Uni-Encoder Meets Multi-Encoders: Representation Before Fusion for Brain Tumor Segmentation with Missing Modalities

Supplementary Material

This supplementary material details data preprocessing and inference settings, reports additional metrics on BraTS 2023 and BraTS 2024, presents extended ablation studies on crop and patch sizes, analyzes model complexity, and discusses limitations of UniME.

A. Data Preprocessing

We follow standard BraTS preprocessing practices used in recent work [16, 24, 45, 68]. We load the four MRI modalities (FLAIR, T1ce, T1, and T2) and the associated segmentation labels from the original BraTS dataset. We do not resample the voxel grid or alter spatial coordinates. We compute a brain mask as the union of non-zero voxels across modalities and apply z-score normalization per modality and per case within this mask.

For comparability with previous work [36, 67, 68] that trains with $128 \times 128 \times 128$ crop size, and because a few volumes have a dimension less than 128, we apply symmetric padding along the affected axis to 128 voxels. The impact of this padding is empirically negligible. For the BraTS 2023 dataset, only one case has a dimension of 127 voxels, so padding does not alter gross tumor morphology or bias evaluation.

BraTS 2024 corresponds to a post-operative glioma segmentation setting and introduces an additional annotation for the resection cavity, which is not present in the standard pre-operative BraTS datasets. To ensure compatibility with prior incomplete-modality segmentation methods and maintain consistent evaluation across BraTS 2023 and BraTS 2024, the resection cavity is merged into the enhancing tumor class during preprocessing. Then, the evaluation regions therefore follow the standard BraTS definition: the whole tumor (WT) includes all abnormal regions, the tumor core (TC) includes necrotic or non-enhancing tumor together with enhancing tumor, and the enhancing tumor (ET) region corresponds to the merged enhancing tumor category. Under this protocol, the enhancing tumor region in BraTS 2024 effectively represents the union of enhancing tumor and resection cavity.

B. Inference Details

At inference, we use a 3D sliding-window strategy with overlap. For each model, the window size equals the training crop size, and the stride is half the window length, yielding 50% overlap. Per-window logits are passed through softmax and accumulated as spatial probability maps. In overlapping regions, probabilities are averaged across overlapping windows, and the final voxel label is obtained via argmax.

To suppress false positives in the enhancing tumor region, we apply a standard post-processing rule [16, 24, 27, 45, 68]. If the predicted ET volume contains fewer than 500 voxels, we set the corresponding ET mask to zero. For completeness, Tab. 8 and Tab. 9 report ET segmentation performance on BraTS 2023 and BraTS 2024 before and after this post-processing.

C. Extended Comparisons

In addition to the DSC comparisons in Tab.1 and Tab.2, we also evaluate boundary quality using the 95th percentile Hausdorff Distance (HD95), which provides a robust measure of boundary accuracy. Specifically, for two 3D segmentation surface point sets ∂A and ∂B, the directed 95th percentile Hausdorff distance is defined as:

$$h_{95}(A, B) = \mathcal{P}_{95\%}\{\min_{b \in \partial B} \|a - b\|_2 : a \in \partial A\}, \quad (8)$$

where $\mathcal{P}_{95\%}$ denotes the 95th percentile and $\|\|_2$ the L_2 norm. The HD95 between A and B is defined as:

$$\text{HD95} = \max(h_{95}(A, B), h_{95}(B, A)). \quad (9)$$

We compute HD95 for the WT, TC, and ET using the DisTorch implementation [42].

Tab. 10 and Tab. 12 report HD95 results on BraTS 2023 and BraTS 2024, respectively. Across both datasets, UniME consistently outperforms competing methods on WT and ET in HD95. Performance on TC is very close to the best-performing method. Overall, UniME accurately characterizes lesion boundaries, as reflected by HD95.

D. Extended Ablation Studies

Effect of Crop Size. Transformers capture long-range dependencies, and previous work in general medical image segmentation indicates that increasing the input resolution can improve segmentation performance [9]. Therefore, we conduct ablations on crop size at 80, 96, and 128. Tab. 13 reports the detailed results.

