Mathematical Biosciences and Engineering

MBE, 20(9): 17384–17406.
DOI: 10.3934/mbe.2023773
Received: 28 June 2023
Revised: 11 August 2023
Accepted: 22 August 2023
Published: 11 September 2023

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

Research article

SDS-Net: A lightweight 3D convolutional neural network with multi-branch attention for multimodal brain tumor accurate segmentation

Qian Wu[1,2,*], Yuyao Pei[2], Zihao Cheng[2], Xiaopeng Hu[3] and Changqing Wang[2]

[1] School of Humanistic Medicine, Anhui Medical University, Hefei 230032, China
[2] School of Biomedical Engineering, Anhui Medical University, Hefei 230032, China
[3] Department of Medical Imaging, First Affiliated Hospital of Anhui Medical University, Hefei 230032, China

* **Correspondence:** Email: wuqian@ahmu.edu.cn; Tel: +8655113339100757.

Abstract: The accurate and fast segmentation method of tumor regions in brain Magnetic Resonance Imaging (MRI) is significant for clinical diagnosis, treatment and monitoring, given the aggressive and high mortality rate of brain tumors. However, due to the limitation of computational complexity, convolutional neural networks (CNNs) face challenges in being efficiently deployed on resource-limited devices, which restricts their popularity in practical medical applications. To address this issue, we propose a lightweight and efficient 3D convolutional neural network SDS-Net for multimodal brain tumor MRI image segmentation. SDS-Net combines depthwise separable convolution and traditional convolution to construct the 3D lightweight backbone blocks, lightweight feature extraction (LFE) and lightweight feature fusion (LFF) modules, which effectively utilizes the rich local features in multimodal images and enhances the segmentation performance of sub-tumor regions. In addition, 3D shuffle attention (SA) and 3D self-ensemble (SE) modules are incorporated into the encoder and decoder of the network. The SA helps to capture high-quality spatial and channel features from the modalities, and the SE acquires more refined edge features by gathering information from each layer. The proposed SDS-Net was validated on the BRATS datasets. The Dice coefficients were achieved 92.7, 80.0 and 88.9% for whole tumor (WT), enhancing tumor (ET) and tumor core (TC), respectively, on the BRTAS 2020 dataset. On the BRTAS 2021 dataset, the Dice coefficients were 91.8, 82.5 and 86.8% for WT, ET and TC, respectively. Compared with other state-of-the-art methods, SDS-Net achieved superior segmentation performance with fewer

parameters and less computational cost, under the condition of 2.52 M counts and 68.18 G FLOPs.

Keywords: brain tumor; magnetic resonance imaging; volume data; lightweight; convolutional neural network; attention mechanism; depthwise separable convolution

1. Introduction

Glioma is one of the most aggressive types of brain tumors, accounting for approximately 80% of all cranial malignancies, with a high recurrence, mortality rate, low cure rate and slightly predominant in males. The common clinical treatments of glioma include radiotherapy, chemotherapy and surgery, which all depend on medical imaging, processing and analysis. Among all medical imaging, Magnetic Resonance Imaging (MRI) has been a crucial technique in the diagnosis and treatment of glioma, since it cannot only clearly display the soft tissue structures based on multimodal images, but provide non-invasive and radiation-free imaging for the human body [1]. Multi-modalities show the different anatomical and functional information of glioma, which are beneficial for the depiction of the type, shape, size and boundary of the tumors, as shown in Figure 1. Accurate delineation MRI of glioma requires true labels of data, including background (label 0), necrotic and non-enhancing tumor (label 1), edema (label 2) and enhancing tumor (label 4). Physicians often delineate tumor regions manually in a slice-by-slice manner from a large amount of MRI images, which is tedious and time-consuming, and the accuracy depends highly on the experience of the experts and threshold setting. Therefore, automatic and accurate segmentation of brain tumors plays a vital role in clinical diagnosis and treatment [2].

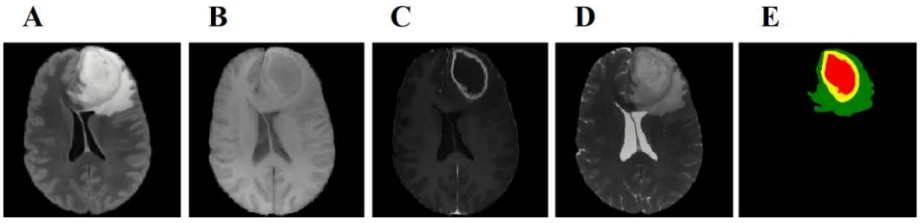

Figure 1. Multimodal MRI images. (A) Flair, (B) T1, (C) T1ce, (D) T2 and (E) Label. The red area is necrotic and non-enhancing, the green area is edema and the yellow area is enhancing tumor.

Convolutional neural networks (CNNs) have achieved impressive achievements in brain tumor segmentation because of their powerful feature extraction capabilities. Fully Convolutional Networks (FCNs) [3] abandon the full connection layer to reduce network parameters and achieve pixel-level segmentation. Especially, since Ronneberger introduced the U-Net [4] based on FCN in 2015, encoder-decoder symmetric networks with skip connection have demonstrated exceptional performance in medical image semantic segmentation tasks [5–8]. In order to meet demands of clinical applications in terms of accuracy and efficiency, a large number of networks based on U-Net have appeared to optimize the performance by improving network structure, adding new modules, expanding to 3D networks and so on. For example, Jungo et al. enhanced the U-Net

architecture by introducing residual flow and improving the pooling flow to combine global and local feature information [9]. Cinar et al. [10] proposed a hybrid architecture that combined the strengths of DenseNet121 and U-Net to achieve more accurate and precise segmentation results for small tumor regions. Wu et al. [11] introduced DE-ResUNet, which connected T1-weighted image and texture features to the dual encoder and used a channel attention mechanism to extract more useful informative features. Jiang et al. [12] proposed a dual-branch decoder network DDU-net. Çiçek et al. proposed the 3D U-Net [13], Mehta et al. [14] proposed a 3D U-Net brain tumor segmentation network and Abdollahi et al. [15] proposed the network V-Net to expand 2D into 3D network, which can extract more spatial contextual information and achieve remarkable results in the segmentation of volumetric medical image data. Zhu et al. [16] proposed a segmentation method that outperformed several state-of-the-art methods on the BraTS benchmarks by fusing deep semantic and edge information from multimodal MRI. Li et al. [17] introduced X-Net, a dual decoder structure that combined CNN and Transformer features, which showed outstanding performance in medical image segmentation tasks. Similarly, Xu et al. [18] also integrated CNN and Transformer, adopted a multi-dimensional statistical feature extraction module and obtained excellent results in brain tumor and heart segmentation tasks.

Attention mechanisms are widely regarded as crucial components in medical image processing. In MRI scans of the brain, where the tumor area is relatively small, these mechanisms play a vital role in focusing on the tumor region and extracting relevant and informative features. By leveraging various attention mechanisms, the accuracy of small tumor segmentation can be significantly improved. For example, Kong et al. [19] presented a 3D FCN with a dual-attention mechanism, incorporating both global attention and local attention, to enhance significantly the segmentation performance for glioma segmentation based on MRI. Attention U-Net [20] added attention gates (AGs) in encoder and decoder to readjust the output characteristics of encoder, a gating signal can be generated by the module to remove irrelevant and noise blurring responses and highlight the significant features transmitted through skipping connections. He et al. [21] proposed a cloud-based medical image segmentation method, which improved segmentation accuracy by introducing an interactive fusion attention module. Squeeze-and-excitation (SE*) was proposed to automatically acquire the importance of each feature channel, and then increased the weight of the relevant features in the 2018 CVPR. Roy et al. [22] introduced three SE* module extended structures in the fully convolutional neural network. Shuffle attention (SA) is a multi-branch attention mechanism [23] that combines spatial attention and channel attention simultaneously. Unlike SE*, which mainly focuses on channel attention, SA overcomes the limitations of using a single attention mechanism and incorporates global attention. Furthermore, SA offers the advantage of reducing computational complexity compared to SE*.

