MBGNet: Mamba-Based Boundary-Guided Multimodal Medical Image Segmentation Network

Ke Xu
College of Software, Xinjiang University
690185697@qq.com

Min Li
School of Computer Science and Technology, Xinjiang University
MinLi927@163.com

Guangjian Liu
College of Software, Xinjiang University
liuguangjian@stu.xju.edu.cn

Chen Chen
College of Software, Xinjiang University
1343432873@qq.com

Cheng Chen
College of Software, Xinjiang University
chenchengoptics@gmail.com

Enguang Zuo
College of Intelligent Science and Technology (Future Technology), Xinjiang University
zeg@xju.edu.cn

Xiaoyi Lv
College of Software, Xinjiang University
xjuwawj01@163.com

Abstract

Multimodal medical image segmentation plays an important role in fields such as medical image diagnosis and biomedical research. Although Mamba performs well in medical image feature extraction, it still faces challenges in capturing fine boundaries in lesions. Therefore, in this paper, a Mamba-based boundary-guided multimodal medical image segmentation network (MBGNet) is proposed. To address Mamba's deficiency in capturing boundary information, we designed a Boundary Information Encoding Module (BIEM). This module employs multiple boundary extraction strategies to capture boundary information across different modalities and uses an external attention mechanism to enhance the interaction and understanding of boundary relationships. Additionally, we designed an Information Guidance Module (IGM) to address information loss during boundary and content fusion. This module uses the boundary segmentation map as a guideline, integrating local features and global context for content segmentation, effectively overcoming information loss. Finally, experimental results on the BraTS2019 and BraTS2020 glioma tumor datasets show that MBGNet achieves DICE coefficients of 88.17% and 88.12%, and Hausdorff 95 distances of 5.21 and 5.08, respectively. These results confirm the superior performance of MBGNet in the segmentation of complex lesion regions, providing a more accurate and reliable method for multimodal medical image analysis.

***Keywords:** Multimodal medical image segmentation, Mamba, Feature extraction, Boundary information guidance.*

1. Introduction

Image segmentation is pivotal in medical image analysis, aiming to accurately distinguish between lesions and background [30]. To accomplish this, multimodal medical image segmentation techniques have been developed, providing comprehensive insights into tissue and pathological states by integrating information from diverse imaging

modalities. Nonetheless, the multimodal glioma medical image segmentation used in this study faces challenges due to inter-modal inconsistencies, such as variations in imaging parameters, resolution, and contrast, which can degrade segmentation performance [29]. The Mamba methodology addresses these issues by employing dynamic weights to adapt across different modalities, thus enhancing segmentation accuracy [27]. Despite its efficacy with multimodal data, Mamba may lack sensitivity in capturing subtle boundary details. To remedy this, integrating Mamba with boundary information is proposed to precisely delineate the fine boundaries of tissues and lesions, thereby improving the accuracy and reliability of segmentation outcomes.

Currently, mainstream methods in medical image segmentation primarily include those based on Convolutional neural networks (CNNs) and Transformers [16]. CNNs effectively recognize local features through convolutional operations. However, they are limited in capturing global contextual information, which affects the overall segmentation performance. In contrast, Transformers have shown significant advantages in modeling global information due to their unique self-attention mechanism. Despite this, the high computational complexity of self-attention poses challenges for efficient segmentation. Recently, the Mamba approach has garnered attention in the image domain [11]. Mamba allows each element in a sequence to interact with previously scanned samples via a unique compressed hidden state, effectively reducing computational complexity from quadratic to linear [28].

Although Mamba offers significant advantages in terms of reduced parametric and computational costs and efficient extraction of image feature information, it struggles with capturing the edge information of lesion regions due to the inherent quality issues in multimodal medical images. This limitation prevents various segmentation models from achieving optimal performance in multimodal image segmentation. Boundary information is crucial not only for the positional localization of lesion regions but also for determining the accuracy of image segmentation. Therefore, ensuring the accurate extraction of edge information while maintaining the efficient segmentation performance of Mamba has become the central focus of this research.

In summary, this study makes the following core contributions:

- In this study, we propose a Mamba-based boundary-guided multimodal medical image segmentation network, named MBGNet. This model retains the global information modeling efficiency inherent in Mamba while incorporating a boundary extraction module to enhance the identification of fine boundaries. This integration significantly improves the segmentation accuracy for small and complex structures in medical images.
- In this study, we construct a Boundary Information Encoding Module (BIEM). This module extracts boundary data from different modalities and facilitates interaction through an external attention mechanism. By effectively fusing multiple boundary information sources, the BIEM enhances the accuracy and reliability of boundary segmentation.
- In this study, we designed an Information Guidance Module (IGM). This module utilizes a complete boundary segmentation map as an external auxiliary input, integrating boundary contours with lesion area features to achieve precise segmentation of lesions.

2. Related Work

2.1. Mamba's Application in Image Segmentation

With the introduction of Mamba into the visual domain, an increasing number of researchers have begun to explore its methods and applications in image segmentation. In 2D image segmentation, Ruan et al. [21] proposed a model employing a U-shaped architecture using SSM (VM-UNet). This model addresses challenges in long-range modeling and computational complexity in medical image segmentation by incorporating Visual State Space (VSS) blocks to capture extensive contextual information. In 3D image segmentation, Xing et al. [24] introduced a 3D medical image segmentation model (SegMamba), which leverages Mamba to capture long-range dependencies within full-volume features at each scale, thereby tackling computational challenges associated with high-dimensional medical images. Compared to transformers, Mamba exhibits lower complexity, prompting researchers to explore more lightweight segmentation models. For instance, Liao et al. [14] proposed a lightweight medical image segmentation model (LightM-UNet) aimed at addressing computational resource limitations of existing UNet models in mobile healthcare applications. In the domain of multimodal image segmentation, Mamba has also demonstrated exceptional performance. Wan et al. [23] introduced a network for multimodal semantic segmentation (Sigma), which employs a Siamese encoder and an innovative Mamba fusion mechanism to efficiently select and segment key information from different modalities, such as RGB, thermal imaging, and depth information, enhancing the model's robustness and reliability under adverse conditions. Given the complexity of multimodal medical image data, current Mamba models have not fully exploited their potential.Therefore, this paper proposes a Mamba-based method and utilizes glioma datasets to address the challenges of multimodal medical image segmentation. This method not only selects specific feature extraction schemes based on the characteristic differences of multimodal glioma data but also compensates for detail loss in image segmentation through guided segmentation.