As shown in Tab. 13, a crop size of 96 yields the best overall performance, balancing context coverage and computational efficiency. Larger crops improve WT accuracy, indicating that UniME benefits from additional

Type																	Avg.
	Flair	○	○	○	●	○	○	●	○	●	●	●	●	●	○	●	
	T1	○	○	●	○	○	●	●	●	○	○	●	●	○	●	●	
	T1ce	○	●	○	○	●	●	○	○	○	●	●	○	●	●	●	
	T2	●	○	○	○	●	○	○	●	●	○	○	●	●	●	●	
ET	HeMIS	4.96	30.62	5.91	18.92	57.34	46.31	18.09	7.63	21.47	65.83	68.14	18.48	68.17	60.23	67.41	37.30
	U-HEVD	35.96	71.10	22.41	36.63	75.89	73.97	41.62	39.12	45.07	76.27	77.08	45.84	77.15	76.75	77.49	58.16
	RFNet	46.89	79.36	38.73	40.97	80.56	80.82	48.62	50.45	55.10	81.10	82.20	55.96	81.27	81.08	81.59	65.65
	mmFormer	46.19	78.64	38.70	41.46	80.75	80.84	47.01	48.96	51.54	81.35	81.82	53.64	81.28	81.23	81.48	64.99
	M^3AE	58.68	82.61	54.02	57.17	83.93	83.13	60.73	60.35	63.09	84.26	84.63	63.52	84.56	83.65	84.62	72.60
	M^2FTrans	53.01	82.12	44.80	48.49	83.06	83.11	54.22	55.29	58.48	83.28	83.56	58.88	83.33	83.28	83.40	69.22
	IM-Fuse	50.57	81.30	48.19	51.15	82.00	82.31	55.00	54.78	57.18	82.51	82.95	58.61	82.25	82.41	82.50	68.91
	M^2SegMamba	54.06	81.32	50.29	54.44	82.95	83.09	59.13	57.92	60.28	82.95	83.48	62.13	83.07	83.40	83.44	70.80
	Ours	**59.16**	**84.15**	**54.24**	**58.69**	**85.55**	**84.29**	**62.16**	**61.07**	**63.83**	**84.71**	**84.89**	**64.90**	**85.40**	**85.41**	**85.51**	**73.60**
ET*	HeMIS	4.96	30.62	5.91	18.92	57.34	46.31	18.89	7.60	21.46	65.83	68.54	18.76	68.56	60.23	67.74	37.44
	U-HEVD	35.95	71.10	22.41	36.63	75.89	73.96	42.02	38.98	45.07	76.27	77.48	46.24	77.50	77.11	78.67	58.35
	RFNet	48.87	80.51	39.40	41.63	81.76	81.62	49.22	51.44	55.80	82.70	84.20	57.05	82.47	83.08	83.59	66.89
	mmFormer	46.59	79.84	39.05	41.98	82.35	82.79	47.09	49.67	52.50	82.90	83.35	54.39	82.88	83.23	83.08	66.11
	M^3AE	58.68	83.95	54.00	57.09	84.79	85.12	60.90	60.25	63.03	84.90	85.80	63.50	85.43	85.23	85.74	73.23
	M^2FTrans	54.58	84.04	45.79	49.81	85.03	85.51	55.34	56.60	60.18	84.87	85.56	60.38	84.90	85.64	85.78	70.93
	IM-Fuse	51.74	82.85	48.80	51.74	84.40	84.30	56.04	56.68	58.71	84.51	85.34	60.94	84.25	84.81	84.90	70.67
	M^2SegMamba	54.84	81.72	51.09	55.24	84.55	83.49	59.48	58.69	60.59	84.53	84.67	62.37	84.27	85.40	85.03	71.73
	Ours	**61.12**	**86.95**	**55.35**	**60.00**	**87.54**	**87.40**	**63.15**	**62.11**	**65.11**	**87.59**	**87.94**	**65.93**	**87.80**	**87.87**	**87.98**	**75.59**

Table 8. **Performance comparison of ET region segmentation (DSC %) on BraTS 2023.** Available and missing modalities are denoted by ● and ○, respectively. ET represents the enhancing tumor region without post-processing, while ET* indicates results after applying post-processing. , , and indicate the first-, second-, and third-best performance, respectively. WT, TC, and ET represent whole tumor, tumor core, and enhancing tumor, respectively.

