

Received 18 August 2024, accepted 11 September 2024, date of publication 16 September 2024,  
date of current version 26 September 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3460797

## RESEARCH ARTICLE

# BTIS-Net: Efficient 3D U-Net for Brain Tumor Image Segmentation

LI LIU<sup>1,2</sup> AND KAIJIAN XIA<sup>3</sup>, (Member, IEEE)

<sup>1</sup>Jiangsu Vocational College of Information Technology, Wuxi, Jiangsu 214153, China

<sup>2</sup>Department of Biomedical Engineering, Faculty of Engineering, University of Malaya, Kuala Lumpur 50603, Malaysia

<sup>3</sup>Changshu Hospital, Soochow University, Changsu, Jiangsu 215500, China

Corresponding author: Kaijian Xia (KJXIA@suda.edu.cn)

Engineering Technology Research and Development Center of Higher Vocational Colleges in Jiangsu Province (Su Jiao Ke Han [2023] No. 11), Excellent Science and Technology Innovation Team of Higher Vocational Colleges in Jiangsu Province (Su Jiao Ke [2023] No. 3), Cultivation Project of Demonstration Virtual Simulation Training Base in Jiangsu Province (Su Jiao Zhi Han [2023] No. 30), "Dual-qualified" Master Teacher Studio of Vocational Education in Jiangsu Province (Su Jiao Shi Han [2022] No. 31), High-level Key Professional Construction Project in Jiangsu Province (Su Jiao Gao [2017] No. 17), Excellent Young Key Teachers of "Blue and Green Project" of Higher Vocational Colleges in Jiangsu Province (Su Jiao Shi Han [2023] No. 27).

**ABSTRACT** Brain tumor segmentation techniques are essential for the precise delineation of tumors and normal brain tissues which is essential for the guidance of surgical intervention and clinical decisions. However, for resource-constrained clinical environments, more efficient and lightweight segmentation models are needed so that they can be applied in real-time for surgical navigation and clinical decision-making. To tackle this issue, the proposed study introduces a very effective 3D U-Net model that is specifically designed for brain tumor image segmentation. This study presents the primary contributions as follows: 3D depth separable convolution is introduced to decrease the number of model training parameters, hence enhancing the overall efficiency of the model. The dilated dense residual block is designed to expand the sensory field, allowing the network to grasp a broader range of features and structures within the input data. As a result, the model's performance and generalization ability to handle complex tasks are improved. The integration of the confusion area segmentation module enhances the model's capability to discern intricate image details and edges, thereby augmenting the overall segmentation efficacy. Evaluation of the proposed BTIS-Net involves experimentation on two widely recognized datasets, namely BraTS 2019 and BraTS 2021. The Dice similarity coefficient, Positive predictive value, and Sensitivity exhibit average improvements of 5.68, 5.38, and 2.14, respectively. Additionally, the Hausdorff distance is reduced by an average of 2.71. The experimental results validate the efficient segmentation performance of the BTIS-Net model, showcasing exceptional outcomes even under resource constraints.

**INDEX TERMS** 3D U-Net, 3D depth separable convolution, brain tumor image segmentation, confusion area segmentation, dilated dense residual block.

## I. INTRODUCTION

The correct and quick brain tumor segmentation(BTS) is of utmost importance in clinical diagnosis and therapy planning, as it represents a significant neurological illness [1]. The importance of precise brain tumor segmentation has increased due to the ongoing advancements in medical imaging technology, particularly the widespread utilization of high-resolution brain magnetic resonance imaging (MRI) technology.

The associate editor coordinating the review of this manuscript and approving it for publication was Binit Lukose<sup>1b</sup>.

Traditional medical image segmentation methods face many challenges when dealing with brain tumors, such as complex tumor morphology, variability between different patients, and image noise. The field of medical image segmentation has shown notable advancements in recent years due to the emergence and widespread adoption of deep learning algorithms [2]. U-Net architecture has demonstrated exceptional performance in the domain of medical images [3]. This project aims to examine the application of 3D U-Net, a deep learning-based method, in segmenting medical images of brain tumors. Deep learning models have the advantage

of independently acquiring representative features that are better suited to different types and shapes of tumors [4]. The objective of this project is to enhance the precision and effectiveness of BTS within the constraints of limited computational resources. Additionally, the aim is to equip healthcare practitioners with more dependable diagnostic tools for real-world production settings.

Traditional brain tumor image segmentation (BTIS) is mainly based on threshold segmentation [5], region growth [6], and edge detection [7]. Nevertheless, the intricate nature of brain tumor images may restrict the effectiveness of conventional techniques when confronted with tumors of varying forms and intensities. To overcome these limitations, researchers have begun to introduce machine learning techniques [8], [9], [10]. Although these methods have achieved some success, their performance is usually limited by hand-designed features that cannot adequately capture complex image information. Recent years have seen significant progress in BTIS thanks to the rise of deep learning techniques [11]. The utilization of deep learning models in medical image segmentation has demonstrated significant achievements due to their robust feature learning capabilities. Deep learning models possess the capability to autonomously acquire knowledge of the characteristics present in an image, hence leading to enhanced efficacy in the segmentation of boundaries and shapes of brain tumors. Meanwhile, some studies have also focused on 3D image segmentation to capture the morphology of brain tumors more accurately by considering spatial relationships. 3D tumor image segmentation has richer information and more accurate spatial localization compared to 2D images, but at the same time, it needs to cope with challenges such as computational complexity and increased data requirements. Some lightweight U-Net models, such as the lightweight U-Net architecture named Skinny proposed by Tarasiewicz et al., achieve efficient brain tumor segmentation with fewer parameters. However, this model is limited to processing 2D slices and fails to fully utilize the spatial information of 3D data [12].

This research presents a very effective 3D U-Net model that is specifically designed for the BTIS job. The following are the main points of this study's contributions:

(1) The implementation of 3D depth separable convolution (3D-DSC) is proposed as a means to decrease the number of trainable parameters in the model and therefore reduce computational complexity, resulting in resource conservation.

(2) The construction of the dilated dense residual block (DDRb) module aims to enhance the receptive field by introducing dense connections, thereby facilitating the transfer of information across multiple layers. This expansion of the receptive field in each layer leads to improved classification performance and generalization ability of the model.

(3) The incorporation of the confusion area segmentation (CAS) module enhances the network's segmentation performance within the confusion area, hence enhancing the overall segmentation performance of the model.

(4) Experimental results show that the BTIS-Net model outperforms the benchmark method (3D U-Net) on the BraTS 2019 and BraTS 2021 datasets, with improvements of 5.68 in Dice similarity coefficient (Dice), 5.38 in Positive predictive value (PPV), and 2.14 in Sensitivity (Sen), and a reduction of 2.71 Hausdorff distance (HD).

## II. RELATED WORK

### A. IMPROVEMENT OF THE U-NET MODEL

Although U-Net has demonstrated considerable success across multiple domains as a very effective model for image segmentation, it is not without its limitations. Firstly, training generally necessitates a substantial quantity of annotated data, which can pose difficulties in specific fields like medical imaging. Furthermore, U-Net is susceptible to overfitting problems when training data are scarce [13]. Moreover, the intricate structure and multitude of parameters of the system necessitate substantial computational resources and memory, hence reducing its compatibility with low-end devices. Furthermore, there may exist artifacts at the boundaries of the segmentation process, which can result in less seamless segmentation outcomes. Furthermore, U-Net's architecture is rigid, lacking adaptability for objects of varying sizes, and it primarily deals with images of a single scale, necessitating supplementary processing for data of several scales [14]. Scientists are consistently enhancing the U-Net model to enhance its suitability for a broader array of activities, hence eliminating these limitations. TABLE 1 tabulates the improvements of the U-Net model in various applications.