2.2. Boundary Segmentation Techniques in Image Segmentation

Boundary segmentation techniques are vital in the field of image segmentation, particularly for tasks requiring fine-grained segmentation. These methods aim to accurately identify and extract boundary information of objects within an image, overcoming challenges such as complex backgrounds, similar regions, and blurred edges. In recent years, extensive research has explored the application of boundary segmentation techniques. For example, Gab Allah et al. [1] proposed the Edge U-Net model, which achieves precise segmentation of brain tumor MRI images by integrating multi-scale boundary-related information and adjacent contextual data during the decoding phase. Similarly, Yang et al. [25] developed a novel 3D network for automatic CT image segmentation, focusing on spatial context modeling and explicit edge segmentation priors, significantly enhancing the accuracy and robustness of abdominal organ segmentation. Bui et al. [4] introduced a multi-scale edge-guided attention network (MEGANet) that addresses challenges in polyp segmentation within colonoscopy images by combining classical edge detection techniques with attention mechanisms. Chen et al. [7] proposed an edge-enhanced semantic segmentation network, which improves the extraction of edge information by sharing parameters between backbone networks and employing specialized loss functions. Despite these advancements, the aforementioned models face limitations when addressing edge blurring in multimodal image data. Different modalities possess distinct feature distributions and noise characteristics, potentially reducing segmentation accuracy. Therefore, this paper proposes the use of different edge detection operators to separately extract boundary information from multimodal data while employing an external attention mechanism for fusion. This method clarifies the complementarity between the interior and boundary of lesions, enhancing segmentation performance.

3. Method

3.1. Preliminaries for Mamba

3.1.1 State Space Models

State Space Models (SSMs) are employed for sequence-to-sequence modeling and are characterized by their dynamic properties, which remain constant over time [12]. Due to their linear complexity, SSMs can implicitly map sequences to a latent state space, effectively capturing the inherent dynamics of the system. Formally, an SSM is defined by the following equations:

$$h'(t) = \boldsymbol{A}h(t) + \boldsymbol{B}x(t) \tag{1}$$

$$y(t) = \boldsymbol{C}h(t) \tag{2}$$

Here, $\mathrm{x(t)}$, $\mathrm{h(t)}$, and $\mathrm{y(t)}$ represent the input, hidden state, and output, respectively, while $\mathrm{h'(t)}$ denotes the time derivative of $\mathrm{h(t)}$. A is the state matrix, and B and C are projection parameters.

Models based on SSMs are typically continuous-time models and require discretization when integrated into deep learning algorithms [19]. SSMs achieve this by introducing a time scale parameter Δ and employing the Zero-Order Hold (ZOH) rule to transform A and B into discrete parameters $\overline{\mathrm{A}}$ and $\overline{\mathrm{B}}$. The equations are as follows:

$$\overline{\mathbf{A}} = e^{\Delta \mathbf{A}} \tag{3}$$

$$\overline{\mathbf{B}} = \Delta \mathbf{A}^{-1}(e^{\Delta \mathbf{A}} - \mathbf{I}) \cdot \Delta \mathbf{B} \tag{4}$$

$$\bar{\mathbf{C}} = \mathbf{C} \tag{5}$$

$$h_t = \overline{\mathbf{A}} h_{t-1} + \overline{\mathbf{B}} x_t \tag{6}$$

$$y_t = \bar{\mathbf{C}} h_t \tag{7}$$

Finally, the model computes the output y through global convolution operations within a structured convolution kernel $\overline{\mathrm{K}}$:

$$\overline{\mathbf{K}} = (\mathbf{CB}, \mathbf{CAB}, \mathbf{CA}^2\overline{\mathbf{B}}, \ldots, \mathbf{CA}^{L-1}\overline{\mathbf{B}}) \tag{8}$$

$$y = \overline{\mathbf{K}} \otimes y_t \tag{9}$$

3.1.2 2D Selective Scan Mechanism

To address the incompatibility between the original one-dimensional input sequence in SSMs and the two-dimensional data in the visual domain, researchers have introduced the 2D Selective Scan (SS2D) mechanism [18]. Figure 1 illustrates the functioning of SS2D. SS2D constructs four independent sequences by scanning image patches in the 2D visual data across four different directions. This four-directional scanning approach ensures that each element in the feature map can incorporate information from all other positions in every direction. Subsequently, each feature sequence is processed using the Selective Scanning State Space Model (S6) [8]. Finally, the processed feature sequences are aggregated to reconstruct the 2D feature map.

3.2. Overview of the Model Architecture

Figure 2 illustrates the architecture of the proposed model. To extract boundary and content information, we divided the four modalities of glioma MRI data (FLAIR, T1ce, T1, and T2) into two groups: one consisting of FLAIR and T1ce, and the other including all four modalities, as shown in the orange and blue boxes on the left side of Figure 2. For boundary information extraction, we prioritized the FLAIR and T1ce modalities because these modalities more clearly depict the boundary contours of the lesion areas. For content information extraction, multimodal

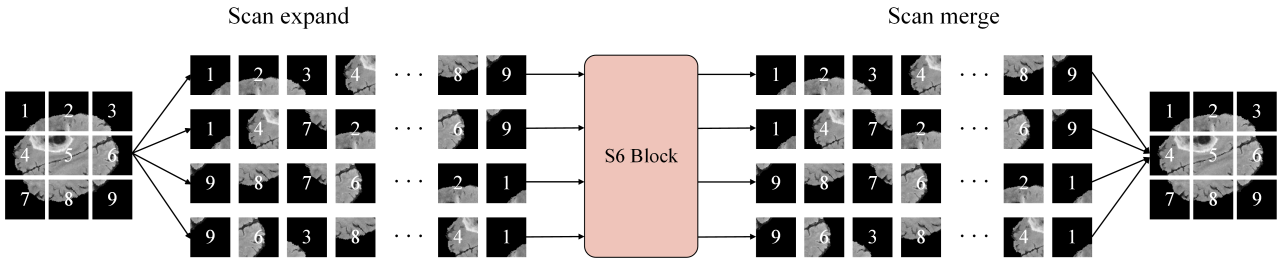

Figure 1. Overview of the 2D Selective Scan mechanism.

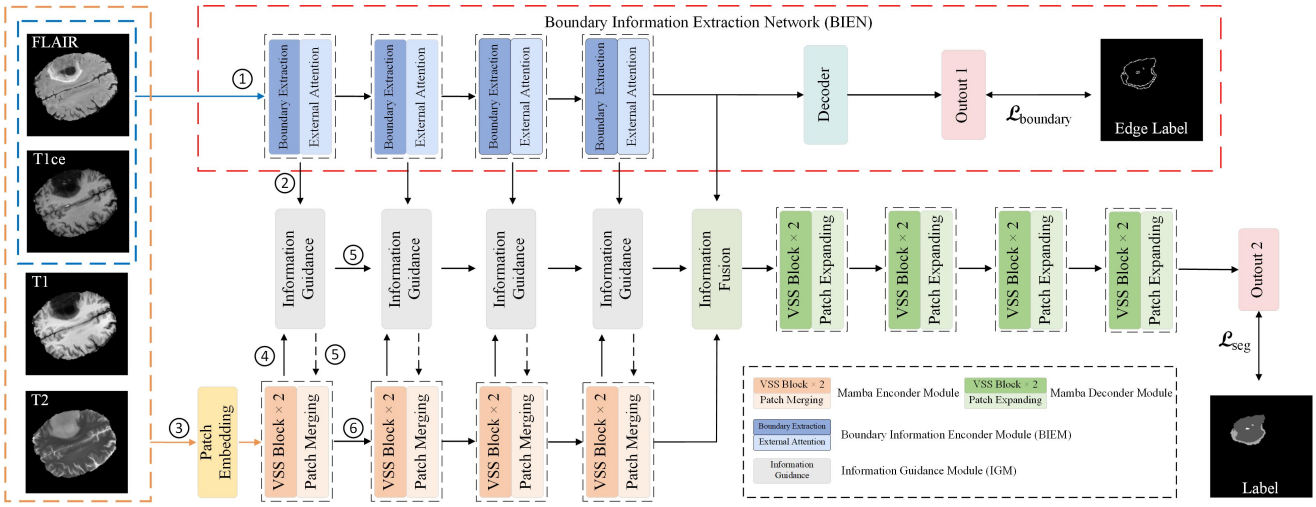

Figure 2. Overview of the MBGNet Model Framework.

image data can compensate for the lack of rich features in single-modality images. Therefore, all four modalities were used as input data for content segmentation.