Type																	Avg.
	Flair	○	○	○	●	○	○	●	○	●	●	●	●	●	○	●	
	T1	○	○	●	○	○	●	●	●	○	○	●	●	○	●	●	
	T1ce	○	●	○	○	●	●	○	○	○	●	●	○	●	●	●	
	T2	●	○	○	○	●	○	○	●	●	○	○	●	●	●	●	
ET	HeMIS	15.76	26.47	8.68	1.12	41.52	34.96	15.65	23.69	26.04	40.02	40.82	24.03	46.25	41.10	44.30	28.69
	U-HEVD	39.93	54.21	47.12	34.97	63.32	62.44	49.10	51.77	46.12	66.85	69.19	51.66	69.06	67.13	70.42	56.22
	RFNet	56.49	71.94	55.26	51.93	74.42	74.59	60.02	59.88	60.31	75.91	76.90	61.92	76.90	75.59	77.52	67.30
	mmFormer	57.57	72.00	55.28	52.76	75.77	75.10	61.45	62.66	62.80	76.61	78.14	64.35	78.07	77.59	79.30	68.63
	M^3AE	63.44	73.61	61.44	60.03	78.32	76.88	64.86	65.43	66.74	77.87	80.41	67.24	80.42	78.70	80.86	71.75
	M^2FTrans	56.97	73.48	58.11	55.70	76.89	76.80	62.07	62.03	62.15	77.78	79.09	64.39	78.87	77.99	79.45	69.45
	IM-Fuse	60.71	74.14	59.02	55.12	77.16	76.97	64.75	64.24	65.17	78.61	80.14	67.72	80.35	77.89	80.44	70.83
	M^2SegMamba	59.61	74.67	60.13	57.43	78.11	77.27	64.40	64.27	65.54	78.90	80.45	66.98	79.50	79.03	81.20	71.17
	Ours	**64.75**	**77.56**	**65.12**	**63.15**	**80.75**	**79.87**	**69.30**	**69.47**	**69.73**	**81.11**	**81.90**	**71.55**	**83.04**	**81.19**	**83.17**	**74.78**
ET*	HeMIS	16.08	26.47	8.62	2.15	41.52	34.96	15.59	24.06	26.38	40.02	41.51	24.71	46.57	41.10	44.98	28.98
	U-HEVD	39.93	54.21	47.12	34.97	63.24	62.44	49.81	51.74	46.06	67.22	69.93	52.31	69.06	67.12	70.79	56.40
	RFNet	57.23	73.53	57.04	53.70	75.06	76.25	61.37	60.60	61.18	77.14	78.59	63.73	77.67	76.20	78.75	68.54
	mmFormer	58.54	73.29	56.63	53.53	77.49	76.55	61.92	63.25	62.57	77.17	78.92	65.42	79.18	78.64	80.24	69.56
	M^3AE	63.55	73.98	61.92	60.08	79.04	77.53	64.91	65.46	66.89	78.30	80.81	67.34	81.03	79.82	82.12	72.19
	M^2FTrans	58.47	75.19	58.51	57.33	78.23	77.71	63.51	63.16	63.76	79.48	80.83	65.97	80.36	79.31	81.25	70.87
	IM-Fuse	61.36	76.02	61.01	55.87	77.99	78.33	64.59	64.87	65.06	79.28	80.43	67.08	80.01	78.94	80.77	71.44
	M^2SegMamba	60.66	75.19	60.81	58.50	78.51	77.79	65.04	64.53	65.59	79.24	80.56	67.66	80.32	79.44	81.33	71.68
	Ours	**66.00**	**78.57**	**66.04**	**64.40**	**80.99**	**80.34**	**69.58**	**69.43**	**69.73**	**81.75**	**82.41**	**70.89**	**82.30**	**81.58**	**82.81**	**75.12**

Table 9. **Performance comparison of ET region segmentation (DSC %) on BraTS 2024.** Available and missing modalities are denoted by ● and ○, respectively. ET represents the enhancing tumor region without post-processing, while ET* indicates results after applying post-processing. , , and indicate the first-, second-, and third-best performance, respectively. WT, TC, and ET represent whole tumor, tumor core, and enhancing tumor, respectively.

global context. However, larger crops degrade performance on smaller, more intricate substructures such as TC and ET. Specifically, TC and ET peak at a crop size of 96, while larger crops reduce accuracy.

