




# SCAU-net: 3D self-calibrated attention U-Net for brain tumor segmentation

Dongwei Liu<sup>1</sup> · Ning Sheng<sup>2</sup> · Yutong Han<sup>1</sup> · Yaqing Hou<sup>3</sup> · Bin Liu<sup>4</sup> · Jianxin Zhang<sup>1,2</sup>  · Qiang Zhang<sup>2,3</sup>

Received: 30 December 2022 / Accepted: 12 July 2023 / Published online: 12 September 2023  
© The Author(s), under exclusive licence to Springer-Verlag London Ltd., part of Springer Nature 2023

## Abstract

Recently, U-Net architecture with its strong adaptability has become prevalent in the field of MRI brain tumor segmentation. Meanwhile, researchers have demonstrated that introducing attention mechanisms, especially self-attention, into U-Net can effectively improve the performance of segmentation models. However, the self-attention has disadvantages of heavy computational burden, quadratic complexity as well as ignoring the potential correlations between different samples. Besides, current attention segmentation models seldom focus on adaptively computing the receptive field of tumor images that may capture discriminant information effectively. To address these issues, we propose a novel 3D U-Net related brain tumor segmentation model dubbed as self-calibrated attention U-Net (SCAU-Net) in this work, which simultaneously introduces two lightweight modules, i.e., external attention module and self-calibrated convolution module, into a single U-Net. More specifically, SCAU-Net embeds the external attention into the skip connection to better utilize encoding features for semantic up-sampling, and it leverages several 3D self-calibrated convolution modules to replace the original convolution layers, which adaptively computes the receptive field of tumor images for effective segmentation. SCAU-Net achieves segmentation results on the BraTS 2020 validation dataset with the dice similarity coefficient of 0.905, 0.821 and 0.781 and the 95% Hausdorff distance (HD95) of 4.0, 9.7 and 29.3 on the whole tumor, tumor core and enhancing tumor, respectively. Similarly, competitive results are obtained on BraTS 2018 and BraTS 2019 validation datasets. Experimental results demonstrate that SCAU-Net outperforms its baseline and achieves outstanding performance compared to various representative brain tumor models.

**Keywords** Brain tumor segmentation · Three-dimensional U-Net · External attention · Self-calibrated convolution

---

Dongwei Liu and Ning Sheng contributed equally to this work.

---

✉ Jianxin Zhang  
jxzhang@dlnu.edu.cn

<sup>1</sup> School of Computer Science and Engineering, Dalian Minzu University, No.18, Liaohu West Road, Jinzhou New District, Dalian 116600, China

<sup>2</sup> Key Lab of Advanced Design and Intelligent Computing (Ministry of Education), Dalian University, No. 10, Xuefu Street, Dalian Economic and Technological Development Zone, Dalian 116622, China

<sup>3</sup> School of Computer Science and Technology, Dalian University of Technology, No.2, Linggong Road, Ganjingzi District, Dalian 116024, China

<sup>4</sup> International School of Information Science and Engineering (DUT-RUISE), Dalian University of Technology, No.321, Tuqian Road, Jinzhou New District, Dalian 116620, China

## 1 Introduction

In modern society, brain-related diseases are becoming an increasingly intractable problem. According to the survey [1], between 2013 and 2017, the incidence of brain tumors is 7.1–16.7 per 100,000 people, and the fatality rate is 4.4 per 100,000 people, with a 5-year relative survival rate of 36%. As one of the most harmful diseases in brain diseases, brain tumor has seriously threatened human health. Brain tumors are abnormal cells that grow in the brain or skull, including benign and malignant tumors [2]. The most common malignant brain tumor is glioma, which can be further divided into high-grade glioma (HGG) and low-grade glioma (LGG) according to the degree of infiltration. Gliomas are often associated with necrotic and non-enhancing tumors, enhancing tumors and edema. Moreover, brain tumors are characterized by blurred margins,

irregular shape and variable size and location and overlapping sections with normal brain tissue. Therefore, accurately segmenting abnormal tissue and characterizing these tumors are challenging. Magnetic resonance imaging (MRI), a non-invasive imaging technique commonly used for brain tumor detection, can provide valuable information about the shape, size and location of brain tumors, which is of great help in improving the diagnosis of brain tumors by clinicians. Given the advantages of using MRI images, the computer-aided diagnosis and treatment of brain tumors based on MRI images has become one of the most popular research topics in medical imaging [3, 4].

In deep neural networks, the attention mechanism can pay more attention to pivotal information and eliminate some redundant ones to highlight helpful information effectively and reasonably allocate computing resources. To further develop the performance of the U-Net model, many researchers focus on the attention mechanism. Liu et al. [5] introduce squeeze-and-excitation (SE) [6] block into V-Net and propose a deep supervised 3D squeeze-and-excitation V-Net (DSSE-V-Net) to segment brain tumors automatically, which obtains highly competitive performance compared with those methods win in the BraTS 2017 Challenge. Furthermore, the self-attention mechanism has achieved great success in various computer vision tasks as a branch of attention mechanism. Zhu et al. [7] leverage a non-local structure to propose an asymmetric non-local neural network for semantic segmentation, which achieves the new state-of-the-art performance on some datasets. Similarly, on brain tumor segmentation tasks, Jia et al. [8] employ the expectation-maximization attention (EMA) module, a variant of the non-local self-attention mechanism, embed it into the U-Net and win second place in the BraTS 2020 Challenge with super competitiveness. Subsequently, the emergence of the transformer [9] significantly promotes the development of the self-attention mechanism, and researchers also focus on how to add it to medical segmentation modules. Chen et al. [10] embed the transformer block into the down-sampling position of U-Net architecture and propose the TransUNet model that combines the advantages of transformer and U-Net, and has excellent segmentation performance. Wang et al. [11] propose the TranBTS model for brain tumor segmentation and achieve a breakthrough accuracy improvement. Valanarasu et al. [12] propose a gated axial-attention module that improves the existing self-attention mechanism by introducing an additional control mechanism in the self-attention module to perform feature learning from both local and global channels. Final experiments show that their method outperforms convolutional and other correlation transformer-based architectures. Besides, the transformer module has also achieved excellent performance in semantic segmentation for scene understanding [13].

Although these works based on self-attention can achieve excellent segmentation results, they usually need to rely on large-scale pre-training and have high computational complexity, making these methods unable to be easily used. Furthermore, self-attention has quadratic computational complexity, which slows down the data processing speed in 3D medical image segmentation. Self-attention can only model self-affinity within a single sample, ignoring potential correlations between different samples in the entire dataset. Last but not least, the current U-Net models associated with self-attention rarely focus on adaptively computing the receptive field of tumor images to capture discriminant information.

Therefore, to address the above-mentioned issue, we utilize a simpler self-attention-like approach: external attention [14] that solves the enormous computational complexity problem well with two linear layers. At the same time, due to the small number of datasets and the high similarity of samples in brain tumor segmentation, it is necessary to pay attention to the potential correlation between the different samples. The external attention can handle this problem well and improve the model's accuracy to a certain extent. In addition, the self-calibrated convolution structure [15] can adaptively build long-range spatial and inter-channel dependencies around each spatial location through a novel self-calibration operation, helping our model to generate more discriminative representations while easily applied to augment standard convolutional layers. Therefore, we embed the 3D external attention and 3D self-calibrated convolutions into a single 3D U-Net structure to reduce the self-attention mechanism's high computational complexity and parameters while improving segmentation accuracy. The main contributions of this work are summarized as follows: (1) We propose a novel 3D SCAU-Net model for MRI brain tumor segmentation task. SCAU-Net introduces lightweight modules to reduce computational complexity and analyzes potential correlations among different samples in the entire dataset while adaptively computing receptive field to gain effective segmentation results. (2) SCAU-Net embeds the 3D external attention into the skip connection to enhance important encoding features for semantic up-sampling, and it leverages 3D self-calibrated convolutions to replace original convolution layers, adaptively capturing long-range spatial and inter-channel dependencies around each position to compute more discriminant features. (3) Extensive experimental results on three brain tumor segmentation benchmarks of BraTS 2018, 2019 and 2020 demonstrate that SCAU-Net outperforms its baseline model. SCAU-Net also achieves competitive performance with representative brain tumor segmentation methods.