In the proposed architecture, the model operates in two stages. The first stage involves the extraction of boundary information. Initially, the data is fed into the Boundary Information Encoding Module (BIEM), after which the image dimensions are restored through the decoder. Subsequently, the output is compared with the boundary labels to train an efficient Boundary Information Extraction Network (BIEN). Through iterative training, this network is progressively optimized to effectively extract the boundaries of smaller and more challenging segmentation content, ultimately generating a boundary segmentation map that can serve as a guiding map, as illustrated in Output1 of Figure 2. Once the boundary information accurately reflects the contours of the lesion areas, the model proceeds to the second stage: boundary-guided content segmentation. In this stage, the parameters of the BIEN remain unchanged. Based on the existing boundary segmentation results, the Mamba encoding module is utilized to extract content information, which is then fed into the Information Guidance Module (IGM) to achieve boundary-guided content segmentation. The segmentation result is shown as Output2 in Figure 2. The flow of this stage follows the numerical order illustrated in Figure 2.

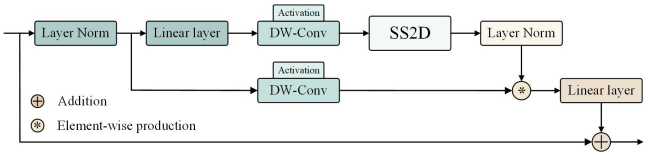

Figure 3. Overview of the 2D Selective Scan mechanism.

As illustrated in Figure 2, both the Mamba encoding module and the Mamba decoding module utilize the Visual State Space (VSS) as the backbone for feature extraction. The internal structure of the VSS module is depicted in Figure 3. Initially, the input data is processed through an initial linear embedding layer and subsequently split into two separate information streams. One information stream passes through a deep convolution layer, followed by an activation function, and then enters the SS2D module. The output from the SS2D module undergoes layer normalization and is then combined with the output from the other information stream. The merged output constitutes the final result of the VSS block.

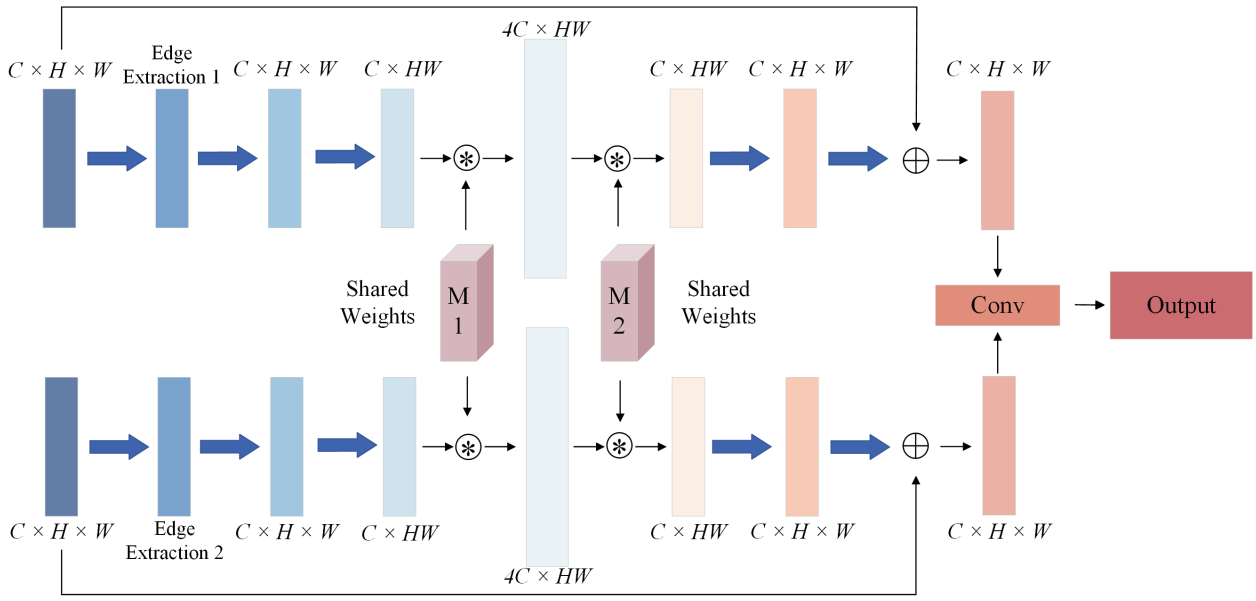

Figure 4. Overview of the External Attention Mechanism.

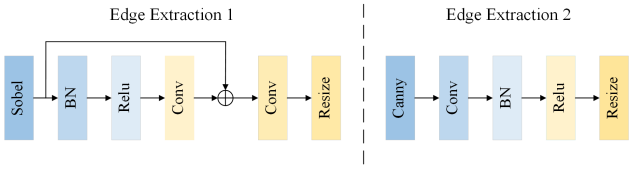

Figure 5. Overview of the Boundary Extraction Module.

3.3. Boundary Information Encoding Module

In multimodal medical image segmentation, boundary information segmentation faces numerous challenges. Firstly, the contrast differences across various modality images may lead to blurred boundaries, complicating accurate identification. Additionally, imaging noise and artifacts can interfere with boundary information extraction, thereby increasing segmentation uncertainty. Collectively, these issues limit the accuracy and reliability of segmentation. To address these challenges, we propose a BIEM designed to extract boundary information from medical images. This module consists of a Boundary Extraction Module and an External Attention Mechanism, as illustrated in Figures 4 and 5, respectively.

The internal structure of the Boundary Extraction Module is illustrated in Figure 4. The model employs both Sobel and Canny edge detection operators to perform edge detection on the images. The Sobel operator calculates the gradient of the image intensity to effectively identify coarse contours of the edges, while the Canny operator, with its multi-stage processing, provides more precise edge detection results. By combining the strengths of both operators, we can capture richer and more accurate edge information.

In implementing the external attention mechanism, this study draws inspiration from the work of Ruan et al. [22, 9, 10]. Their research implicitly considers potential associations between different samples to capture global features within the dataset, thereby enhancing both the quality of feature representations and the model's generalization ability. This concept aligns closely with the objective of enhancing inter-modal interactions in multimodal image processing. Thus, we extend this approach to the field of multimodal medical image processing, aiming to improve model performance and robustness by strengthening inter-modal interactions.

