Mathematical Biosciences and Engineering

http://www.aimspress.com/journal/mbe

MBE, 20(12): 20510–20527.
DOI: 10.3934/mbe.2023907
Received: 25 September 2023
Revised: 02 November 2023
Accepted: 06 November 2023
Published: 13 November 2023

Research article

MAU-Net: Mixed attention U-Net for MRI brain tumor segmentation

Yuqing Zhang[1,2], **Yutong Han**[1,2] **and Jianxin Zhang**[1,2,3,*]

[1] School of Computer Science and Engineering, Dalian Minzu University, Dalian 116600, China
[2] Institute of Machine Intelligence and Biocomputing, Dalian Minzu University, Dalian 116600, China
[3] SEAC Key Laboratory of Big Data Applied Technology, Dalian Minzu University, Dalian 116600, China

* **Correspondence:** Email: jxzhang0411@163.com; Tel: +8641187557005.

Abstract: Computer-aided brain tumor segmentation using magnetic resonance imaging (MRI) is of great significance for the clinical diagnosis and treatment of patients. Recently, U-Net has received widespread attention as a milestone in automatic brain tumor segmentation. Following its merits and motivated by the success of the attention mechanism, this work proposed a novel mixed attention U-Net model, i.e., MAU-Net, which integrated the spatial-channel attention and self-attention into a single U-Net architecture for MRI brain tumor segmentation. Specifically, MAU-Net embeds Shuffle Attention using spatial-channel attention after each convolutional block in the encoder stage to enhance local details of brain tumor images. Meanwhile, considering the superior capability of self-attention in modeling long-distance dependencies, an enhanced Transformer module is introduced at the bottleneck to improve the interactive learning ability of global information of brain tumor images. MAU-Net achieves enhancing tumor, whole tumor and tumor core segmentation Dice values of 77.88/77.47, 90.15/90.00 and 81.09/81.63% on the brain tumor segmentation (BraTS) 2019/2020 validation datasets, and it outperforms the baseline by 1.15 and 0.93% on average, respectively. Besides, MAU-Net also demonstrates good competitiveness compared with representative methods.

Keywords: brain tumor segmentation; transformer; attention mechanism; U-Net

1. Introduction

Glioblastoma, as the most prevalent malignant tumor in central nervous system [1], can be classified into high-grade and low-grade tumors. High-grade gliomas grow rapidly and easily form abnormal tissue, resulting in low survival rates and extremely high surgical risks. In contrast, low-grade gliomas grow at a slower pace, and surgical resection can not only halt their progression but also reduce the

pressure on brain tissue. Therefore, early diagnosis of gliomas is crucial in improving the patient's condition. Glioblastoma commonly is evaluated and diagnosed using multimodal images obtained from magnetic resonance imaging (MRI) [2]. Since manual segmentation of tumors by medical experts is a time-consuming and subjective process [3], there has been a growing interest in computer-aided MRI brain tumor segmentation. Nevertheless, achieving precise automated brain tumor segmentation remains a formidable challenge due to the substantial variations in tumor location, size and morphology among individual patients.

In recent years, there has been rapid development of deep learning in various fields, including computer vision and medical image analysis. Despite the successful application of various deep learning methods [4,5] or improved algorithms [6] to MRI brain tumor segmentation tasks, the U-Net model [7] remains the mainstream approach for this task. The U-Net architecture mainly consists of downsampling layers for feature encoding, upsampling layers for feature recovery, and skip connections for information preservation. Its structure is concise and flexible, and can achieve higher segmentation accuracy with less data. To exploit the inherent spatial information of volumetric data and enhance the contextual perception of 3D images, researchers extend the original 2D U-Net and propose a 3D U-Net [8] that is more suitable for MRI brain tumor image segmentation. Moreover, the potential of attention mechanisms to enhance the performance of deep learning models has been acknowledged, and many researchers have explored their integration into the U-Net to improve the brain tumor segmentation accuracy. Among them, Sun et al. [9] introduced additive attention in the fusion of encoder features and features at different levels, as well as at the skip connections, enabling more effective learning of crucial brain tumor features. Zhang et al. [10] proposed the use of attention gate (AG) modules to automatically focus on features of different shapes and sizes, integrating them with residual blocks in the U-Net for more precise brain tumor image segmentation. Meanwhile, Akbar et al. [11] introduced a multipath residual attention block in the encoder, combining fused attention and dilated convolutions, effectively enhancing the network's ability to capture features of small tumors. Furthermore, several other segmentation networks have also incorporated attention modules [12,13], yielding relatively impressive segmentation outcomes. While the aforementioned studies primarily focus on reinforcing the attention on local features in MRI brain tumor images, the expression of global long-distance dependency features in brain tumor images is not addressed by them. Drawing inspiration from the success of the Transformer [14] in various natural language processing and computer vision tasks, Wang et al. [15] put forward the integration of Transformer modules into the 3D U-Net to enhance MRI brain tumor image segmentation and achieve significant improvements in accuracy. Subsequently, Jiang et al. [16] fused the Swin Transformer with convolutional neural networks (CNN) in both the encoder and decoder stages, while additionally incorporating an enhanced Transformer at the bottleneck to facilitate detailed feature extraction of brain tumors, and experimental results also demonstrate its effectiveness. In addition, the segmentation performance of brain tumor images was also improved by Xu et al. [17] through the integration of CNN and Transformer into a hybrid feature extraction network.

Influenced by the achievements of the attention mechanism in MRI brain tumor segmentation and acknowledging the constraints of a solitary attention module in capturing both local and global feature information of brain tumors comprehensively, this study explores a mixed attention-based approach for MRI brain tumor segmentation within the mainstream 3D U-Net framework. The proposed method integrates both spatial-channel attention and self-attention mechanisms into the network,

aiming to enhance the capacity for extracting local detailed features and facilitating the interaction of global contextual information. Specifically, the learning of local semantic features in the down-sampling regions is enhanced by introducing the Shuffle Attention (SA) module [18] after each convolutional block in the encoder stage. Furthermore, an enhanced Transformer module (ETrans) [19] is introduced at the bottleneck of the 3D U-Net model to strengthen the learning of global semantic information and broaden the receptive field for brain tumors. Consequently, a novel hybrid attention-based 3D U-Net model, named MAU-Net, is established. Experimental validation on the BraTS2019 and BraTS2020 datasets confirms the effectiveness of the proposed MAU-Net method for brain tumor segmentation, as demonstrated by the results of ablation experiments and comparative experiments, which also highlight its competitiveness compared to other representative methods. In summary, this paper makes the following key contributions: 1) We propose a novel model called MAU-Net for brain tumor segmentation tasks. MAU-Net not only pays better attention to the global contextual information, but also learns more intricate local semantic features, thereby enhancing the overall performance of the network model. 2) The original 3D U-Net architecture is extended by MAU-Net, which incorporates the ETrans module at the bottleneck to enable the interaction of distant information. Additionally, the SA module is introduced after each consecutive convolution block in the encoder, thereby bolstering the extraction capability of local information. 3) Extensive evaluation experiments are conducted on the BraTS2019 and BraTS2020 benchmark datasets for brain tumor segmentation. The experimental results demonstrate that MAU-Net surpasses its baseline and also exhibits competitive performance compared to other representative segmentation methods.