Effect of Patch Size. Patch size governs the Uni-Encoder's ability to model fine structures. Reducing the patch size significantly increases the sequence length, enabling the Uni-Encoder to model finer structures, but simultaneously leads to rapidly rising computational complexity. We conduct ablations on patch size at 8, 12, and 16. Tab. 14 reports the results.

A patch size of 8 achieves the best overall performance, indicating that finer tokenization better captures the fine-grained structure of the brain tumor. Increasing the patch size to 12 or 16 degrades accuracy across all regions, as larger patches lose critical spatial detail. Therefore, we select the smallest patch size feasible under our current memory budget. If the patch size were reduced to 4, the token sequence length would approach 14k tokens, significantly increasing computational over-

Type	Flair	T1	T1ce	T2	HeMIS	U-HEVD	RFNet	mmFormer	M^{3}AE	M^{2}FTrans	LS3M	IM-Fuse	M^{2}SegMamba	Ours	Avg.
WT	○	○	○	●	81.36	48.40	31.14	19.61	8.62	14.08	13.46	21.11	13.25	**7.40**	65.58
	○	○	●	○	75.92	39.78	27.21	28.91	12.09	14.69	13.10	13.63	12.33	**9.57**	29.03
	○	●	○	○	73.24	52.47	24.27	20.74	12.34	18.20	13.39	14.30	13.59	**9.28**	24.27
	○	●	○	○	71.03	42.74	22.79	21.05	8.82	11.74	11.96	9.64	7.63	**5.75**	20.97
	○	●	○	●	72.54	28.02	23.29	18.39	7.50	10.49	9.90	10.36	8.93	**6.26**	16.72
	○	●	●	○	69.19	34.52	22.82	18.45	11.07	12.81	11.44	12.20	9.53	**8.28**	17.05
	●	○	○	●	63.02	22.85	14.66	12.75	6.50	8.30	7.79	8.22	5.90	**5.30**	11.16
	●	○	○	○	65.25	27.84	21.27	16.80	7.85	9.77	9.61	12.03	8.21	**6.14**	16.20
	○	●	○	○	68.50	27.30	20.20	12.36	6.27	8.84	8.31	11.34	6.89	**5.37**	10.93
	●	○	●	○	65.00	25.11	18.23	11.62	5.64	9.55	9.51	6.53	6.40	**4.83**	8.45
	●	●	○	○	56.92	16.90	15.00	12.78	5.85	10.14	7.23	7.48	6.04	**5.34**	6.39
	●	●	○	●	53.89	17.82	16.58	11.75	6.55	8.27	7.65	9.40	6.14	**5.15**	6.21
	●	○	●	●	59.12	18.42	18.51	13.85	5.75	10.05	9.50	8.77	6.76	**4.87**	6.76
	○	●	●	●	59.66	19.56	21.90	20.46	7.27	10.36	9.19	9.97	8.79	**5.88**	8.79
	●	●	●	●	49.03	13.73	16.72	16.20	6.21	10.18	8.47	8.90	6.36	**5.13**	6.36
					65.58	29.03	20.97	17.05	7.89	11.16	10.03	10.93	8.45	**6.30**	
TC	○	○	○	●	98.31	72.26	56.41	35.48	**10.63**	23.65	25.00	49.18	31.94	13.06	80.84
	○	○	●	○	89.32	57.41	46.73	44.03	**6.38**	22.66	12.59	17.18	11.06	8.96	44.11