In the field of brain tumor segmentation, the currently existing 3D encoder-decoder networks have obtained excellent results. However, these networks often have a large number of parameters, which results in high computational cost. To maintain good segmentation efficiency while reducing computation and parameters, 3D lightweight networks are proposed for precisely segmenting brain glioma, such as MobileNets [24], ShuffleNet [25] and GhostNet [26], etc. Zhou et al. [27] presented an effective 3D residual neural network (ERV-Net) with lower computational memory consumption. Reddy et al. [28] proposed a lightweight CNN framework that extracted features from brain MRI images and then classified them as normal or abnormal. Luo et al. [29] used a lightweight and efficient hierarchical decoupled convolutional network (HDCNet), which replaced 3D convolution

with a new hierarchical decoupled convolutional module, to achieve efficient exploration of spatial contextual information of images at multiple scales and perspectives. Zhang et al. [30] introduced a hierarchical multi-scale segmentation network (HMNet) with lightweight components, which further reduced the complexity and computational overhead of the network.

In the current field of brain tumor segmentation, there are still some issues. First, a large number of parameters and computational cost are required by traditional CNN networks to directly extract feature information from high-resolution brain MRI volumetric data, which can lead to difficulties in running on resource-limited devices. Second, due to the small size and diverse shapes of glioma regions, as well as the blurred boundary between them and healthy tissue, tumor area segmentation is extremely challenging. Therefore, we are committed to developing a lightweight network that can maintain high accuracy while achieving real-time performance in the brain tumor segmentation task. In this paper, a 3D lightweight network SDS-Net is proposed, which cleverly simplifies parameters and computational resources to obtain high accuracy in glioma segmentation while maintaining low computational complexity. SDS-Net demonstrates the ability to rapidly and accurately segment brain tumors, and can be easily embedded in resource-limited devices to provide a more convenient and efficient solution for clinical diagnosis and treatment. Additionally, we have developed a brain tumor segmentation and visualization software, further enhancing the convenience and efficiency of brain tumor segmentation in clinical applications. Our major contributions are given as follows:

- A novel 3D lightweight encoder-decoder network, named SDS-Net, is introduced for multimodal brain tumor MRI image segmentation. SDS-Net adopts a lightweight backbone structure and extensively explores the rich multimodal and boundary features of brain tumor images, significantly reducing network computation and parameters while maintaining excellent segmentation performance.
- The effective 3D lightweight backbone blocks, known as lightweight feature extraction (LFE) and lightweight feature fusion (LFF) modules, have been developed. The LFE module can effectively extract the local feature information of multimodal MRI images, and the LFF module combines high-level abstract features from the encoder with features from the decoder to achieve a more comprehensive feature representation. The two lightweight modules together make our proposed SDS-Net more lightweight, and enable it to exhibit efficient performance and benefit on resource-limited devices.
- To achieve more precise segmentation results, 3D SA and 3D SE modules are separately incorporated into the encoder and decoder of SDS-Net. By aggregating essential spatial and channel features from different modalities and refining the learning of edge features, the expression capability of crucial features and edges is significantly enhanced, leading to more accurate results in brain tumor segmentation.
- A segmentation and visualization system of brain MRI images is designed to assist doctors in diagnosis and subsequent treatment, which can be used in medical institutions such as hospitals and clinics to provide more accurate diagnosis and treatment plans.

The rest of the paper is structured as follows. Section 2 describes the overall architecture of the proposed network and provides detailed information about the experimental setup, including data preprocessing and evaluation indicators. Section 3 is the quantitative and qualitative analysis of the experimental results. Section 4 draws the main conclusions of this paper.

2. Methods

2.1. Dataset and preprocessing

In this work, several experiments are conducted on MICCAI Brain Tumor Segmentation (BRATS) Challenge datasets [31–33] and the results of the proposed network are qualitatively evaluated on the segmentation and visualization system. The BRATS 2020 contains MRI images of 369 patients with different severity of gliomas and BRATS 2021 contains 1251 patients. Each patient dataset contains four modalities T1, T2, Flair and T1C, in NIFTI format, with a size of 155 × 240 × 240 and the same voxel size of 1.0 × 1.0 × 1.0 mm^3. The training dataset consists of three different types of brain tumor labels, which were hand drawn by medical experts, i.e., the whole tumor (WT) (label 1, 2 and 4), the enhancing tumor (ET) (label 4) and the tumor core (TC) (labels 1 and 4).

Since there are differences in the intensity values from different patients, the z-score standardization operation is required to achieve a common image intensity level. The normalization equation (Eq (1)) can be used to calculate the normalized intensity value X^* of the sample data, where X represents the original intensity value of the data, while μ and σ denote the mean value and standard deviation of the data, respectively.

$$X^* = \frac{X-\mu}{\sigma} \tag{1}$$

Cropping is performed at the center of the image with a size of 128 × 128 × 128 to remove irrelevant regions and address the problem of class imbalance. Four MRI modalities are fused into 4 × 128 × 128 × 128 block as the input of the network, 128 × 128 × 128 corresponds to the input volume size and the number 4 represents four MRI modalities of Flair, T1, T1ce and T2. Preprocessing steps not only eliminate the class imbalance problem but also reduce training time and memory requirements, the results of preprocessing are shown in Figure 2.

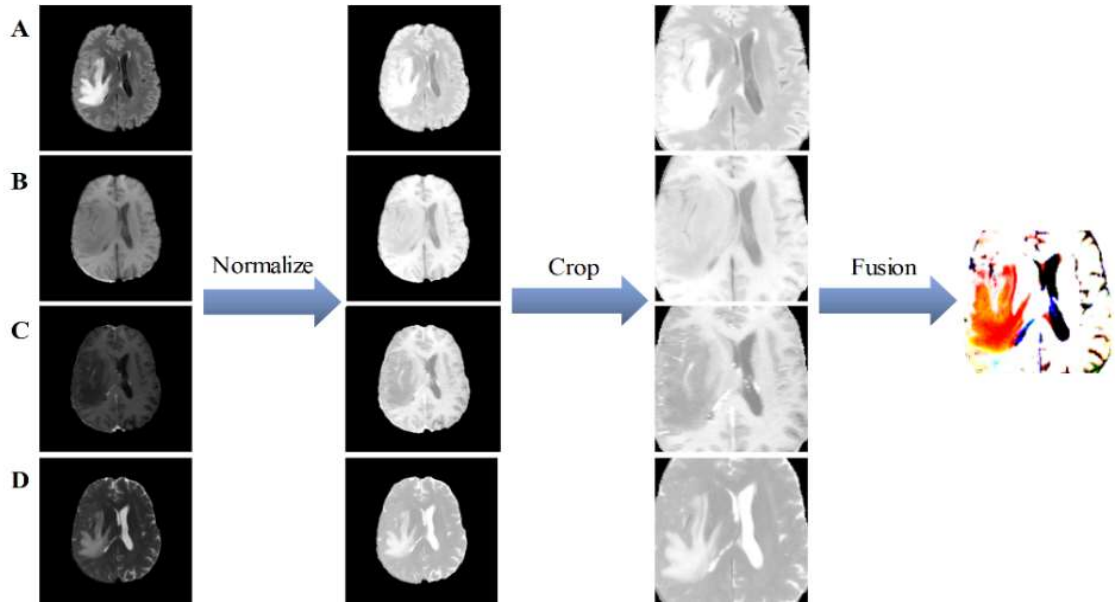

Figure 2. The flow chart of image preprocessing. (A) Flair, (B) T1, (C) T1ce and (D) T2.

2.2. Network structure

The proposed 3D lightweight network SDS-Net for brain tumor segmentation is structured as shown in Figure 3, mainly includes four key components of lightweight feature extraction (LFE) module, lightweight feature fusion (LFF) module, 3D SA module and 3D SE module. In Figure 3, the gray and gray-green blocks are 3D feature extraction blocks that consist of LFE modules with residual connection and 3D SA modules. The yellow-green and blue blocks are respectively LFF modules composed of depthwise separable convolution (DSC) merging blocks and 3D SE module. GN is used for feature normalization in the encoder and decoder, which is superior to Batch Normalization when dealing with small batch size, effectively reducing the risk of overfitting. SDS-Net combines the low-resolution feature map in the encoder with the high-resolution feature map in the decoder through skip connections to capture features at different scales, while can also help to preserve local details. The number of channels is set to 16, 32, 64, 128 and 256, and the kernel size of the convolution block is 3 × 3 × 3. The SA module is employed to enhance the ability of SDS-Net to recognize the critical features, and the SE module is used to improve the fine feature expression capability of SDS-Net; we will describe the two modules in detail below. Furthermore, in order to avoid the issues of gradient vanishing and gradient explosion, the convolution operation incorporates GN and ReLU functions to accelerate the convergence speed of the training process [34]. Table 1 shows the detailed description of the SDS-Net structure. The encoder of SDS-Net extracts features by expanding the channel dimension of the feature map and decreasing the resolution. Its decoder gradually reduces the channel size and restores resolution. To effectively combine low-level detail features with high-level semantic features, a skip connection is used to establish corresponding connections and expand path feature maps.