## 2 Method

In this section, we first introduce the structure of SCAU-Net for brain tumor segmentation. Then, the 3D external attention module and the 3D self-calibrated convolution module are described in detail. Finally, we present our loss function adopted for the SCAU-Net model.

### 2.1 Overall architecture

In this work, we expand the external attention module and the self-calibrated convolution module to 3D modules and propose a single 3D U-Net architecture, namely, SCAU-Net. SCAU-Net can effectively attain context-related information, generate more discriminative representations and reduce the computational complexity. Figure 1 demonstrates the overall architecture of SCAU-Net.

As shown in Fig. 1, the SCAU-Net follows the traditional U-Net architecture and consists of an encoder part and a decoder part, which are connected by skip connections. SCAU-Net model receives input data size of  $4 \times 128 \times 128 \times 128$ , in which the input image size is  $128 \times 128 \times 128$ , and the channel number is 4. The SCAU-Net contains four encoding blocks to capture context features and four decoding blocks to recover spatial information and input image size. In the encoder part, we perform four down-sampling operations using two 3D convolution layers with kernels of size 3 and a 3D Max-pooling with stride 2 to extract features from MRI images, and after each down-sampling, the size of the image is halved, and the number of channels is doubled. Furthermore, the size of the image at the bottom of the network becomes  $512 \times 8 \times 8 \times 8$ . At the same time, we employ upsample layer with `scale_factor` of 2 to up-sampling operations that double the image size and halve the number of channels to recover the spatial information of the feature map and the original size of the input. In SCAU-Net, the external attention (EA) module is embedded into the fourth skip connection to better utilize the feature information of low-level CNNs. The EA module is a very lightweight self-attention module that achieves performance and accuracy equivalent to self-attention through two simple linear layers. This not only greatly improves the traditional self-attention module that is difficult to consider the correlations between all samples but also greatly reduces the computational complexity. The EA module will be explained further in Sect. 2.2.

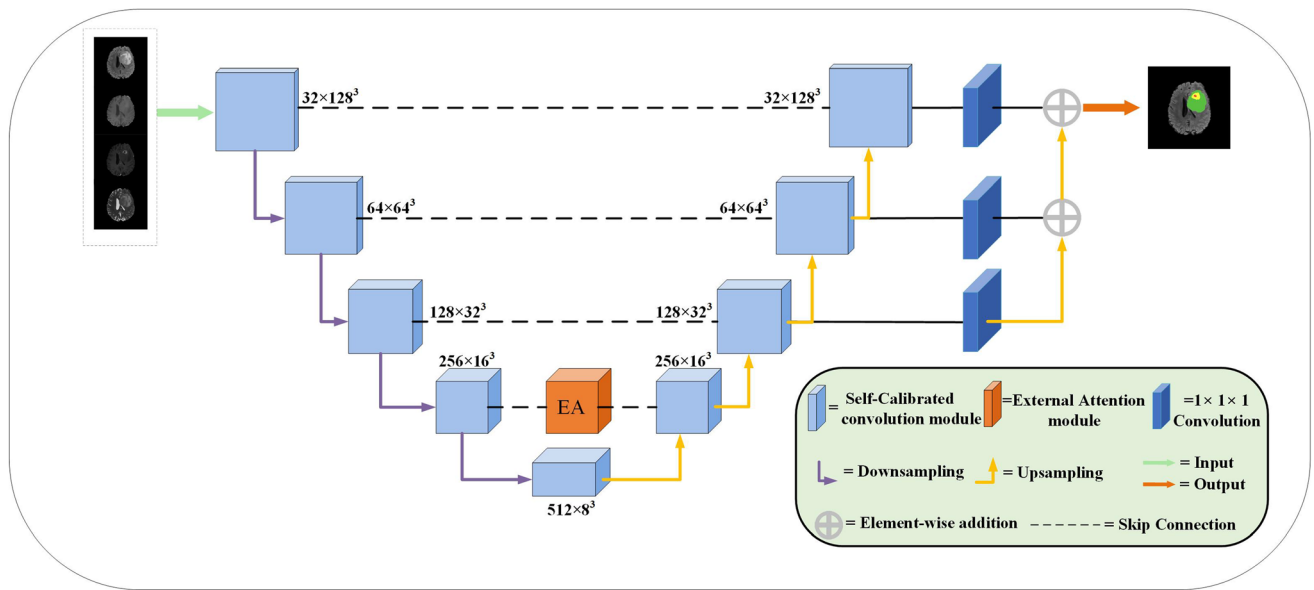
It is worth noting that, different from the traditional 3D U-Net, we utilize the 3D self-calibrated convolution for replacing some original 3D convolutions in up-sampling and down-sampling. The self-calibrated convolution is a relatively lightweight convolution module that significantly

expands the receptive field of each convolution layer through internal channels, thereby enriching the output functions of each encoder. And the self-calibrated convolution will be explained further in Sect. 2.3. The output of each encoder block is down-sampled to the next encoder block and put into the horizontal connection of the corresponding layer. Notably, the output of the fourth-level encoding block is simultaneously input to the EA module to contextual information among all pixels and mine potential relationships across the whole datasets. Finally, at the end of the decoder, we fuse the outputs of the last three decoder blocks through the  $1 \times 1 \times 1$  convolution and up-sampling operations as the final segmentation result.

### 2.2 The 3D external attention module

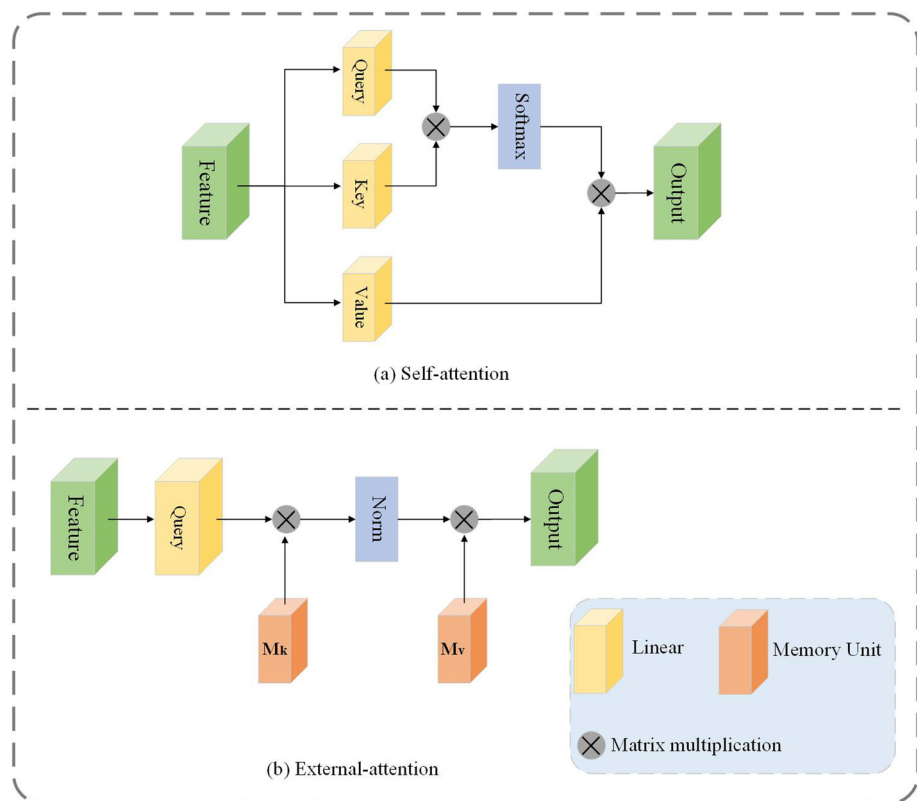
The self-attention with its ability to capture long-range dependencies can improve the performance of various computer vision tasks [16–20]. However, self-attention has disadvantages of computational quadratic complexity and ignoring potential correlations between different samples. When the self-attention captures the long-range dependencies within a single sample, it updates features at each position by computing a weighted sum of features using pair-wise affinities across all positions, which leads to quadratic computational complexity and ignores potential correlations between different samples. Therefore, Guo et al. [14] propose a novel attention mechanism named external attention which is based on two external, learnable, shared memories to learn the most discriminative features across the whole dataset. The external attention can be implemented easily by using only two cascaded linear layers and two normalization layers, which reduces the computational complexity of external attention to a linear level. Therefore, in this work, we embed the external attention module at the horizontal connection position of the U-Net network. Besides, since the MRI brain tumor images are in 3D format, we add the dimension of depth  $D$  to the original 2D external attention module (input image resolution size is  $H \times W$ ) and extend it to the 3D module ( $H \times W \times D$ ) to fit the input data. The overall architecture of the 3D external attention module is demonstrated in Fig. 2.