	○	●	○	○	86.38	68.95	43.31	31.20	11.13	32.15	16.33	19.30	26.43	**10.18**	39.35
	○	●	○	○	88.90	62.95	39.22	50.53	**10.08**	31.46	20.65	21.37	16.71	13.85	34.10
	○	●	●	○	87.78	48.52	46.36	38.67	5.66	17.75	19.38	20.78	20.44	**5.29**	**7.60**
	●	○	○	●	82.19	52.11	42.76	29.58	5.76	24.18	13.27	12.73	10.74	**5.51**	25.46
	●	○	○	○	78.90	36.71	24.39	29.93	9.39	27.10	12.10	16.42	11.89	**8.97**	16.85
	○	●	○	○	81.77	37.04	42.11	37.29	9.94	24.49	18.08	29.53	18.70	**9.88**	21.40
	●	○	●	○	85.02	42.19	38.92	29.81	**9.55**	21.29	15.61	29.66	17.29	9.94	16.34
	●	●	○	○	81.12	45.15	31.91	22.04	**4.98**	25.61	17.61	12.84	13.44	4.99	7.83
	●	●	○	●	70.33	28.82	28.13	26.41	**4.85**	27.36	12.42	13.07	10.82	5.35	80.84
	○	●	●	●	71.74	26.35	32.06	29.75	9.39	26.95	12.10	22.70	12.32	**7.04**	44.11
	●	●	●	●	74.80	35.46	36.98	31.16	5.48	25.37	20.66	19.92	15.93	**4.48**	39.35
					76.24	26.82	45.90	41.28	**5.52**	23.46	19.44	18.67	16.42	5.65	34.10
					59.74	20.87	35.02	34.37	5.25	28.40	17.51	17.67	10.91	**4.36**	**7.60**
					80.84	44.11	39.35	34.10	**7.60**	25.46	16.85	21.40	16.34	7.83	
ET	○	○	○	●	98.93	66.63	45.19	26.45	**9.31**	16.26	17.44	39.33	26.63	9.34	80.84
	○	○	●	○	89.82	51.51	37.43	36.05	**4.72**	17.33	8.83	13.81	8.46	6.02	37.59
	○	●	○	○	86.73	66.72	31.76	20.69	10.27	21.89	11.42	13.30	19.73	**7.86**	30.43
	○	●	○	○	89.58	56.39	30.66	40.66	**8.95**	26.26	15.37	17.38	14.27	10.56	25.58
	○	●	●	○	89.28	39.35	38.64	29.88	4.41	14.41	16.24	16.90	16.27	**4.24**	**9.31**
	●	○	○	●	82.50	44.72	33.09	22.64	4.46	19.03	10.30	10.43	8.40	**4.38**	6.39
	●	○	○	○	79.26	30.40	17.13	20.53	8.58	19.38	8.76	12.19	10.58	**7.40**	16.85
	○	●	○	○	82.64	30.17	30.42	24.98	8.83	14.89	11.01	21.97	15.36	**7.57**	12.28
	●	○	●	○	85.13	35.30	30.33	21.31	8.40	14.43	11.98	23.77	14.90	**7.21**	16.85
	●	●	○	○	81.07	36.84	24.30	17.35	3.80	20.33	13.49	10.74	11.17	**3.51**	13.49
	●	●	○	●	68.63	22.95	20.99	19.05	3.67	21.38	8.68	10.24	9.20	**3.52**	**6.17**
	○	●	●	●	70.17	20.39	22.42	20.80	8.34	17.04	8.75	17.11	11.78	**6.17**	13.49
	●	●	●	●	74.03	27.13	28.81	23.82	3.98	20.73	15.68	15.48	12.93	**3.73**	15.68
					77.09	20.17	37.50	31.99	**4.35**	17.44	14.08	15.31	13.08	4.53	14.08
					57.81	15.16	27.79	27.51	3.77	21.74	12.10	14.76	9.53	**3.58**	12.10
					80.84	37.59	30.43	25.58	6.39	18.84	12.28	16.85	13.49	**5.97**	16.85