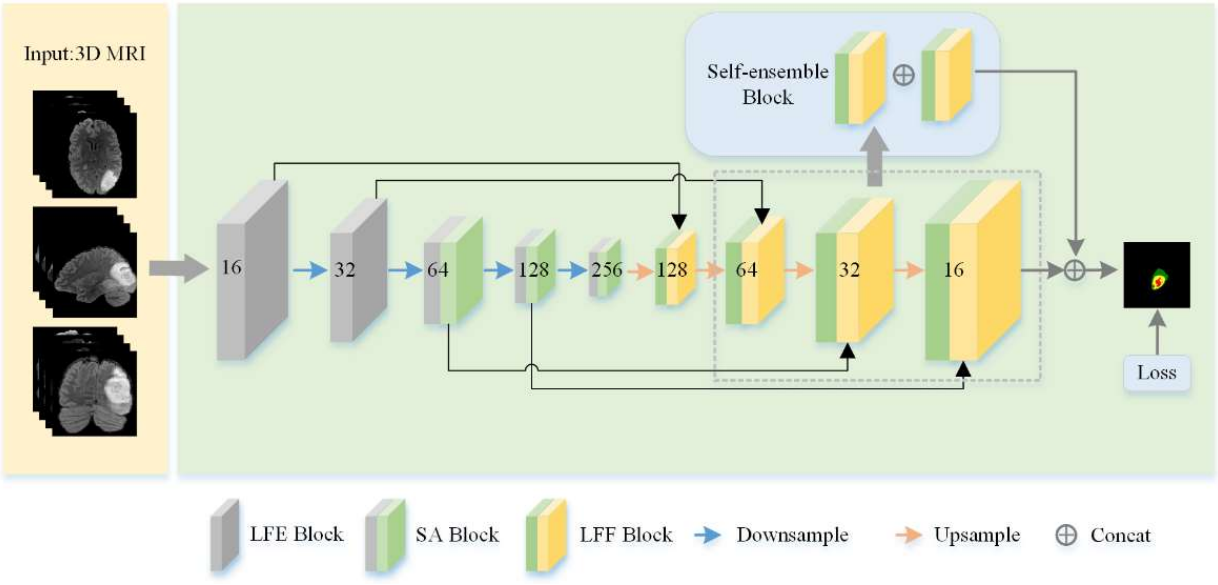

Figure 3. The overall framework of the SDS-Net network.

Table 1. Details of the proposed SDS-Net.

Layer	Input	Type	Stride	Output	Params
LFE1	128 × 128 × 128 × 4	3 × 3 × 3	1	128 × 128 × 128 × 16	9,456
Maxpool1	128 × 128 × 128 × 16	2 × 2 × 2	2	64 × 64 × 64 × 16	0
LFE 2	64 × 64 × 64 × 16	3 × 3 × 3	1	64 × 64 × 64 × 32	43,616
Maxpool2	64 × 64 × 64 × 32	2 × 2 × 2	2	32 × 32 × 32 × 32	0
SA1	32 × 32 × 32 × 32			32 × 32 × 32 × 64	55,456
Maxpool3	32 × 32 × 32 × 64	2 × 2 × 2	2	16 × 16 × 16 × 64	0
SA2	16 × 16 × 16 × 64			16 × 16 × 16 × 128	221,504
Maxpool4	16 × 16 × 16 × 128	2 × 2 × 2	2	8 × 8 × 8 × 128	0
SA3	8 × 8 × 8 × 128			8 × 8 × 8 × 256	885,376
UpSampling1	8 × 8 × 8 × 256	2 × 2 × 2		16 × 16 × 16 × 256	0
LFF1	16 × 16 × 16 × 256	3 × 3 × 3	1	16 × 16 × 16 × 128	106,496
UpSampling2	16 × 16 × 16 × 128	2 × 2 × 2		32 × 32 × 32 × 128	0
LFF2	32 × 32 × 32 × 128	3 × 3 × 3	1	32 × 32 × 32 × 64	913,664
UpSampling3	32 × 32 × 32 × 64	2 × 2 × 2		64 × 64 × 64 × 64	0
LFF3	64 × 64 × 64 × 64	3 × 3 × 3	1	64 × 64 × 64 × 32	229,504
UpSampling4	64 × 64 × 64 × 32	2 × 2 × 2		128 × 128 × 128 × 32	0
LFF4	128 × 128 × 128 × 32	3 × 3 × 3	1	128 × 128 × 128 × 16	57,920
SE	128 × 128 × 128 × 16			128 × 128 × 128 × 3	681
Total Params					2,523,673

2.2.1. 3D SA module

Class imbalance is a challenging issue, especially for smaller tumors, as verified by the BRATS datasets, where the tumor area is only 1.5% of MRI images and the ET represents only 11% of the WT [35]. In order to eliminate the impact of background on segmentation results in brain MRI images and better concentrate on tumor regions, attention mechanism is introduced into medical image segmentation. The multi-branch architecture and group features have been used to build CNN networks, such as SKNet [36], ResNet [37] and AlexNet [38]. SA, a multi-branch attention mechanism, is employed in this study to enhance the performance of the SDS-Net network, specifically in the form of a 3D SA module. The feature map is divided into smaller sub-features by grouping along the channel dimension. Then, each of these sub-features is then processed separately to construct channel attention and spatial attention.

As shown in Figure 4, the 3D SA module groups the features $X \in R^{c \times h \times w \times d}$, where c, h, w and d represent the number of channels, height, width and depth of the feature map, respectively. The feature map is first divided into several sub-features along the channel dimension, and then the spatial and channel attention mechanisms are constructed using shuffle units for each sub-feature to concatenate the spatial and channel dependencies of the features. In the channel attention mechanism, the global average pooling is used to obtain the global information of the channel. Subsequently, a linear transformation (using weights and biases) is applied, and the output is passed through the Sigmoid activation function to achieve more complex feature learning. Finally, the sub-features x_c' and x_s' are aggregated to better extract complementary information between feature maps, and the "Channel Shuffle" [25] operation is used to communicate feature information to solve the problem of information disfluency between groups. The result of experiments shows SDS-Net can further improve the recognition ability of glioma key features by using the 3D SA module, and achieve better segmentation results of tumor sub-regions while reducing computation parameters slightly.

$$x_c' = x_c \times f\left[F_C\left(F_{gp}(x_c)\right)\right] \tag{2}$$

$$x_s' = x_s \times f\left[F_C\left(GN(x_s)\right)\right] \tag{3}$$

where, x_c and x_s represent the channel path feature map and spatial path feature map, respectively. x_c' and x_s' are the corresponding outputs of the two paths. F_{gp} denotes the average pooling operation and f is the sigmoid activation function. F_C represents the linear transformation function and GN refers to the Group Normalization operation.

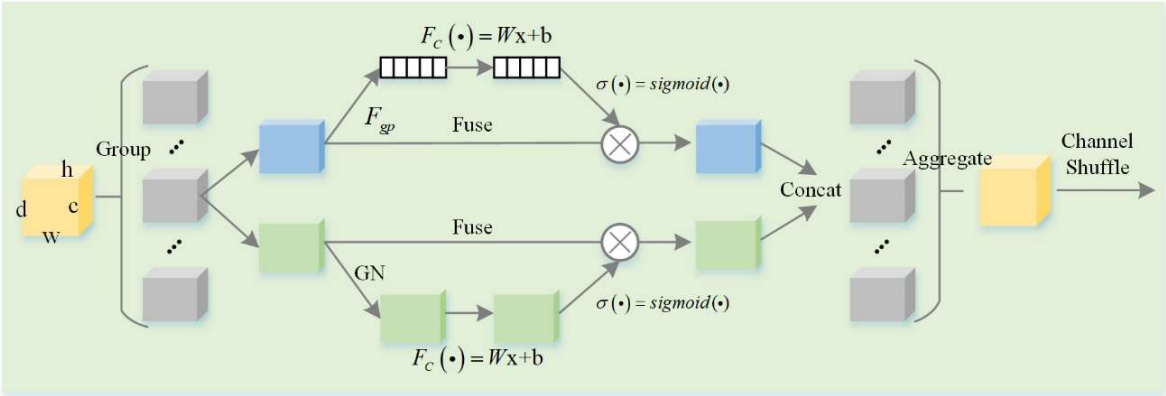

Figure 4. The structure of the 3D SA module.