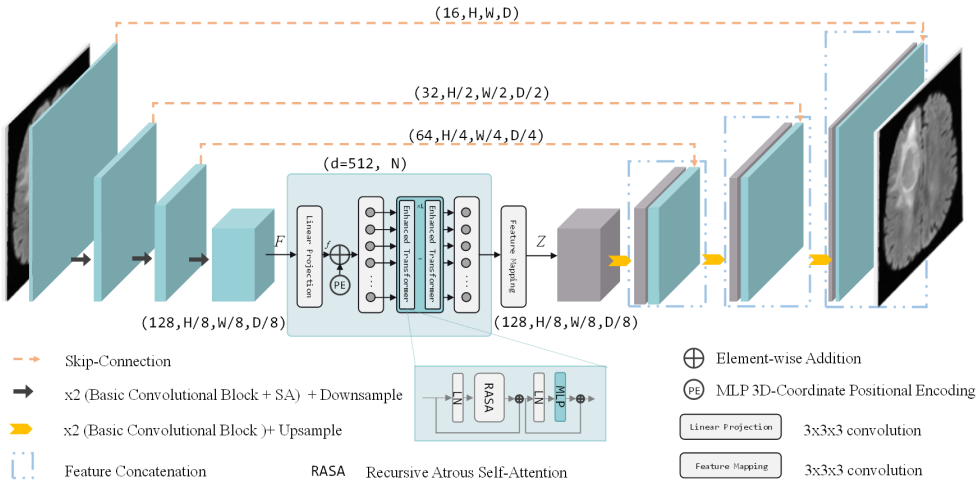

Figure 1. The overall architecture of brain tumor segmentation network based on mixed attention.

2. Methods

2.1. Overall framework

The proposed MAU-Net network is depicted in Figure 1, illustrating its overall architecture. The network model is composed of four layers, comprising of an encoder path, four bottleneck blocks, a decoder path and three skip connections. The input data size of the network is $4 \times 128 \times 128 \times 128$,

indicating an image size of 128 × 128 × 128 with four channels. As each data in the training dataset contains four modalities, the four channels are fed into the network for learning during training. First, to address the limitation of ordinary convolution blocks in the encoder stage, which may fail to fully capture contextual information, SA modules are incorporated after each ordinary convolution block in the 3D U-Net encoder stage to enhance the semantic feature learning capability of local regions. As the data progresses through the down-sampling layers, the size is reduced by half while the number of channels is doubled. Second, in order to address the inadequate learning of global semantic information in CNNs, ETrans is introduced at the bottleneck block to model long-distance dependencies in the global space. Ultimately, high-resolution segmentation results are generated by progressively recovering features through the repeated stacking of ordinary convolution blocks and up-sampling layers in the decoder stage. The specifics of the added SA modules and ETrans modules will be elaborated upon in subsequent sections.

2.1.1. SA

In the context of CNNs, ordinary convolutions have limited receptive fields, which can lead to the neglect of important local detailed features. However, in the case of brain tumor segmentation tasks, capturing the detailed characteristics of tumor images becomes particularly crucial. Attention mechanisms offer a solution by enabling neural networks to focus accurately on all relevant elements of the input, thereby enhancing the performance of deep neural networks. Therefore, in order to augment the local feature extraction capability of the encoder stage and obtain more detailed information features, SA is introduced after consecutive convolution blocks. Additionally, since MRI brain tumor images are in 3D format, the 2D SA module is extended to a 3D form to achieve better segmentation results. The specific architecture diagram of the 3D SA module is depicted in Figure 2.

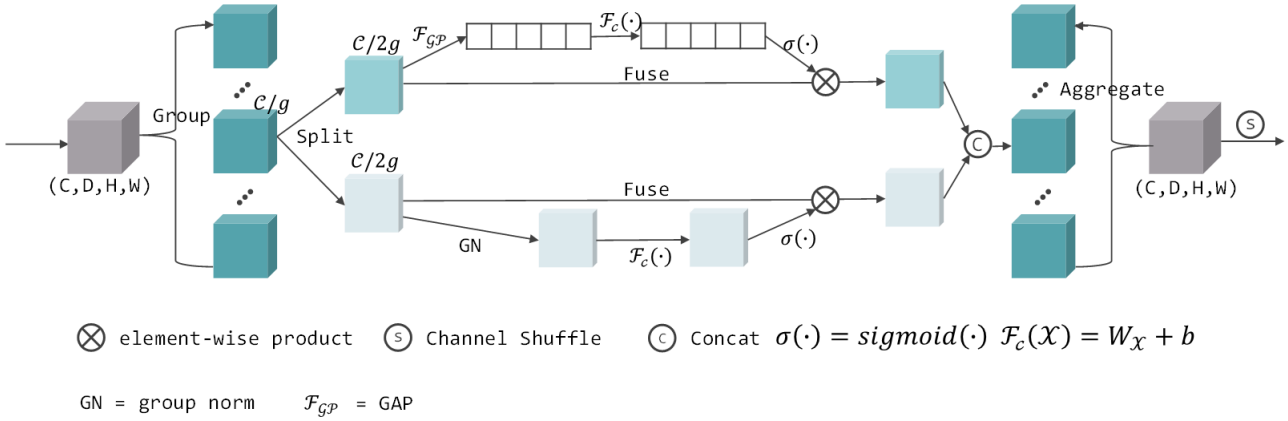

Figure 2. The structure of the SA module.

In the realm of computer vision, two frequently utilized attention mechanisms are channel attention and spatial attention. Nevertheless, exclusively relying on either channel or spatial attention may lead to the neglect of crucial channel or spatial information. To overcome this limitation, prior research endeavors such as GCNet [20] and CBAM [21] have amalgamated spatial attention and channel attention into a unified module, resulting in notable enhancements. However, these approaches still encounter computational burdens. Li et al. [22] proposed Spatial Group-wise

Enhance (SGE) to group the inputs into multiple sub-features according to the dimension of the channel features to represent different semantics, and applied spatial mechanisms to each feature group by scaling the feature vector at all positions using an attention mask. Ma et al. [23] proposed ShuffleNet v2 in which "channel shuffle" operations were defined to facilitate information communication between different branches. Drawing inspiration from aforementioned works, Zhang et al. proposed the SA module, which is highly efficient and lightweight while effectively combining both types of attention mechanisms.