Table 10. **Performance comparison (HD95 mm) on BraTS 2023.** Lower is better. Available and missing modalities are denoted by ● and ○, respectively. , , and indicate the first-, second-, and third-best performance, respectively. WT, TC, and ET represent whole tumor, tumor core, and enhancing tumor, respectively.

Setting	Crop Size	Patch Size	Heads	L	d_{embed}	Params (M) Trans	Params (M) CNN	Params (M) Total	FLOPs (GMacs) Trans	FLOPs (GMacs) CNN	FLOPs (GMacs) Total	Memory (GiB) Trans	Memory (GiB) CNN	Memory (GiB) Other	Memory (GiB) Total	Average Region DSC (%) WT	Average Region DSC (%) TC	Average Region DSC (%) ET	Overall (%) Mean
Scale	96	8	12	12	864	111.01	11.19	122.20	252.30	208.28	460.61	1.57	2.43	0.29	4.29	89.50	80.77	69.72	80.00
	96	8	12	16	864	146.92	11.19	158.11	335.37	208.28	543.69	2.08	2.43	0.29	4.80	90.38	**84.51**	**75.59**	**83.49**
	96	8	16	24	1056	325.75	11.85	337.61	714.21	209.42	923.68	4.01	2.44	0.29	6.74	**90.50**	84.09	75.08	83.22
Crop Size	80	8	12	16	864	146.29	11.19	157.48	174.06	120.53	294.61	1.43	1.43	0.17	3.03	90.17	84.22	74.54	82.97
	96	8	12	16	864	146.92	11.19	158.11	335.37	208.28	543.69	2.08	2.43	0.29	4.80	90.38	**84.51**	**75.59**	**83.49**
	128	8	12	16	864	148.97	11.19	160.16	1064.20	493.70	1557.99	4.18	5.71	0.69	10.58	**90.87**	84.25	74.64	83.25
Patch Size	96	8	12	16	864	146.92	11.19	158.11	335.37	208.28	543.69	2.08	2.43	0.29	4.80	**90.38**	**84.51**	**75.59**	**83.49**
	96	12	12	16	864	150.07	11.19	161.27	84.59	204.11	288.74	1.02	2.43	0.29	3.74	89.99	83.89	74.33	82.74
	96	16	12	16	864	158.00	11.19	169.19	36.01	203.09	239.14	0.78	2.43	0.29	3.50	89.71	83.21	73.74	82.22

Table 11. **Model complexity with different configurations.** , , and indicate the first, second, and third best results, respectively. **Params** refers to the total number of parameters, measured in millions (M). **FLOPs** indicates floating-point operations, measured in giga multiply-accumulate operations (GMacs). **Memory** usage, measured in gigabytes (GiB), represents the estimated memory consumed by model parameters, buffers, and activations during a single forward pass. Average Region DSC and Overall DSC are reported based on evaluations on the BraTS 2023 dataset.

head. Since changing crop size and patch size substantially increases GPU memory usage and computation, we provide a detailed complexity analysis below.

E. Model Complexity

As UniME models cross-modal information from high-resolution inputs with a single ViT encoder, one might expect longer sequences and the attendant $\mathcal{O}(n^2)$ memory cost. In practice, the token sequence length remains modest. The Uni-Encoder tokenizes the channel-stacked volume into a sequence whose length is determined by the crop and patch sizes. For patch size 8, crop size 96, and 4 register tokens, the sequence length is 1732. By contrast, prior designs [38, 67, 68] process features from four modality-specific CNNs concatenated after $4\times$ downsampling, yielding $8 \times 8 \times 8$ grids per modality; the fused sequence length is 2048. Consequently, UniME's effective input sequence length is comparable to existing approaches.

Tab. 15 compares UniME's efficiency against some existing methods [16, 38, 67, 68]. We reproduce the reported hyperparameters and training settings from the

