like approaches, while RfNet and Scratch are representative single-stage methods. Our method consistently achieves strong performance on the ET and TC regions, while showing slightly lower scores on WT. This trend is in line with

observations from other metrics, and further highlights the model's effectiveness in capturing tumor subregions that are critical for clinical decision-making. Despite the relatively lower WT performance, our approach maintains competitive overall accuracy, demonstrating the robustness of the proposed design. Future work will focus on enhancing the balance across different tumor regions without compromising the model's strengths.

In addition, compared with the aforementioned methods—including graph-like models and single-stage baselines based on CNNs and Transformers—our approach achieves competitive results. It is worth noting that, despite its relatively simple and lightweight design, our method effectively captures cross-modal relationships through a genuine graph-based structure. This strong performance, achieved without the need for elaborate fusion strategies or architectural complexity, highlights the efficiency and generalizability of our approach. Moreover, its modular nature allows for straightforward integration into existing frameworks, demonstrating its practicality in real-world multimodal segmentation tasks.

Method	**ET**	**TC**	**WT**	**Avg**
ResUNet+	92.7	92.3	90.1	91.7
mmF	77.6	85.8	89.6	84.3
M3AE	75.5	84.5	90.1	83.4
MMC	80.1	**87.4**	89.0	85.5
Ours	**80.8**	87.3	**90.3**	**86.1**

Table 5. Comparison of results under the full-modality setting. ResUNet+ is a specialized method specifically designed for this scenario.

Compared with full-modality-specific models such as ResUNet+ [2], our method exhibits lower performance under the full-modality setting, as shown in Tab 5. A performance gap under full-modality is inevitable when designing for missing modalities—this trade-off is intrinsic to the problem. To mitigate it, we adopt several targeted strategies: (1) separate encoders ensure modality-specific feature extraction; (2) modality dropout is applied only during graph construction, preserving full-modality inputs at the encoder stage; and (3) a modality-specific decoder refines the output.

These designs ensure that our model maintains strong performance in the full-modality setting while remaining robust and flexible in missing-modality scenarios. In fact, compared with other methods tailored for incomplete inputs, our method achieves superior results under complete modality inputs, demonstrating its well-balanced design.

References

[1] Yuhang Ding, Xin Yu, and Yi Yang. Rfnet: Region-aware fusion network for incomplete multi-modal brain tumor segmentation. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 3975–3984, 2021. 1

[2] Sedat Metlek and Halıt Çetıner. Resunet+: A new convolutional and attention block-based approach for brain tumor segmentation. *IEEE Access*, 11:69884–69902, 2023. 2

[3] Yansheng Qiu, Delin Chen, Hongdou Yao, Yongchao Xu, and Zheng Wang. Scratch each other's back: Incomplete multi-modal brain tumor segmentation via category aware group self-support learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 21317–21326, 2023. 1

[4] Heran Yang, Jian Sun, and Zongben Xu. Learning unified hyper-network for multi-modal mr image synthesis and tumor segmentation with missing modalities. *IEEE Transactions on Medical Imaging*, 42(12):3678–3689, 2023. 1

[5] Matthew D Zeiler and Rob Fergus. Visualizing and understanding convolutional networks. In *European conference on computer vision*, pages 818–833. Springer, 2014. 1

[6] Zechen Zhao, Heran Yang, and Jian Sun. Modality-adaptive feature interaction for brain tumor segmentation with missing modalities. In *International conference on medical image computing and computer-assisted intervention*, pages 183–192. Springer, 2022. 1