BraTS-UMamba: Adaptive Mamba UNet with Dual-Band Frequency based Feature Enhancement for Brain Tumor Segmentation

Haoran Yao[1*], Hao Xiong[2*], Dong Liu[1], Hualei Shen[1(✉)], and Shlomo Berkovsky[2]

[1] College of Computer and Information Engineering, Henan Normal University, Xinxiang, China
shenhualei@htu.edu.cn

[2] Centre for Health Informatics, Australian Institute of Health Innovation, Macquarie University, Sydney, Australia

Abstract. Brain tumor segmentation (BraTS) of 3D Magnetic Resonance Imaging (MRI) aims to facilitate clinical analysis of brain cancer. Existing BraTS segmentation works tend to exploit convolutional neural networks (CNNs) or vision transformers (ViTs), yet CNNs have a restricted receptive field that focuses on local context only and ViTs suffer from high computational overheads due to quadratic complexity. Recently, Mamba has shown superior performance over ViTs in long-range dependency modeling, offering linear computational complexity and lower memory consumption. However, these methods primarily learn feature representation in the spatial domain, overlooking valuable heuristics embedded in the frequency domain. Inspired by this, we propose BraTS-UMamba, a novel Mamba-based U-Net designed to enhance brain tumor segmentation by capturing and adaptively fusing bi-granularity based long-range dependencies in the spatial domain while integrating both low- and high-band spectrum clues from the frequency domain to refine spatial feature representation. We further enhance segmentation through an auxiliary brain tumor classification loss. Extensive experiments on two public benchmark datasets demonstrate the superiority of our BraTS-UMamba over state-of-the-art methods.

Keywords: Brain tumor segmentation · Mamba · Frequency domain.

1 Introduction

Accurate brain tumor segmentation of 3D Magnetic Resonance Imaging (MRI) images is crucial for diagnosis, treatment planning and management of brain cancer [15, 25]. Many convolutional neural network (CNN) based methods have been proposed for brain tumor segmentation. The works of [12, 20] proposed fully

[*] H. Yao and H. Xiong contributed equally to this paper and are considered co-first authors.
✉ Corresponding author.

convolutional network (FCN) [9] based methods for brain tumor segmentation. Ronneberger *et al.* [13] introduced U-Net, an U-shaped architecture with skip connections to enhance localization ability and gradient flow, which has become widely used for medical image segmentation. Based on U-Net, nnU-Net [6] developed a dynamic adaptation mechanism, allowing automatic optimization of the network depth, achieving promising brain tumor segmentation results. Zhou *et al.* [26] further optimized U-Net by integrating 3D ShuffleNet as the encoder, creating a computationally efficient segmentation model.

However, CNN-based methods can hardly capture global information owing to their fixed receptive field and pooling operations [11]. Therefore, attention mechanism [19] and vision transformers (ViTs) [3] have been adopted to overcome this. For instance, [18, 28] leveraged attention mechanisms to strengthen the representation of global dependencies and improve the segmentation accuracy. The works of [24, 16] leveraged ViTs to model global context within and across different brain MRI sequences, and S^2CA-Net [25] employed ViTs to facilitate the feature extraction of brain tumors' shape and scale. The hybrid CNN-Transformer architectures were also explored to integrate transformers with local information and enhance brain tumor boundary representation [7, 15]. Due to their quadratic complexity, ViTs face the challenge of high computational overheads.

Recently, Mamba based methods [4, 8, 10, 21–23, 27] have been proposed to perform state-space sequence modeling with linear computational complexity and lower memory consumption, exhibiting superior performance compared to attention and ViTs. Of these, [22, 27] performed sequence modeling along different directions, such as forward and backward, to enhance the feature representation learning. Besides, U-Mamba [10] adopted a hybrid CNN-Mamba architecture to extract local details and long-distance dependencies concurrently. Although Mamba based approaches showed strong performance in vision tasks, it has not been applied to brain tumor segmentation. Moreover, small tumors often have blurred boundaries, making it hard for Mamba to capture local details. In addition, CNN-, ViT- and Mamba-based methods tend to utilize spatial features, ignoring valuable information in the frequency domain [11] that can help enhance the representation of spatial features.