Type																	Avg.
	Flair	○	○	○	●	○	○	●	○	●	●	●	●	●	○	●	
	T1	○	○	●	○	○	●	●	●	○	○	●	●	○	●	●	
	T1ce	○	●	○	○	●	●	○	○	○	●	●	○	●	●	●	
	T2	●	○	○	○	●	○	○	●	●	○	○	●	●	●	●	
WT	HeMIS	63.79	66.71	63.09	55.90	63.31	63.99	57.63	60.44	56.19	59.15	53.67	51.97	52.88	58.26	49.79	58.45
	U-HEVD	71.87	71.24	47.93	46.59	60.43	50.65	33.42	44.79	41.75	47.63	39.89	31.48	41.51	44.79	35.42	47.29
	RFNet	26.03	23.50	21.29	22.30	26.05	20.82	16.19	21.27	17.19	17.41	16.53	17.84	18.76	21.76	19.13	20.41
	mmFormer	14.43	17.47	12.12	12.86	14.21	12.84	9.55	10.49	11.34	12.87	11.28	9.51	12.79	11.29	11.40	12.30
	M^3AE	10.78	13.18	13.24	8.83	10.00	12.03	9.12	10.34	8.39	7.95	8.35	8.23	7.51	9.73	8.26	9.73
	M^2FTrans	17.12	14.41	12.75	11.87	11.42	12.62	10.48	11.02	11.47	9.93	9.64	9.83	9.87	10.67	9.59	11.51
	LS3M	12.91	16.19	13.40	11.68	13.01	12.66	9.69	10.68	9.57	10.34	8.92	9.18	9.71	11.30	9.38	11.24
	IM-Fuse	15.95	14.85	12.68	11.13	11.07	11.99	8.71	10.01	8.60	8.47	8.69	8.29	7.68	10.02	8.64	10.45
	M^2SegMamba	11.62	13.03	12.50	10.52	9.25	11.02	8.68	10.62	9.18	8.79	8.29	9.21	8.51	8.97	7.66	9.86
	Ours	**8.04**	**9.53**	**9.47**	**7.66**	**7.51**	**8.90**	**5.73**	**7.08**	**5.82**	**6.43**	**6.12**	**5.27**	**5.07**	**7.71**	**5.29**	**7.04**
TC	HeMIS	79.77	87.53	85.83	78.90	79.94	86.65	81.50	76.02	71.32	78.52	74.79	70.11	66.15	77.27	65.36	77.31
	U-HEVD	97.01	99.31	78.18	77.27	86.67	80.35	57.57	70.33	69.32	74.48	64.60	54.94	66.59	71.10	59.78	73.83
	RFNet	55.27	36.69	29.48	39.12	50.36	30.15	30.32	39.22	37.16	32.01	29.33	34.15	38.03	39.28	36.49	37.14
	mmFormer	21.84	19.20	14.66	22.61	23.16	15.49	15.03	13.79	19.25	19.07	16.91	16.02	22.51	17.36	19.41	18.42
	M^3AE	12.75	**9.27**	**14.18**	**14.19**	**6.90**	9.34	13.63	12.40	12.22	**7.96**	**6.79**	11.70	7.04	**6.70**	**6.49**	**10.11**
	M^2FTrans	25.75	14.33	14.80	22.10	12.23	13.50	17.60	14.07	18.63	12.10	11.06	15.11	11.52	9.68	10.75	14.88
	LS3M	16.85	19.64	18.12	19.56	16.71	14.60	15.56	14.59	13.10	12.51	10.72	12.98	10.98	13.58	11.65	14.74
	IM-Fuse	33.49	19.47	18.26	22.81	15.23	12.41	14.63	17.08	14.59	9.39	8.99	13.27	10.34	12.34	8.73	15.40
	M^2SegMamba	16.80	12.89	14.24	17.89	10.07	**9.11**	13.58	13.12	13.60	9.68	9.54	13.35	9.98	8.39	9.00	12.08
	Ours	**12.53**	15.01	14.46	18.13	8.15	9.41	**11.90**	**11.71**	**11.99**	9.95	7.04	**10.70**	**6.57**	7.70	6.54	10.79
ET	HeMIS	80.05	87.34	85.75	77.74	79.23	86.29	81.95	75.56	70.77	77.76	73.31	69.64	64.81	76.60	63.78	76.70
	U-HEVD	96.37	98.56	77.26	76.82	84.14	79.10	53.57	68.30	68.65	71.29	61.53	52.10	63.89	68.43	56.61	71.78
	RFNet	49.65	32.06	25.88	36.18	45.81	25.92	26.04	34.95	32.60	28.11	25.23	30.17	32.83	35.74	31.77	32.86
	mmFormer	18.30	16.50	12.20	16.93	19.97	12.54	12.52	11.90	16.27	15.76	13.52	13.44	19.67	14.11	16.28	15.33
	M^3AE	11.66	**8.39**	13.19	**12.90**	6.97	8.22	13.27	11.74	11.36	7.66	6.42	11.26	6.74	6.74	5.95	9.50
	M^2FTrans	17.54	9.82	12.87	14.70	9.16	10.69	11.19	10.78	12.70	7.74	7.48	9.95	7.97	7.76	7.26	10.51
	LS3M	14.09	14.63	15.52	17.13	13.08	10.36	12.99	12.92	11.62	10.13	8.80	11.82	8.89	11.14	9.85	12.20
	IM-Fuse	27.45	15.20	14.13	19.07	12.34	10.20	13.60	14.01	12.55	**7.40**	7.32	11.58	8.72	10.61	7.42	12.77
	M^2SegMamba	13.42	10.55	13.26	15.72	8.19	8.09	12.66	11.63	12.36	7.53	8.00	11.87	7.92	6.90	7.04	10.34
	Ours	**10.73**	12.20	**11.24**	13.77	**6.34**	**7.25**	**10.04**	**10.67**	**10.60**	8.36	**5.74**	**9.84**	**5.46**	**6.05**	**5.83**	**8.94**

Table 12. **Performance comparison (HD95 mm) on BraTS 2024.** Lower is better. Available and missing modalities are denoted by ● and ○, respectively. , , and indicate the first-, second-, and third-best performance, respectively. WT, TC, and ET represent whole tumor, tumor core, and enhancing tumor, respectively.