As shown in Figure 5, the data from the two modalities, X_1 and X_2, are fed into two separate branches as inputs. Both X_1 and X_2 belong to $\mathbb{R}^{C\times H\times W}$. First, the boundary extraction module processes these inputs, yielding boundary features that integrate information from different modalities. Subsequently, convolution operations reshape the inputs into $X \in \mathbb{R}^{C\times HW}$. The memory unit M1 then expands the feature map to $X \in \mathbb{R}^{4C\times HW}$, followed by memory unit M2 restoring it to $X \in \mathbb{R}^{C\times HW}$.

$$X_1' = M2(M_1(Conv(Sobel(X_1)))) \tag{10}$$

$$X_2' = M2(M_1(Conv(Sobel(X_2)))) \tag{11}$$

In this context, Conv denotes a 1×1 convolution, and the memory units M_1 and M_2 share parameters. These units are designed to map input features to a higher-dimensional space, facilitating the learning of global feature representations. The shared parameters allow the external attention mechanism to compute and apply correlations between images, achieving bidirectional enhancement and fusion of

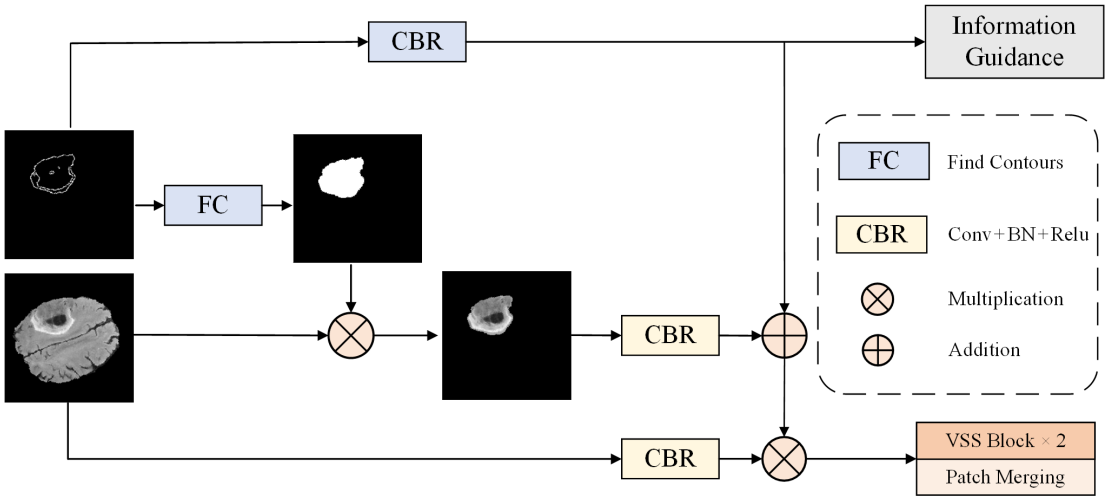

Figure 6. Overview of the Information Guidance Module.

features, thereby improving the model's performance in processing multimodal information. After these operations, the feature map is restored to the original image dimensions and is connected with the original image through residual connections. Finally, convolution operations are applied to fuse the concatenated information.

$$X_1'' = X_1 + Conv(X_1') \tag{12}$$

$$X_2'' = X_2 + Conv(X_2') \tag{13}$$

$$Output = Conv(Concat(X_1', X_2')) \tag{14}$$

In this manner, the BIEM effectively integrates features from diverse information sources, thereby enhancing the model's capacity to understand and analyze boundary information.

3.4. Information Guidance Module

In multimodal medical image processing, content segmentation often suffers from decreased accuracy due to the complex shapes of target regions and blurred boundary information, which consequently impacts diagnostic precision. Therefore, effectively utilizing boundary information is crucial for improving segmentation accuracy and diagnostic reliability. Boundary information provides structural cues that assist the model in accurately localizing target regions. Based on this, we propose a boundary Information Guidance Module (IGM).

The working principle of the IGM is illustrated in Figure 6. We collectively refer to convolution, batch normalization, and ReLU activation as the CBR module. Initially, the boundary map Y is processed through the CBR module to generate Output1, which is then passed to the subsequent Information Guidance Module:

$$Output_1 = \mathrm{CBR}(\mathrm{Y}) \tag{15}$$

Simultaneously, the Find Contours operation is employed to extract contour information from the image, resulting in a lesion region contour map Y' composed of 0 and 1. This contour map can be viewed as an attention map, aiding the model in focusing on critical regions:

$$Y' = FC(Y) \tag{16}$$

Next, the contour map Y' is element-wise multiplied with the original image Z to produce an image Y'', where non-lesion regions are set to 0. This enhances the segmentation effect of the lesion regions.

$$Y'' = Y' * Z \tag{17}$$

Then, the CBR operation is applied to this image, and the result is added to Output1. Finally, the summed result is element-wise multiplied with the CBR processed original image to obtain the information-guided output Output_2, which is then fed back into the Mamba encoding module:

$$I = Output_1 + CBR(Y'') \tag{18}$$

$$I' = I + CBR(Z) \tag{19}$$

$$Output_2 = I * I' \tag{20}$$

Through this process, the module effectively utilizes boundary information to guide the model's focus on lesion areas, thereby enhancing the accuracy of segmentation and responsiveness to critical regions.

3.5. Loss Function

This study incorporates two loss functions. The first, termed $\mathrm{L}_{\mathrm{boundary}}$, is designed to extract boundary information by combining binary cross-entropy with Dice loss. The second, termed $\mathrm{L}_{\mathrm{seg}}$, integrates cross-entropy and Dice loss to regularize content segmentation.

$$L_{boundary} = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(\hat{y}_i) + (1- y_i \log(1-\hat{y}_i) + \left(1 - \frac{2|y \cap \hat{y}|}{|y| + |\hat{y}|}\right) \tag{21}$$

$$L_{seg} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} y_{i,c} \log(\hat{y}_{i,c}) + \left(1 - \frac{2|y \cap \hat{y}|}{|y| + |\hat{y}|}\right) \tag{22}$$

In $L_{boundary}$, N represents the number of samples, y_i denotes the ground truth label of the i-th sample, and $\hat{y}_i$ is the predicted label. Both of these values are binary, indicating whether the sample is a boundary. On the other hand, L_{seg} addresses a multi-class classification problem, where C represents the total number of categories. For the i-th sample, $y_{i,c}$ denotes its ground truth label for the C-th category, while $\hat{y}_{i,c}$ is the corresponding predicted label.

4. Experimental Details

4.1. Datasets and Evaluation Metrics

In this study, we utilized the BraTS 2019 and BraTS 2020 glioma tumor MRI datasets [17, 2, 3] to evaluate the performance of our model. Each sample includes MRI images from the FLAIR, T1, T1 contrast-enhanced, and T2. The label information covers four main regions: healthy tissue, necrotic and non-enhancing tumor regions, edema regions, and enhancing tumor regions. Specifically, the whole tumor (WT) region encompasses all tumor areas, including necrotic non-enhancing tumor, edema, and enhancing tumor. Conversely, the tumor core (TC) region comprises the necrotic parts of the non-enhancing tumor and the enhancing tumor (ET) region.