In this paper, we propose *BraTS-UMamba* that exploits Mamba in a U-Net shaped encoder-decoder architecture for effective brain tumor segmentation. At each encoder layer, we design a novel Adaptive Mamba (AdM) module to capture bi-granularity based global features that describe long-range dependencies from different perspectives. Unlike other Mamba methods using hard fusion, our AdM module adaptively fuses these bi-granularity based features. To consider small-sized tumors, we also equip AdM modules with multi-scale convolutions to extract local details from multiple scales. Besides, brain tumor segmentation emphasizes the completeness and clear boundary of the segmentation maps. Hence, we further refine these spatial features by leveraging frequency domain information. In the frequency domain, high-band spectrum focuses on edges and texture variations, while low-band spectrum represents global structures and continu-

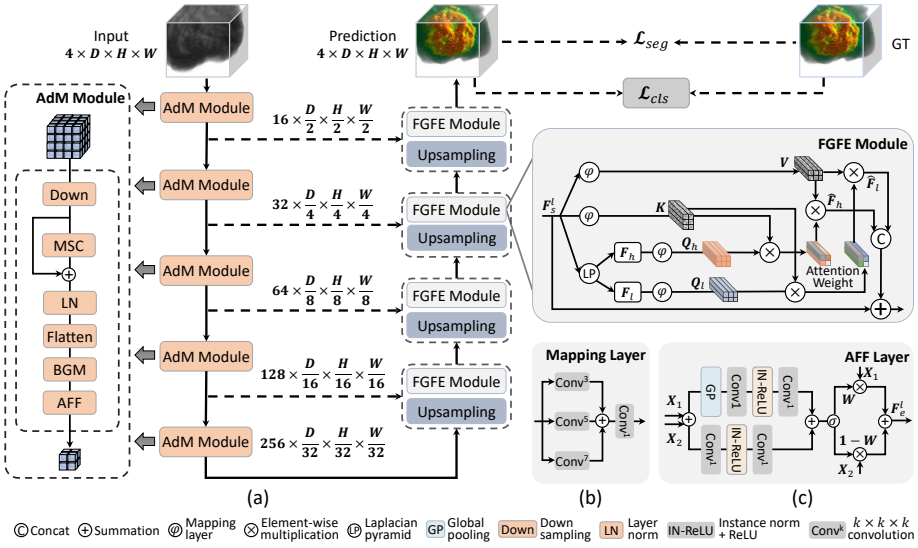

Fig. 1. Overview of the proposed BraTS-UMamba - a UNet style architecture. Its key components include adaptive Mamba (AdM) module, frequency guidance based feature enhancement (FGFE) module and two training losses with the main loss $\mathcal{L}_{seg}$ and an auxiliary one $\mathcal{L}_{cls}$.

ous regions. We introduce the Frequency Guidance based Feature Enhancement (FGFE) module to complement spatial features by selecting informative features from both of low- and high-band spectrum. We also apply an auxiliary brain tumor classification loss to enhance segmentation accuracy. Experimental results show that BraTS-UMamba outperforms several competitive baselines on two challenging brain tumor segmentation datasets. Also, we report an ablation study highlighting the contribution of the components of BraTS-UMamba.

2 Method

Fig. 1 illustrates the architecture of BraTS-UMama. It consists of three key components: 1) adaptive Mamba (AdM) module in each encoder layer to capture both local and global features at multiple scales, 2) Frequency Guidance based Feature Enhancement (FGFE) module in each decoder layer to enhance spatial features using clues from both low- and high-band spectrum, and 3) the output is simultaneously constrained by two losses: the main one for brain tumor segmentation and the auxiliary for brain tumor classification.