TABLE 1. Improvement details of the U-Net model in various applications.

Method	Solved problem	Solution strategy
Zhang et al.[15]	Improving model generalization ability	i. Diversified dense connection methods ii. Diverse data enhancement methods
Zhou et al.[16] Ibtehaz et al.[17] Zhang et al.[18]	Reducing model complexity	i. Inception module ii. The global pooling layer iii. Deep supervision iv. Dense skip connection v. Full-scaled skip connection vi. Skip connection using addition
Oktay et al.[19] Zhu et al.[20] Jin et al.[21]	Concentrating on effective features and omitting irrelevant features	i. Residual module ii. Dense module iii. Dense skip connection iv. Full-scaled skip connection v. Feature pyramid vi. Bidirectional feature network
Milletari et al.[22] Dolz et al.[23]	Enhancing feature fusion	i. Residual module ii. Dense module iii. Dense skip connection iv. Full-scaled skip connection
Isensee et al.[24]	Speeding up convergence	i. Residual module ii. Dense skip connection iii. Full-scaled skip connection iv. Bridged U-Net v. Data normalization
Jin et al.[25]	Enlarging receptive field	i. Deformable convolution ii. Dilated convolution

As shown in TABLE 1, to address the problems in U-Net, researchers have launched a series of useful improvement explorations. The model achieves a significant performance increase in the image segmentation problem by efficiently focusing on critical features and disregarding irrelevant information with the use of the attention mechanism, directed connectivity, and sparse connectivity. Furthermore, certain studies have effectively expanded the model's sensory field by using methods like residual structure and dilated convolution [26]. This enhances the model's capacity to handle global input and addresses the issue of vanishing or expanding gradient. These improvement studies provide powerful solutions for the optimization of U-Net networks in several aspects.

### B. APPLICATION OF THE U-NET MODEL IN BRAIN TUMOR SEGMENTATION

Compared to segmentation tasks for other categories of medical images, segmenting brain tumor regions is more challenging, with each type of tumor region presenting unique difficulties. For instance, in the case of the whole tumor (WT) region, there is a tendency to inadvertently include background areas, leading to issues of over-segmentation. The tumor core (TC) region is prone to adhesion with other areas. Additionally, in the case of the enhanced tumor (ET) region, small tumor portions are often punctate or discontinuous, making the segmentation task particularly challenging.

The BTS method based on U-Net employs an encoder-decoder strategy, utilizing an MRI brain image segmentation network for processing. This approach involves classifying each pixel of the entire brain image to complete the segmentation task, making it a classic semantic segmentation network model. In practical applications, the network performs downsampling and upsampling through convolution and deconvolution operations to achieve segmentation of the brain image. One notable benefit of this approach resides in its capacity to operate without any limitations imposed on the dimensions of the brain image. Moreover, it can directly provide the segmentation result for the entire image, leading to its widespread application in MRI brain image segmentation in recent years. The architecture of MRI BTS based on U-Net is illustrated in FIGURE 1.

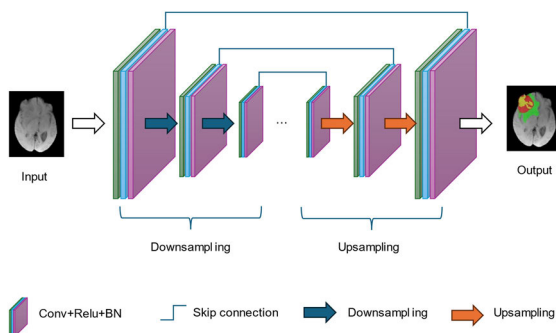


FIGURE 1. MRI BTS architecture diagram based on U-Net.

TABLE 2 shows the performance of various improved U-Net models for MRI brain tumor segmentation on the BraTS series datasets. As observed from the table, these methods generally achieve satisfactory accuracy in WT segmentation. However, compared to WT segmentation, the segmentation results for the core tumor and ET are slightly less impressive. The main reason for this lies in the relatively smaller proportion of core and ET in the entire image, which may lead to insufficient preservation of relevant features during the downsampling process. Nevertheless, by extracting multiscale feature information and effectively integrating semantic and positional information, the accuracy of the FCN in MRI BTS can be significantly improved. Additionally, compared to convolutional neural network segmentation methods that use image block inputs, the U-Net-based BTS method achieves semantic-level dense prediction in an end-to-end manner. This approach avoids the potential issues of repetitive storage and convolution calculations that may arise from using image blocks. As a result, it significantly improves BTS efficiency while maintaining effective segmentation accuracy.

TABLE 2. BTS accuracy based on the improved U-Net method.

Method	Dataset	WT	CT	ET
Dong et al.[27]	BraTS 2015	0.86	0.86	0.65
Liu et al.[28]	BraTS 2015	0.87	0.86	0.79
Kong et al.[29]	BraTS 2017	0.92	0.80	0.76
Beers et al.[30]	BraTS 2017	0.88	0.73	0.73
Isensee et al.[24]	BraTS 2017	0.86	0.78	0.65
Kamnitsas et al.[31]	BraTS 2017	0.89	0.79	0.73
Feng et al.[32]	BraTS 2019	0.91	0.84	0.79

## III. BTIS-NET

### A. MODEL STRUCTURE

The framework of the BTIS-Net model is shown in FIGURE 2. Overall, the BTIS-Net model introduces the DDRB module and the CAS module based on the U-Net network. The network structure of the U-Net based on the DDRB module is still the traditional U-Net structure, which is a 4-layer network with symmetric left and right branches. The left branch performs 3 DDRB convolution operations and 3 downsampling operations on the input image. The fourth layer performs 1 DDRB convolution operation and 1 upsampling operation. The third layer of the right branch copies the information of the third layer on the left (its symmetric layer), connects the third layer information obtained by upsampling the right branch, and then performs a DDRB convolution operation on the connection result. By analogy, the network output layer ( $1 \times H_1 \times W_1 \times D_1$ ) is finally obtained to output the first segmentation result.

Next, the introduced CAS module, i.e. the three-layer anti-attention network of the rightmost branch of the network, will optimize the segmentation result. The core of the CAS model is the anti-attention mechanism. Therefore, after obtaining the initial segmentation result, the CAS module needs to