Specifically, the SA module initially divides the input equally into multiple sub-features along the channel dimension and processes them in parallel. For each sub-feature, it is further divided into two branches along the channel dimension, and attention operations are performed on both the spatial and channel dimensions before merging them. Subsequently, all sub-features are aggregated, and "channel shuffle" operations are employed to facilitate information communication between different sub-features. It should be noted that, in relation to each sub-feature of the two branches, the branch that undergoes the channel attention operation can be represented as:

$$s = F_{gp}(X_{k1}) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} X_{k1}(i, j) \tag{2.1}$$

$$X_{k1}^{'} = \sigma(F_c(s)) \cdot X_{k1} = \sigma(W_1 s + b_1) \cdot X_{k1} \tag{2.2}$$

First, the input X_{k1} undergoes global average pooling (GAP) to generate channel-wise statistical information, denoted as s. Subsequently, a gate mechanism with a *sigmoid* activation function is employed to achieve a precise and compact feature output. In this process, parameters W_1 and b_1 are utilized for scaling and biasing operations. Meanwhile, the branch that undergoes spatial attention operation is represented as follows:

$$X_{k2}^{'} = \sigma(W_2 \cdot GN(X_{k2}) + b_2) \cdot X_{k2} \tag{2.3}$$

In this branch, the input X_{k2} undergoes GroupNorm (GN) processing to acquire spatial statistical information. Subsequently, $F_c(\cdot)$ is applied to enhance the representation of this information. Finally, adaptive output is obtained by utilizing a *sigmoid* activation function. The correlation between spatial and channel attention is fully exploited by the SA module, thereby reducing the computational burden. The integration of the SA module into the encoder stage of the network effectively enhances the feature learning of local semantic regions.

2.1.2. ETrans

The significance of global information in brain tumor segmentation cannot be overstated. Acknowledging the advantages of the self-attention mechanism in capturing long-distance dependencies in images, the integration of Transformer modules enhances the interaction of global information in brain tumor images, resulting in more precise tumor segmentation. However, the capability of the Transformer based on ordinary self-attention is limited in small-scale networks. To address this limitation, the ETrans module with enhanced self-attention layers is incorporated at the bottleneck of the model to bolster the feature representation of global context. Additionally, to accommodate the 3D model architecture, the ETrans module is modified from a 2D to a 3D form. The specific structure of the expanded ETrans module is depicted in Figure 3.

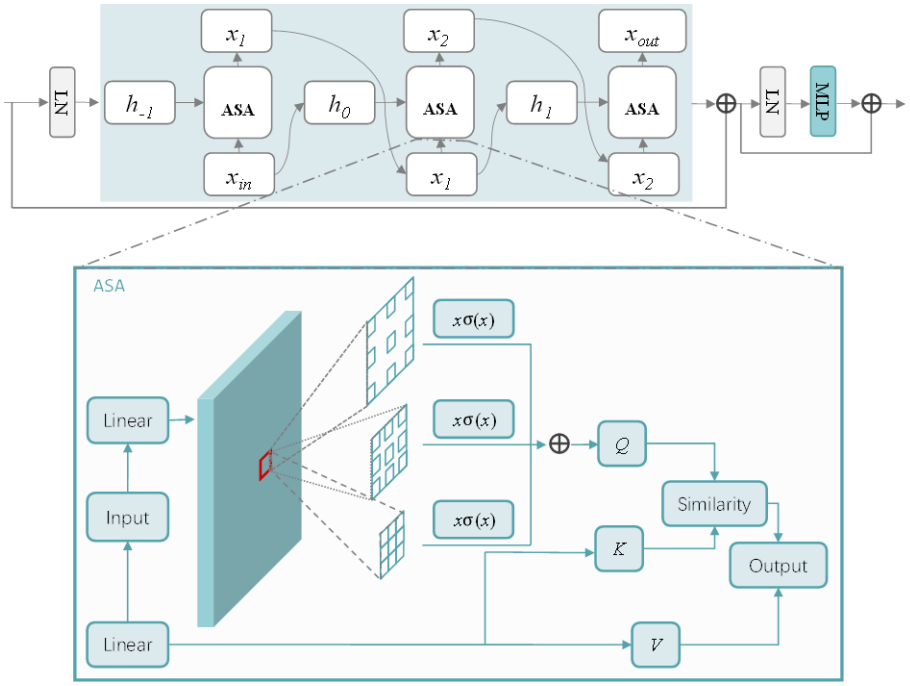

Figure 3. The structure of the enhanced transformer module.

The effectiveness of multiple-scale features in segmentation has been well established [24]. Simultaneously, dilated convolutions [25] provide a viable approach for capturing multi-scale contextual information while preserving the same parameter count as standard convolutions. Consequently, ETrans leverages weight-shared dilated convolutions in computing similarity mappings to enhance model performance. Furthermore, recursive operations are introduced to deepen the network and further improve its segmentation capability without increasing the parameter usage. Notably, the integration of multi-scale information in the generation of similarity weights can be succinctly summarized as follows:

$$\hat{Q} = Conv(X, W_q^{k=1}, r = 1, g = 1) \tag{2.4}$$

$$Q = \sum_{r\in\{1,3,5\}} SiLU(Conv(\hat{Q}, W_q^{k=3}, r, g = d)) \tag{2.5}$$

$$SiLU(m) = m \odot sigmoid(m) \tag{2.6}$$

where X, Q are the input features and the final query value, respectively, W_q^k is the convolution kernel weight and k, r, g, d are the convolution kernel size, dilation rate, convolution group number and channel number, respectively. By examining Eqs (2.4)–(2.6) and Figure 3, it becomes apparent that a linear projection is initially conducted through a $1 \times 1 \times 1$ convolution. Subsequently, three convolutions with different dilation rates but shared kernels are applied to capture multi-scale contextual information. To minimize parameter costs, the number of convolution groups is set to match the number of channels in the features. Ultimately, the parallel features from various scales are weighted and aggregated with Sigmoid Linear Unit (SiLU) [26,27] utilized to determine the weights

for each scale. This design allows the utilization of multi-scale information in the similarity calculations between any pair of spatial positions for queries and keys in self-attention. Simultaneously, the self-attention layer of ETrans is constructed by incorporating atrous self-attention (ASA) as a nonlinear activation function and employing a recursive approach, as depicted in the following formula:

$$x_{t+1} = ASA(F(X_t, h_{t-1})) \tag{2.7}$$

$$h_{t-1} = X_{t-1} \tag{2.8}$$

$$x_t = ASA(F(X_{t-1}, h_{t-2})) \tag{2.9}$$

where t, h represent the number of steps and hidden layer features, respectively, the initial hidden layer feature $h_{-1} = 0$. $F(X, h) = W_f X + U_f h$ is a linear function combining input features and hidden layer features and W_F,U_F are projection weights, both of which are one. To restrict computational costs, the default depth of recursion is set to two. Introducing an enhanced Transformer in the architecture effectively models global dependency relationships, while the inclusion of the SA module in the encoder stage enhances the capture capability of local features. The combination of finer-grained local information with global information contributes to improving the segmentation accuracy of brain tumors.