2.2.2. 3D SE module

In order to alleviate the problem of edge detail loss by convolution, and further improve the network's fine feature expression capability, the 3D SE module [39,40] is added to the decoder. The structure is illustrated in Figure 5 and described by the following equation, primarily consisting of 1 × 1 × 1 convolution and up-sampling operation. First, the 3D SE module converts the features of different layers in the decoder into 3 channels through 1 × 1 × 1 convolution, and then the output feature map is subjected to an up-sampling operation to recover the image size. This approach helps

to fuse the feature information from different layers and reduce the loss of partial information, thereby further enhancing the segmentation performance.

$$\check{y}_s = C_n(y_s) + Up(\check{y}_{s-1}), \qquad s = 1,2,3 \cdots X \tag{4}$$

where $\check{y}_s$ represents the output of each layer in the SE module, y_s denotes the output of each layer in the decoder, C_n refers to the 1 × 1 × 1 convolution with a channel size of 3, Up is the up-sampling operation and X is set to 3 in this experiment.

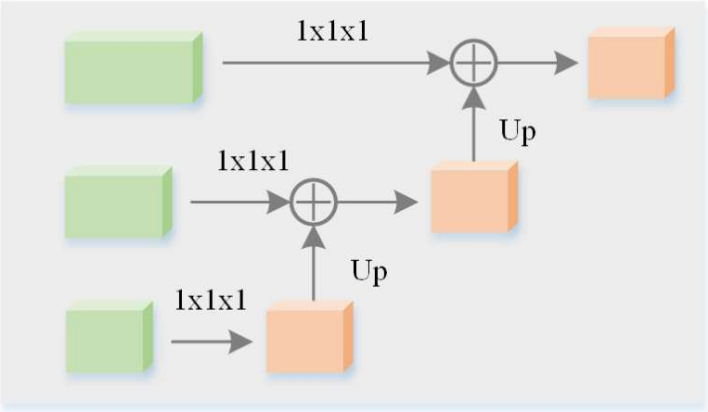

Figure 5. The structure of the 3D SE module.

2.2.3. 3D Lightweight backbone blocks

As shown in Figure 6(A), to reduce the parameters and computation of the network, DSC [24] and traditional convolution are combined to construct the lightweight backbone blocks, namely LFE and LFF modules. LFE module first performs two 3 × 3 × 3 convolution operations, comparing with a 5 × 5 × 5 convolution, using two stacked smaller convolutional kernels can result in less computational consumption, while keeping the same receptive field and perceptual background. DSCs with different convolution kernels are used for each channel's feature map and decompose the traditional convolution into two operations of deep convolution and point-by-point convolution. Deep convolution (DC) only applies one convolution in each channel of the input data to obtain a set of convoluted output feature maps, thus significantly reducing the number of parameters. Pointwise convolution followed DC applies a 1 × 1 × 1 convolution in the output feature graph and combines all the feature graphs of the deep convolution. The computation of each pointwise convolution with 1 × 1 × 1 kernel size is very small, resulting in a significant reduction in the complexity of the SDS-Net network. Finally, the output feature maps are fused by residual connection to enhance the reusability of feature information in the network.

The structure of the LFF module is depicted in Figure 6(B), where a 3 × 3 × 3 convolution operation is first performed, the size of the feature map remains unchanged and the number of channels is reduced by half, then the features are connected to the corresponding layer of the encoder by skip connection. The fusion of the features transmitted from the encoder effectively enhances and optimizes the local feature representation, thereby improving the ability to perceive details. The feature map is processed by multi-channels with the same DSC as the LFE module. The major

advantage of the three-dimensional lightweight backbone blocks is the use of fewer parameters and computation, while introducing more local features and enhancing the network's feature representation capacity. The experimental results indicate that by incorporating the LFE and LFF modules, SDS-Net achieves precise segmentation of gliomas with limited computational memory. The specific details are shown in the following equation:

$$x_1' = f\left[C_{2n}\left(C_{2n}(x_1)\right)\right] \tag{5}$$

$$y_1 = f\left[P_{2n}\left(D_{2n}(x_1')\right) + x_1'\right] \tag{6}$$

$$x_2 = Cat\left[f\left[C_{n/2}(x_2)\right], y\right] \tag{7}$$

$$y_2 = f\left[P_n\left(D_n(x_2)\right)\right] \tag{8}$$

where, x_1 and x_2 are the inputs of the LFE and LFF modules, respectively, while y_1 and y_2 are their corresponding outputs. C, D and P represent traditional convolution, deep convolution and pointwise convolution, respectively. Additionally, n denotes the number of channels, f is the ReLU activation function and Cat indicates the concatenation of two feature maps.

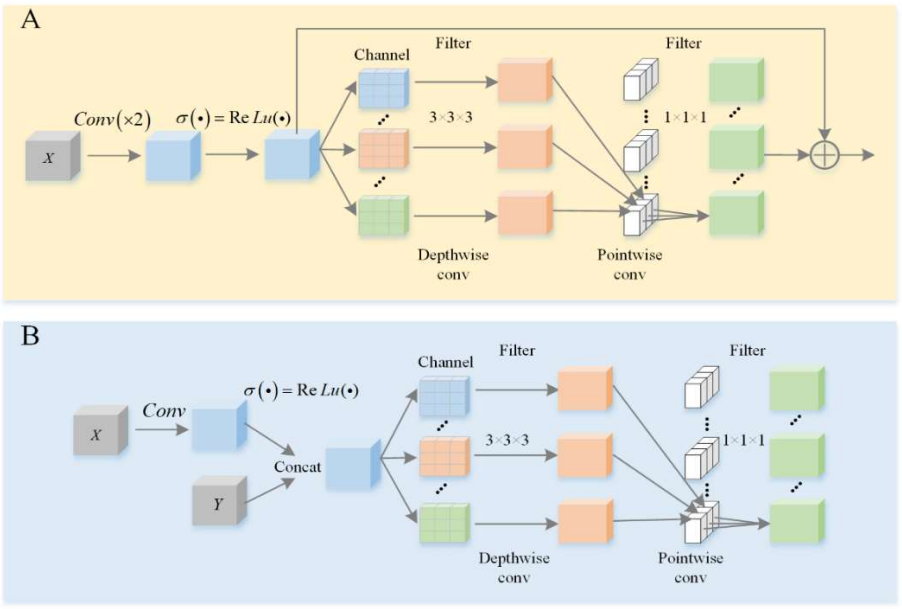

Figure 6. The structures of the LFE module and LFF module are shown in (A) and (B) respectively.

2.3. Evaluation metrics

The segmentation performance of SDS-Net is evaluated quantitatively by Dice coefficient and Hausdorff distance. The use of these metrics can help researchers more comprehensively and accurately evaluate the performance of segmentation networks, while providing more accurate auxiliary tools for tumor diagnosis and treatment in clinical medicine.

Dice coefficient (Eq (9)) is one of the commonly used metrics to evaluate segmentation accuracy. We compare the predicted region with the actual region, and a higher value indicates that the

prediction result is more similar to the ground truth. Here, True Positive (TP), False Positive (FP) and False Negative (FN) represent the number of correctly predicted tumor pixels, non-tumor pixels in tumor tissue, and pixels incorrectly predicted as healthy tissue, respectively. Dice coefficient ranges from 0 to 1, and the value 1 indicates that the prediction is consistent with the ground truth.

$$Dice = \frac{2TP}{2TP+FP+FN} \tag{9}$$

Hausdorff distance (Eq (10)) is the similarity of the boundary between the predicted region and the ground truth, where X and Y represent the predicted tumor region and ground truth, respectively, and $d(X,Y)$ is the Euclidean distance between x and y points on the predicted and ground truth.

$$d(X,Y) = max\left\{\underset{x\in X}{max}\ \underset{y\in Y}{min}\, d(x,y), \underset{y\in Y}{max}\ \underset{x\in X}{min}\, d(x,y)\right\} \tag{10}$$

2.4. Implementation details

The proposed network SDS-Net is built in the PyTorch framework with the Adam optimizer. The learning rate is set to 0.0003 and the batch size is fixed at 1. The network is trained using an NVIDIA GeForce RTX 3060 (12G) with a training period of 200 epochs, and the early stop is set to 20 to prevent overfitting. The network is optimized using a weighted mixed loss of Binary cross-entropy loss and Dice loss.