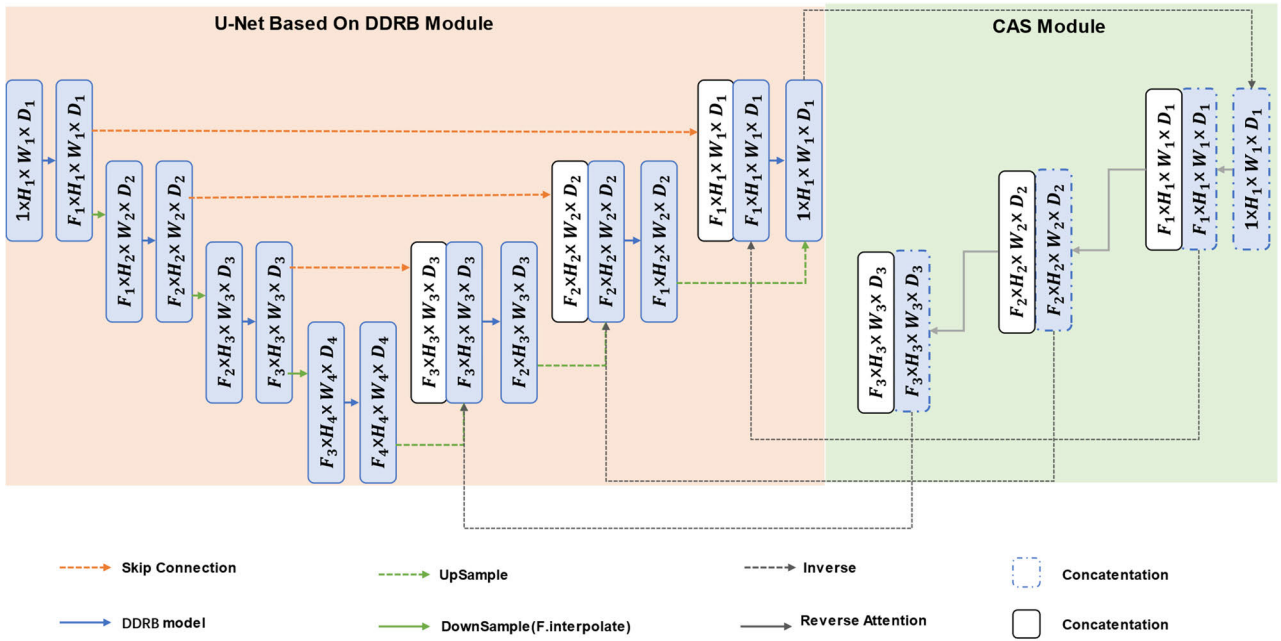


FIGURE 2. BTIS-Net model structure.

perform an anti-attention operation on it. As shown in the network structure of FIGURE 2, the rightmost branch of the proposed BTIS-Net performs a total of 3 anti-attention operations. First, the first forward segmentation result is negated to obtain  $1 \times H_1 \times W_1 \times D_1$ , and an anti-attention operation is performed on it to obtain  $F_1 \times H_1 \times W_1 \times D_1$ . Two operations will be performed here. One is to connect  $F_1 \times H_1 \times W_1 \times D_1$  with the top layer ( $F_1 \times H_1 \times W_1 \times D_1$ ) of the right branch (decoder) of U-Net and then perform the next round of anti-attention operation. The other is to negate the current anti-attention operation to obtain  $F_1 \times H_1 \times W_1 \times D_1$  and pass it to the corresponding layer of the right branch. This information will be used as the information for the second forward training of the U-Net network. And so on, a total of 3 rounds of anti-attention operations are performed.

### B. 3D DEPTH SEPARABLE CONVOLUTION

Standard 3D convolution performs a convolution operation on the entire input data stereo, which means that each filter moves over all depths, heights, and widths simultaneously, resulting in a large and computationally intensive number of parameters. In contrast, 3D-DSC partitions the convolution into two distinct stages: first performing a convolution operation on each depth channel, and then combining these outputs using a  $1 \times 1 \times 1$  convolution kernel. This decomposition reduces the amount of computation and number of parameters [33]. The standard 3D convolution and 3D-DSC structures are shown in FIGURE 3.

FIGURE 3 illustrates three parameters: the kernel's spatial dimension ( $K$ ), the input channel count ( $M$ ), and the output channel count ( $N$ ). Additionally, it depicts two

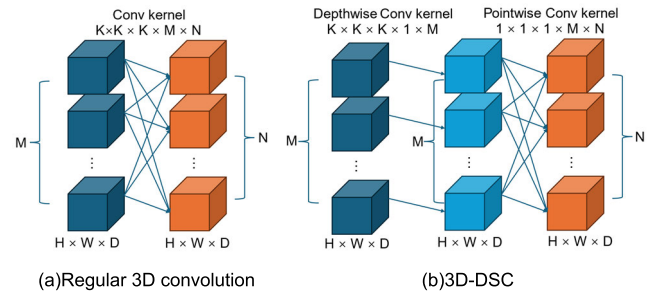


FIGURE 3. Standard 3D convolution with 3D-DSC structure.

TABLE 3. Comparison of No. multiplication and No. parameters for 3D convolution and 3D-DSC.

Method\Metric	No. of Multiplication	No. of Parameters
3D Convolution	$K^3 \cdot M \cdot H \cdot W \cdot D \cdot N$	$K^3 \cdot M \cdot N$
3D-DSC	$M \cdot H \cdot W \cdot D (K^3 + N)$	$M (K^3 + N)$
3D Convolution /3D-DSC	$(K^3 \cdot N) / (K^3 + N)$	$(K^3 \cdot N) / (K^3 + N)$

tensors,  $X$  and  $Y$ , where  $H$  represents the height,  $W$  denotes the width, and  $D$  signifies the depth of the feature map within the tensor. TABLE 3 displays a comparison of the computational and parametric quantities obtained for the two convolutional approaches.

3D-DSC requires approximately 19 times fewer parameters and computations when  $K = 3$  and  $N = 64$ . Reducing the number of multiplications leads to lower computational complexity, thereby resulting in a faster network.



### C. DDRB MODULE

To better understand the contextual information of the image, reduce information loss, and expand the receptive field as much as possible, this paper constructs a DDRB module. This module consists of three components. Component 1 consists of three densely connected  $3 \times 3 \times 3$  convolutional layers with dilation rates of 1, 2, and 3. This component can expand the receptive field of the convolution kernel and capture a wider range of contextual information. Component 2 is a  $1 \times 1 \times 1$  convolutional layer, which enhances the propagation and reuse of features by directly connecting each layer to all subsequent layers. Component 3 is residual learning, which simplifies the learning process and helps the network train better by introducing residual learning. The structural diagram of the DDRB module is shown in FIGURE 4. The parameter  $h$  in the figure represents the growth rate, which is used to regulate the number of new features added to the accumulated information in the previous layer by each expanded convolutional layer.

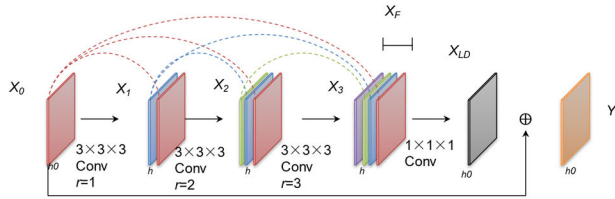


FIGURE 4. DDRB module structure.

To enhance the absorption of contextual features, we introduce a dense connection among individual convolutional layers, where the current layer directly links to all subsequent layers. This dense connection strengthens the transfer and utilization of feature information. Assuming that  $X_0$  contains  $h_0$  feature maps, and each expanding convolutional layer generates  $h$  feature maps (the  $h$  value interval is usually between 8 and 48), then the final output will contain  $h_0 + 3h$  feature maps, which incorporate multi-scale contextual information.

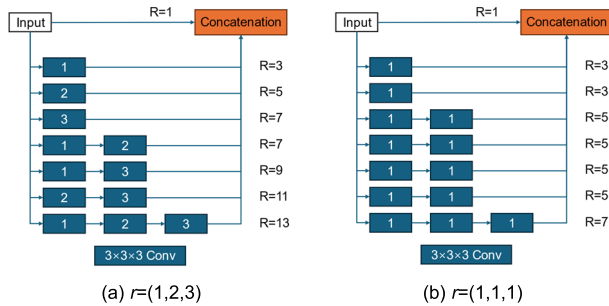


FIGURE 5. Dense connections at different dilation rates.

FIGURE 5(a) shows the process of using convolution with progressively increasing dilation rate in combination with dense connection to capture features in the DDRB module. The sampling results obtained after various dense connections and different dilation convolution layers are used as

output. In contrast, the structure of regular dense blocks [34] is shown in FIGURE 5(b). The comparison between the two shows that the module used in this study can capture features over a wider range and also mitigates the problem of feature redundancy that may be caused by the dense connection.