2.2. Loss function

Due to the significant imbalance between healthy tissues and tumor tissues in brain tumor MRI images, the task of segmenting brain tumors encounters a severe category imbalance issue. Consequently, the combination of dice loss L_{DC} and cross entropy loss L_{CE} is adopted as the loss function of the network model. The combined loss function can be defined as follows:

$$Loss = \partial L_{DC} + (1 - \partial) L_{CE} \tag{2.10}$$

where ∂ is the balance parameter from zero to one.

The dice loss function is extensively employed for medical image segmentation and facilitating effective learning from class-imbalanced samples [28]. Conversely, the cross-entropy loss is employed to tackle the issue of multitask imbalance by minimizing the discrepancy between training samples and balancing metric [29]. The combination of these two loss functions contributes to alleviating the problem of class imbalance to a certain extent. L_{DC} and L_{CE} can be defined as:

$$L_{DC} = 1 - \frac{2\sum_{i=0}^{N} y_i \hat{y}_l}{\sum_{i=0}^{N} (y_i + \hat{y}_l)} \tag{2.11}$$

$$L_{CE} = -\sum_{i=0}^{N} y_i log \hat{y}_l \tag{2.12}$$

where N denotes the total number of categories, y_i denotes the one-hot coding of category i and $\hat{y}_l$ denotes the correct prediction probability of category i.

3. Experiments and results

3.1. Experimental settings

The experiments are conducted using the PyTorch deep learning framework on NVIDIA A10 TENSOR GPU * 4, each equipped with 24GB of Graphics Double Data Rate 6 (GDDR6) memory. The model employs the Adam optimizer with an initial learning rate of 0.001. A momentum of 0.95 and a weight decay of $1e^{-5}$ are used. During training, a batch size of 16 is set, and the training iterations are performed for a total of 8,000 times.

3.2. Datasets and preprocessing

The publicly available MRI brain tumor segmentation dataset used in experiments are BraTS2019/2020 [30,31]. The data contained within the BraTS dataset is a collection of brain tumor mpMRI scan images acquired under standard clinical conditions and gathered by multiple universities and hospitals using various scanning devices such as Philips (1.5, 3). Subsequently, standard preprocessing procedures are undergoing, including conversion co-registration to a consistent anatomical template, resampling to achieve uniform isotropic resolution ($1mm^3$) and skull-stripping. The BraTS2019 training dataset comprises samples from 335 patients with glioma, which include 259 cases of high-grade glioma and 76 cases of low-grade glioma. The BraTS2019 validation dataset consists of 125 sample data. In contrast, the BraTS2020 training dataset has been expanded to 369 samples, while the validation dataset remains at 125 cases, as in the 2019 validation dataset. Each case in the BraTS2019/2020 training dataset contains five files, including brain MRI images in four modes (Flair, T1, T1ce and T2) and the ground truth segmentation map. Each image has dimensions of 240 mm × 240 mm × 155 mm and is associated with annotations for four distinct categories: Normal tissue (label 0), the non-enhancing tumor core (label 1), edema (label 2) and the enhancing tumor (label 4). The labeling for segmentation is manually annotated by experts. The evaluation of the segmentation performance focus on three brain tumor lesion regions: Whole tumor (WT), tumor core (TC) and enhancing tumor (ET). The whole tumor comprises labels 1, 2 and 4, the tumor core consists of labels 1 and 4 and the enhancing tumor only includes label 4. For both datasets, the ground truth segmentation labels for the training dataset are provided by the organizers of BraTS, while the labels for the validation dataset are not publicly available. To ensure the authenticity and accuracy of the experimental results, the segmentation results are submitted to the online BraTS platform where the evaluation results for the validation dataset are obtained.

Figure 4 shows two MRI brain tumor image samples extracted from the BraTS2020 training dataset. The upper image corresponds to a case of high-grade glioma, while the lower image portrays a case of low-grade glioma. Additionally, within the figure, the label 1 area is indicated by the red region, the label 2 area corresponds to the green region and the label 4 area is represented by the yellow region.

To mitigate the impact of device noise, enhance image contrast and alleviate overfitting, the Z-score is employed to standardize the brain tumor dataset [32]. This involves processing each image by utilizing the grayscale mean and standard deviation. The specific calculation formula is as follows:

$$\acute{z} = \frac{z - \mu}{\delta} \tag{3.1}$$

where z is the input image, $\acute{z}$ is the normalized image, μ is the average value of the input image and

δ is the standard deviation of the input image. To accommodate the relatively small proportion of tumor regions in input images, a cropping technique is applied, reducing the size of all images to $128 \times 128 \times 128$ for input purposes. Furthermore, various strategies are implemented to enhance the images, including random rotation and scaling, random elastic deformation, random flipping and intensity transformation. These strategies are aimed at improving the overall quality and robustness of the images.

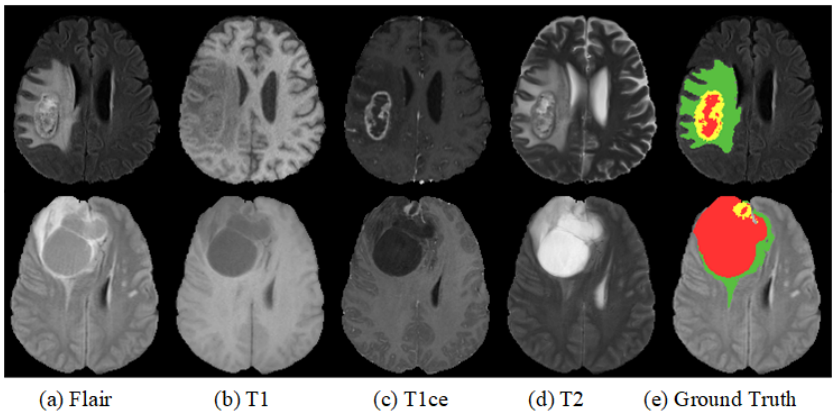

Figure 4. MRI images of two brain tumors with four different modes and Ground Truth. From left to right: (a) Flair, (b) T1, (c) T1ce, (d) T2 and (e) Ground Truth. Each color represents a tumor class, namely, red for necrotic and non-enhancing, yellow for enhancing tumor and green for edema.