In this study, we employed the Dice Similarity Coefficient (DSC) and Hausdorff Distance 95 (HD95) to evaluate the segmentation performance of the model on the WT, ET, and TC regions. The DSC is used to measure the overlap between the segmentation result P1 and the ground truth T1. The calculation formula is as follows:

$$DSC(P1, T1) = \frac{2|P1 \cap T1|}{|P1| + |T1|} \tag{23}$$

Here, P1 represents the predicted segmentation result, while T1 denotes the ground truth. The Hausdorff Distance between two surfaces, A and B, can be computed as follows:

$$H(A, B) = \max(\mathrm{d}(A, B), d(B, A)) \tag{24}$$

The calculation formulas for d(A,B) and d(B,A) are as follows:

$$\mathrm{d(A, B)} = \min(||\mathrm{b} - \mathrm{a}||) \tag{25}$$

$$\mathrm{d(B, A)} = \min(||\mathrm{b} - \mathrm{a}||) \tag{26}$$

Here, d(A, B) and d(B, A) represent the one-way Hausdorff distance from set A to set B, and from set B to set A, respectively. $||\mathrm{X} - \mathrm{Y}||$ denotes the Euclidean distance between point sets X and Y.

4.2. Comparative Experiment

To validate the efficacy of our model, we compared the experimental results with various mainstream medical image segmentation models, including U-Net and its variants (U-Net[20], ResUNet++[13], TransUNet[6], Swin-Unet[5]), as well as Mamba-based segmentation models (VM-UNet [21], Swin-UMamba [15], LightM-UNet[14], VM-UNET-V2[26]).

Table 1 illustrates the performance of our model on the BraTS 2019 and 2020 datasets, while Figures 7 and 8 provide a visual representation of the segmentation results for various models. As shown in Table 1 for the BraTS 2020 dataset, the proposed segmentation model demonstrates superior performance in terms of the Dice and HD95 metrics. Notably, the model achieves the best performance in tumor regions WT, TC, and ET, with improvements of 5.29%, 0.79%, and 0.28%, respectively, over the second-best competitor. This indicates that our model effectively aligns predicted results with ground-truth segmentation labels. Additionally, in terms of the HD95 metric, our model shows outstanding performance in the three lesion regions, with improvements of 0.31, 0.42, and 0.32, respectively, compared to the second-best model. This success can be attributed to the boundary-guided segmentation approach we proposed, which enhances the model's focus on edge information in lesion areas, thereby improving the HD95 metric.

Figures 7 and 8 present the comparative segmentation results of MBGNet and other segmentation models across different scenarios, intuitively demonstrating the performance superiority of MBGNet. To more clearly highlight the differences in segmentation performance, regions with significant discrepancies are enlarged and marked with red boxes. In Figure 7, we compare the segmentation results of MBGNet with the traditional U-Net and its derived versions. From Sample 1 and Sample 2, it can be observed that MBGNet exhibits high sensitivity to small lesion areas, accurately capturing these regions while maintaining consistency with the ground truth labels. In Sample 3 and Sample 4, MBGNet showcases its remarkable capability to analyze the shapes and boundaries of complex lesion structures, delivering clear delineation of individual lesion regions while ensuring both integrity and precision.

Figure 8 illustrates the performance comparison between MBGNet and segmentation models based on the Mamba architecture in other scenarios. As observed in Sample 1 and Sample 2, MBGNet demonstrates an excellent ability to model boundary details, accurately capturing intricate

Table 1. Performance Metrics of Various Models on the BraTS 2019/2020 Dataset.

Datasets	Models	DICE				HD95			
		WT	TC	ET	Average	WT	TC	ET	Average
BraTS2019	Unet	80.94	77.13	67.56	75.21	8.87	9.77	5.63	8.09
	ResUNet++	80.51	79.48	71.19	77.06	7.91	7.01	6.56	7.16
	TransUnet	79.76	79.05	69.94	76.25	8.57	7.83	6.07	7.49
	Swin-Unet	85.09	80.59	72.40	79.36	8.03	7.06	6.12	7.07
	VM-UNet	85.27	86.34	80.45	84.02	5.95	6.15	5.01	5.70
	Swin-UMamba	85.67	86.56	80.89	84.37	5.78	5.84	4.92	5.51
	LightM-UNet	83.45	85.90	79.67	82.99	5.89	6.07	5.14	5.70
	VM-UNET-V2	87.89	86.78	81.23	85.30	5.76	5.93	5.08	5.59
	MBGNet	**90.57**	**89.23**	**84.72**	**88.17**	**5.50**	**5.67**	**4.48**	**5.21**
BraTS2020	Unet	81.73	77.31	68.02	75.69	8.03	9.76	5.50	7.76
	ResUNet++	81.27	80.51	71.20	77.66	8.01	7.53	6.62	7.39
	TransUnet	80.13	79.76	69.94	76.61	8.42	7.64	5.91	7.32
	Swin-Unet	85.19	80.74	73.01	79.65	8.00	7.09	5.94	7.01
	VM-UNet	82.58	88.17	81.81	84.18	5.92	6.14	4.57	5.54
	Swin-UMamba	85.93	86.09	83.19	85.07	5.65	5.98	4.97	5.53
	LightM-UNet	83.14	86.17	83.76	84.35	6.02	6.54	5.03	5.86
	VM-UNET-V2	86.09	85.73	83.48	85.10	5.87	6.32	5.14	5.77
	MBGNet	**91.38**	**88.96**	**84.04**	**88.12**	**5.34**	**5.65**	**4.25**	**5.08**

Bolding indicates the best performing model and underlining indicates the second best performing model. The same applies to subsequent tables.

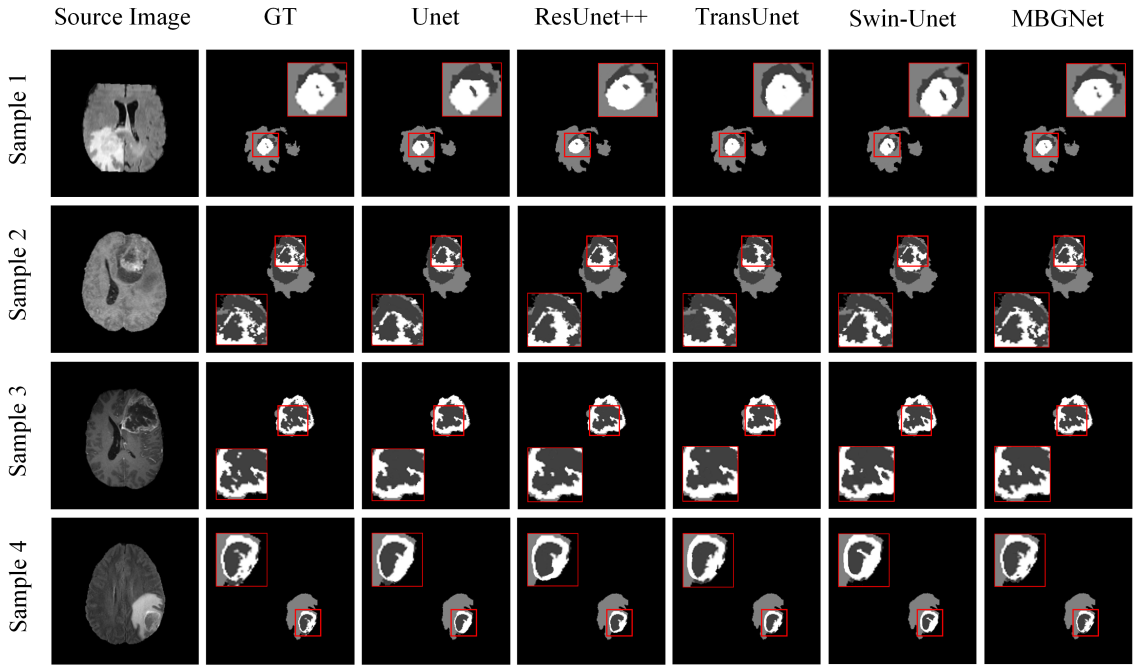

Figure 7. Visual Results of Various Segmentation Models on the BraTS2019 Dataset.