A  $1 \times 1 \times 1$  convolution can be employed to convert the high-dimensional feature map  $X_T$  into a lower-dimensional space, hence enhancing computational efficiency and parameter utilization. The corresponding equation is shown below:

$$X_{LD} = H_1(X_T) \quad (1)$$

This process reduces the channel dimension of the  $X_T$  from  $h_0 + 3h$  to  $h_0$  and also introduces the idea of residual learning by element-by-element summation of the inputs and outputs of the module through a shortcut connection between them. This step further enhances the flow of information and refines the output feature representation. As a result, the final output representation of the DDRB module is as follows:

$$Y = X_{LD} + X_0 \quad (2)$$

### D. CAS MODULE

In the study of brain tumor image segmentation, confusion areas refer to areas where the network is uncertain or inaccurate when predicting the classification of certain pixels. These areas are usually located at the boundary between tumor and normal tissue, low-contrast areas, areas with complex internal structures of tumors, and areas affected by data noise and artifacts. In these areas, the network's prediction probabilities for different categories are similar, making it difficult to determine the attribution of pixels. To solve the problem of these confusion areas, this study constructed a CAS module based on the inverse attention mechanism [35] to improve the network architecture, thereby improving the accuracy and robustness of the segmentation results. The working principle of the CAS module is shown in FIGURE 6. Through the inverse attention mechanism, the module can more effectively process and distinguish these areas with high uncertainty, thereby improving the overall segmentation performance.

The red circles in FIGURE 6 outline the overlapping areas of different tumor sites, which are the confused areas. The reverse attention mechanism accurately divides these areas as follows:

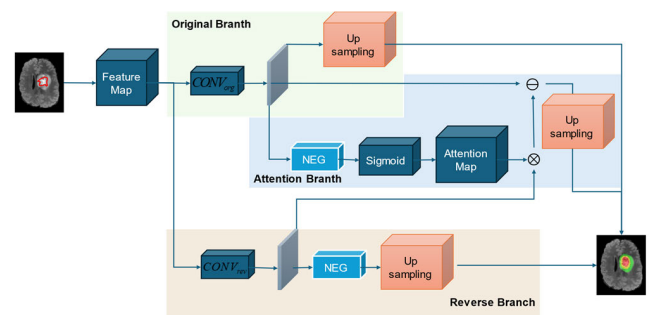


FIGURE 6. Schematic diagram of CAS module.

The original branch (first branch) uses the original U-Net network to learn the probability distribution of pixels belonging to each category. The reverse branch (second branch) learns the probability distribution of pixels not belonging to each category. This branch uses backpropagation to obtain the opposite prediction results of the initial branch. The attention branch (third branch) is used to learn the weights of the combination between the initial prediction and the reverse prediction. Specifically, this branch subtracts the prediction results of the initial branch from the prediction results of the reverse branch to obtain a difference map. Then, an attention mechanism is used to adjust the results of reverse learning so that it more accurately covers the confused areas in the initial learning results. To prevent reverse learning from being less effective than initial learning, we introduce the prediction results of reverse learning on the confused areas of the initial learning results, thereby using the reverse attention mechanism to enhance the segmentation performance. In this way, the reverse attention mechanism can more effectively process and distinguish areas with high uncertainty, thereby improving the overall segmentation performance. Equation (3) provides the mathematical expression:

$$I_{ra}(i, j) = \text{Sigmoid}(-F_{CONV_{org}}(i, j)) \quad (3)$$

where  $(i, j)$  denotes the pixel position coordinates and  $F_{CONV_{org}}$  denotes the response map of the  $CONV_{org}$ . Due to the effects of negative class (NEG) inversion and sigmoid function, the effect of the reverse attention mechanism is that the negative response (smaller response) in the original prediction of U-Net will be highlighted and its reverse attention will be considered; while the positive response (larger response) will be suppressed, become very small, and its reverse attention will be ignored.

#### IV. EXPERIMENTAL RESULTS AND DISCUSSION

To verify the effectiveness and efficiency of the proposed model, we conducted multiple sets of experiments on the BraTS 2019 and BraTS 2021 public data sets. Experiments include resource-saving performance, segmentation performance, confusion region segmentation, ablation experiments, Pareto front analysis, and statistical analysis. These experiments comprehensively evaluate the performance of the proposed model in different aspects and verify its significant advantages in resource utilization and segmentation performance.

##### A. EXPERIMENTAL BACKGROUND DESCRIPTION

The datasets used in this article are BraTS 2019 and BraTS 2021. Both datasets are from the BraTS series, and the main task is to divide brain tumors into three parts: TC, ET, and WT. When testing the model, we selected these two versions of the dataset to fully evaluate the performance of the improved UNet model on different datasets. The specific dataset details are shown in Table 4. All experimental results are based on the performance evaluation of the model on the validation set.

TABLE 4. Data set details.

Year	Total	Training	Validation	Testing
2019	626	335	125	166
2021	2040	1251	219	570

Table 4 shows the difference in data volume between the two datasets. BraTS 2021 contains more patient samples than BraTS 2019. In addition, BraTS 2021 also contains additional clinical data and multimodal scans, further enriching the dataset. The reasons for selecting these two datasets in this study are as follows: First, by testing the model on the BraTS 2019 and BraTS 2021 datasets, we can evaluate the generalization ability of the model under different data conditions. The BraTS 2021 dataset has a larger sample size and more diverse tasks, providing a more challenging testing environment for the generalization ability of the model. Second, the improvement in annotation quality in BraTS 2021 helps analyze the performance of the model on high-quality annotated data and enables us to evaluate the impact of annotation quality on model performance. Finally, the BraTS 2021 dataset provides more diverse samples and tasks, tests the robustness of the model under more challenging conditions, and helps us understand the adaptability of the model under diverse data.

This study utilizes the following evaluation metrics: Dice, PPV, Sen, and HD. The Dice coefficient is utilized to evaluate the extent of overlap between expected segmentation outcomes and actual segmentation outcomes, with values ranging from 0 to 1, where 1 signifies full overlap. The PPV quantifies the model's capacity to precisely forecast positive samples, with values ranging from 0 to 1, where 1 signifies flawless accuracy. The Sen is a measure of its capacity to accurately detect positive samples. It is measured on a scale of 0 to 1, with a value of 1 indicating full recognition. The HD quantifies the maximum distance between the segmentation outcome and the actual outcome, where smaller HD values indicate a higher degree of proximity between the segmentation outcome and the actual outcome. The calculation formulas of the four indicators are as follows:

$$\text{Dice} = \frac{2 \times |A \cap B|}{|A| + |B|} \quad (4)$$

$$\text{PPV} = \frac{TP}{TP + FP} \quad (5)$$

$$\text{Sen} = \frac{TP}{TP + FN} \quad (6)$$

$$\text{HD} = \max(\max_{a \in A} \min_{b \in B} \|a - b\|, \max_{b \in B} \min_{a \in A} \|b - a\|) \quad (7)$$

where TP is a true positive example, FP is a false positive example, FN is a false negative example,  $|A \cap B|$  is the intersection area of the segmentation result and the true label,  $|A|$  and  $|B|$  are the areas of the segmentation result and the true label respectively, and  $\|a - b\|$  represents the distance between point  $a$  and point  $b$ .