The external attention module is a modified module based on the self-attention mechanism. Before showing the external attention module, let us briefly introduce the self-attention mechanism. As shown in Fig. 2(a), given an input feature map  $F \in \mathbb{R}^{N \times d}$ , where  $N$  is the number of elements (or pixels in images) and  $d$  is the number of feature dimensions, self-attention linearly projects the input to a query matrix  $Q \in \mathbb{R}^{N \times d'}$ , a key matrix  $K \in \mathbb{R}^{N \times d'}$  and a value matrix  $V \in \mathbb{R}^{N \times d}$ . The self-attention calculation is



**Fig. 1** The overall architecture of SCAU-Net. SCAU-Net is an end-to-end network that integrates the external attention (EA) module and 3D self-calibrated convolution (SC) modules into a single 3D U-Net model

**Fig. 2 a** Self-attention structure and **b** 3D external attention structure



mainly divided into three steps. First, the similarity between the query and each key is calculated to obtain the weights. Then, these weights are normalized using the softmax function. Finally, the weighted sum operation is performed on the normalized weights and the

corresponding values to obtain the final attention. Therefore, the self-attention can be formulated as follows:

$$A = \text{softmax}(QK^T), F_{\text{out}} = AV \tag{1}$$

where  $A \in \mathbb{R}^{N \times N}$  is the attention matrix.

From the formula, we can conclude that the use of self-attention has a significant drawback of the high computational complexity of  $O(dN^2)$ . The quadratic complexity in the number of input pixels makes it very time-consuming and memory consumption for complex pixel image calculations, which leads to previous brain tumor segmentation work rarely involving self-attention to complete the corresponding segmentation task. However, the strong feature representation ability of self-attention is also worth learning, and we should probably seek another solution to use the self-attention mechanism in the brain tumor segmentation task. Then, Guo et al. [14] propose a novel attention module named external attention that can easily compensate for this shortcoming of self-attention. As shown in Fig. 2(b), the external attention module computes attention between the input pixels and an external memory unit  $M \in \mathbb{R}^{S \times d}$ , by:

$$A = \text{Norm}(FM^T), F_{\text{out}} = AM \tag{2}$$

where  $M$  is a learnable parameter independent of the input, which acts as a memory of the whole training dataset.  $A$  is the attention map inferred from this learned dataset-level prior knowledge and updates the input features from  $M$  by the similarities in  $A$ . In the specific implementation, we replace the matrices  $K$  and  $V$  in the original self-attention mechanism with two different memory units  $M_k$  and  $M_v$  that constantly update themselves based on similarities in attention. Therefore, the final external attention calculation formula can be defined as below:

$$A = \text{Norm}(FM_k^T)M_v \tag{3}$$

The final external attention calculation is linear in the number of pixels, and the computational complexity of it is  $O(dSN)$ , where  $d$  and  $S$  are hyper-parameters. From the visualization of brain tumors, we know that only a few pixels are meaningful for segmentation in brain tumors. Thus, we can utilize the two memory units of external attention to learn the required pixel value features in the input features to improve the segmentation accuracy. The 3D external attention can significantly reduce the time and computational complexity and mine the potential connections between different samples to improve the segmentation accuracy.

### 2.3 The 3D self-calibrated convolution

In order to better improve the performance of the model, we utilize several 3D self-calibrated convolution modules for replacing the original convolution layers in up-sampling and down-sampling. A traditional convolutional layer  $f$  generally consists of a group of filter sets  $K = [k_1, k_2, \dots, k_C]$ , where  $k_i$  denotes the  $i$ -th set of filters

with size  $C$ , and it can transform the input  $X = [x_1, x_2, \dots, x_C] \in \mathbb{R}^{C \times H \times W \times D}$  to the output  $Y = [y_1, y_2, \dots, y_C] \in \mathbb{R}^{\hat{C} \times \hat{H} \times \hat{W} \times \hat{D}}$ . Therefore, the output feature map at channel  $i$  is calculated as follows:

$$y_i = k_i * X = \sum_{j=1}^C k_i^j * X_j \tag{4}$$

where “\*” denotes convolution and  $k_i = [k_i^1, k_i^2, \dots, k_i^C]$ . In this way, each output feature map results from summing by executing Eq. (4) multiple times through all channels. However, the pattern of such convolutional filter learning is similar, which causes the network model to generate feature maps that are less discriminative. Therefore, we introduce the self-calibrated convolution module to solve this problem very well. As shown in Fig. 3, we illustrate the self-calibrated convolution module from the following steps:

Firstly, the input feature map  $X \in \mathbb{R}^{C \times H \times W \times D}$  is split into two features  $X_u \in \mathbb{R}^{\frac{C}{2} \times H \times W \times D}$  for the path above and  $X_d \in \mathbb{R}^{\frac{C}{2} \times H \times W \times D}$  for the path below. Secondly, given a group of filter sets  $K$  with the shape  $(C, \hat{C}, k_h, k_w, k_d)$ , where  $C, \hat{C}, k_h, k_w$  and  $k_d$  denote the input channel number, the output channel number, the spatial height, width and depth, respectively. Then divide  $K$  into four parts:  $K_1, K_2, K_3$  and  $K_4$ , each part has shape  $(\frac{C}{2}, \frac{\hat{C}}{2}, k_h, k_w, k_d)$ , are responsible for performing different functions. Without loss of generality, we assume that the number of input and output channels is equal, i.e.,  $C = \hat{C}$  and that  $C$  is divisible by 2. Furthermore, to efficiently gather rich informative contextual information from each spatial location, the convolutional feature transformation is performed in two different scale-spaces: an original scale-space (the feature map shares the same resolution as the input) and a small latent space after down-sampling (small resolution latent space for self-correction operation). The down-sampled features have larger receptive fields, so the transformed embeddings in the smaller latent space are used as a reference to guide the feature transformation process in the original feature space.

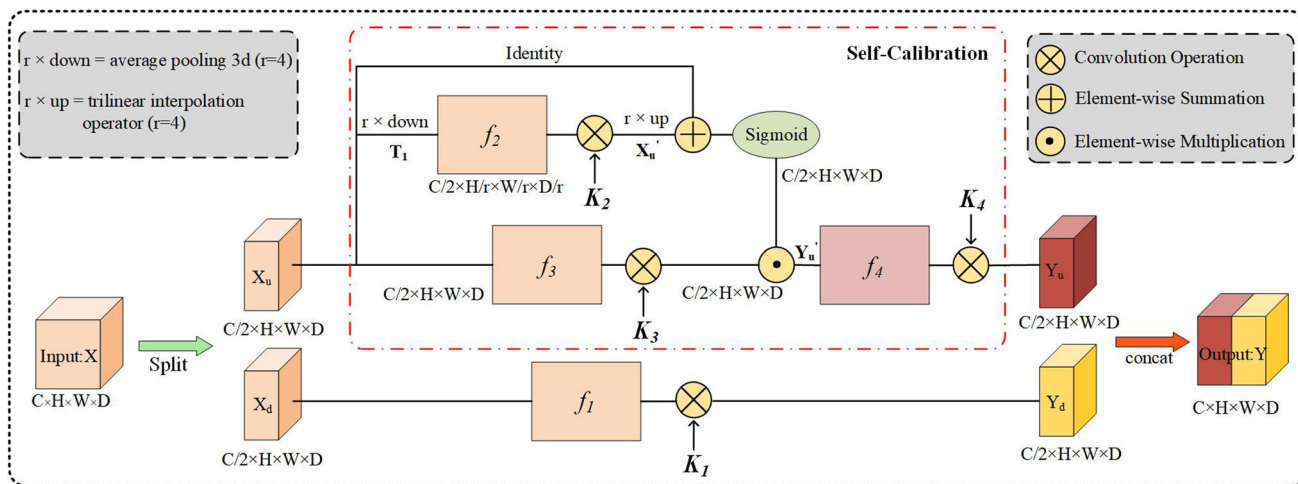
Then, the operation process in the self-calibrated convolution scale-space can be formulated as follows:

$$T_1 = \text{AvgPool}_r(X_u) \tag{5}$$

The feature  $X_u$  is down-sampled by average pooling 3D with filter size  $r \times r \times r$  and stride  $r$  ( $r=4$  in this work). Further,  $T_1$  performs an up-sampling operation based on  $K_2$  to get feature  $X_u'$ :

$$X_u' = \text{Up}(f_2(T_1)) = \text{Up}(T_1 * K_2) \tag{6}$$

where  $\text{Up}(\cdot)$  is a trilinear interpolation operator to conduct



**Fig. 3** The 3D self-calibrated convolutions. As can be seen, the original filters are separated into four portions, each of which is in charge of a different functionality

the up-sampling operation. At this time, the middle step of the self-calibration operation can be written as follows:

$$Y_u' = f_3(X_u) \cdot \sigma(X_u + X_u') \tag{7}$$

where  $f_3(X_u) = X_u * K_3$ ,  $\sigma$  means the sigmoid function, and “ $\cdot$ ” denotes element-wise multiplication. Therefore, the final output after self-calibration operation can be formulated as follows:

$$Y_u = f_4(Y_u') = Y_u' * K_4 \tag{8}$$

Moreover, in the original scale feature space, feature  $Y_d$  is extracted by  $K_1$  convolution on feature  $X_d$ . Finally, the two scale-space output features  $Y_u$  and  $Y_d$  are concatenated to obtain the final output feature  $Y$ . Compared with the traditional convolution operation, the self-calibrated convolution module can generate more discriminative representations without introducing additional parameters. Therefore, it is a wise choice to introduce the self-calibrated convolution module into the task of brain tumor segmentation.

### 2.4 Combined loss function

The problem of data imbalance is a longstanding challenge in brain tumor segmentation tasks. In brain tumor MRI images, the average area occupied by healthy tissue is 98.46%, while the area occupied by edema, enhanced tumor and non-enhanced tumor only are 1.02%, 0.29% and 0.23%, respectively. The considerable difference in the occupancy rate of different regions in brain tumor MRI significantly impacts its segmentation accuracy. In order to address this problem, we adopt a combination of dice loss and cross-entropy loss as the loss function in SCAU-Net.

The combined loss function  $L_{DCE}$  can be defined as follows:

$$L_{DCE} = \alpha L_{DC} + (1 - \alpha)L_{CE} \tag{9}$$

where  $\alpha$  is the balance parameter varying from 0 to 1,  $L_{DC}$  is the dice loss and  $L_{CE}$  is the cross-entropy loss.

## 3 Experimental datasets and results

### 3.1 Datasets

We adopt the BraTS 2018–2020 datasets from the Multi-modal Brain Tumor Segmentation Challenge (BraTS) [21–23] to evaluate the effectiveness of SCAU-Net. The BraTS 2018 dataset consists of a training dataset and a validation dataset. The training dataset contains 285 glioma patients, which includes 210 high-grade glioma (HGG) cases and 75 low-grade glioma (LGG) cases, and the validation dataset includes 66 patient cases of unknown grades. The BraTS 2019 dataset grows to 259 HGG and 76 LGG cases in its training dataset and 125 patient cases of unknown grade in the validation dataset. In the BraTS 2020 dataset, some changes have taken place in the form of the dataset. The training dataset is expanded to 369 cases without separating HGG and LGG, while the validation dataset keeps 125 cases of unknown grade. Since the BraTS 2020 dataset has the largest amount of data and sufficient pathological information, we pay more attention to the BraTS 2020 dataset and conduct core experiments on it. In these three datasets, each patient’s MRI scan consists of four modalities, i.e., native T1-weighted (T1), post-contrast T1-weighted (T1CE/T1Gd), T2-weighted (T2) and T2 fluid-attenuated inversion recovery (T2-FLAIR), and all of them co-registered to a common anatomical template

(SRI [24]) and resampled to  $1\text{mm}^3$ . The basic labels include four types, termed healthy parts (label 0), necrotic and non-enhancing tumors (NCR/NET—label 1), edema around the tumor (ED—label 2) and GD-enhancing tumors (ET—label 4). The various sub-regions considered for the segmentation evaluation are the whole tumor (combined area of labels 1, 2 and 4), the tumor core (combined area of labels 1 and 4) and the enhancing tumor (label 4). Additionally, for these datasets, the ground truth of training datasets is provided by the BraTS organizers. However, labels of validation datasets are unavailable to the public. Therefore, the model's predicted results are evaluated via submission to the official BraTS website, ensuring the evaluation result's fairness and authority. Figure 4 shows an example of the 2020 training dataset.

### 3.2 Data preprocessing

In the BraTS dataset, all patient data are collected based on multiple institutions using different scanners and protocols. Due to the uncertainty of brain tumor morphology, location, blurring of boundaries and manual annotation deviation, the preprocessing of brain tumor images is particularly important. In this work, we perform the following preprocessing and data enhancement on each dataset: (i) Due to the severe class imbalance of brain tumors, we first crop the collected MRI brain tumor data from  $240 \times 240 \times 155$  pixels to  $128 \times 128 \times 128$  pixels; (ii) the z-score normalization applies the average and standard deviation to process each image; (iii) random rotation is with the angle between  $[-10^\circ$  and  $+10^\circ]$  and (iv) random intensity shifts between  $[-0.1, 0.1]$ .

### 3.3 Evaluation metrics and settings

We employ the dice similarity coefficient (DSC) and 95% Hausdorff distance (HD95) metrics widely used in the medical segmentation community to evaluate the given model, and they are defined as follows:

$$\text{Dice Score} = \frac{2\text{TP}}{\text{FP} + 2\text{TP} + \text{FN}} \quad (10)$$

where TP, FP, TN and FN denote the number of true-positive, false-positive, true-negative and false-negative voxels, respectively. The value range is 0–1, and the value closer to 1 indicates more similar contours of brain tumors.

$$\text{HD}(T, P) = \max\{\sup_{t \in T} \inf_{p \in P} d(t, p), \sup_{p \in P} \inf_{t \in T} d(t, p)\} \quad (11)$$

where  $t$  and  $p$  refer to the points on the ground truth regions  $T$  and the predicted regions  $P$ , respectively.  $d(t, p)$  denotes the distance between points  $t$  and  $p$ .

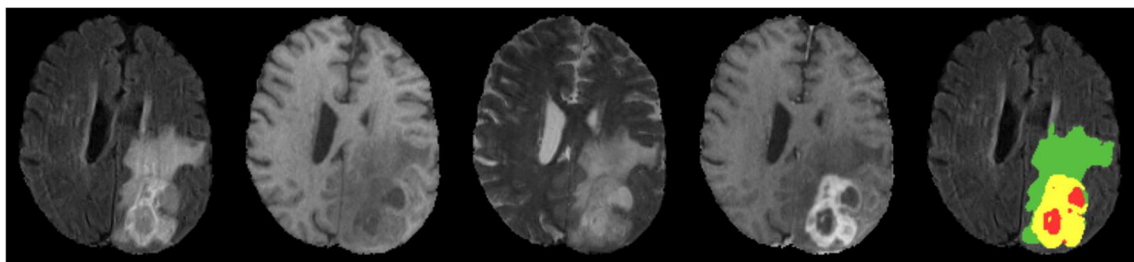
In this work, we utilize the Adam optimizer with an initial learning rate of 0.001. Momentum is 0.95, the weight decay is  $1\text{e-}5$ , the batch size is set to 4 and the epochs are 500 to ensure the best experimental results. Our model is implemented using the PyTorch deep learning framework on an NVIDIA GeForce GTX 3090 GPU with 24 GB of memory.