2.1 Adaptive Mamba

Mamba based networks [21, 27] that build up long-distance dependencies from different directions have shown promising performance in vision tasks. Unlike

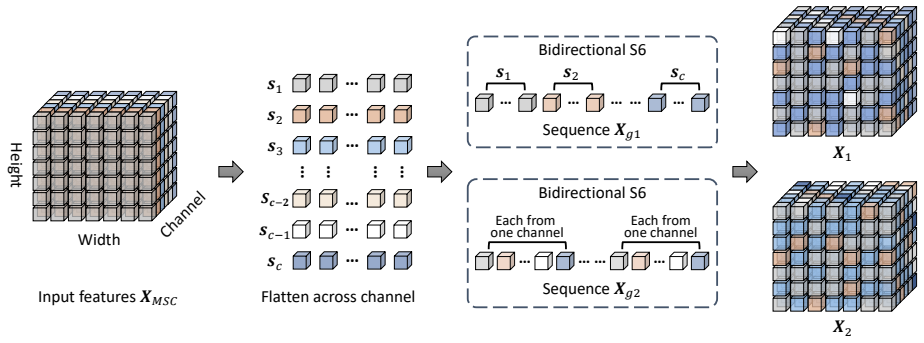

Fig. 2. Details of the Bi-granularity Mamba (BGM) layer. We leave out the depth dimension for ease of illustration.

prior works, we propose the *Adaptive Mamba* (AdM) module that extracts global features from different perspectives and adopts a soft way to fuse them adaptively. We also integrate the AdM module with convolutions of multiple receptive fields to capture multi-scale local features. As shown in Fig. 1, the AdM module in the l-th encoder layer takes features from previous layer as input and produces the output feature $\boldsymbol{F}_e^l$, where $l = 1, 2, \cdots, 5$. This process can be written as:

$$\boldsymbol{X}_{MSC} = MSC(\boldsymbol{F}_e^{l-1}),\ \boldsymbol{X}_{BGM} = BGM(\boldsymbol{X}_{MSC}),\ \boldsymbol{F}_e^l = AFF(\boldsymbol{X}_{BGM}), \quad (1)$$

where feature $\boldsymbol{F}_e^{l-1}$ from previous layer sequentially goes through the multi-scale convolution layer $MSC(\cdot)$, bi-granularity Mamba layer $BGM(\cdot)$, and adaptive feature fusion layer $AFF(\cdot)$. Here, the MSC layer aims to extract multi-scale local features from $\boldsymbol{F}_e^{l-1}$ using 4 parallel convolutions with kernel sizes $1 \times 1 \times 1$, $3 \times 3 \times 3$, $5 \times 5 \times 5$, and $7 \times 7 \times 7$.

Bi-granularity Mamba (BGM) Layer. Suppose $\boldsymbol{X}_{MSC}$ has c channels, the BGM layer first flattens it into c subsequences across the channel direction and obtains subsequences: $\boldsymbol{s}_1, \boldsymbol{s}_2, \cdots, \boldsymbol{s}_c$. Based on these subsequences, we follow [22] to reconstruct two sequences $\boldsymbol{X}_{g1}$ and $\boldsymbol{X}_{g2}$ with different granularities, as shown in Fig. 2. Then, we have:

$$\boldsymbol{X}_1 = S6(\boldsymbol{X}_{g1}) + S6(\boldsymbol{X}_{g1}^{'}), \quad (2)$$

$$\boldsymbol{X}_2 = S6(\boldsymbol{X}_{g2}) + S6(\boldsymbol{X}_{g2}^{'}), \quad (3)$$

where $\boldsymbol{X}_{g1}^{'}$, $\boldsymbol{X}_{g2}^{'}$ are the reversed sequences of $\boldsymbol{X}_{g1}$, $\boldsymbol{X}_{g2}$, and $S6(\cdot)$ denotes the selective scan space state model (S6) [4] that allows each element in a sequence to interact with all previously scanned elements via a compressed hidden state.