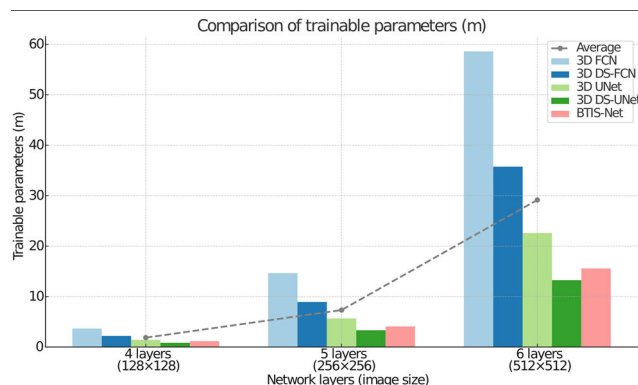
The experimental environment is configured with a Windows 10 operating system and equipped with Intel Core i7-8700 processor (64-bit). The memory is 32 GB, and the graphics card is NVIDIA RTX 4090 with 24 GB GDDR6X video memory. The network structure is implemented using the PyTorch framework and accelerated by CUDA 11.0.

### B. RESOURCE-SAVING PERFORMANCE EXPERIMENT

BTIS-Net incorporates 3D-DSC as a substitute for the conventional 3D convolution operation. This approach effectively decreases the quantity of model parameters and processing requirements, resulting in resource conservation. To assess the efficacy of this approach, this subsection conducts a comparative analysis of the trainable parameter count between the 3D-DSC models that were introduced and those that were not, utilizing two models, namely 3D FCN and 3D U-Net. In addition, as inspired by UNet++ [39], it is known that U-Net with a different number of layers is suitable for different types of medical images. Usually, a 4-layer network is used to train images of  $128 \times 128$  size, a 5-layer network is used to train images of  $256 \times 256$  size, and a 6-layer network is used to train images of  $512 \times 512$  size. Therefore, during the experiments we compared the number of trainable parameters obtained on different network layers and different size images. TABLE 5 and FIGURE 7 display the count of trainable parameters produced by each model on the BraTS 2019 dataset, using it as an illustrative example.

**TABLE 5.** Trainable parameter count comparison (m).

Model\Network layers(Image size)	4 Layer (128×128)	5 Layer (256×256)	6 Layer (512×512)
3D FCN	3.66	14.64	58.59
3D DS-FCN	2.20	8.93	35.74
3D U-Net	1.42	5.65	22.58
3D DS-U-Net	0.83	3.32	13.24
BTIS-Net	1.16	4.08	15.57



**FIGURE 7.** Comparison of the number of trainable parameters obtained by each model (m).

Based on the experimental results in TABLE 5 and FIGURE 7, it can be observed that the trainable parameter

quantities of different models vary significantly under different image sizes. At the 4-layer ( $128 \times 128$ ) image size, the 3D FCN model has the highest trainable parameters, reaching 3.66 million, while the 3D DS-U-Net model has the lowest, at only 0.83 million. The BTIS-Net model has 1.16 million trainable parameters at this image size, slightly higher than the 3D DS-U-Net. This indicates that, with a relatively small network scale and image size, the addition of the reverse attention mechanism has a limited impact on the number of parameters, allowing BTIS-Net to maintain a low parameter count.

At the 5-layer ( $256 \times 256$ ) and 6-layer ( $512 \times 512$ ) image sizes, the trainable parameters of each model increase significantly. The 3D FCN model's parameter count increases the fastest, reaching 58.59 million at the 6-layer ( $512 \times 512$ ) size. In contrast, the BTIS-Net model has 4.08 million and 15.57 million trainable parameters at the 5-layer and 6-layer image sizes, respectively, still lower than the 3D FCN and 3D DS-FCN models, but slightly higher than the 3D DS-U-Net. This suggests that although the addition of the reverse attention mechanism in BTIS-Net increases the parameter count, it remains within a controllable range and is still much lower than traditional 3D FCN and 3D DS-FCN models. Overall, BTIS-Net effectively utilizes the reverse attention mechanism to enhance model performance and accuracy while maintaining a relatively low parameter count.

### C. IMAGE SEGMENTATION PERFORMANCE EXPERIMENT

This study primarily introduces two types of comparison models—U-Net improved models and non-U-Net models—to validate the segmentation performance of the proposed BTIS-Net model for brain tumor images. The U-Net and its improved models are 3D-U-Net [36], ResU-Net [37], RU-Net2+ [38], UNet++ [39], TransAttUNet [40]. Non-U-Net models are 3D-FCN [41], V-Net [22], DenseVoxNet [42], 3D ConvNet [43], GoogleNet [44], VGGNet [45]. In TABLE 6 and TABLE 7, we can see how each model performed regarding segmentation on BraTS2019 and BraTS2021. The best results are made bold. At the same time, to more intuitively compare the performance of each model on the four indicators, respectively give the segmentation performance comparison charts on the BraTS 2019 and BraTS 2021 datasets.

Drawing from the insights gleaned from TABLE 6, it is evident that the BTIS-Net model attains Dice metrics of 79.12%, 90.96%, and 83.79% for ET, WT, and TC respectively. Additionally, it attains PPV metrics of 87.82%, 94.76%, and 83.29%, Sensitivity metrics of 84.16%, 83.94%, and 84.1%, and HD metrics of 3.61, 5.64, and 5.28. Through a comparative analysis of the experimental outcomes derived from alternative models, it can be deduced that the BTIS-Net model exhibited superior segmentation performance when applied to the BraTS 2019 dataset. This is followed by UNet++ and TransAttUNet. In addition, the experimental performance obtained by other non-U-Net models is much

**TABLE 6.** Segmentation performance comparison on the BraTS 2019 dataset.

Method/Index	Dice(%)↑			PPV(%)↑			Sen(%)↑			HD(mm)↓		
	ET	WT	TC	ET	WT	TC	ET	WT	TC	ET	WT	TC
Non-U-Net models												
3D-FCN	70.08	82.46	75.62	75.58	84.96	77.42	82.64	81.56	82.75	6.72	10.26	9.08
V-Net	68.64	84.67	74.98	74.34	87.57	75.58	82.28	82.15	81.92	6.56	8.87	8.92
DenseVoxNet	70.57	85.33	75.35	79.37	88.73	76.05	82.13	82.04	82.09	6.27	9.90	8.83
3D ConvNet	72.3	86.25	75.76	80.9	89.65	76.26	81.73	81.42	80.87	6.12	10.10	9.25
GoogleNet	69.63	83.44	73.82	78.43	86.44	79.12	81.62	80.31	81.76	7.72	10.84	12.14
VGGNet	68.74	84.19	74.63	77.34	87.79	78.23	80.81	80.58	81.72	7.68	11.25	12.86
U-Net improved models												
3D-U-Net	73.96	86.23	78.45	80.46	89.33	77.15	82.84	81.66	83.81	5.87	9.68	8.45
ResU-Net	72.22	86.78	79.31	81.02	89.88	78.61	82.69	83.55	83.64	5.24	7.893	8.20
RU-Net2+	71.69	88.95	81.73	84.49	92.65	80.53	84.03	83.78	83.87	4.51	6.04	6.84
UNet++	78.27	89.52	82.84	86.87	93.82	82.54	<b>84.29</b>	<b>85.03</b>	83.18	5.32	6.89	6.94
TransAttUNet	<b>79.5</b>	90.44	81.06	<b>88.2</b>	93.34	83.16	83.95	84.74	<b>84.78</b>	4.52	7.97	7.43
BTIS-Net	79.12	<b>90.96</b>	<b>83.79</b>	87.82	<b>94.76</b>	<b>83.29</b>	84.16	83.94	84.10	<b>3.61</b>	<b>5.64</b>	<b>5.28</b>