### 3.4 Experimental results

We extensively evaluate SCAU-Net on three brain tumor segmentation benchmarks, i.e., BraTS 2018–2020 datasets. We first perform ablation experiments on the BraTS 2020 dataset to demonstrate the effects of 3D external attention and 3D self-calibrated convolution modules. Then, we show the computational complexity and parameters between our proposed network and the self-attention network. Finally, the experiment results compared with representative works on BraTS 2018–2020 validation datasets are reported and analyzed to prove the effectiveness of SCAU-Net.

#### 3.4.1 Ablation experiments

Firstly, we conduct ablation experiments on the BraTS 2020 training dataset and validation dataset to evaluate the effectiveness of the 3D external attention module (EA) and 3D self-calibrated convolution module (SC) for brain tumor



**Fig. 4** An example of the brain MRI data from the BraTS 2020 training dataset. From left to right: T2-FLAIR, T1, T1Gd, T2 and ground truth. Each color represents a tumor class (label): red (label

1)—necrosis and non-enhancing, green (label 2)—edema and yellow (label 4)—enhancing tumor (color figure online)

segmentation. On the training dataset, we utilize fivefold cross-validation to test the segmentation performance of the 3D external attention module and the 3D self-calibrated convolution module. Moreover, the segmentation results are listed in Table 1. In our work, the 3D U-Net is taken as the baseline model, the 3D external attention module is added to the first skip connection, the second skip connection, the third skip connection and the fourth skip connection, respectively, and we call them EA<sup>1</sup>, EA<sup>2</sup>, EA<sup>3</sup> and EA<sup>4</sup>. In order to capture long-range dependencies and comprehensively consider related issues such as model memory consumption, we perform ablation experiments on the four skip connection parts in turn to determine the optimal embedding locations.

As shown in Table 1, we clearly see that the model of baseline + EA<sup>4</sup> + SC achieves the optimal segmentation performance, which obtains dice similarity coefficient (DSC) of 0.909, 0.845 and 0.817 on the whole tumor (WT), tumor core (TC) and enhancing tumor (ET) segmentation, respectively. And it is superior to the baseline with 0.8%, 2.3% and 2.1% on the WT, TC and ET, respectively. Therefore, in the following experiments, our model utilizes this method by default. Besides, we embed the SC module into the baseline, and 0.1%, 1.2% and 1.2% improve its accuracy in these three regions compared with the baseline, demonstrating the SC module's effectiveness in brain tumor segmentation task. At the same time, we embed self-attention (SA) into the fourth horizontal connection of the baseline to form the baseline + SA<sup>4</sup> model, which achieves DSC of 0.903, 0.835 and 0.797 on the WT, TC and ET segmentation, respectively. Then, in order to find the optimal embedding location for the EA module, we perform four sets of experiments at the positions of skip connections. Furthermore, the experimental results show that no matter EA module is added to which skip connection position of U-Net, its performance is better than the baseline. In particular, when it is added to the fourth skip connection, baseline + EA<sup>4</sup> achieves the optimal segmentation performance, and its DSC values are 0.904,

0.838 and 0.811 on the WT, TC and ET segmentation, respectively. Compared with the baseline + SA<sup>4</sup> model, the model formed by embedding the external attention module on the baseline has achieved comparable or even superior segmentation performance in the segmentation of the three tumor sub-regions. Among them, in the fourth layer embedded in the EA module (baseline + EA<sup>4</sup>), it surpasses baseline + SA<sup>4</sup> by 1.4% on ET segmentation, which is enough to show that the EA module has better performance in brain tumor segmentation. Therefore, ablation experiment results show that using EA and SC modules benefits the brain tumor segmentation task. In addition, Table 1 shows the FLOPs and parameters of the baseline and its added modules. We can see that SCAU-Net only increases the FLOPs and parameters a small compared to baseline but achieves an effective performance boost. Meanwhile, it can be seen that the EA module added to the fourth layer of the skip connection has fewer parameters than the SA module, but it achieves higher performance.

Moreover, we also visualize the brain tumor segmentation results of the four models as shown in Fig. 5. In this figure, the five columns from left to right demonstrate MRI images in ground truth (GT), segmentation results of U-Net (baseline), ExU-Net (baseline + EA<sup>4</sup>), ScU-Net (baseline + SC) and SCAU-Net (baseline + EA<sup>4</sup> + SC), respectively. And we apply various colors to represent different tumor classes for the brain tumor segmented images, the yellow regions represent enhancing tumor, the green regions indicate edema and the red regions are necrosis and non-enhancing. As shown from Fig. 5, the SCAU-Net model achieves the optimal brain tumor segmentation results.

Then, to further verify the effectiveness and robustness of our module, we perform the ablation experiment on the BraTS 2020 validation dataset. We utilize 369 MRI images on the BraTS 2020 training dataset to train our models and then submit 125 predicted results of the validation dataset to the BraTS online website. In this experiment, we further

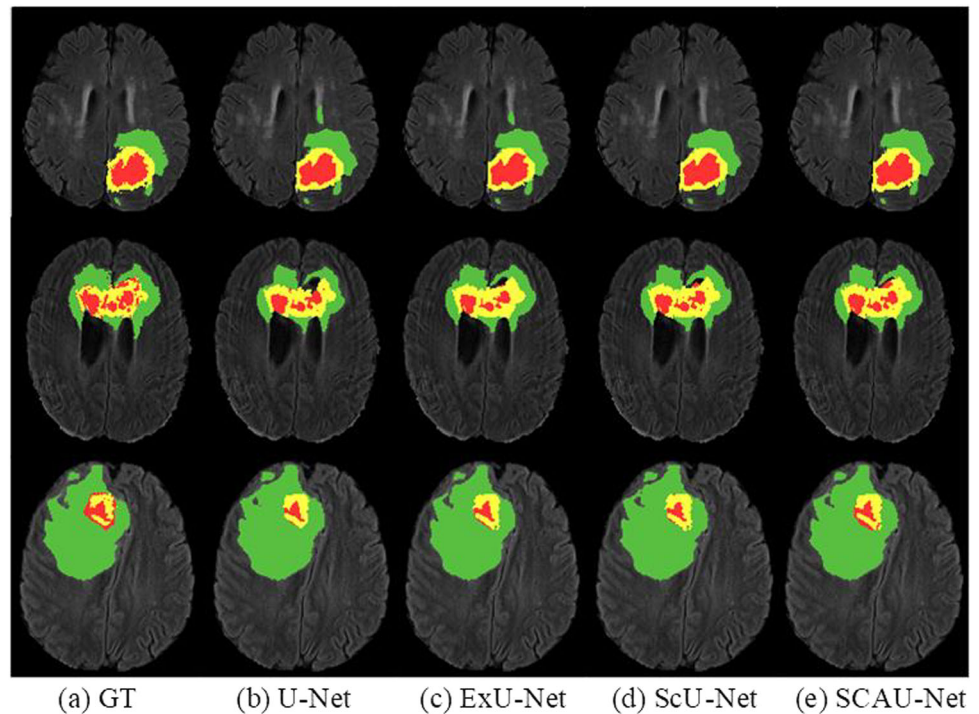
**Table 1** Ablation DSC results on the BraTS 2020 training dataset

Methods	WT	TC	ET	FLOPs (GMac)	Parameters (M)	Time (h)
Baseline	0.901	0.822	0.796	693.19	28.36	13.2
Baseline + SC	0.902	0.834	0.808	696.01	33.07	19.5
Baseline + EA <sup>1</sup>	0.899	0.825	0.801	706.42	28.37	23.7
Baseline + EA <sup>2</sup>	0.903	0.825	0.802	697.77	28.38	23.2
Baseline + EA <sup>3</sup>	0.902	0.831	0.806	695.08	28.41	22.8
Baseline + EA <sup>4</sup>	0.904	0.838	0.811	694.14	28.53	22.4
Baseline + SA <sup>4</sup>	0.903	0.835	0.797	693.73	29.05	25.3
Baseline + EA <sup>4</sup> + SC	<b>0.909</b>	<b>0.845</b>	<b>0.817</b>	696.68	33.24	31.6

Bold font represents the best result on the corresponding accuracy indicator



**Fig. 5** Examples of segmentation results on the BraTS 2020 training dataset. From left to right: MRI images of **a** ground truth (GT), and segmentation results of **b** U-Net (Baseline), **c** ExU-Net (baseline + EA<sup>4</sup>), **d** ScU-Net (baseline + SC) and **e** SCAU-Net (baseline + EA<sup>4</sup> + SC). The visualized color of tumors: enhancing tumor (yellow), edema (green) and necrotic and non-enhancing tumor (red) (color figure online)



delve into the related issues of the external attention module. The experimental results are shown in Table 2.