Adaptive Feature Fusion (AFF) Layer. As shown in Fig. 1 (c), we utilize the attention mechanism to estimate the fusion weight for $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$, and

pass the initial feature integration $\hat{\boldsymbol{X}}=\boldsymbol{X}_1+\boldsymbol{X}_2$ to the AFF layer to generate the fused feature $\boldsymbol{F}_e^l$. We simultaneously estimate pixel-wise and channel-wise attention scores $\boldsymbol{Z}_p$ and $\boldsymbol{Z}_c$ as:

$$\boldsymbol{Z}_p = PA(\hat{\boldsymbol{X}}), \boldsymbol{Z}_c = CA(\hat{\boldsymbol{X}}), \tag{4}$$

$$\boldsymbol{W} = \sigma(\boldsymbol{Z}_p \oplus \boldsymbol{Z}_c), \tag{5}$$

$$\boldsymbol{F}_e^l = \boldsymbol{W} \odot \boldsymbol{X}_1 + (1 - \boldsymbol{W}) \odot \boldsymbol{X}_2, \tag{6}$$

where $PA(\cdot)$ refers to pixel attention with convolutions and ReLU to obtain $\boldsymbol{Z}_p$, and channel attention $CA(\cdot)$ combines global pooling with convolutions and ReLU to obtain $\boldsymbol{Z}_c$. Then, the sigmoid activation function $\sigma(\cdot)$ squashes the attention scores into fusion weight $\boldsymbol{W}$ within the [0,1] range. Here, $\oplus$ and $\odot$ denote broadcasting summation and element-wise multiplication, respectively.

2.2 Frequency Guidance based Feature Enhancement

To improve BraTS-UMamba's ability to represent brain tumor boundaries, fine-grained textures and global layout, we introduce the *Frequency Guidance based Feature Enhancement* (FGFE) module using both low- and high-frequency information to complement spatial features.

We design a cross-domain attention fusion to alleviate feature redundancy and select informative features for complementary fusion of features from spatial domain and frequency domain. Given a spatial feature $\boldsymbol{F}_s^l$ in the l-th decoder layer, we utilize the 3D Laplacian pyramid decomposition [2] to decompose it into the high- and low-frequency components $\boldsymbol{F}_h$ and $\boldsymbol{F}_l$ that are passed through a mapping layer to produce the query matrices $\boldsymbol{Q}_l$ and $\boldsymbol{Q}_h$. Meanwhile, $\boldsymbol{F}_s^l$ passes through another two mapping layers to generate the key and value matrices $\boldsymbol{K}, \boldsymbol{V}$ that are shared by both $\boldsymbol{Q}_l$ and $\boldsymbol{Q}_h$. Then, we have:

$$\hat{\boldsymbol{F}}_l = softmax\left(\frac{\boldsymbol{Q}_l\boldsymbol{K}^T}{\sqrt{d}}\right)\boldsymbol{V}, \hat{\boldsymbol{F}}_h = softmax\left(\frac{\boldsymbol{Q}_h\boldsymbol{K}^T}{\sqrt{d}}\right)\boldsymbol{V}. \tag{7}$$

Afterwards, we concatenate $\hat{\boldsymbol{F}}_l$, $\hat{\boldsymbol{F}}_h$ along channel dimension and add it back to $\boldsymbol{F}_s^l$, as shown in Fig. 1.

2.3 Loss Function

Brain tumors are typically small, while normal brain tissue occupies most part of the brain. During learning, the dominant part of the normal tissue may distract the focus and introduce biases. To mitigate this, we add an auxiliary brain tumor classification loss to the brain tumor segmentation loss.