3. Results and discussion

3.1. Comparative experiments with the state-of-the-art methods

We compare the proposed SDS-Net with the various advanced networks on the BRATS 2020 dataset, including CNN, variants of CNN combined with Transformer and other lightweight methods. The quantitative comparison experimental results are shown in Table 2. Through this comparative analysis, the lightweight networks SDS-Net in this paper has demonstrated superior performance in the task of brain tumor segmentation. Specifically, compared to traditional CNN networks [13,41], our proposed SDS-Net achieves high accuracy while having lower computational complexity, using only 2.52 M parameters and 68.18 G FLOPs. This makes SDS-Net more efficient to run on resource-limited devices. Compared to the variant of CNN combined with Transformer [42,43], SDS-Net can fully leverage the local features from multimodal and the high-quality spatial and channel information, thereby better capturing image features during the segmentation process. For TC regions located between other sub-tumor regions and the difficulty to distinguish the correct boundary, our proposed method achieves the best Dice coefficient of 88.9% in the challenging region. Compared to other lightweight methods [30,44,45], Zhang et al. [30] introduced a hierarchical multi-scale network, achieving an excellent parameter count of 0.80 M. Chen et al. [45] proposed a novel 3D dilated multi-fiber network that demonstrates remarkable computational efficiency with 27.04 G FLOPs. Although SDS-Net has slightly higher parameter and computational complexity than these two networks, our method still achieves satisfactory results in the brain tumor segmentation task. With the flexible design of the network structure, SDS-Net achieves higher Dice coefficient in the WT, ET and TC regions, with values of 92.7, 80.0 and 88.9%, respectively.

Table 2. Comparisons of the segmentation performance in the BRATS 2020 dataset.

Methods	FLOPs	Parameters	Dice			Hausdorff		
			WT	ET	TC	WT	ET	TC
Nuechterlein et al. [44]	76.51 G	3.63 M	0.883	0.737	0.814	7.10	31.30	14.62
Zhang et al. [30]	129.40 G	0.80 M	0.901	0.781	0.823	5.95	21.34	7.06
Isensee et al. [41]	1902.15 G	16.32 M	0.907	0.814	0.848	6.94	5.85	5.07
Wang et al. [42]	333.00 G	32.99 M	0.901	0.787	0.855	4.96	17.95	9.77
Xing et al. [43]	71.77 G	32.99 M	0.920	0.800	0.864	4.57	5.27	5.32
Chen et al. [45]	27.04 G	3.88 M	0.901	0.754	0.824	–	–	–
SDS-Net (Ours)	68.18 G	2.52 M	0.927	0.800	0.889	17.86	12.93	11.35

We also compared the proposed SDS-Net with other networks on the BRATS 2021 dataset, and the detailed results of these comparative experiments are shown in Table 3. Peiris et al. [46] enhanced the generalization and robustness of the segmentation network by employing dual adversarial learning and introducing uncertain information. In comparison, the proposed SDS-Net exhibits superior performance in segmentation accuracy and uses 1/13 of its parameters. Liu et al. [40] combined residual blocks and spatial group-wise attention blocks, achieving the best Dice coefficients in ET and TC segmentation tasks, with 83.3 and 86.8%, respectively. Jiang et al. [47] proposed a novel medical image segmentation network SwinBTS combined with Transformer, which showed excellent performance in the brain tumor segmentation task, achieving 91.8% Dice score on the WT region segmentation. In our experiments, we carefully considered the balance between network performance and resource consumption. SDS-Net achieved Dice scores of 91.8, 82.5 and 86.8% in WT, ET and TC, respectively. In addition, the proposed network uses 2.52 M parameters and 68.18 G FLOPs. Therefore, the lightweight network proposed in this paper has more prominent advantages overall, achieving better segmentation results with fewer parameters and computational complexity.

Table 3. Comparisons of the segmentation performance in the BRATS 2021 dataset.

Methods	FLOPs	Parameters	Dice			Hausdorff		
			WT	ET	TC	WT	ET	TC
Çiçek et al. [13]	1902.15 G	16.32 M	0.880	0.757	0.782	25.03	14.11	13.76
Liu et al. [40]	–	–	0.916	0.833	0.868	5.94	19.28	7.57
Peiris et al. [46]	776.11 G	19.17 M	0.908	0.814	0.854	5.37	21.83	8.56
Jiang et al. [47]	–	–	0.918	0.832	0.848	3.65	16.03	14.51
SDS-Net (Ours)	68.18 G	2.52 M	0.918	0.825	0.868	21.07	13.13	11.99

3.2. Ablation study

3.2.1. Evaluating the effectiveness of different modules

We perform ablation experiments to evaluate the results of different SDS-Net modules on the BRATS 2021 dataset. The U-Net network structure has shown promising results in brain tumor segmentation. First, we use the modified U-Net as the benchmark network, where the encoder includes 5 convolution modules and 4 max pooling layers for feature extraction and resolution reduction. The decoder consists of 4 de-convolution modules and 4 tri-linear interpolation up-sampling operations to restore the edge information of the image. To fairly show the efficiency of each SDS-Net module, the SDS-Net has the same input size, network layers number and training parameters as the benchmark network. We gradually add feature extraction module SE, attention mechanism SA module and lightweight backbone blocks to the benchmark network, and then compare the computational complexity and segmentation precision through ablation experiments.

The results are illustrated in Table 4. First, to improve the feature fusion in the decoder, we add the 3D SE module to the benchmark. As can be seen from the Table 4, the Dice coefficient of WT, ET and TC is increased by 0.4, 0.3 and 0.9%, respectively, while the Hausdorff distance is decreased by 0.47, 1.65 and 1.13 mm, respectively. The evaluation metrics in the three tumor regions are all improved, but the computational cost of the network is slightly increased. Second, in order to reduce the computation burden, the lightweight 3D SA module is also added to the benchmark network, the last three convolutional modules in the encoder are replaced with triple 3D SA modules, which not only reduce the number of parameters, but also enhance the network's ability to extract key features, resulting in the optimal Dice coefficient and the lowest Hausdorff distance for the WT and ET regions. The experiments show that replacing by the 3D SA modules can light the FLOPs to 190.27 G and greatly reduce computation parameters to 4.80 M, while maintaining the segmentation precision. Finally, the lightweight blocks with the combination of DSCs and traditional convolution, two LFE modules are adopted to the encoder and four LFF modules are used to the decoder, respectively, which further achieve the more lightweight network with parameter amount 2.52 M and computational cost 68.18 G FLOPs. Compared with the benchmark network, our SDS-Net has significant improvements in all metrics of the tumor regions, with Dice coefficient enhancing by 0.6, 0.7 and 1.2% for WT, ET and TC, respectively, while the Hausdorff distance is decreased by 0.92, 1.74 and 2.47 mm for WT, ET and TC, respectively.

Figure 7 shows the results of ablation experiments for different modules, and the Dice coefficients are shown using boxplots and bar graph. From the figure, it can be observed that the proposed SDS-Net presents fewer instances of abnormal and scattered data, indicating that it has high stability and consistency. Although the performance of SDS-Net in the ET region is slightly lower than baseline + SE + SA method, overall, it shows significant effectiveness in improving segmentation accuracy. To show the segmentation effect more clearly, Figure 8 shows the visual segmentation results of the ablation experiment. In the second row, we can observe that the baseline network exhibits limited performance in accurately segmenting the individual edema region in the upper right corner. However, with the gradual introduction of functional modules into the baseline network, resulting in SDS-Net, the segmentation performance for the individual region gradually approaches the ground truth. Therefore, this further verifies the effectiveness of each model and demonstrates the performance of our SDS-Net.

Table 4. The results of ablation study with different modules.