3.3. Evaluation metric

The Dice similarity coefficient (DSC), which is frequently utilized in medical image segmentation, is employed as the evaluation metric for model [33]. The DSC is utilized to quantify the similarity between the segmentation results and the ground truth images, serving as a measure of set similarity. It ranges from zero to one, with values closer to one indicating a higher degree of similarity between the two samples. The specific calculation formula is as follows:

$$Dice\ score = \frac{2TP}{(FP + 2TP + FN)} \tag{3.2}$$

Among them, TP, FP, TN and FN represent the number of brain tumor voxels predicted correctly, normal brain tissue voxels predicted as brain tumors, normal brain tissue voxels predicted correctly and brain tumor voxels predicted as normal brain tissue, respectively.

3.4. Experiments and results

To validate the effectiveness of the proposed MAU-Net model for brain tumor segmentation, a series of experiments are conducted on the BraTS2019 and BraTS2020 datasets. First, ablation experiments are performed to assess the model's performance. Second, comparative experiments are carried out on the BraTS2019 and BraTS2020 validation datasets, comparing the results with other representative segmentation networks to demonstrate the generalization capability of MAU-Net. Finally, the visual

segmentation results of MAU-Net on the BraTS2020 training dataset compared with other approaches are given to further illustrate the performance of the new model.

3.4.1. Ablation experiment results

In order to fully verify the segmentation performance of the MAU-Net model, ablation experiments were performed on the datasets of BraTS2019 and BraTS2020. In the ablative experiment of the validation dataset, the model was trained using MRI images from the entire BraTS2019/2020 training datasets, and 125 predictions were submitted through the official online platform for the validation datasets to obtain corresponding DSC values. In the ablative experiment of the BraTS 2020 training dataset, the data images from the entire training dataset are partitioned into two datasets with an 8:2 ratio for training and validation, respectively. A five-fold cross-validation approach was employed to ensure result reliability. Ablation experiments are performed by incorporating SA modules and ETrans modules into the baseline method, which is the three-dimensional U-Net. The ablation results of the BraTS2019 validation dataset are presented in Table 1.

Table 1. Ablation DSC values on BraTS 2019 validation dataset.

Methods	ET	WT	TC
U-Net	0.7703	0.8911	0.7953
SA-U-Net	0.7721	0.8975	0.8044
MAU-Net	0.7788	0.9015	0.8109

As shown in Table 1, the DSC values of the three-dimensional U-Net baseline for the enhancing tumor, the whole tumor and the tumor core are 77.03, 89.11 and 79.53%, respectively. By incorporating the SA module after the consecutive convolutional blocks in the encoder stage, there is an improvement of 0.18, 0.64 and 0.91% in accuracy for the enhancing tumor, the whole tumor and the tumor core, respectively, which proves that the SA module has a certain positive impact on segmentation performance. Building upon this, the ETrans module is further added at the bottleneck, resulting in the MAU-Net segmentation network, and the DSC values obtained on the enhancing tumor, the whole tumor and the tumor core can reach 77.88, 90.15 and 81.09%, respectively. Compared to the inclusion of only the SA module, the performance improvement of 0.67, 0.40 and 0.65% is obtained in the three lesion areas, respectively, and the accuracy of 0.85, 1.04 and 1.56% is observed respectively compared with the baseline, particularly in the tumor core. These results highlight the effectiveness of the proposed MAU-Net model in brain tumor segmentation, attributed to the enhanced local feature extraction capability and the improved interaction ability of global context provided by the two added modules.

To further demonstrate the generalization and effectiveness of the proposed method, the same ablation experiments are carried out on the BraTS2020 validation dataset. The ablation results obtained are presented in Table 2.

From Table 2, it can be observed that the baseline model, U-Net, achieves DSC results of 76.42, 89.52 and 80.36% for the enhancing tumor, the whole tumor and the tumor core, respectively. After incorporating the SA module in the encoder stage, there is an improvement of 0.5, 0.44 and 0.22% in DSC values for the enhancing tumor, the whole tumor and the tumor core, respectively. Furthermore,

the introduction of the ETrans module at the bottleneck results in varying degrees of accuracy improvement in the three segmentation regions, further confirming the effectiveness of the integrated modules. The final MAU-Net model exhibits promising performance on the BraTS2020 validation dataset, achieving accuracies of 77.47, 90.00 and 81.63% on the enhancing tumor, the whole tumor and the tumor core, respectively, which is an improvement of 1.05, 0.48 and 1.27%, respectively compared to the baseline. In summary, the ablation experiments on the BraTS2020 validation dataset further validate the accuracy enhancement of the proposed model for brain tumor segmentation.

Table 2. Ablation DSC values on BraTS 2020 validation dataset.

Methods	ET	WT	TC
U-Net	0.7642	0.8952	0.8036
SA-U-Net	0.7692	0.8996	0.8058
MAU-Net	0.7747	0.9000	0.8163

Furthermore, ablative experiments are conducted on the BraTS 2020 training dataset to comprehensively validate the effectiveness of the added modules. The experimental findings are presented in Table 3.

Table 3. Ablation DSC values on BraTS 2020 training dataset.

Methods	ET	WT	TC
U-Net	0.8126	0.9034	0.8613
SA-U-Net	0.8193	0.9061	0.8649
MAU-Net	0.8350	0.9185	0.8746

Table 3 reveals that the incorporation of SA modules independently into the U-Net framework yields superior segmentation results across all three evaluation metrics, surpassing the baseline. Moreover, the integration of two distinct attention mechanisms in the MAU-Net model leads to DSC values of 83.50, 91.85 and 87.46% for the enhancing tumor, the whole tumor and the tumor core segmentation, respectively. The new model exhibits varying degrees of improvement in the three tumor segmentation regions when compared to the baseline and SA-U-Net models, providing compelling evidence for the significant enhancement in segmentation performance achieved by the MAU-Net approach.

3.4.2. Compared experiment results

In this section, to ascertain the competitiveness of the proposed model, comparisons are conducted between the segmentation experimental results of other representative models and the proposed model on the validation datasets of BraTS2019 and BraTS2020. The comparison results are exhibited in Tables 4 and 5.

Table 4. DSC values of compared experiments on BraTS 2019 validation dataset.