We have the ground truth brain tumor mask $\boldsymbol{M} \in \{0,1,2,3\}^{D\times H\times W}$ and the network output $\boldsymbol{P} \in \mathbb{R}^{D\times H\times W}$ containing predicted probabilities that are utilized to generate the predicted segmentation mask $\hat{\boldsymbol{S}} \in \{0,1,2,3\}^{D\times H\times W}$. The segmentation loss $\mathcal{L}_{seg}$ is supervised by the Dice loss $Dice(\boldsymbol{M}, \hat{\boldsymbol{S}})$. For auxiliary

brain tumor classification loss $\mathcal{L}_{cls}$, we divide the ground truth mask $\boldsymbol{M}$ into 3D patches of size $16 \times 16 \times 16$ along the depth, height and width dimensions. Each patch is assigned the ground truth classification label $\boldsymbol{C}_p = 1$ if it contains brain tumor or 0 otherwise. Likewise, we divide the network output $\boldsymbol{P}$ into 3D patches $\hat{\boldsymbol{M}}_p$ of size $16 \times 16 \times 16$ and compute brain tumor classification loss $\mathcal{L}_{cls}$ using the binary cross-entropy loss $BCE(\hat{\boldsymbol{M}}_p, \boldsymbol{C}_p)$. The network is trained via both losses with a trade-off parameter λ (set to 0.6) as:

$$\mathcal{L}_{total} = \mathcal{L}_{seg} + \lambda \mathcal{L}_{cls}. \tag{8}$$

3 Evaluation

3.1 Experimental Setting

Datasets. We evaluate the performance of all methods on two datasets, **MSD BTS** [17] and **BraTS2023-GLI** [1], which contain 484 and 1,251 MRI scans, respectively. These MRI scans were captured by four modalities: T1-weighted (T1), T1-weighted contrast-enhanced (T1ce), T2-weighted (T2), and fluid attenuated inversion recovery (FLAIR). The datasets also provide brain tumor ground truth masks annotated by clinical experts, marking three tumor regions: necrotic tumor core (NCR), peritumoral edematous/invaded tissue (ED), and enhancing tumor (ET). The performance is evaluated on three regions: the ET region, tumor core (TC) region for ET and NCR, and the whole tumor (WT) region for ET, NCR, and ED.

Baselines and Evaluation Metrics. We compare our method against seven recent segmentation baselines including attention-based (EoFormer [15], M^{2}FTrans [16], SDV-TUNet [28]), transformer-based (S^{2}CA-Net [25], UNETR++ [14], SWinUNETR-V2 [5]), and the latest Mamba-based segmentation method SegMamba [22] that also builds long-range dependencies of a different granularity. To our knowledge, we are the first Mamba based method for brain tumor segmentation. In the evaluation, we use the Dice Similarity Coefficient (DSC) and 95% Hausdorff distance (HD95).

Implementation Details. BraTS-UMamba was implemented using PyTorch on a workstation equipped with an NVIDIA 4090 GPU. Our model was trained for 1,000 epochs using the Adam optimizer with batch size of 4. Data augmentation techniques, including rotation, scaling, elastic deformation, and random cropping, were applied during training. The brain volume was divided into patches of $128 \times 128 \times 128$ with an overlapping step 96 for training or inference, and the 5-fold cross validation was utilized for evaluation.

3.2 Comparison with SOTA Methods

As shown in Tables 1 and 2, our proposed BraTS-UMamba consistently outperforms all the baselines on two datasets. For instance, BraTS-UMamba achieves

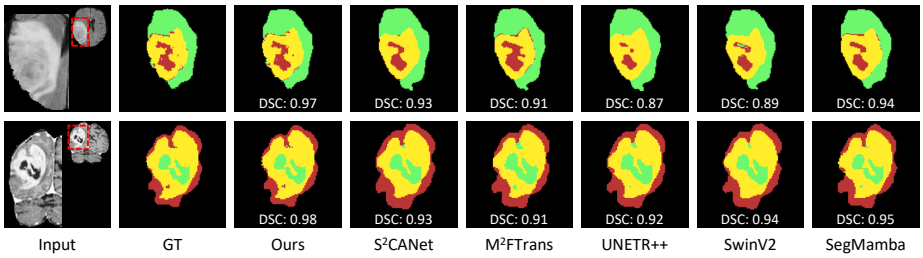

Fig. 3. Visual comparisons of BraTS-UMamba and several leading baselines on **MSD BTS** (top row) and **BraTS2023-GLI** (bottom row). The green, yellow and red regions denote necrotic core (NCR), enhancing tumor (ET), and edema (ED).