Methods	FLOPs	Parameters	Dice			Hausdorff		
			WT	ET	TC	WT	ET	TC
Baseline	202.89 G	9.45 M	0.912	0.818	0.856	21.99	14.87	14.46
Baseline + SE	202.95 G	9.45 M	0.916	0.821	0.865	21.52	13.22	13.33
Baseline + SE+ SA	190.27 G	4.80 M	0.918	**0.830**	0.861	**20.02**	15.13	14.74
SDS-Net (Ours)	**68.18 G**	**2.52 M**	**0.918**	0.825	**0.868**	21.07	**13.13**	**11.99**

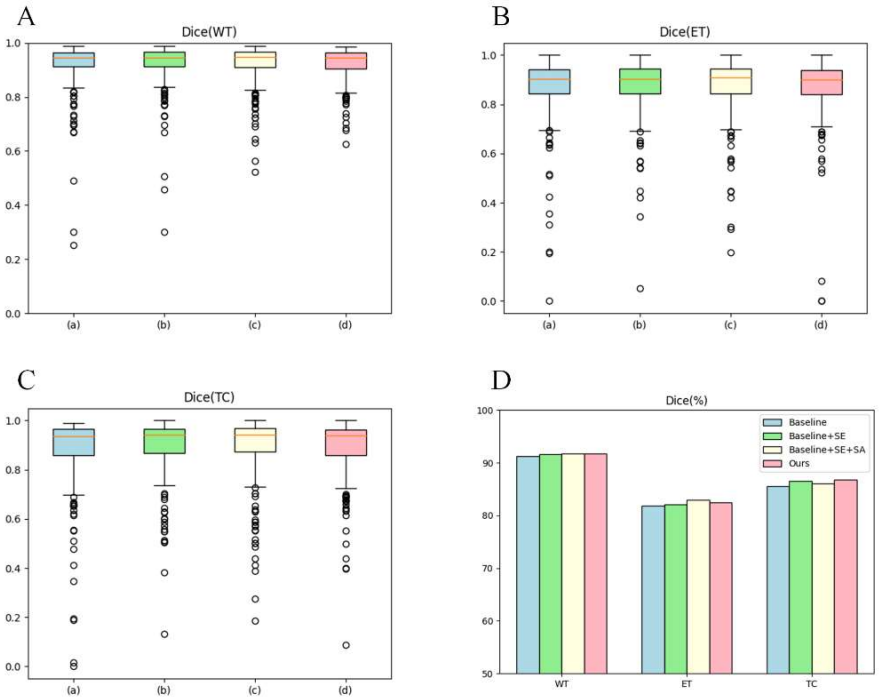

Figure 7. Boxplots and bar graph of the Dice of ablation experiments for different modules. (A)–(C) are the Dice of WT, ET and TC, respectively. Within the (A)–(C) figures, (A)–(D) represent baseline, baseline + SE, baseline + SE +SA and SDS-Net. (D) is the bar chart comparison result of different networks.

3.2.2. Evaluating the effectiveness of different attention mechanisms

To evaluate the effectiveness of different attention mechanisms in SDS-Net, ablation experiments are performed on the BRATS 2021 dataset. Three different attention mechanisms, including the CBAM module, the 3D SE* module and the 3D SA module, are explored and the performances are compared in the segmentation tasks. Table 5 illustrates the results of the ablation experiments. From Table 5, 3D CBAM obtained the second place in Hausdorff distance of the WT region, but it does not perform well in other metrics. In the ET region, the 3D SE* achieved the optimal results, with a highest Dice coefficient of 82.7% and a lowest Hausdorff distance of 13.00.

3D SA almost outperformed the other two different attention mechanisms in all metrics, except for a slightly lower metric of the ET region than 3D SE*, the Dice coefficients of WT, ET and TC brain tumor regions are 91.8, 82.5 and 86.8% respectively, and the optimal Hausdorff distance is achieved in the WT and TC regions. Thus, we choose 3D SA as the attention module for SDS-Net encoder. Figure 9 visualizes the segmentation results obtained by different attention mechanisms of the network SDS-Net, the visualization shows the results obtained from 3D SA are closer to the ground truth than CBAM and 3D SE*. The 3D SA module incorporates an ultra-lightweight attention mechanism that combines spatial attention and channel attention; unfortunately, there is no notable change of network parameters and computational cost comparing to the other two attention mechanisms.

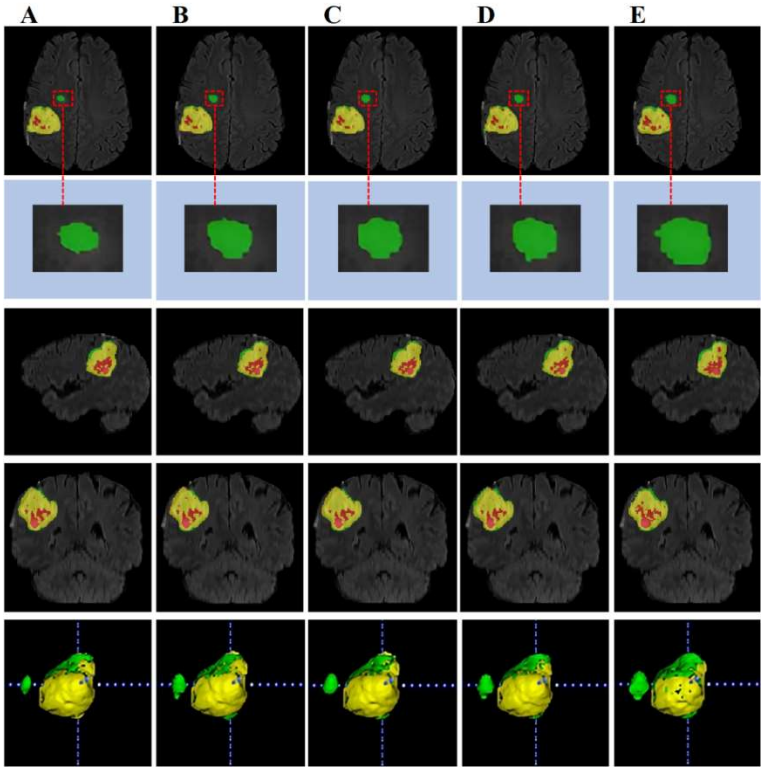

Figure 8. Visualization of segmentation results. (A)–(E) are the segmentation results of the different networks in Table 4, as well as the ground truth. The red area is necrotic and non-enhancing, the green area is edema and the yellow area is enhancing tumor.

Table 5. The results of ablation study with attention mechanisms.

Methods	FLOPs	Parameters	Dice			Hausdorff		
			WT	ET	TC	WT	ET	TC
SDS-Net + CBAM	68.18 G	2.53 M	0.901	0.798	0.845	22.03	13.52	13.43
SDS-Net + SE*	68.18 G	2.53 M	0.915	**0.827**	0.859	23.95	**13.00**	13.62
SDS-Net + SA	**68.18 G**	**2.52 M**	**0.918**	0.825	**0.868**	**21.07**	13.13	**11.99**

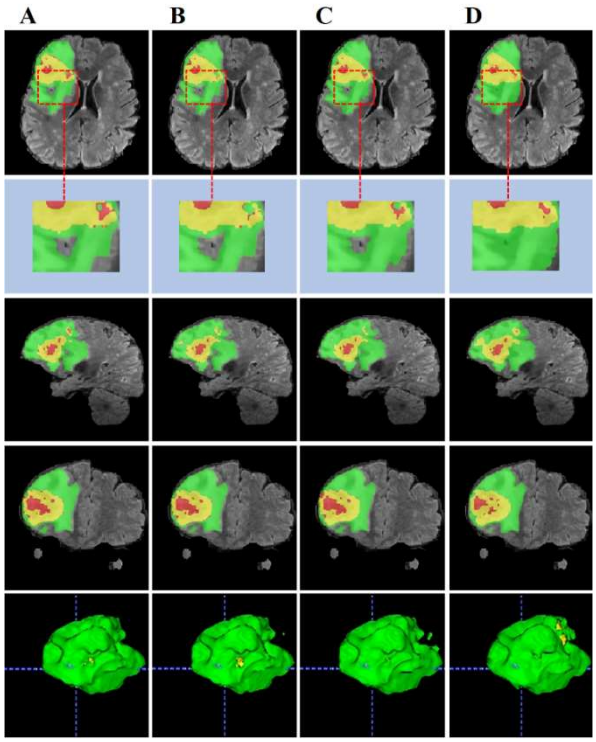

Figure 9. Visualization of segmentation results. (A-D) are the segmentation results of CBAM block, SE* block and SA block respectively added to the network in Table 5, as well as the ground truth. The red area is necrotic and non-enhancing, the green area is edema and the yellow area is enhancing tumor.