Methods	ET	WT	TC
Oktay et al.[34]	0.7596	0.8881	0.7720
Wang et al.[15]	0.7746	0.8933	0.8012
Akbar et al.[11]	0.7420	0.8848	0.8098
Zhang et al.[10]	0.7090	0.8700	0.7770
Tong et al.[35]	0.7510	0.8850	0.7760
Xue et al.[36]	0.7500	0.9000	**0.8300**
MAU-Net (ours)	**0.7788**	**0.9015**	0.8109

Table 5. DSC values of compared experiments on BraTS 2020 validation dataset.

Methods	ET	WT	TC
Sun et al.[37]	0.7230	0.8920	0.7880
Wang et al.[15]	0.7683	0.8988	0.8116
Akbar et al.[11]	0.7291	0.8858	0.8019
Sun et al.[9]	0.7064	0.8875	0.7194
Cheng et al.[38]	**0.7800**	0.8940	0.8140
Jiang et al.[16]	0.7736	0.8906	0.8030
Peng et al.[39]	0.7600	**0.9000**	0.8000
MAU-Net (ours)	0.7747	**0.9000**	**0.8163**

Table 4 reveals that the MAU-Net model achieves DSC values of 77.88, 90.15 and 81.09% for the enhancing tumor, the whole tumor and the tumor core, respectively. Notably, MAU-Net demonstrates superior performance in segmenting the whole tumor and the enhancing tumor, with DSC values of 90.15 and 77.88%, respectively. Among the other representative methods compared, Oktay et al. [34] introduced a new AG module integrated into the skip connection of the U-Net architecture to construct a novel segmentation model. Akbar et al. [11] proposed a method that combines attention and atrous convolution using a multipath residual attention block for brain tumor segmentation. Wang et al. [15] enhanced the feature representation ability of global information by incorporating an ordinary Transformer into the U-Net network. Zhang et al. [10] explored the effectiveness of a segmentation network approach that simultaneously integrates attention gates and residual blocks. Tong et al. [35] constructed a dual three-way CNN system, extracting comprehensive feature information from multimodal inputs and employing a three-branch classification block for segmentation, with each branch trained separately. Compared with the model methods of Oktay et al., Akbar et al., Zhang et al. and Tong et al., the MAU-Net model exhibits significant advantages in the accuracy of brain tumor segmentation. Furthermore, when compared to the reproduced TransBTS model by Wang et al., the MAU-Net model achieves an improvement of 0.42, 0.82 and 0.97% in DSC values for the three different tumor regions, respectively. This can be attributed to the presence of two distinct attention mechanisms within the MAU-Net. Additionally, a comparison is made with the multipath codec network proposed by Xue et al. [36], which processes various image modalities using multiple codecs and incorporates a Squeeze-and-Excitation block to assign weights to different

modalities. This could account for the superior performance of this method over the MAU-Net model in tumor core region segmentation. Overall, the comparative experimental results demonstrate that the proposed MAU-Net model enhances the accuracy of brain tumor segmentation to a certain extent and exhibits competitive performance compared to other representative models.

The comparative experimental results on the BraTS2020 validation dataset in Table 5 reveal notable findings. The MAU-Net model achieves DSC values of 77.47, 90.00 and 81.63% for the enhancing tumor, the whole tumor and the tumor core, respectively. Among the other segmentation methods compared, Sun et al. [37] constructed the ResU-Net model utilizing the residual unit module, while Sun et al. [9] not only incorporated the additive attention mechanism to guide the scale feature in the encoder but also introduced it to the skip connection for adaptive learning of crucial feature information. Comparatively, the MAU-Net model exhibits clear advantages over the models proposed by Sun et al. [37], Sun et al. [9] and Akbar et al. [11] in all three segmentation evaluation regions. Jiang et al. [16] proposed SwinBTS, which enhances the context information learning by integrating the Swin Transformer into the encoder and decoder of the U-shaped network. In comparison, the MAU-Net model outperforms SwinBTS by 0.11, 0.94 and 1.33% in the enhancing tumor, the whole tumor and the tumor core, respectively. Furthermore, the MAU-Net model approach improves the average value by 0.41% compared to the reproduced TransBTS segmentation results of Wang et al. [15]. Peng et al. [39] proposed AD-Net, an automatically weighted dilated convolutional network that employs channel feature separation to learn multimodal brain tumor features. It also utilizes a deeply supervised training technique for efficient fitting. While the MAU-Net model achieves comparable accuracy to the AD-Net for the whole tumor, it surpasses AD-Net by more than one percentage point in the other two brain tumor segmentation regions. These methods are sufficient to demonstrate the superiority of the network with the integrated mixed attention mechanism proposed by us in terms of brain tumor segmentation accuracy. Additionally, Cheng et al. [38] trained a patch-based 3D U-Net model with soft dice loss, generalized dice loss and multi-class cross-entropy loss, and proposed a label drop operation to address the significant class imbalance problem. The MAU-Net model slightly falls behind Cheng et al.'s model in the enhancing tumor segmentation, but still maintains a narrow lead in the whole tumor and tumor core regions. The aforementioned comparison results further validate the effectiveness and competitiveness of the novel model proposed in this paper for brain tumor segmentation tasks.

3.4.3. Visualization analysis

To provide a more intuitive and clear demonstration of the segmentation effect of the proposed model, we additionally visualize the segmentation results by comparing the MAU-Net method with other models. The visualization results are depicted in Figure 5, which displays three representative cases selected to showcase the segmentation results. From left to right, the images include the Flair image, the ground truth and the segmentation results of the 3D U-Net, TransBTS and MAU-Net models. The segmentation results of the ground truth, 3D U-Net, TransBTS and MAU-Net are superimposed on the Flair image. The visual image results clearly demonstrate that the proposed model method achieves more accurate tumor area segmentation. Additionally, to provide a better presentation of the distribution characteristics of DSC values and conduct statistical analysis, we further present a box plot of the DSC values obtained by MAU-Net on the BraTS 2020 validation dataset in Figure 5. It can be observed that our method still achieves a certain level of effectiveness in

this medical image segmentation task.