Table 1. Evaluation with the MSD BTS dataset (best result in boldface).

Methods	DSC (%) ↑				HD95 (mm) ↓			
	ET	WT	TC	Avg.	ET	WT	TC	Avg.
Eoformer [15]	74.28	88.29	80.95	81.17	5.98	8.59	7.10	7.23
M^2FTrans [16]	77.31	89.53	82.87	83.24	6.12	7.10	5.80	6.34
SDV-TUNet [28]	73.42	87.69	79.59	80.23	5.96	7.09	7.52	6.86
S^2CA-Net [25]	77.35	89.40	82.68	83.14	5.62	7.49	7.54	6.88
UNETR++ [14]	75.21	88.69	81.87	81.93	6.17	8.27	7.87	7.44
SwinUNETR-V2 [5]	75.92	88.69	82.28	82.50	5.45	8.03	6.96	6.81
SegMamba [22]	76.82	89.62	82.74	83.06	5.31	7.37	6.45	6.38
Our method	**80.63**	**90.61**	**84.08**	**85.11**	**3.92**	**4.93**	**5.14**	**4.66**

Table 2. Evaluation with the BraTS2023-GLI dataset (best result in boldface).

Methods	DSC (%) ↑				HD95 (mm) ↓			
	ET	WT	TC	Avg.	ET	WT	TC	Avg.
Eoformer [15]	83.11	90.68	87.60	87.13	4.60	7.50	5.92	6.01
M^2FTrans [16]	84.17	91.11	87.77	87.68	4.38	7.84	5.73	5.98
SDV-TUNet [28]	83.96	90.50	87.36	87.27	3.41	5.61	5.29	4.77
S^2CA-Net [25]	83.91	91.60	87.91	87.81	4.46	6.94	6.04	5.82
UNETR++ [14]	83.62	91.63	87.82	87.69	4.21	5.92	5.50	5.21
SwinUNETR-V2 [5]	84.07	91.85	88.12	88.01	3.92	5.86	5.33	5.04
SegMamba [22]	84.65	92.12	88.34	88.37	4.22	4.92	5.80	4.98
Our method	**85.74**	**92.90**	**90.71**	**89.78**	**3.11**	**4.09**	**3.80**	**3.66**

an average DSC of 85.11% and an average HD95 of 4.66 on the MSD BTS dataset, outperforming the second-best method M^2FTrans by 2.25% and 26.50%, respectively. For the BraTS2023-GLI dataset, BraTS-UMamba surpasses the second-best results by 1.60% and 23.27% in terms of average DSC and HD95 scores, respectively. The results highlight BraTS-UMamba's strong potential for brain tumor segmentation. In Fig. 3, we also visually compare BraTS-UMamba and several leading baselines, demonstrating that our method can produce brain tumor segmentation with better connectivity and clearer boundary.

Table 3. Ablation study of investigating key components in our model on the MSD BTS dataset. The best result is indicated in boldface.

	baseline	S1	S2	S3	S4	ours
DSC (%) ↑	81.74	82.63	83.65	83.50	84.07	**85.11**
HD95 (mm) ↓	7.42	6.52	6.11	5.89	5.61	**4.66**

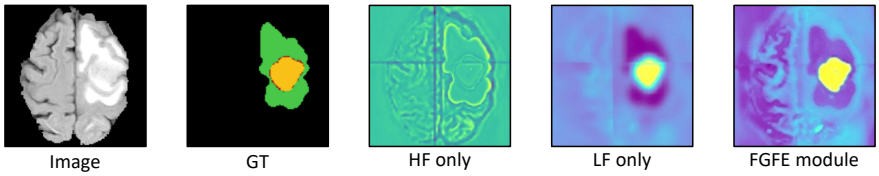

Fig. 4. Effects of feature maps guided by high-frequency component only (HF only), low-frequency component only (LF only), and both of low-, high-frequency (FGFE module).