boundary features and achieving a high degree of consistency with the ground truth lesion labels. In Sample 3 and Sample 4, even in cases where certain lesion regions in the original images are difficult to discern, MBGNet is still able to accurately segment these areas and clearly distinguish between different lesions. This highlights its robust capability to extract lesion features from low-quality images. These results indicate that MBGNet exhibits strong robustness and reliability when processing various types of lesions. Furthermore, its adaptability enables it to meet the demands of segmentation tasks in diverse and complex scenarios.

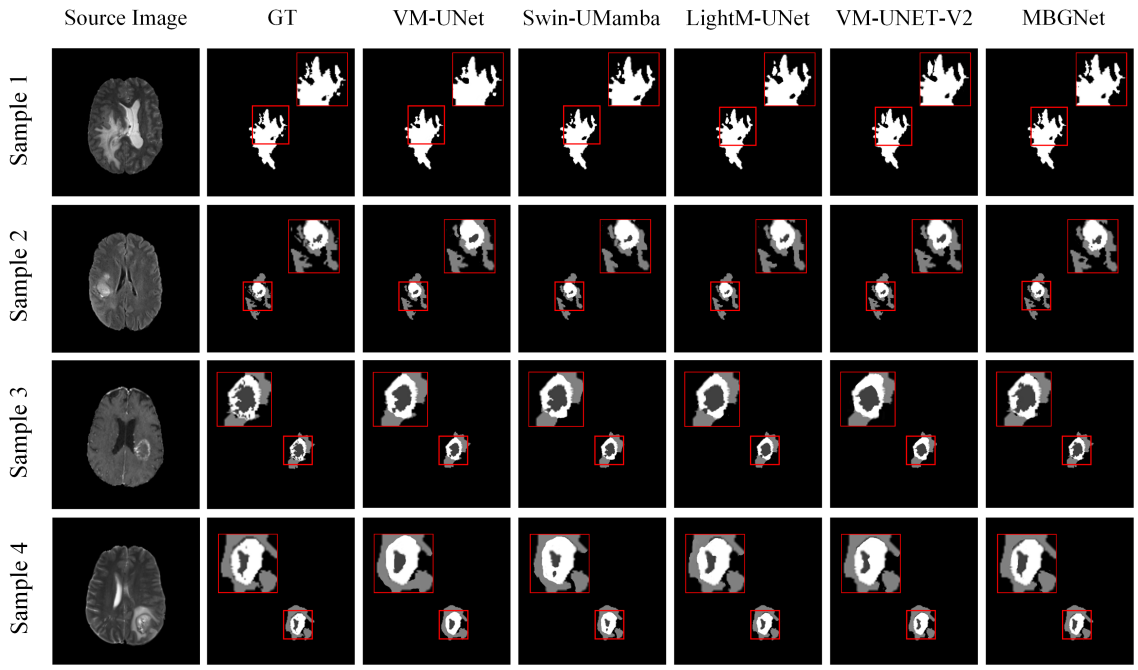

Figure 8. Visual Results of Various Segmentation Models on the BraTS2020 Dataset.

4.3. Ablation Study

To validate the efficacy of the proposed boundary-guided segmentation approach, this section conducts detailed ablation studies on each individual module. By analyzing the performance of these modules one by one, we can assess their impact on the overall model performance, thereby gaining a deeper understanding of the advantages of the boundary-guided segmentation method. The comparisons in this subsection include the entire Boundary Information Extraction Network (BIEN) from the first phase, the Boundary Information Encoding Module (BIEM), and the Information Guidance Module (IGM).

As indicated in Table 2, the model performance decreases on both datasets when all modules are removed, as well as when only BIEN, BIEM, or IGM is retained. The specific reasons are as follows: Firstly, the absence of BIEN impairs the model's ability to locate and distinguish complex structures, diminishing its robustness against morphological variations and irregular boundaries, ultimately affecting the overall segmentation quality. Secondly, the lack of BIEM results in the loss of boundary information extraction and interaction between multi-modal boundary information. This challenge hampers the model's ability to identify clear segmentation boundaries and weakens the correlation between different modalities. Lastly, when IGM is removed, content segmentation lacks the guidance of boundary information, potentially leading to the omission of subtle lesion features, resulting in less precise segmentation of the lesion areas. In summary, these modules play a crucial role in the overall performance of the model, and their absence significantly negatively impacts the segmentation outcomes.

Figures 9 and 10 present the visualized results of ablation experiments conducted on different datasets using MBGNet. To more intuitively highlight the differences in segmentation performance, regions with significant discrepancies are enlarged and marked with red boxes. From the comparative results shown in Figures 9 and 10, it can be observed that when all modules are removed, the segmentation performance of the model declines significantly, barely maintaining a rudimentary segmentation framework with noticeably reduced accuracy. Upon the introduction of the BIEN module, the model's boundary capture capabilities are substantially enhanced, enabling clear delineation of lesion boundaries and avoiding boundary ambiguity or confusion between different lesion types. When the BIEM module is incorporated, the model's ability to segment poorly expressed lesion areas in the original images is significantly improved. This improvement is attributed to the multi-modal interaction mechanism within the BIEM, which effectively enhances the model's understanding and grasp of global segmentation regions. With the addition of the IGM module, the model further improves its segmentation performance in complex regions. This improvement is due to the guidance provided by the boundary segmentation map within IGM, which enables the model to focus more effectively on critical regions requiring segmentation while accurately distinguishing between different lesion areas. Finally,

Table 2. Performance Metrics of Various Modules on the BraTS 2019/2020 Dataset.