3.3. Convergence of loss functions

The Binary cross-entropy (BCE) loss (Eq (11)) is commonly used in binary classification tasks. This loss function is calculated on a pixel-by-pixel basis, comparing the probability of the actual voxel value (P_i) with the predicted value (G_i). When P_i is equal to 1, the loss function is log (G_i); when P_i is equal to 0, the loss function is log $(1 - G_i)$, where N is the number of voxels.

$$BCE = -\left[\sum_{i=1}^{N} P_i \, log \, G_i + (1 - P_i) \, log(1 - G_i)\right] \tag{11}$$

Another frequently used loss function (Eq (12)) is the Dice loss function, which aims to optimize the similarity between the actual voxel value (P_i) and the predicted mask value (G_i).

$$Dice = 1 - \frac{2\sum_{i=1}^{N} P_i G_i}{\sum_{i=1}^{N} P_i + \sum_{i=1}^{N} G_i} \tag{12}$$

In this experiment, the BCE-Dice loss is taken to be the loss function (Eq (13)), which is a weighted hybrid of the Binary cross-entropy and Dice loss function. The convergence of the loss function is an important indicator for measuring the performance of CNNs. Faster convergence speed cannot only significantly reduce the training complexity but also enhance the segmentation performance. Figure 10 shows the convergence trends of the loss functions for 3D U-Net, Benchmark network and our SDS-Net trained on BRATS 2021 dataset. We can observe that SDS-Net converges faster comparing to other networks.

$$BCEDice = 0.5BCE + Dice \tag{13}$$

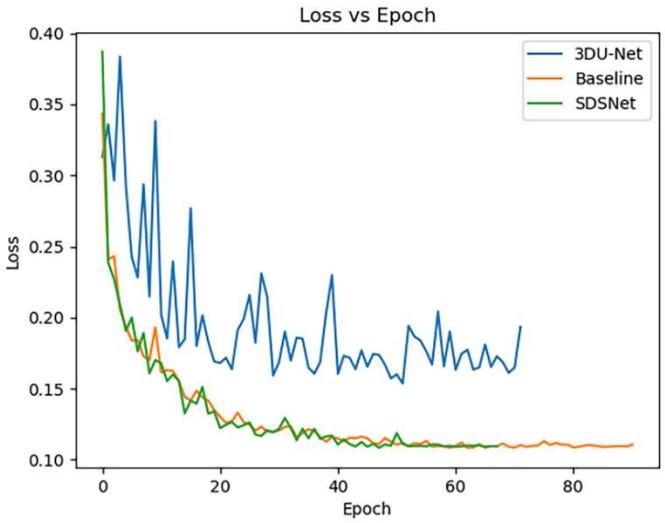

Figure 10. The compared loss results in the training stage for 3D U-Net, Benchmark and SDS-Net.

3.4. Computational complexity

The common evaluation metrics to measure network performance and complexity are FLOPs and the number of parameters, respectively. Table 6 presents the number of parameters and FLOPs for the 3D U-Net, Benchmark network and SDS-Net with the same brain image size of 128 × 128 × 128. SDS-Net achieves 2.52 M parameters and 68.18 G FLOPs, while, 3D U-Net and Benchmark network get 16.32, 9.45 M parameters and 1902.15, 202.89 G FLOPs, respectively. In comparison, our SDS-Net uses fewer parameters and lower computational cost. The number of parameters and FLOPs of the 3D conventional convolution can be calculated as Eqs (14) and (15).

$$Parameters = C_{in} \times C_{out} \times k_H \times k_w \times k_D + C_{out} \tag{14}$$

$$FLOPs = 2 \times O_h \times O_w \times O_d \times Parameters \tag{15}$$

The number of parameters and FLOPs for the 3D DSC can be calculated as Eqs (16) and (17).

$$Parameters^* = (C_{in} \times k_H \times k_w \times k_D + 1) + (C_{in} + 1) \times C_{out} \tag{16}$$

$$FLOPs^* = 2 \times O_h \times O_w \times O_d \times Parameters^* \tag{17}$$

where C_{in} and C_{out} represent the number of input and output channels of the feature map respectively. The kernel size is $k_H \times k_w \times k_D$. The depth, height and width of the output feature map are O_h, O_w and O_d, respectively.

Table 6. Comparison of the number of parameters and FLOPs.

Network	Parameters(M)	FLOPs(G)
3D U-Net	16.32 M	1902.15 G
Baseline	9.45 M	202.89 G
SDS-Net (Ours)	**2.52 M**	**68.18 G**

3.5. Segmentation and visualization system

CNN has become one of the most popular researches in data processing and analysis, such as image registration, segmentation and fusion. Especially in the medical image processing, deep learning technology has been widely used in the automatic detection and diagnosis of diseases, among which brain tumor segmentation is a crucial task. CNN typically requires large amounts of parameters and computational resources, which may pose challenges in medical devices with limited computing resources. Therefore, lightweight CNN has become increasingly vital to medical data application. In this study, a lightweight segmentation network is proposed that not only achieves better segmentation results with fewer parameters and computational complexity, but can also be applied to medical devices to assist doctors in tumor diagnosis and treatment plans.

Figure 11. The visualization of the segmentation system.

The segmentation and visualization system consists of the image acquisition module, image pre-processing module, 3D image segmentation module and image visualization module. Figure 11 shows the interface of the brain tumor segmentation and visualization system, which includes information of the patient, 2D slices of all MRI modalities and the segmentation results of the tumor area. Three feature types are drawn with different colors to differentiate various labels, while the visualization module can calculate the area size of the segmentation result and fuse the tumor area with the original image. The system is developed by PyCharm 2021.2.2 and PyQt5 with Windows 10, 16 GB memory and RTX3060 GPU, which can be used in medical institutions such as hospitals and clinics. 3D image segmentation module loads the lightweight segmentation network SDS-Net, so it can work stably, fast, accurately and conveniently display the segmentation results of brain MRI in the same group.

4. Conclusions

In this paper, we introduce SDS-Net, a lightweight and efficient 3D network designed for brain tumor MRI segmentation, which consists of four key components of LFE module, LFF module, 3D SA module and 3D SE module. Encoder block consists of LFE module with residual connection and 3D SA module, decoder block with DSC is composed of LFF module and 3D SE module. Specially, SA is placed in the encoder and SE fine feature expression is added to the decoder of SDS-Net; by the operations, SDS-Net can efficiently concentrate on the region of interest in the images and extract the information rich image features. SDS-Net combines DSC and traditional convolution in the LFE and LFF modules, which further helps to decrease the network parameters and computer cost, making the network more lightweight and efficient. GN for feature normalization and skip connection is also used in the encoder and decoder. In addition, BCE-Dice loss function helps to classify pixels more accurately and makes the network converge faster. Finally, ablation experiments explore the effect of the addition of different modules to SDS-Net, including using SE module, lightweight backbone blocks and replacing the last three feature extraction modules by 3D SA modules. Moreover, we compare SDS-Net with the advanced approaches on the BRATS datasets, and through quantitative analysis and visual comparison of SDS-Net segmentation results with the labels, it can be found that SDS-Net has better computational efficiency, lower parameter amount and the best segmentation performance.

However, this study also has certain limitations: First, while multi-modal image data is currently processed using early fusion methods, further research is needed to determine whether the later fusion of multi-modal images is superior. Second, in the future, we plan to migrate the proposed method to the field of ischemic stroke lesion segmentation.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

This research was funded by the Anhui University Scientific Research Project under Grant (2022AH050660).

Conflict of interest

The authors declare there is no conflict of interest.

References

1. Z. Zhou, Z. He, Y. Jia, AFPNet: A 3D fully convolutional neural network with atrous-convolution feature pyramid for brain tumor segmentation via MRI images, *Neurocomputing*, **402** (2020), 235–244. https://10.1016/j.neucom.2020.03.097