3.3 Ablation Study

We investigate the effects of the key components of BraTS-UMamba. In Table 3, S1 refers to our method that only keeps the Adaptive Mamba module, while 'baseline' replaces the adaptive feature fusion (AFF) layer in S1 with a simple concatenation. The adaptive fusion (S1) enhances performance against the hard fusion of baseline. Based on S1, we gradually add the low-frequency component guided feature enhancement (S2), high-frequency component guided feature enhancement (S3), and the FGFE module (S4). Although S2 and S3 improve the accuracy further, S4 considering both low- and high-frequency information shows a substantial enhancement. We visualize the feature maps after considering the frequency domain information in Fig. 4, which illustrates that low- and high-frequency features can capture continuous regions and boundaries effectively. Finally, 'ours' that incorporates the auxiliary brain tumor classification loss function achieves the best segmentation performance.

4 Conclusions

In this work, we propose the BraTS-UMamba model for brain tumor segmentation that devises an adaptive Mamba to fuse bi-granularity global features in a soft manner and exhibits a higher flexibility than the hard fusion. By leveraging the frequency domain information, it enables our network to capture clearer tumor boundary and ensures a better connectivity of the segmentation masks. Besides, an auxiliary brain tumor classification loss combined with the traditional segmentation loss enhances the segmentation accuracy. Experimental results on two datasets validate the effectiveness and accuracy of BraTS-UMamba.

Acknowledgments. This work was partially supported by the Provincial Science and Technology Research Project of Henan (No. 232102211024) and the Postgraduate Education Reform and Quality Improvement Project of Henan (No. YJS2024AL095).

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

References

1. Brain tumor segmentation challenge (BraTS 2023). `https://www.synapse.org/Synapse:syn51156910/wiki/622351`, last accessed: 2025/02/15
2. Dippel, S., Stahl, M., Wiemker, R., Blaffert, T.: Multiscale contrast enhancement for radiographies: Laplacian pyramid versus fast wavelet transform. IEEE Transactions on Medical Imaging **21**(4), 343–353 (2002)
3. Dosovitskiy, A., et al.: An image is worth 16x16 words: Transformers for image recognition at scale. In: International Conference on Learning Representations (2021)
4. Gu, A., Dao, T.: Mamba: linear-time sequence modeling with selective state spaces. arXiv preprint arXiv:2312.00752 (2023)
5. He, Y., Nath, V., Yang, D., Tang, Y., Myronenko, A., Xu, D.: SwinUNETR-V2: stronger swin Transformers with stagewise convolutions for 3D medical image segmentation. In: Greenspan, H., et al. (eds.) MICCAI 2023, LNCS, vol. 14223, pp. 416–426. Springer, Cham (2023)
6. Isensee, F., Jäger, P.F., Full, P.M., Vollmuth, P., Maier-Hein, K.H.: nnU-Net for brain tumor segmentation. In: Crimi, A., Bakas, S. (eds.) BrainLes 2020, LNCS, vol. 12659, pp. 118–132. Springer, Cham (2021)
7. Lin, J., et al.: CKD-TransBTS: clinical knowledge-driven hybrid transformer with modality-correlated cross-attention for brain tumor segmentation. IEEE Transactions on Medical Imaging **42**(8), 2451–2461 (2023)
8. Liu, J., et al.: Swin-UMamba: Mamba-based UNet with ImageNet-based pretraining. In: Linguraru, M.G., et al. (eds.) MICCAI 2024, LNCS, vol. 15009, pp. 615–625. Springer, Cham (2024)
9. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 3431–3440 (2015)
10. Ma, J., Li, F., Wang, B.: U-Mamba: enhancing long-range dependency for biomedical image segmentation. arXiv preprint arXiv:2401.04722 (2024)
11. Ma, R., Zhang, Y., Zhang, B., Fang, L., Huang, D., Qi, L.: Learning attention in the frequency domain for flexible real photograph denoising. IEEE Transactions on Image Processing **33**, 3707–3721 (2024)
12. Ribalta Lorenzo, e.a.: Segmenting brain tumors from FLAIR MRI using fully convolutional neural networks. Computer Methods and Programs in Biomedicine **176**, 135–148 (2019)
13. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W., Frangi, A. (eds.) MICCAI 2015, LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015)
14. Shaker, A.M., Maaz, M., Rasheed, H., Khan, S., Yang, M.H., Khan, F.S.: Unetr++: delving into efficient and accurate 3D medical image segmentation. IEEE Transactions on Medical Imaging **43**(9), 3377–3390 (2024)