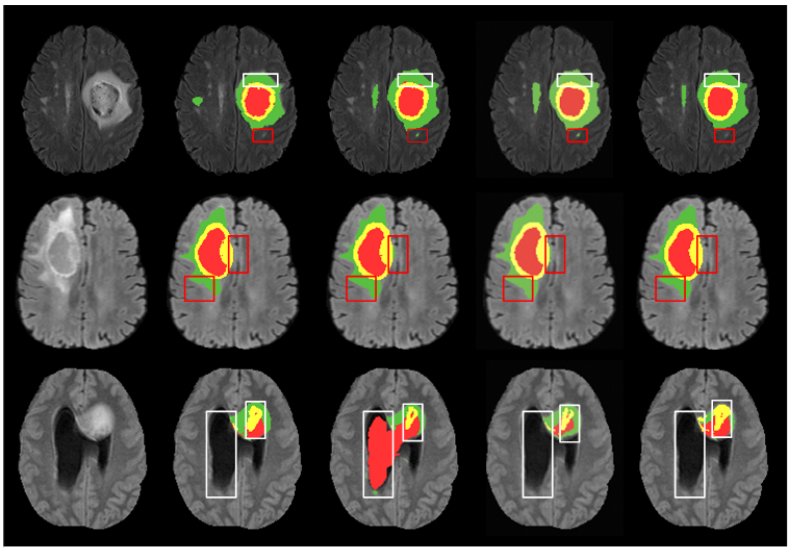

Figure 5. Examples of visualization results on BraTS2020 training datas. From left to right: (a) Flair, (b) Ground Truth, (c) 3D U-Net, (d) TransBTS and (e) MAU-Net results overlaid on Flair image. Each color represents a tumor class, namely, red for necrotic and non-enhancing, yellow for enhancing tumor and green for edema.

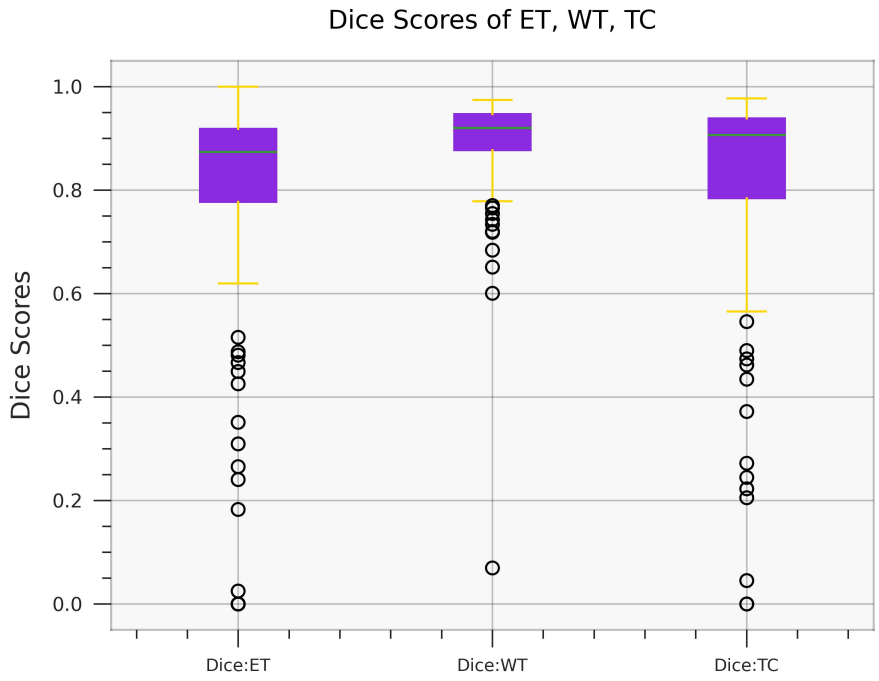

Figure 6. Box plot of DSC values obtained on the BraTS 2020 validation dataset.

4. Conclusions

In this paper, we introduced MAU-Net, a novel 3D U-shaped network that leverages mixed attention mechanisms to achieve effective brain tumor segmentation. Expanding on the three-dimensional U-Net architecture, MAU-Net enhances the model's capability to extract local information by incorporating Shuffle Attention modules after the continuous convolution block in the encoder stage. Moreover, enhanced Transformer modules were introduced at the bottleneck to improve the model's ability to interact with global information. A series of ablation experiments conducted on the publicly available BraTS2019/2020 dataset demonstrated the superiority of the proposed model over the baseline, resulting in improved segmentation accuracy of brain tumors to a certain extent. Additionally, when compared with representative methods in the field, MAU-Net exhibits higher average DSC values in the segmentation results, thereby emphasizing its competitiveness in brain tumor segmentation tasks. In future work, we aim to explore the application of the MAU-Net model to other typical medical images.

Acknowledgements

This work was supported in part by the National Natural Science Foundation of China under Grant 61972062, the Applied Basic Research Project of Liaoning Province under Grant 2023JH2/101300191 and 2023JH2/101300193, and the Fundamental Research Funds for the Central Universities under Grant 04442023128.

References

1. P. Y. Wen, R. J. Packer, The 2021 WHO classification of tumors of the central nervous system: clinical implications. *Neuro-oncology*, **21** (2021), 1215–1217. https://doi.org/10.1093/neuonc/noab120
2. Z. K. Jiang, X. G. Lyu, J. X. Zhang, Q. Zhang, X. P. Wei, Review of deep learning methods for MRI brain tumor image segmentation, *J. Image Graphics*, **25** (2020), 215–228.
3. S. Pereira, A. Pinto, V. Alves, C. A. Silva, Brain tumor segmentation using convolutional neural networks in MRI images, *IEEE Trans. Med. Imaging*, **35** (2016), 1240–1251. https://doi.org/10.1109/TMI.2016.2538465
4. Z. Zhu, X. He, G. Qi, Y. Li, B. Cong, Y. Liu, Brain tumor segmentation based on the fusion of deep semantics and edge information in multimodal MRI, *Inf. Fusion*, **91** (2023), 376–387. https://doi.org/10.1016/j.inffus.2022.10.022
5. R. Ranjbarzadeh, A. B. Kasgari, S. J. Ghoushchi, S. Anari, M. Naseri, M. Bendechache, Brain tumor segmentation based on deep learning and an attention mechanism using MRI multi-modalities brain images, *Sci. Rep.*, **11** (2021), 10930.
6. R. Ranjbarzadeh, P. Zarbakhsh, A. Caputo, E. B. Tirkolaee, M. Bendechache, Brain tumor segmentation based on an optimized convolutional neural network and an improved chimp optimization algorithm, *SSRN*, **2022** (2022), forthcoming. https://dx.doi.org/10.2139/ssrn.4295236