Datasets	Modules				DICE				HD95			
	BIEN	BIEM	IGM	Mamba	WT	TC	ET	Average	WT	TC	ET	Average
BraTS2019	×	×	×	×	84.47	85.23	80.04	83.24	6.17	6.54	5.23	5.98
	✓	×	×	×	86.61	86.18	83.06	85.28	5.68	5.78	4.59	5.35
	×	✓	×	×	86.32	86.91	83.45	85.56	5.95	5.91	4.76	5.54
	×	×	✓	×	88.14	87.44	82.70	86.09	6.03	6.12	4.82	5.65
	×	×	×	✓	87.09	87.17	82.95	85.73	5.84	5.88	4.73	5.48
	✓	✓	✓	✓	**90.57**	**89.23**	**84.72**	**88.17**	**5.50**	**5.67**	**4.48**	**5.21**
BraTS2020	×	×	×	×	85.23	86.14	80.95	84.10	6.01	6.25	5.09	5.78
	✓	×	×	×	85.23	86.34	83.23	84.93	5.47	5.76	4.36	5.19
	×	✓	×	×	88.67	88.59	82.45	86.57	5.58	5.89	4.57	5.34
	×	×	✓	×	89.98	87.78	83.67	87.14	5.62	5.93	4.72	5.42
	×	×	×	✓	89.81	88.16	83.42	87.13	5.53	5.87	4.40	5.26
	✓	✓	✓	✓	**91.38**	**88.96**	**84.04**	**88.12**	**5.34**	**5.65**	**4.25**	**5.08**

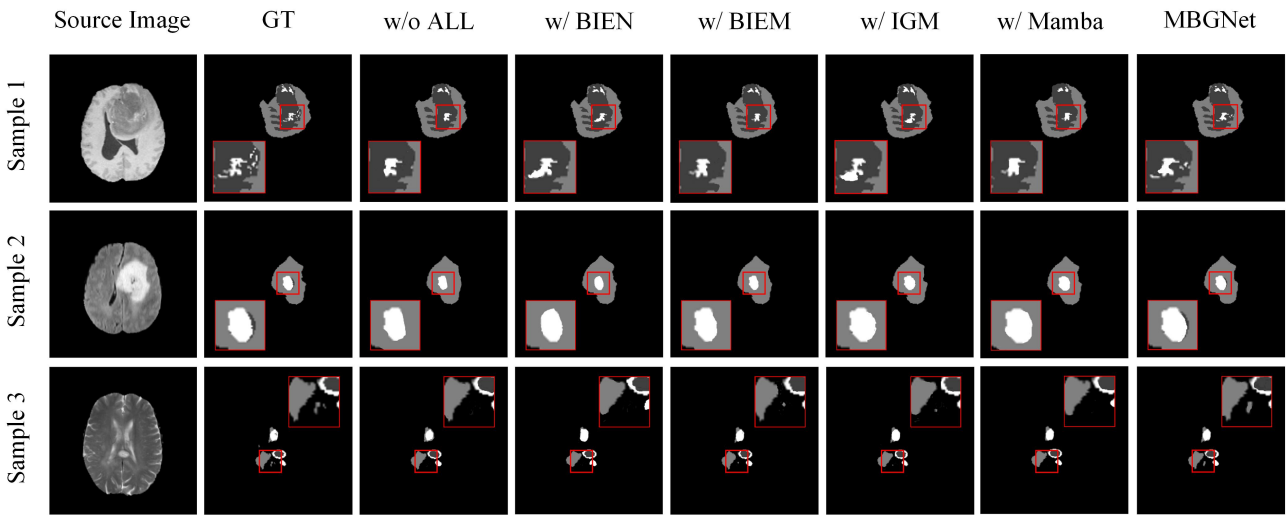

Figure 9. The ablation study results for the modules on the BraTS2019 dataset.

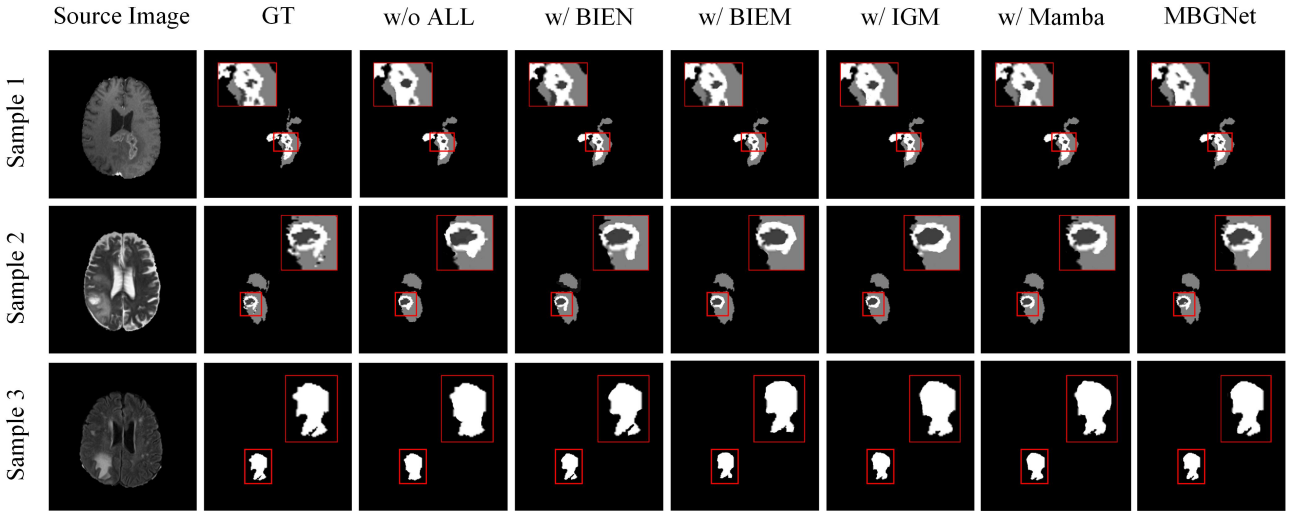

Figure 10. The ablation study results for the modules on the BraTS2020 dataset.

when all modules are integrated, the model demonstrates significant improvements in boundary information extraction, content segmentation, and regional delineation. This indicates that the modules exhibit strong synergistic effects, with each module contributing significantly to the overall performance enhancement. These findings validate the effectiveness of the module designs and underscore the superiority of the overall architecture.

5. Conclusion

In this study, we aim to enhance the accuracy and boundary identification capability in multimodal glioma image segmentation. To address these objectives, we focus on two pivotal issues: the precise extraction of boundary information and the prevention of information loss during content segmentation. Inaccurate boundary extraction can lead to blurred segmentation boundaries, while information loss affects the integrity of the segmented regions. To tackle these challenges, we employed Mamba as the backbone for feature extraction of multi-modal information and integrated a boundary extraction module to enhance the precision of boundary information. Additionally, we designed a BIEM that fuses multiple boundary information through an external attention mechanism, enabling the model to focus more intently on key boundary details within the image. Concurrently, to combat the problem of content information loss, we developed an IGM. This module uses the complete boundary segmentation map as a guidance map in content segmentation and, by integrating existing multi-modal information, effectively compensates for information that might be lost during the segmentation process. Our approach was systematically evaluated on the BraTS2019 and BraTS2020 datasets. The results demonstrate that, compared to current medical image segmentation models, our method exhibits superior performance in both boundary identification and overall segmentation accuracy.

In future research, we plan to further expand the application scope of the segmentation model MBGNet, with a particular focus on exploring its performance and applicability across various multimodal medical imaging datasets. Specifically, we aim to investigate the potential of MBGNet in handling more complex multimodal data scenarios, in order to validate its robustness and effectiveness in different types of medical imaging and disease diagnosis tasks.

6. Acknowledgement

This work was supported by the "Tianshan Innovation Team Program" of the Autonomous Region (Grant No.20231103582) and the Research Project of the Fundamental Research Funds for Universities in Xinjiang Uygur Autonomous Region (Grant No.XJEDU2023P012).