**TABLE 7.** Segmentation performance comparison on the BraTS 2021 dataset.

Method/Index	Dice(%)↑			PPV(%)↑			Sen(%)↑			HD(mm)↓		
	ET	WT	TC	ET	WT	TC	ET	WT	TC	ET	WT	TC
Non-U-Net models												
3D-FCN	71.44	82.46	78.75	80.63	86.91	79.85	82.88	82.63	83.20	7.12	9.83	10.26
V-Net	70.08	81.14	76.83	78.27	85.56	77.40	80.31	78.48	79.62	8.83	10.21	9.54
DenseVoxNet	72.26	86.77	77.08	79.54	87.43	78.86	81.52	80.93	80.55	6.84	9.42	9.79
3D ConvNet	71.24	83.20	76.65	80.28	84.11	77.69	80.59	79.23	79.58	6.11	9.56	8.18
GoogleNet	70.03	81.92	74.32	77.30	83.29	76.53	78.97	78.85	77.79	9.33	11.42	10.85
VGGNet	69.91	80.28	73.23	75.32	84.12	75.42	79.83	79.64	78.72	10.06	10.87	11.29
U-Net improved models												
3D-U-Net	70.73	84.85	80.62	81.82	87.92	79.42	81.95	82.86	81.32	6.43	9.84	7.17
ResU-Net	73.98	86.33	82.49	82.33	88.48	81.57	82.87	82.12	83.58	7.65	8.58	7.89
RU-Net2+	74.83	85.58	85.31	82.78	83.59	82.28	83.98	82.44	82.65	6.94	9.29	8.06
UNet++	76.39	88.46	<b>87.05</b>	86.29	89.17	<b>84.55</b>	84.59	85.33	83.80	5.86	8.43	6.55
TransAttUNet	75.17	87.84	85.14	87.75	85.20	83.83	83.37	83.98	82.73	4.94	7.17	6.03
BTIS-Net	<b>78.65</b>	<b>89.87</b>	86.48	<b>88.36</b>	<b>90.23</b>	83.91	<b>84.81</b>	<b>85.49</b>	<b>84.75</b>	<b>4.39</b>	<b>6.42</b>	<b>5.88</b>

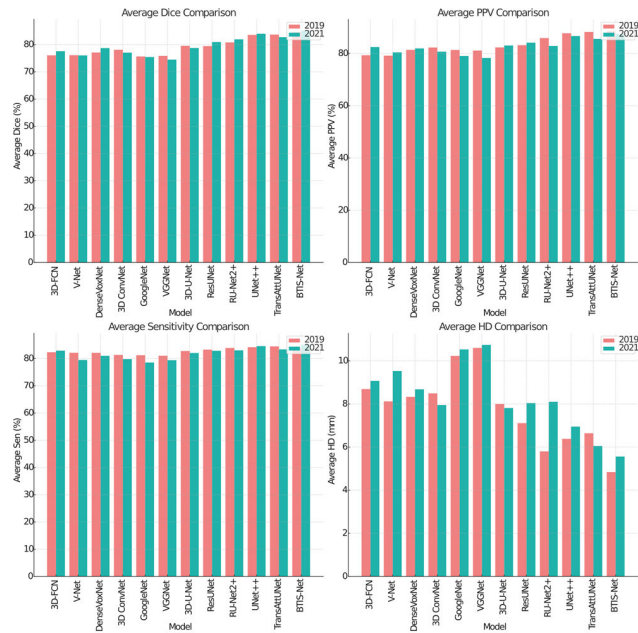
worse than U-Net and its improved models. Among the many non-U-Net models, the best performance was obtained by the 3D ConvNet model. However, compared to the BTIS-Net model, it obtains a decrease in Dice metrics by 13.12%, 7.44%, and 10.93% on ET, WT, and TC, respectively, a decrease in PPV metrics by 11.93%, 7.36%, and 6.08%, respectively, and a decrease in Sen metrics by 3.98%, 4.00%, and 2.83%, respectively, and an increase in HD metrics by 112.74%, 99.47%, 143.56%. This also confirms that the U-Net model is favorable for medical image segmentation tasks.

Observing TABLE 7, it can be concluded that overall, the models perform better on the BraTS 2021 dataset than the BraTS 2019. This is mainly because the BraTS 2021 dataset has a richer dataset, which helps the models to learn better features. In addition, the BTIS-Net model still shows excellent performance. Furthermore, the BTIS-Net model demonstrates enhanced performance in comparison to previous enhanced U-Net models. This finding also demonstrates that as the dataset expands and gets more comprehensive, the BTIS-Net model exhibits enhanced capability in extracting data features, hence optimizing the overall segmentation performance.

To comprehensively demonstrate the segmentation performance of each model on different datasets, to more clearly compare their advantages and disadvantages, and to verify their versatility and robustness, we conducted a comparative analysis of the performance of each model on the BraTS 2019 and BraTS 2021 datasets. Through this comparison, we can evaluate the consistency of the performance of the model under different dataset conditions. The performance comparison of each model on these two datasets is shown in FIGURE 8.

Overall, first of all, the Dice values of most models on the BraTS 2021 dataset are higher than those on the BraTS 2019 dataset, especially the models such as 3D-U-Net, ResUNet, and TransAttUNet have higher segmentation accuracy on the BraTS 2021 dataset. Secondly, the PPV values of all models on the BraTS 2021 dataset are generally better than those on the BraTS 2019 dataset, showing that the models have stronger prediction capabilities for positive samples on the BraTS 2021 dataset. Third, the Sen values on the two datasets are not much different, but they are slightly improved on the BraTS 2021 dataset, indicating that the models have stronger recognition capabilities for target areas on the BraTS 2021 dataset.





**FIGURE 8.** Comparison of segmentation performance of various models on BraTS 2019 and BraTS 2021 dataset.

Finally, the Hausdorff distance of most models on the BraTS 2021 dataset is low, which means that the segmentation results are closer to the actual situation in terms of spatial distance.

In terms of excellent model analysis, the proposed BTIS-Net has the best or near-best performance in all indicators on both datasets, especially on HD, indicating that its segmentation results are very accurate and have small spatial errors. BTIS-Net achieves the highest values on Dice and PPV, showing that its prediction accuracy and positive sample recognition ability are the strongest; it performs close to the best on Sen, indicating that its recognition of the target area is relatively comprehensive; it achieves the minimum value on HD, indicating that its segmentation results are closest to the actual situation in space. This excellent performance may be attributed to the optimization of BTIS-Net in the model structure, combining multi-scale feature fusion and deep extraction of contextual information, and enhancing the model's recognition ability for complex structures. In addition, TransAttUNet performs very well on both datasets in terms of indicators such as Dice, PPV, and Sen, especially on Dice and PPV. This may be attributed to its combination of Transformer and attention mechanism, which enhances feature extraction and modeling capabilities. Experimental results show that although the performance of the models on different datasets varies, BTIS-Net and TransAttUNet always perform well on both datasets, especially BTIS-Net, which performs best in all indicators.

#### D. CAS PERFORMANCE EXPERIMENT

This study presents the introduction of the CAS module, which aims to enhance the precision of model segmentation

within the confusion area. This study performed a visualization comparison of the predicted masks on the validation sets of the BraTS 2019 and BraTS 2021 datasets. The segmentation results of the two images in the two datasets for each model are shown in FIGURE 9.

By comparing and analyzing the segmentation results of the four images shown in FIGURE 9, it can be seen that, whether in BraTS 2019 or BraTS 2021, in terms of overall segmentation accuracy, BTIS-Net shows its accurate segmentation of WT, TC, and ET in the segmented images. The model can accurately distinguish and segment tumor areas in different parts, especially in the detailed segmentation of the overlapping areas of WT and TC, ET and TC. In the image, the boundaries segmented by BTIS-Net are clearer and the regional definition is more accurate. TransAttUNet also performs well in the segmented images, especially in the ET and WT parts. Its segmentation boundary clarity in the ET part is close to BTIS-Net, but it is slightly inferior in the segmentation of TC, which is manifested as some blurred boundaries or inaccurate segmentation of overlapping areas.