7. O. Ronneberger, P. Fischer, 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*, (2015), 234–241. https://doi.org/10.1007/978-3-319-24574-4_28
8. Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, O. Ronneberger, 3D U-Net: learning dense volumetric segmentation from sparse annotation, in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece*, (2016), 424–432. https://doi.org/10.1007/978-3-319-46723-8_49
9. J. K. Sun, R. Zhang, L. J. Guo, Multi-scale feature fusion and additive attention guide brain tumor MR image segmentation, *J. Image Graphics*, **28** (2023), 1157–1172.
10. J. Zhang, Z. Jiang, J. Dong, Y. Hou, B. Liu, Attention gate resU-Net for automatic MRI brain tumor segmentation, *IEEE Access*, **8** (2020), 58533–58545. https://doi.org/10.1109/ACCESS.2020.2983075
11. A. S. Akbar, C. Fatichah, N. Suciati, Single level UNet3D with multipath residual attention block for brain tumor segmentation, *J. King Saud Univ. Comput. Inf. Sci.*, **34** (2022), 3247–3258. https://doi.org/10.1016/j.jksuci.2022.03.022
12. D. Liu, N. Sheng, T. He, W. Wang, J. Zhang, J. Zhang, SGEResU-Net for brain tumor segmentation, *Math. Biosci. Eng.*, **19** (2022), 5576–5590. https://doi.org/10.3934/mbe.2022261
13. D. Liu, N. Sheng, Y. Han, Y. Hou, B. Liu, J. Zhang, et al., SCAU-net: 3D self-calibrated attention U-Net for brain tumor segmentation, *Neural Comput. Appl.*, **35** (2023), 23973–23985.
14. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, et al., Attention is all you need, *Adv. Neural Inf. Process. Syst.*, **30** (2017).
15. W. Wang, C. Chen, M. Ding, H. Yu, S. Zha, J. Li, Transbts: Multimodal brain tumor segmentation using transformer, in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France*, (2021), 109–119.
16. Y. Jiang, Y. Zhang, X. Lin, J. Dong, T. Cheng, J. Liang, SwinBTS: A method for 3D multimodal brain tumor segmentation using swin transformer, *Brain Sci.*, **12** (2022), 797. https://doi.org/10.3390/brainsci12060797
17. Y. Xu, X. He, G. Xu, G. Qi, K. Yu, L. Yin, et al., A medical image segmentation method based on multi-dimensional statistical features, *Front. Neurosci.*, **16** (2022), 1009581. https://doi.org/10.3389/fnins.2022.1009581
18. Q. L. Zhang, Y. B. Yang, Sa-net: Shuffle attention for deep convolutional neural networks, in *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, (2021), 2235–2239. https://doi.org/10.1109/ICASSP39728.2021.9414568
19. C. Yang, Y. Wang, J. Zhang, H. Zhang, Z. Wei, Z. Lin, et al., Lite vision transformer with enhanced self-attention, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, (2022), 11998–12008. https://doi.org/10.48550/arXiv.2112.10809
20. Y. Cao, J. Xu, S. Lin, F. Wei, H. Hu, Gcnet: Non-local networks meet squeeze-excitation networks and beyond, in *Proceedings of the IEEE/CVF international conference on computer vision workshops*, (2019), 1971–1980.

21. S. Woo, J. Park, J. Lee, Cbam: Convolutional block attention module, in *Proceedings of the European conference on computer vision (ECCV)*, (2018), 3–19.
22. X. Li, X. Hu, J. Yang, Spatial group-wise enhance: Improving semantic feature learning in convolutional networks, preprint, arXiv:1905.09646.
23. N. Ma, X. Zhang, H. T. Zheng, J. Sun, Shufflenet v2: Practical guidelines for efficient cnn architecture design, in *Proceedings of the European conference on computer vision (ECCV)*, (2018), 116–131. https://doi.org/10.48550/arXiv.1807.11164
24. H. Zhao, J. Shi, X. Qi, Pyramid scene parsing network, in *Proceedings of the IEEE conference on computer vision and pattern recognition*, (2017), 2881–2890.
25. L. C. Chen, G. Papandreou, F. Schroff, H. Adam, Rethinking atrous convolution for semantic image segmentation, preprint, arXiv:1706.05587.
26. D. Hendrycks, K. Gimpel, Gaussian error linear units (gelus), preprint, arXiv:1606.08415.
27. P. Ramachandran, B. Zoph, Q. V. Le, Searching for activation functions, preprint, arXiv:1710.05941.
28. F. Milletari, N. Navab, S. A. Ahmadi, V-net: Fully convolutional neural networks for volumetric medical image segmentation, in *2016 Fourth International Conference on 3D Vision (3DV)*, (2016), 565–571. https://doi.org/10.1109/3DV.2016.79
29. F. Isensee, P. F. Jäger, P. M. Full, P. Vollmuth, K. H. Maier-Hein, nnU-Net for brain tumor segmentation, in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 6th International Workshop*, 2021. https://doi.org/10.1007/978-3-030-72087-2_11
30. S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J. S. Kirby, et al., Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features, *Sci. Data*, **4** (2017), 1–13.
31. S. Bakas, M. Reyes, A. Jakab, S. Bauer, M. Rempfler, A. Crimi, et al., Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge, preprint, arXiv:1811.02629.
32. S. S. Sastry, *Advanced Engineering Mathematics*, Jones & Bartlett Learning.
33. L. R. Dice, Measures of the amount of ecologic association between species, *Ecology*, **26** (1945), 297–302. https://doi.org/10.2307/1932409
34. O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, Attention u-net: Learning where to look for the pancreas, preprint, arXiv:1804.03999.
35. J. Tong, C. Wang, A dual tri-path CNN system for brain tumor segmentation, *Biomed. Signal Process. Control*, **81** (2023), 104411. https://doi.org/10.1016/j.bspc.2022.104411
36. Y. Xue, M. Xie, F. G. Farhat, O. Boukrina, A. M. Barrett, J. R. Binder, et al., A multi-path decoder network for brain tumor segmentation, in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 5th International Workshop, BrainLes 2019, Held in Conjunction with MICCAI 2019, Shenzhen, China*, (2020), 255–265.

37. J. Sun, Y. Peng, D. Li, Y. Guo, Segmentation of the multimodal brain tumor images used Res-U-Net, in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 6th International Workshop, BrainLes 2020, Held in Conjunction with MICCAI 2020, Lima, Peru*, (2021), 263–273.

38. K. Cheng, C. Hu, P. Yin, Q. Su, G. Zhou, X. Wu, et al., Glioma sub-region segmentation on Multi-parameter MRI with label dropout, in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 6th International Workshop, BrainLes 2020, Held in Conjunction with MICCAI 2020, Lima, Peru*, (2021), 420–430.

39. Y. Peng, J. Sun, The multimodal MRI brain tumor segmentation based on AD-Net, *Biomed. Signal Process. Control*, **80** (2023), 104336. https://doi.org/10.1016/j.bspc.2022.104336

AIMS Press