15. She, D., Zhang, Y., Zhang, Z., Li, H., Yan, Z., Sun, X.: Eoformer: edge-oriented Transformer for brain tumor segmentation. In: Greenspan, H., et al. (eds.) MICCAI 2023, LNCS, vol. 14223, pp. 333–343. Springer, Cham (2023)
16. Shi, J., Yu, L., Cheng, Q., Yang, X., Cheng, K.T., Yan, Z.: M^2FTrans: modality-masked fusion Transformer for incomplete multi-modality brain tumor segmentation. IEEE Journal of Biomedical and Health Informatics **28**(1), 379–390 (2023)
17. Simpson, A.L., et al.: A large annotated medical image dataset for the development and evaluation of segmentation algorithms. arXiv preprint arXiv:1902.09063 (2019)
18. Ullah, Z., Usman, M., Jeon, M., Gwak, J.: Cascade multiscale residual attention CNNs with adaptive ROI for automatic brain tumor segmentation. Information Sciences **608**, 1541–1556 (2022)
19. Vaswani, A., et al.: Attention is all you need. In: Proceedings of the 31st International Conference on Neural Information Processing Systems, pp. 6000–6010. Curran Associates, Long Beach (2017)
20. Wang, J., et al.: DFP-ResUNet: convolutional neural network with a dilated convolutional feature pyramid for multimodal brain tumor segmentation. Computer Methods and Programs in Biomedicine **208**, 106208 (2021)
21. Wang, J., Chen, J., Chen, D., Wu, J.: LKM-UNet: large kernel vision Mamba UNet for medical image segmentation. In: Linguraru, M.G., et al. (eds.) MICCAI 2024, LNCS, vol. 15008, pp. 360–370. Springer, Cham (2024)
22. Xing, Z., Ye, T., Yang, Y., Liu, G., Zhu, L.: SegMamba: long-range sequential modeling Mamba for 3D medical image segmentation. In: Linguraru, M.G., et al. (eds.) MICCAI 2024, LNCS, vol. 15008, pp. 578–588. Springer, Cham (2024)
23. Xu, Z., et al.: Polyp-Mamba: polyp segmentation with visual Mamba. In: Linguraru, M.G., et al. (eds.) MICCAI 2024, LNCS, vol. 15008, pp. 510–521. Springer, Cham (2024)
24. Zhang, Y., et al.: mmFormer: multimodal medical transformer for incomplete multimodal learning of brain tumor segmentation. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds.) MICCAI 2022, LNCS, vol. 13435, pp. 107–117. Springer, Cham (2022)
25. Zhou, L., Jiang, Y., Li, W., Hu, J., Zheng, S.: Shape-scale co-awareness network for 3D brain tumor segmentation. IEEE Transactions on Medical Imaging **43**(7), 2495–2508 (2024)
26. Zhou, X., Li, X., Hu, K., Zhang, Y., Chen, Z., Gao, X.: ERV-Net: an efficient 3D residual neural network for brain tumor segmentation. Expert Systems with Applications **170**, 114566 (2021)
27. Zhu, L., Liao, B., Zhang, Q., Wang, X., Liu, W., Wang, X.: Vision Mamba: efficient visual representation learning with bidirectional state space model. arXiv preprint arXiv:2401.09417 (2024)
28. Zhu, Z., Sun, M., Qi, G., Li, Y., Gao, X., Liu, Y.: Sparse dynamic volume TransUNet with multi-level edge fusion for brain tumor segmentation. Computers in Biology and Medicine **172**, 108284 (2024)