2. R. Cao, X. Pei, N. Ge, C. Zheng, Clinical target volume auto-segmentation of esophageal cancer for radiotherapy after radical surgery based on deep learning, *Technol. Cancer Res. Treat.*, **20** (2021), 15330338211034284. https://10.1177/15330338211034284
3. J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (2015), 3431–3440. https://10.1109/CVPR.2015.7298965
4. O. Ronneberger, P. Fischer, T. Brox, U-Net: Convolutional networks for biomedical image segmentation, in *International Conference on Medical lmage Computing and Computer-Assisted Intervention*, (2015), 234–241. https://10.1007/978-3-319-24574-4_28
5. A. Hatamizadeh, Y. Tang, V. Nath, D. Yang, A. Myronenko, B. Landman, et al., UNETR: Transformers for 3D medical image segmentation, in *2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, (2022), 1748–1758. https://10.1109/WACV51458.2022.00181
6. A. Hou, L. Wu, H. Sun, Q. Yang, H. Ji, B. Cui, et al., Brain segmentation based on UNet++ with weighted parameters and convolutional neural network, in *2021 IEEE International Conference on Advances in Electrical Engineering and Computer Applications (AEECA)*, (2021), 644–648. https://10.1109/AEECA52519.2021.9574279
7. S. Li, J. Liu, Z. Song, Brain tumor segmentation based on region of interest-aided localization and segmentation U-Net, *Int. J. Mach. Learn. Cybern.*, **13** (2022), 2435–2445. https://10.1007/s13042-022-01536-4
8. N. Sheng, D. Liu, J. Zhang, C. Che, J. Zhang, Second-order ResU-Net for automatic MRI brain tumor segmentation, *Math. Biosci. Eng.*, **18** (2021), 4943–4960. https://10.3934/mbe.2021251
9. A. Jungo, R. Mckinley, R. Meier, U. Knecht, L. Vera, J. Pérez-Beteta, et al., Towards uncertainty-assisted brain tumor segmentation and survival prediction, in *International MICCAI Brainlesion Workshop*, (2018), 474–485. https://10.1007/978-3-319-75238-9_40
10. N. Cinar, A. Ozcan, M. Kaya, A hybrid DenseNet121-UNet model for brain tumor segmentation from MR Images, *Biomed. Signal Process. Control*, **76** (2022), 103647. https://10.1016/j.bspc.2022.103647
11. L. Wu, S. Hu, C. Liu, MR brain segmentation based on DE-ResUnet combining texture features and background knowledge, *Biomed. Signal Process. Control*, **75** (2022), 103541. https://10.1016/j.bspc.2022.103541
12. M. Jiang, F. Zhai, J. Kong, A novel deep learning model DDU-net using edge features to enhance brain tumor segmentation on MR images, *Artif. Intell. Med.*, **121** (2021), 102180. https://10.1016/j.artmed.2021.102180
13. Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, O. Ronneberger, 3D U-Net: Learning dense volumetric segmentation from sparse annotation, in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, (2016), 424–432. https://10.1007/978-3-319-46723-8_49
14. R. Mehta, T. Arbel, 3D U-Net for brain tumour segmentation, in *International MICCAI Brainlesion Workshop*, (2019), 254–266. https://10.1007/978-3-030-11726-9_23
15. A. Abdollahi, B. Pradhan, A. Alamri, VNet: An end-to-end fully convolutional neural network for road extraction from high-resolution remote sensing data, *IEEE Access*, **8** (2020), 179424–179436. https://10.1109/ACCESS.2020.3026658

16. 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://10.1016/j.inffus.2022.10.022
17. Y. Li, Z. Wang, L. Yin, Z. Zhu, G. Qi, Y. Liu, X-Net: A dual encoding–decoding method in medical image segmentation, *Vis. Comput.*, **39** (2023), 2223–2233. https://10.1007/s00371-021-02328-7
18. 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). https://10.3389/fnins.2022.1009581
19. D. Kong, X. Liu, Y. Wang, D. Li, J. Xue, 3D hierarchical dual-attention fully convolutional networks with hybrid losses for diverse glioma segmentation, *Knowl. Based Syst.*, **237** (2022), 107692. https://10.1016/j.knosys.2021.107692
20. O. Oktay, J. Schlemper, L. L. Folgoc, M. J. Lee, M. P. Heinrich, K. Misawa, et al., Attention U-Net: Learning where to look for the pancreas, preprint, arXiv:1804.03999. https://10.48550/arXiv.1804.03999
21. X. He, G. Qi, Z. Zhu, Y. Li, B. Cong, L. Bai, Medical image segmentation method based on multi-feature interaction and fusion over cloud computing, *Simul. Model. Pract. Theory*, **126** (2023), 102769. https://10.1016/j.simpat.2023.102769
22. A. G. Roy, N. Navab, C. Wachinger, Concurrent spatial and channel 'squeeze & excitation'in fully convolutional networks, in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, (2018), 421–429. https://10.1007/978-3-030-00928-1_48
23. 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://10.1109/ICASSP39728.2021.9414568
24. A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, et al., Mobilenets: Efficient convolutional neural networks for mobile vision applications, preprint, arXiv:1704.04861. https://10.48550/arXiv.1704.04861
25. X. Zhang, X. Zhou, M. Lin, J. Sun, Shufflenet: An extremely efficient convolutional neural network for mobile devices, in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, (2018), 6848–6856. https://10.1109/CVPR.2018.00716
26. K. Han, Y. Wang, Q. Tian, J. Guo, C. Xu, C. Xu, Ghostnet: More features from cheap operations, in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, (2020), 1580–1589. https://10.1109/CVPR42600.2020.00165
27. X. Zhou, X. Li, K. Hu, Y. Zhang, Z. Chen, X. Gao, ERV-Net: An efficient 3D residual neural network for brain tumor segmentation, *Expert Syst. Appl.*, **170** (2021), 114566. https://10.1016/j.eswa.2021.114566
28. K. R. Reddy, R. Dhuli, A novel lightweight CNN architecture for the diagnosis of brain tumors using MR images, *Diagnostics*, **13** (2023), 312. https://10.3390/diagnostics13020312
29. Z. Luo, Z. Jia, Z. Yuan, J. Peng, HDC-Net: Hierarchical decoupled convolution network for brain tumor segmentation, *IEEE J. Biomed. Health Inform.*, **25** (2021), 737–745. https://10.1109/JBHI.2020.2998146
30. R. Zhang, S. Jia, M. J. Adamuand, W. Nie, Q. Li, T. Wu, HMNet: Hierarchical multi-scale brain tumor segmentation network, *J. Clin. Med.*, **12** (2023), 538. https://10.3390/jcm12020538

31. U. Baid, S. Ghodasara, S. Mohan, M. Bilello, E. Calabrese, E. Colak, et al., The RSNA-ASNR-MICCAI BraTS 2021 benchmark on brain tumor segmentation and radiogenomic classification, preprint, arXiv:2107.02314. https://10.48550/arXiv.2107.02314
32. B. H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, et al., The multimodal brain tumor image segmentation benchmark (BRATS), *IEEE Trans. Med. Imaging*, **34** (2015), 1993–2024. https://10.1109/TMI.2014.2377694
33. 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), 170117. https://10.1038/sdata.2017.117
34. Y. Cao, W. Zhou, M. Zang, D. An, Y. Feng, B. Yu, MBANet: A 3D convolutional neural network with multi-branch attention for brain tumor segmentation from MRI images, *Biomed. Signal Process. Control*, **80** (2023), 104296. https://10.1016/j.bspc.2022.104296
35. P. Wang, A. C. S. Chung, Relax and focus on brain tumor segmentation, *Med. Image Anal.*, **75** (2022), 102259. https://10.1016/j.media.2021.102259
36. X. Li, W. Wang, X. Hu, J. Yang, Selective kernel networks, in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, (2019), 510–519. https://10.1109/CVPR.2019.00060
37. K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in *2016 IEEE Conference on Computer Vision and Pattern Recognition(CVPR)*, (2016), 770–778. https://10.1109/CVPR.2016.90
38. A. Krizhevsky, I. Sutskever, G. E. Hinton, ImageNet classification with deep convolutional neural networks, *Commun. ACM*, **60** (2017), 84–90. https://10.1145/3065386
39. Y. X. Zhao, Y. M. Zhang, C. L. Liu, Bag of tricks for 3D MRI brain tumor segmentation, in *International MICCAI Brainlesion Workshop*, (2019), 210–220. https://10.1007/978-3-030-46640-4_20
40. 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://10.3934/mbe.2022261
41. F. Isensee, P. F. Jäger, P. M. Full, P. Vollmuth, K. H. Maier-Hein, nnU-Net for brain tumor segmentation, in *International MICCAI Brainlesion Workshop*, (2021), 118–132. https://10.1007/978-3-030-72087-2_11
42. W. Wang, C. Chen, M. Ding, H. Yu, S. Zha, J. Li, TransBTS: Multimodal brain tumor segmentation using transformer, in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, (2021), 109–119. https://10.1007/978-3-030-87193-2_11
43. Z. Xing, L. Yu, L. Wan, T. Han, L. Zhu, NestedFormer: Nested modality-aware transformer for brain tumor segmentation, in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, (2022), 140–150. https://10.1007/978-3-031-16443-9_14
44. N. Nuechterlein, S. Mehta, 3D-ESPNet with pyramidal refinement for volumetric brain tumor image segmentation, in *International MICCAI Brainlesion Workshop*, (2018), 245–253. https://10.1007/978-3-030-11726-9_22
45. C. Chen, X. Liu, M. Ding, J. Zheng, J. Li, 3D dilated multi-fiber network for real-time brain tumor segmentation in MRI, in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, (2019), 184–192. https://10.1007/978-3-030-32248-9_21

46. H. Peiris, Z. Chen, G. Egan, M. Harandi, Reciprocal adversarial learning for brain tumor segmentation: A solution to BraTS challenge 2021 segmentation task, in *International MICCAI Brainlesion Workshop*, (2022), 171–181. https://10.1007/978-3-031-08999-2_13
47. 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). https://10.3390/brainsci12060797

AIMS Press