It is observed from Table 2 that our proposed model, i.e., baseline + EA<sup>4</sup> + SC also achieves the best results on this dataset, which is consistent with our results on the training dataset. Moreover, our model outperforms baseline by 1.3%, 1.9% and 2.0% on WT, TC and ET, respectively. Especially, the segmentation of WT has also been significantly improved on the BraTS 2020 validation dataset. The experiment results further show that the attention mechanism can make the model pay more attention to the tumor area. Simultaneously, the baseline + SC model exceeds the baseline with 0.4%, 0.5% and 1.1% on WT, TC and ET, respectively. The baseline + EA<sup>4</sup> gains 0.4%, 1.4% and 1.2% accuracy over the baseline model on WT, CT and ET, respectively. These results also demonstrate that both the EA module and the SC module are helpful for brain tumor segmentation tasks. At the same time, compared with the baseline + SA<sup>4</sup> model, the baseline + EA<sup>4</sup> model improves the segmentation results of the three sub-tumor regions by a total of 0.6%, which is enough to show that the EA module is superior to the SA module. Furthermore, we give the standard deviation of DSC on the corresponding three segmented regions (WT, TC and ET). It can be seen that the standard deviation on the DSC is slight, which shows the stability of the SCAU-Net model segmentation results.

Finally, we compare the computational complexity and parameters of SCAU-Net with two other self-attention

networks. As shown in Table 3, the computational complexity and parameters of SCAU-Net are 696.68 GMac and 33.24 M, respectively. Compared with these two self-attention networks, our method's computational complexity and parameters are significantly reduced. In particular, compared to TransBTS [11], our computation and parameters are reduced by approximately 16% and 69%, respectively.

### 3.4.2 Comparison results with state-of-the-art methods

To verify the generalization and effectiveness of our proposed SCAU-Net, we also conduct experiments on the BraTS 2018, 2019 and 2020 validation datasets. The following three tables show our comparative results on these three datasets.

Compared results on BraTS 2018 validation dataset with state-of-the-art methods are given in Table 4. We can obviously find that SCAU-Net achieves DSC values of 90.9%, 85.1% and 80.4% on WT, TC and ET, respectively. And it is worth noting that our SCAU-Net gets the same highest value in DSC values on the WT as Isensee et al. [25], which comes second in BraTS 2018 Challenge. Compared with works in [26–28], our model's DSC value on the WT exceeds 0.3%. On the segmentation results of TC and ET, Myronenko et al. [29] achieve the optimal DSC accuracy. Meanwhile, SCAU-Net is slightly lower than Brügger et al. [27] on DSC values of TC and ET. Brügger et al. [27] embed the reversible block into the traditional U-Net structure and propose a partially

**Table 2** Ablation DSC results on the BraTS 2020 validation dataset

Methods	DSC		
	WT	TC	ET
Baseline	0.892 ± 0.108	0.802 ± 0.239	0.761 ± 0.276
Baseline + SC	0.896 ± 0.098	0.807 ± 0.176	0.772 ± 0.264
Baseline + EA <sup>1</sup>	0.896 ± 0.105	0.807 ± 0.222	0.767 ± 0.289
Baseline + EA <sup>2</sup>	0.897 ± 0.102	0.805 ± 0.215	0.766 ± 0.291
Baseline + EA <sup>3</sup>	0.896 ± 0.101	0.811 ± 0.221	0.768 ± 0.284
Baseline + EA <sup>4</sup>	0.896 ± 0.104	0.816 ± 0.198	0.773 ± 0.265
Baseline + SA <sup>4</sup>	0.896 ± 0.099	0.814 ± 0.204	0.769 ± 0.266
Baseline + EA <sup>4</sup> + SC	<b>0.905 ± 0.101</b>	<b>0.821 ± 0.177</b>	<b>0.781 ± 0.261</b>

Bold font represents the best result on the corresponding accuracy indicator

**Table 3** Comparison of FLOPs and parameters of SCAU-Net and self-attention

Methods	FLOPs (GMac)	Parameters (M)
U-Net (baseline)	693.19	28.36
Baseline+Non-local [16]	806.25	38.48
TransBTS [11]	823.89	106.77
SCAU-Net (ours)	696.68	33.24

reversible U-Net structure, which significantly reduces memory consumption and improves the performance of brain tumor segmentation. Moreover, our model also achieves comparable performance on the TC and ET compared with OM-Net [28] which solves class imbalance better. And compared with [30], our method achieves perfect surpassing on the WT, TC and ET segmentation. As for HD95 results, SCAU-Net obtains the HD95 value of 4.04, 5.78 and 2.53 on the WT, TC and ET segmentation, respectively. It is obvious that our SCAU-Net outperforms

**Table 4** Compared results with state-of-the-art methods on BraTS 2018 validation dataset

Methods	DSC			HD95		
	WT	TC	ET	WT	TC	ET
Elhamzi et al. [30]	0.88	0.77	0.65	11.69	13.38	6.89
Kao et al. [31]	0.905	0.803	0.768	4.32	7.56	3.81
Isensee et al. [25]	<b>0.909</b>	0.814	0.788	5.83	7.20	2.74
Myronenko et al. [29]	0.907	<b>0.860</b>	<b>0.817</b>	4.52	6.85	3.82
Chen et al. [26]	0.906	0.845	0.801	4.66	6.44	3.06
Brügger et al. [27]	0.906	0.857	0.806	5.61	7.83	3.35
Zhou et al. [28]	0.906	0.854	0.803	4.62	6.39	3.44
SCAU-Net (ours)	<b>0.909</b>	0.851	0.804	<b>4.04</b>	<b>5.78</b>	<b>2.53</b>

Bold font represents the best result on the corresponding accuracy indicator

all state-of-the-art methods in terms of HD95 values. In a word, our SCAU-Net model can achieve superior performance compared with state-of-the-art methods on the BraTS 2018 validation dataset.

Then, we conduct the compared experiments on the BraTS 2019 validation datasets, and the experimental results are given in Table 5. In this table, SCAU-Net obtains DSC values of 90.7%, 82.0% and 78.2% on WT, TC and ET, respectively. It is clear that SCAU-Net outperforms other state-of-the-art methods in DSC values of the ET. Zhao et al. [32] combine the different kinds of tricks into a single 3D U-Net, such as random patch size training, warming-up learning, semi-supervised learning and result fusing, which comes to the second place in the BraTS 2019 Challenge for the tumor segmentation. Compared with Zhao et al. [32], our model's result is slightly lower than theirs on WT segmentation. But, it still achieves the optimal segmentation performance on WT segmentation compared to the other seven methods [15, 33–38]. For the HD95 metrics results, SCAU-Net gains HD95 values of 4.11, 6.11 and 2.96 on the WT, TC and ET segmentation, respectively. Furthermore, we can see that our SCAU-Net obtains the best result on the WT and ET segmentation. Compared with the state-of-the-art methods, SCAU-Net is only inferior to Zhao et al. [32] on TC, but it exceeds the other seven methods. Therefore, the results of these comparisons again demonstrate that embedding the external attention module and self-calibrated convolution module into the U-Net is extremely suitable and effective for brain tumor segmentation tasks.