In the processing of confusing areas, BTIS-Net shows in the image that it is very superior in processing confusing areas (such as the overlapping area of TC and ET, and the intersecting area of TC and WT). Through the confusing area segmentation module, BTIS-Net can better distinguish these overlapping and intersecting areas, thereby improving the precision and accuracy of segmentation. In the image, it can be observed that BTIS-Net's segmentation effect in these confusing areas is significantly better than other models. In addition, RU-Net2+ and UNet++ also show good segmentation capabilities when processing confusing areas, but compared with BTIS-Net, there are still some cases of mis-segmentation and missed segmentation, especially in overlapping areas.

In terms of the accuracy of segmentation boundaries, BTIS-Net shows its superior performance in the image, especially in small tumor areas and edge areas. This accuracy is particularly evident in the processing of confusing areas, which can effectively avoid mis-segmentation between different areas. Although models such as 3D-U-Net and DenseVoxNet perform well in overall segmentation, they are slightly insufficient in boundary accuracy and small-area segmentation, and some blurred boundaries can be seen in the image.

Based on the above analysis, we can conclude that BTIS-Net has the following advantages: First, it has high-precision segmentation. BTIS-Net shows high-precision and high-consistency segmentation effects when segmenting WT, TC, and ET, especially in boundary clarity and segmentation of small areas. Second, effective confusion area processing. By introducing the confusion area segmentation module, BTIS-Net can better distinguish the overlapping and intersecting areas of different parts, thereby improving the accuracy of segmentation. Third, boundary accuracy. BTIS-Net excels in the accuracy of segmentation boundaries, avoiding mis-segmentation and missed segmentation

**TABLE 8. Ablation experiment results.**

Method	Dice(%)↑			PPV(%)↑			Sen(%)↑			HD(mm)↓			Params (m)
	ET	WT	TC	ET	WT	TC	ET	WT	TC	ET	WT	TC	
3D U-Net	73.96	86.23	78.45	80.46	89.33	77.15	82.84	81.66	83.81	5.87	9.68	8.45	5.65
3D U-Net+3D-DS	73.13	85.62	77.73	80.10	88.48	77.32	80.80	81.23	82.93	5.92	9.74	9.16	<b>3.32</b>
3D U-Net+3D-DS+DDRB	78.54	88.78	82.84	86.94	91.85	82.20	84.11	82.84	83.97	4.75	6.93	6.67	3.36
3D U-Net+3D-DS+DDRB+CAS	<b>79.12</b>	<b>90.96</b>	<b>83.79</b>	<b>87.82</b>	<b>94.76</b>	<b>83.29</b>	<b>84.16</b>	<b>83.94</b>	<b>84.10</b>	<b>3.61</b>	<b>5.64</b>	<b>5.28</b>	3.65

between different regions, and the boundaries in the image are clearly defined. Through the analysis of the segmented images, it can be seen that BTIS-Net shows excellent performance in processing complex brain tumor segmentation tasks, especially in the segmentation of confusion areas. This superiority is due to its confusion area segmentation module and precise boundary processing capabilities, which give it obvious advantages in practical applications.

### E. ABLATION STUDY

This subsection conducts ablation experiments to comprehensively validate the performance of BTIS-Net and compare the effects of alternative structures on its performance. TABLE 8 presents the outcomes of the ablation experiments conducted on the BraTS 2019 dataset. The best results are made bold. Here the input image is used as  $256 \times 256$  and the value of each metric is the average of the three regions WT, TC, and ET.

TABLE 8 shows that 3D DS added into the 3D U-Net model decreases Dice, PPV, Sen, and HD metrics for ET, WT, and TC and increases HD metrics. Nevertheless, the number of parameters experienced a significant decrease from 5.65 to 3.32, representing a reduction of 41%. This discovery implies that 3D-DSC can reduce the number of parameters in the model, but it does not improve the model's segmentation ability. Continuing to introduce the DDRB module based on the above methodology, it is observed that the Dice, PPV, and Sen metrics are significantly improved and the HD metrics have a significant decrease. In addition, there is a very slight increase in Params. For the model's segmentation performance, this proves that the DDRB module works. This is because the module introduces dense connections between the convolutional layers that maximize the absorption of contextual features of the image. However, there is also a small increase in the parameters due to the multiple convolutional operations embedded in the module.

Continuing to introduce the confusion area division module CAS based on the above method, it is observed that the three metrics of Dice, PPV, and Sen have been slightly improved and the HD metrics have been somewhat decreased. In addition, Params increased by 10%. The efficacy of the CAS module is demonstrated by its capacity to enhance the entire image by augmenting the confusion area. Additionally, the module augments the model's parameter count by employing a three-layer inverse attention network.