References

[1] A. M. G. Allah, A. M. Sarhan, and N. M. Elshennawy. Edge u-net: Brain tumor segmentation using mri based on deep u-net model with boundary information. *Expert Systems with Applications*, 213:118833, 2023. 3

[2] S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J. S. Kirby, J. B. Freymann, K. Farahani, and C. Davatzikos. Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features. *Scientific data*, 4(1):1–13, 2017. 7

[3] S. Bakas, M. Reyes, A. Jakab, S. Bauer, M. Rempfler, A. Crimi, R. T. Shinohara, C. Berger, S. M. Ha, M. Rozycki, et al. Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the brats challenge. *arXiv preprint arXiv:1811.02629*, 2018. 7

[4] N.-T. Bui, D.-H. Hoang, Q.-T. Nguyen, M.-T. Tran, and N. Le. Meganet: Multi-scale edge-guided attention network for weak boundary polyp segmentation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 7985–7994, 2024. 3

[5] H. Cao, Y. Wang, J. Chen, D. Jiang, X. Zhang, Q. Tian, and M. Wang. Swin-unet: Unet-like pure transformer for medical image segmentation. In *European conference on computer vision*, pages 205–218. Springer, 2022. 7

[6] J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, Y. Wang, L. Lu, A. L. Yuille, and Y. Zhou. Transunet: Transformers make strong encoders for medical image segmentation. *arXiv preprint arXiv:2102.04306*, 2021. 7

[7] L. Chen, Z. Qu, Y. Zhang, J. Liu, R. Wang, and D. Zhang. Edge enhanced gciffnet: A multiclass semantic segmentation network based on edge enhancement and multiscale attention mechanism. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2024. 3

[8] W. Dong, H. Zhu, S. Lin, X. Luo, Y. Shen, X. Liu, J. Zhang, G. Guo, and B. Zhang. Fusion-mamba for cross-modality object detection. *arXiv preprint arXiv:2404.09146*, 2024. 3

[9] M.-H. Guo, Z.-N. Liu, T.-J. Mu, and S.-M. Hu. Beyond self-attention: External attention using two linear layers for visual tasks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(5):5436–5447, 2022. 5

[10] B. Hu, P. Zhou, H. Yu, Y. Dai, M. Wang, S. Tan, and Y. Sun. Leanet: Lightweight u-shaped architecture for high-performance skin cancer image segmentation. *Computers in Biology and Medicine*, 169:107919, 2024. 5

[11] J. Huang, L. Yang, F. Wang, Y. Wu, Y. Nan, A. I. Aviles-Rivero, C.-B. Schönlieb, D. Zhang, and G. Yang. Mambamir: An arbitrary-masked mamba for joint medical image reconstruction and uncertainty estimation. *arXiv preprint arXiv:2402.18451*, 2024. 2

[12] T. Huang, X. Pei, S. You, F. Wang, C. Qian, and C. Xu. Localmamba: Visual state space model with windowed selective scan. *arXiv preprint arXiv:2403.09338*, 2024. 3

[13] D. Jha, P. H. Smedsrud, M. A. Riegler, D. Johansen, T. De Lange, P. Halvorsen, and H. D. Johansen. Resunet++: An advanced architecture for medical image segmentation.

In *2019 IEEE international symposium on multimedia (ISM)*, pages 225–2255. IEEE, 2019. 7

[14] W. Liao, Y. Zhu, X. Wang, C. Pan, Y. Wang, and L. Ma. Lightm-unet: Mamba assists in lightweight unet for medical image segmentation. *arXiv preprint arXiv:2403.05246*, 2024. 2, 7

[15] J. Liu, H. Yang, H.-Y. Zhou, Y. Xi, L. Yu, C. Li, Y. Liang, G. Shi, Y. Yu, S. Zhang, et al. Swin-umamba: Mamba-based unet with imagenet-based pretraining. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 615–625. Springer, 2024. 7

[16] J. Ma, F. Li, and B. Wang. U-mamba: Enhancing long-range dependency for biomedical image segmentation. *arXiv preprint arXiv:2401.04722*, 2024. 2

[17] B. H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, Y. Burren, N. Porz, J. Slotboom, R. Wiest, et al. The multimodal brain tumor image segmentation benchmark (brats). *IEEE transactions on medical imaging*, 34(10):1993–2024, 2014. 7

[18] X. Pei, T. Huang, and C. Xu. Efficientvmamba: Atrous selective scan for light weight visual mamba. *arXiv preprint arXiv:2403.09977*, 2024. 3

[19] Y. Qiao, Z. Yu, L. Guo, S. Chen, Z. Zhao, M. Sun, Q. Wu, and J. Liu. Vl-mamba: Exploring state space models for multimodal learning. *arXiv preprint arXiv:2403.13600*, 2024. 3

[20] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*, pages 234–241. Springer, 2015. 7

[21] J. Ruan and S. Xiang. Vm-unet: Vision mamba unet for medical image segmentation. *arXiv preprint arXiv:2402.02491*, 2024. 2, 7

[22] J. Ruan, S. Xiang, M. Xie, T. Liu, and Y. Fu. Malunet: A multi-attention and light-weight unet for skin lesion segmentation. In *2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 1150–1156. IEEE, 2022. 5

[23] Z. Wan, Y. Wang, S. Yong, P. Zhang, S. Stepputtis, K. Sycara, and Y. Xie. Sigma: Siamese mamba network for multi-modal semantic segmentation. *arXiv preprint arXiv:2404.04256*, 2024. 2

[24] Z. Xing, T. Ye, Y. Yang, G. Liu, and L. Zhu. Segmamba: Long-range sequential modeling mamba for 3d medical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 578–588. Springer, 2024. 2

[25] Z. Yang, D. Lin, D. Ni, and Y. Wang. Recurrent feature propagation and edge skip-connections for automatic abdominal organ segmentation. *Expert Systems with Applications*, 249:123856, 2024. 3

[26] M. Zhang, Y. Yu, S. Jin, L. Gu, T. Ling, and X. Tao. Vm-unet-v2: rethinking vision mamba unet for medical image segmentation. In *International Symposium on Bioinformatics Research and Applications*, pages 335–346. Springer, 2024. 7

[27] L. Zhu, B. Liao, Q. Zhang, X. Wang, W. Liu, and X. Wang. Vision mamba: Efficient visual representation learning with bidirectional state space model. *arXiv preprint arXiv:2401.09417*, 2024. 2

[28] Q. Zhu, Y. Cai, Y. Fang, Y. Yang, C. Chen, L. Fan, and A. Nguyen. Samba: Semantic segmentation of remotely sensed images with state space model. *Heliyon*, 2024. 2

[29] Z. Zhu, X. He, G. Qi, Y. Li, B. Cong, and Y. Liu. Brain tumor segmentation based on the fusion of deep semantics and edge information in multimodal mri. *Information Fusion*, 91:376–387, 2023. 2

[30] Z. Zhu, X. Ma, W. Wang, S. Dong, K. Wang, L. Wu, G. Luo, G. Wang, and S. Li. Boosting knowledge diversity, accuracy, and stability via tri-enhanced distillation for domain continual medical image segmentation. *Medical Image Analysis*, 94:103112, 2024. 1