Finally, we also perform the experiments on the BraTS 2020 validation dataset to further evaluate the SCAU-Net model. As shown in Table 6, SCAU-Net achieves DSC values of 90.5%, 82.1% and 78.1% on WT, TC and ET, respectively. Moreover, comparing it with other typical brain tumor segmentation methods, SCAU-Net achieves the optimal result on WT and TC segmentation. On the segmentation results of the ET, our result is slightly lower

**Table 5** Compared results with state-of-the-art methods on BraTS 2019 validation dataset

Methods	DSC			HD95		
	WT	TC	ET	WT	TC	ET
Li et al. [33]	0.886	0.813	0.771	6.23	7.41	6.03
Chen et al. [34]	0.903	0.793	0.756	4.49	8.19	4.77
Zhao et al. [32]	<b>0.910</b>	<b>0.835</b>	0.754	4.57	<b>5.58</b>	3.84
Cheng et al. [35]	0.902	0.824	0.777	5.41	7.26	5.28
Guo et al. [36]	0.903	0.833	0.773	7.10	7.68	4.44
Liu et al. [15]	0.901	0.833	0.775	5.86	6.61	3.70
Zhou et al. [37]	0.871	0.718	0.727	6.7	9.3	6.3
Ahmad et al. [38]	0.903	0.823	0.719	4.69	7.08	3.31
SCAU-Net (ours)	0.907	0.820	<b>0.782</b>	<b>4.11</b>	6.11	<b>2.96</b>

Bold font represents the best result on the corresponding accuracy indicator

**Table 6** Compared results with state-of-the-art methods on BraTS 2020 validation dataset

Methods	DSC			HD95		
	WT	TC	ET	WT	TC	ET
Tang et al. [39]	0.889	0.784	0.698	4.5	10.1	34.3
Cheng et al. [40]	0.894	0.814	0.780	7.1	12.7	24.4
Sundaresan et al. [41]	0.890	0.770	0.770	4.4	15.3	29.4
Wang et al. [11]	0.890	0.814	<b>0.785</b>	6.5	10.5	<b>16.7</b>
Guan et al. [42]	0.850	0.690	0.680	8.4	31.6	47.4
Fang et al. [43]	0.876	0.769	0.670	9.4	12.5	50.8
Huang et al. [44]	0.860	0.772	0.700	6.7	15.1	39.1
SCAU-Net (ours)	<b>0.905</b>	<b>0.821</b>	0.781	<b>4.0</b>	<b>9.7</b>	29.3

Bold font represents the best result on the corresponding accuracy indicator

than Wang et al. [11] who first utilize the transformer in 3D CNN for global feature modeling. However, it ranks second place in all ET segmentation results. Based on the improvement of the U-Net structure, Tang et al., Cheng et al. and Sundaresan et al. [39–41] propose variational-autoencoder regularized 3D MultiResUNet, modified 3D U-Net and triplanar ensemble network, respectively, and both achieve excellent segmentation results. Furthermore, Wang et al., Guan et al., Fang et al. and Huang et al. [11, 42–44] introduce attention modules such as SE, non-local, CBAM and self-attention into the brain tumor segmentation model, and the experimental results further demonstrate the strong effectiveness of the attention mechanism on the brain tumor task. For the HD95 metrics result, we obtain as excellent as on the DSC metrics that achieve the superior result on the WT and TC segmentation. Moreover, our result ranks third on the ET among seven typical methods. In general, the compared results on the BraTS 2020 validation dataset prove again the

competitive performance of our SCAU-Net and the effectiveness for brain tumor segmentation tasks.

## 4 Discussion

Many researchers have proved that brain tumor segmentation methods based on attention mechanisms can significantly improve segmentation accuracy. However, few researchers have considered the computational complexity and parameters of self-attention mechanisms in high-dimensional data such as 3D medical images. For example, in H2NF-Net [8], training of the ensemble network model requires more than 11 G of memory and utilizes 4 NVIDIA Tesla P40 GPUs and 4 NVIDIA Geforce GTX 2080Ti GPUs to train the cascaded model and single model, respectively. In our work, we only use one NVIDIA GeForce GTX 3090 GPU with 24 GB of memory to train the proposed model and achieve competitive results. Moreover, the self-attention networks ignore potential connections between different samples in the dataset and can not generate more discriminative features due to fixed receptive fields. Therefore, in this paper, we utilize the 3D external attention module to solve the enormous computational complexity problem and neglect of samples' latent relations in self-attention embedding medical segmentation. Meanwhile, the 3D self-calibrated convolution module is introduced to adaptively compute the receptive field of tumor images to capture discriminant information and further improve the accuracy of segmentation. Finally, the experimental results show that our method can effectively reduce the computational complexity and parameters of self-attention on medical image segmentation tasks while achieving more competitive results. As shown in Table 3, it can be clearly seen that our method's computational complexity and parameters are significantly reduced compared to adding a self-attentive module at the same position in the U-Net network.

Overall, for the experimental results on the BraTS 2018–2020 datasets, it can be found that the highest average DSC value of 0.855 and the best HD95 are available on the BraTS 2018 dataset, which may be attributed to the fact that the BraTS 2018 dataset contains fewer samples compared to the BraTS 2019 and 2020 datasets. As the number of samples increases on the BraTS 2019 and 2020 datasets, the segmentation effect on the WT and TC regions is still excellent. But, the segmentation effect in the ET region decreases. In particular, the HD95 value in the ET region is large on the BraTS 2020 dataset, which reflects that our model's performance on the ET region segmentation needs to be improved.

## 5 Conclusions

In this paper, we mainly explore the effectiveness of external attention and self-calibrated convolution for brain tumor segmentation tasks. The proposed SCAU-Net model integrates the 3D external attention module and the 3D self-calibrated convolution module with a primeval and single U-Net architecture. The external attention module is embedded in the skip connection of the U-Net model to replace self-attention and achieve high segmentation accuracy. Meanwhile, we also embed the self-calibrated convolution module into the up-sampling and down-sampling of the U-Net model to further improve the model performance. Extensive experiments and compared results demonstrate that our model can achieve excellent segmentation results and strong competitiveness without increasing the burden of the machine. In the future, we will explore more advanced attention/self-attention modules for constructing more robust brain tumor segmentation networks. In addition, we will also attempt to evaluate SCAU-Net on other typical medical image segmentation applications.

**Acknowledgements** This work was partially supported by the National Natural Science Foundation of China under Grant 61972062, the NSFC-Liaoning Province United Foundation under Grant U1908214 and the Young and Middle-aged Talents Program of the National Civil Affairs Commission.

**Availability of data and materials** The datasets during the current study are available in the: [https://www.cbica.upenn.edu/MICCAI\\_BraTS2020\\_TrainingData](https://www.cbica.upenn.edu/MICCAI_BraTS2020_TrainingData)[https://www.cbica.upenn.edu/MICCAI\\_BraTS2020\\_ValidationData](https://www.cbica.upenn.edu/MICCAI_BraTS2020_ValidationData)[https://www.cbica.upenn.edu/sbia/Spyridon.Bakas/MICCAI\\_BraTS/2019/MICCAI\\_BraTS\\_2019\\_Data\\_Training.zip](https://www.cbica.upenn.edu/sbia/Spyridon.Bakas/MICCAI_BraTS/2019/MICCAI_BraTS_2019_Data_Training.zip)[https://www.cbica.upenn.edu/sbia/Spyridon.Bakas/MICCAI\\_BraTS/2019/MICCAI\\_BraTS\\_2019\\_Data\\_Validation.zip](https://www.cbica.upenn.edu/sbia/Spyridon.Bakas/MICCAI_BraTS/2019/MICCAI_BraTS_2019_Data_Validation.zip).