Crop Size	Average Region DSC (%)			Overall
	WT	TC	ET	Mean
80	90.17	84.22	74.54	82.97
96	90.38	**84.51**	**75.59**	**83.49**
128	**90.87**	84.25	74.64	83.25

Table 13. **Ablation on the crop size**. , , and indicate the first, second, and third best results, respectively.

Patch Size	Average Region DSC (%)			Overall
	WT	TC	ET	Mean
8	**90.38**	**84.51**	**75.59**	**83.49**
12	89.99	83.89	74.33	82.74
16	89.71	83.21	73.74	82.22

Table 14. **Ablation on the patch size**. , , and indicate the first, second, and third best results, respectively.

released code. All experiments were conducted on a single NVIDIA RTX Pro 6000 GPU with 96 GiB of memory. To clearly quantify differences in training efficiency, we report the time required to complete 250 iterations using their recommended batch size and crop size.

Despite its larger parameter count, UniME attains the highest throughput. This significant efficiency gain is primarily due to algorithmic optimizations, including the use of FlashAttention [11], mixed-precision training, and PyTorch graph compilation. These optimizations mitigate the computational overhead introduced by the Transformer layers, ensuring that increased model complexity does not compromise practical training efficiency. Moreover, UniME exhibits affordable GPU memory consumption during training (23.80 GiB), underscoring efficient use of computational resources.

F. Limitations

Although UniME attains state-of-the-art results on BraTS 2023 and BraTS 2024, several limitations remain. **Two-stage Objective Difference.** Stage 1 uses pixel-wise reconstruction loss, whereas Stage 2 optimizes semantic segmentation. This difference may induce a suboptimal inductive bias. Although we mitigate this via (i) using a lightweight decoder in Stage 1 to steer the Uni-Encoder toward semantic features, and (ii) reintroducing a CNN decoder in Stage 2 to emphasize semantic expression, the objective gap may still limit further gains. Future work will explore better representation learning strategies for medical image scenarios, including missing-modality cases.

Methods	Configuration			Complexity				Average Region DSC (%)			Overall
	Batch Size	Crop	Optimizer	Params (M)	Memory (GiB)	Elapsed Time (s)	Throughput (samples/s)	WT	TC	ET	Mean
RFNet	2	80	Adam	8.99	19.08	88.36	5.66	87.60	79.20	66.89	77.90
mmFormer	4	128	Adam	36.65	57.64	293.80	3.40	87.48	78.49	66.11	77.36
IM-Fuse	2	128	RAdam	64.39	56.37	354.19	1.41	88.74	80.59	70.67	80.00
M^2SegMamba	1	128	Adam	70.48	51.72	234.12	1.07	88.99	82.11	71.73	80.94
Ours	4	96	AdamW	158.11	23.80	79.52	12.58	**90.38**	**84.51**	**75.59**	**83.49**

Table 15. **Efficiency comparison across methods.** , , and indicate the first-, second-, and third-best performance, respectively. Params refers to the total number of parameters, measured in millions (M). Memory usage, measured in gigabytes (GiB), represents the peak GPU memory consumed during training. Elapsed Time indicates the time for 250 iterations during training, measured in seconds (s). Throughput measures the number of samples processed per second. Average Region DSC and Overall DSC are reported based on evaluations on the BraTS 2023 dataset.

Dataset Limitations. We currently evaluate only on BraTS 2023 and BraTS 2024 datasets. Ideally, generalization should be validated on more datasets. However, suitable alternatives are limited. More specifically, many multimodal datasets exhibit challenges such as unaligned modalities in ISLES [26], incomplete modality coverage in CHAOS [29] and AMOS [28], and predominantly single-modality data in MSD [2] (with brain cases originating from BraTS).

In alignment with previous research [16, 17, 24, 36, 38, 45, 67–69], we have thus adopted the BraTS series as our primary evaluation benchmark. Although we recognize the importance of diverse datasets for comprehensive validation, our current assessments leverage the most suitable and widely accepted resources available. Future research efforts will seek additional datasets tailored to our specific task requirements and further investigate two-stage heterogeneous architectures in broader vision tasks.