## Declarations

**Conflict of interest** The authors declare no conflicts of interest.

## References

1. Miller KD, Ostrom QT, Kruchko C, Patil N, Tihan T, Cioffi G, Fuchs HE, Waite KA, Jemal A, Siegel RL et al. (2021) Brain and other central nervous system tumor statistics, 2021. *CA Cancer Journal Clin* 71(5):381–406
2. Liu J, Li M, Wang J, Wu F, Liu T, Pan Y (2014) A survey of mri-based brain tumor segmentation methods. *Tsinghua Sci Technol* 19(6):578–595
3. Chahal PK, Pandey S (2021) A hybrid weighted fuzzy approach for brain tumor segmentation using mr images. *Neural Comput Appl*, pp 1–15
4. Amin J, Anjum MA, Gul N, Sharif M (2022) A secure two-qubit quantum model for segmentation and classification of brain tumor using mri images based on blockchain. *Neural Comput Appl*, pp 1–14
5. Liu P, Dou Q, Wang Q, Heng PA (2020) An encoder-decoder neural network with 3d squeeze-and-excitation and deep supervision for brain tumor segmentation. *IEEE Access* 8:34029–34037
6. Hu J, Shen L, Sun G (2018) Squeeze-and-excitation networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp 7132–7141
7. Zhu Z, Xu M, Bai S, Huang T, Bai X (2019) Asymmetric non-local neural networks for semantic segmentation. In: *Proceedings of the IEEE/CVF international conference on computer vision*, pp 593–602
8. Jia H, Cai W, Huang H, Xia Y (2021) H2nf-net for brain tumor segmentation using multimodal mr imaging: 2nd place solution to brats challenge 2020 segmentation task. In: *Brainlesion: glioma, multiple sclerosis, stroke and traumatic brain injuries-6th international workshop*, Springer, pp 58–68
9. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I (2017) Attention is all you need. *Adv Neural Inf Process Syst*, 30
10. Chen J, Lu Y, Yu Q, Luo X, Adeli E, Wang Y, Lu L, Yuille AL, Zhou Y (2021) Transunet: Transformers make strong encoders for medical image segmentation. Preprint at [arXiv:2102.04306](https://arxiv.org/abs/2102.04306)
11. Wang W, Chen C, Ding M, Yu H, Zha S, Li J (2021) Transbts: Multimodal brain tumor segmentation using transformer. In: *International conference on medical image computing and computer-assisted intervention*, pp 109–119. Springer
12. Valanarasu JMJ, Oza P, Hacihaliloglu I, Patel VM (2021) Medical transformer: gated axial-attention for medical image segmentation. In: *Medical Image computing and computer assisted intervention—MICCAI 2021: 24th international conference, Strasbourg, France, 27 Sep–1 Oct, 2021, Proceedings, Part I 24*, pp 36–46. Springer
13. Liu H, Zhang J, Yang K, Hu X, Stiefelbogen R (2022) Cmx: Cross-modal fusion for rgb-x semantic segmentation with transformers. *arXiv preprint arXiv:2203.04838*
14. Guo MH, Liu ZN, Mu TJ, Hu SM (2022) Beyond self-attention: external attention using two linear layers for visual tasks. *IEEE Trans Pattern Anal Mach Intell* 45(5):5436–5447
15. Liu JJ, Hou Q, Cheng MM, Wang C, Feng J (2020) Improving convolutional networks with self-calibrated convolutions. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp 10096–10105
16. Wang X, Girshick R, Gupta A, He K (2018) Non-local neural networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp 7794–7803
17. Fu J, Liu J, Tian H, Li Y, Bao Y, Fang Z, Lu H (2019) Dual attention network for scene segmentation. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp 3146–3154
18. Zhang H, Goodfellow I, Metaxas D, Odena A (2019) Self-attention generative adversarial networks. In: *International conference on machine learning*, pp 7354–7363. PMLR
19. Yuan Y, Huang L, Guo J, Zhang C, Chen X, Wang J (2018) Ocnet: Object context network for scene parsing. Preprint at [arXiv:1809.00916](https://arxiv.org/abs/1809.00916)
20. Bello I, Zoph B, Vaswani A, Shlens J, Le QV (2019) Attention augmented convolutional networks. In: *Proceedings of the IEEE/CVF international conference on computer vision*, pp 3286–3295
21. Menze BH, Jakab A, Bauer S, Kalpathy Cramer J, Farahani K, Kirby J, Burren Y, Porz N, Slotboom J, Wiest R et al (2014) The multimodal brain tumor image segmentation benchmark (brats). *IEEE Trans Med Imag* 34(10):1993–2024
22. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby JS, Freymann JB, Farahani K, Davatzikos C (2017) Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features. *Sci Data* 4(1):1–13

23. Bakas S, Reyes M, Jakab A, Bauer S, Rempfler M, Crimi A, Shinohara RT, Berger C, Ha SM, Rozycki M, et al (2018) Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the brats challenge. Preprint at [arXiv:1811.02629](https://arxiv.org/abs/1811.02629)
24. Rohlfing T, Zahr NM, Sullivan EV, Pfefferbaum A (2010) The sri24 multichannel atlas of normal adult human brain structure. *Human Brain Map* 31(5):798–819
25. Isensee F, Kickingereder P, Wick W, Bendszus M, Maier-Hein KH (2018) No new-net. In: International MICCAI brainlesion workshop, pp 234–244. Springer
26. Chen C, Liu X, Ding M, Zheng J, Li J (2019) 3d dilated multi-fiber network for real-time brain tumor segmentation in mri. In: International conference on medical image computing and computer-assisted intervention, pp 184–192. Springer
27. Brügger R, Baumgartner CF, Konukoglu E (2019) A partially reversible u-net for memory-efficient volumetric image segmentation. In: International conference on medical image computing and computer-assisted intervention, pp 429–437. Springer
28. Zhou C, Ding C, Wang X, Lu Z, Tao D (2020) One-pass multi-task networks with cross-task guided attention for brain tumor segmentation. *IEEE Trans Image Process* 29:4516–4529
29. Myronenko A (2018) 3d mri brain tumor segmentation using autoencoder regularization. In: International MICCAI brainlesion workshop, pp 311–320. Springer
30. Elhamzi W, Ayadi W, Atri M (2022) A novel automatic approach for glioma segmentation. *Neural Comput Appl* 34(22):20191–20201
31. Kao PY, Ngo T, Zhang A, Chen JW, Manjunath B (2018) Brain tumor segmentation and tractographic feature extraction from structural mr images for overall survival prediction. In: International MICCAI brainlesion workshop, pp 128–141. Springer
32. Zhao YX, Zhang YM, Liu CL (2020) Bag of tricks for 3d mri brain tumor segmentation. In: International MICCAI brainlesion workshop, pp 210–220. Springer
33. Li X, Luo G, Wang K (2020) Multi-step cascaded networks for brain tumor segmentation. In: International MICCAI brainlesion workshop, pp 163–173. Springer
34. Chen M, Wu Y, Wu J (2020) Aggregating multi-scale prediction based on 3d u-net in brain tumor segmentation. In: International MICCAI brainlesion workshop, pp 142–152. Springer
35. Cheng X, Jiang Z, Sun Q, Zhang J (2020) Memory-efficient cascade 3d u-net for brain tumor segmentation. In: International MICCAI brainlesion workshop, pp 242–253. Springer
36. Guo D, Wang L, Song T, Wang G (2020) Cascaded global context convolutional neural network for brain tumor segmentation. In: International MICCAI brainlesion workshop, pp 315–326. Springer
37. Zhou T, Canu S, Vera P, Ruan S (2021) Latent correlation representation learning for brain tumor segmentation with missing mri modalities. *IEEE Trans Image Process* 30:4263–4274
38. Ahmad P, Jin H, Qamar S, Zheng R, Saeed A (2021) Rd2a: densely connected residual networks using aspp for brain tumor segmentation. *Multimedia Tools Appl* 80(18):27069–27094
39. Tang J, Li T, Shu H, Zhu H (2021) Variational-autoencoder regularized 3d multiresunet for the brats 2020 brain tumor segmentation. In: International MICCAI brainlesion workshop, pp 431–440. Springer
40. Cheng K, Hu C, Yin P, Su Q, Zhou G, Wu X, Wang X, Yang W (2021) Glioma sub-region segmentation on multi-parameter mri with label dropout. In: International MICCAI brainlesion workshop, pp 420–430. Springer
41. Sundaresan V, Griffanti L, Jenkinson M (2021) Brain tumour segmentation using a triplanar ensemble of u-nets on mr images. In: International MICCAI brainlesion workshop, pp 340–353. Springer
42. Guan X, Yang G, Ye J, Yang W, Xu X, Jiang W, Lai X (2022) 3d agse-vnet: an automatic brain tumor mri data segmentation framework. *BMC Med Imag* 22(1):1–18
43. Fang Y, Huang H, Yang W, Xu X, Jiang W, Lai X (2022) Nonlocal convolutional block attention module vnet for gliomas automatic segmentation. *Int J Imag Syst Technol* 32(2):528–543
44. Huang H, Yang G, Zhang W, Xu X, Yang W, Jiang W, Lai X (2021) A deep multi-task learning framework for brain tumor segmentation. *Front Oncol* 11:690244

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.