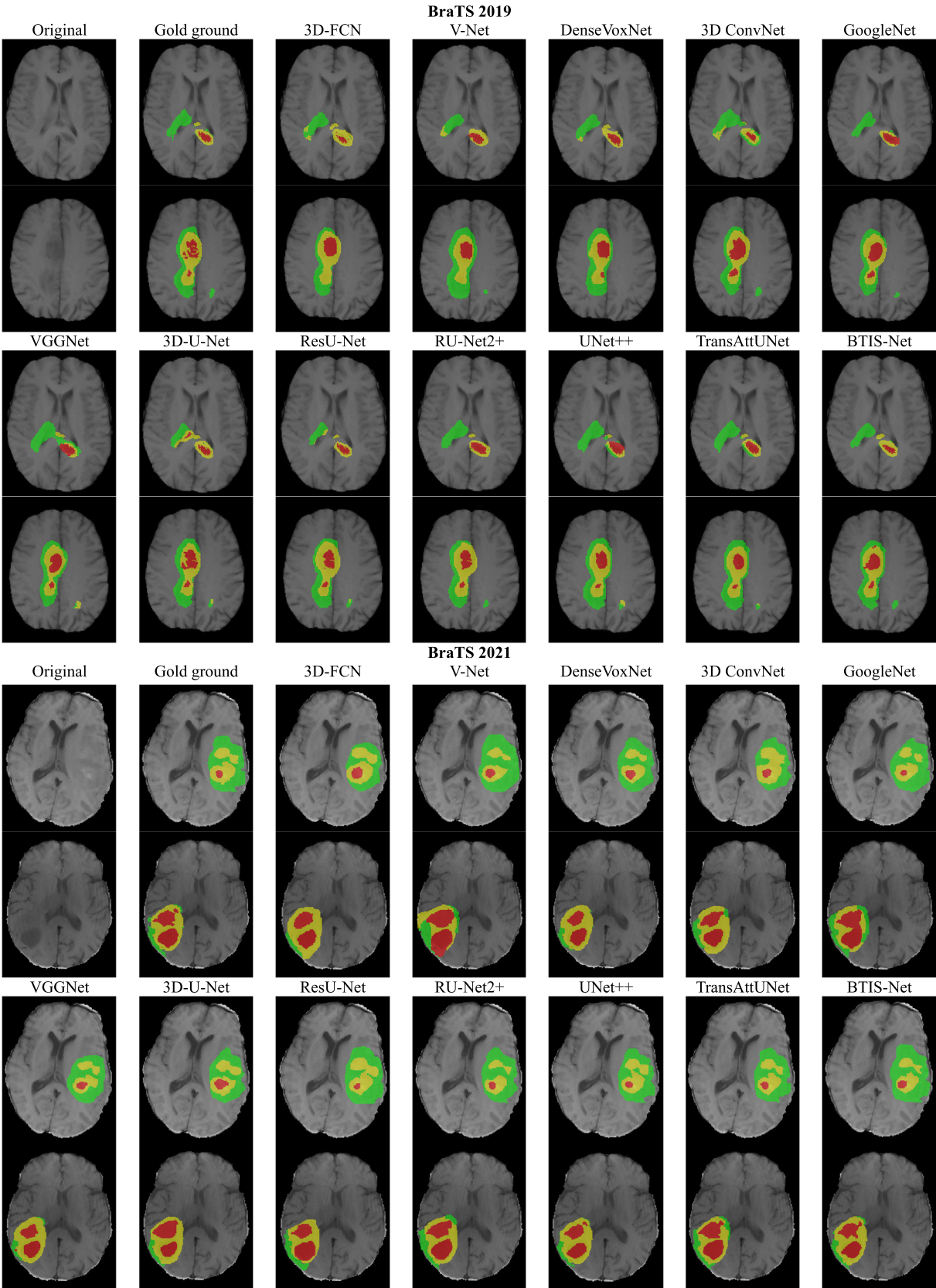
### F. PARETO FRONT ANALYSIS

To evaluate the comprehensive performance of different segmentation models on the BraTS 2019 and BraTS 2021 datasets, especially the trade-off between segmentation accuracy and model complexity. Based on the data shown in Tables 6 and 7, and calculating the average values of the three-part segmentation of ET, TC, and WT, the Pareto frontier graphs shown in FIGURE 10 and FIGURE 11 are drawn. This figure shows the relationship between the segmentation performance of different models on the BraTS 2019 dataset and the BraTS 2021 dataset and the number of trainable parameters.

Pareto frontier analysis of models on the BraTS 2019 and BraTS 2021 datasets can help us understand the trade-offs between multiple performance indicators and the number of trainable parameters of different models. The following is a detailed description of the Pareto frontier analysis on these two datasets:

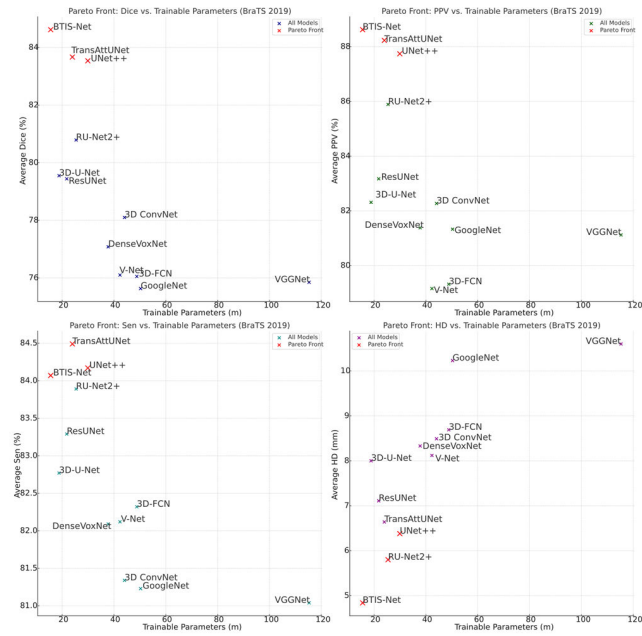
(1) According to the analysis of the BraTS 2019 dataset, BTIS-Net performs best in Dice, PPV, and HD indicators, with the highest segmentation performance and the least trainable parameters. In particular, it is the only optimal non-dominated solution in Dice, PPV, and HD indicators. In terms of sensitivity indicators, in addition to BTIS-Net, TransAttUNet also performs well, showing different trade-offs between sensitivity and trainable parameters. Therefore, BTIS-Net performs well in all indicators and is particularly suitable for application scenarios that require high-performance segmentation with limited computing resources.

(2) According to the analysis of the BraTS 2021 data set, BTIS-Net performs well on the four indicators of Dice, PPV, Sen, and HD, has the highest segmentation performance and the fewest trainable parameters, and is the only optimal non-dominated algorithm on all indicators. solution model. UNet++ and TransAttUNet also perform well on some metrics but require relatively more trainable parameters. In comparison, VGGNet and GoogleNet have lower performance and require the most trainable parameters. This shows that BTIS-Net has significant advantages when processing the BraTS 2021 dataset, especially achieving high-performance segmentation while maintaining low model complexity, and is suitable for applications that require high segmentation performance when computing resources are limited. Application scenarios.

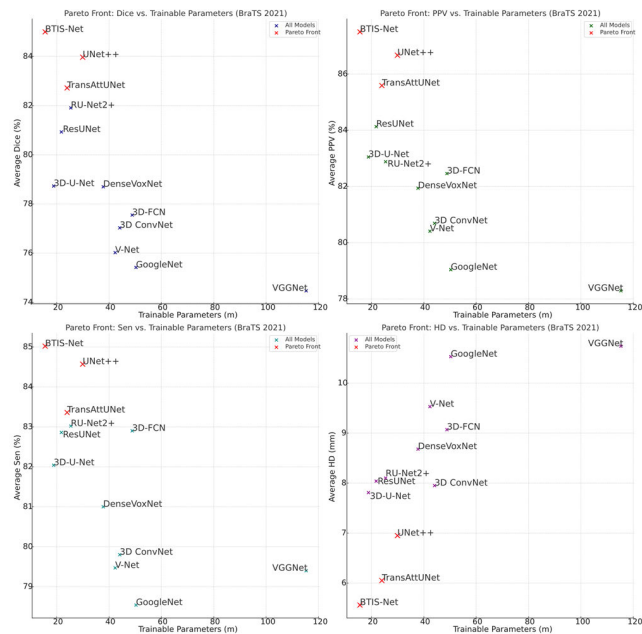


**FIGURE 9.** Segmentation results of different models on the BraTS 2019 and 2021 datasets. The figure shows the segmentation results of the introduced comparison model and the proposed model in two images of the validation set of the BraTS 2019 and 2021 datasets. The green area in the figure represents the whole tumor (WT), the yellow area represents the tumor core (TC), and the red area represents the enhanced tumor (ET).





**FIGURE 10.** Comparison of trainable parameters and segmentation performance of various models on the BraTS 2019 dataset.



**FIGURE 11.** Comparison of trainable parameters and segmentation performance of various models on the BraTS 2021 dataset.

(3) According to the Pareto frontier analysis of the BraTS 2019 and BraTS 2021 data sets, BTIS-Net performs optimally on all three indicators of Dice, PPV, and HD, with the highest segmentation performance and the fewest trainable parameters. On the sensitivity metric, in addition to BTIS-Net, TransAttUNet also performs well, showing different trade-offs between sensitivity and trainable parameters. Overall, BTIS-Net shows significant advantages in processing

**TABLE 9.** Statistical significance analysis results of each model on the BraTS 2019 data set.

Model	Dice Rank	PPV Rank	Sen Rank	HD Rank
3D-FCN	2.7	1.8	5.0	8.9
V-Net	2.8	1.3	4.9	6.9
DenseVoxNet	4.8	4.6	5.5	7.9
3D ConvNet	5.9	6.8	2.7	9.1
GoogleNet	1.9	3.8	2.2	11.2
VGGNet	2.9	3.9	1.5	11.8
3D-U-Net	7.9	6.0	7.2	7.1
ResUNet	7.2	7.8	7.9	5.0
RU-Net2+	8.9	9.0	9.3	2.0
UNet++	10.6	10.3	10.9	3.5
TransAttUNet	10.6	10.9	11.5	3.5
BTIS-Net	11.8	1.8	5.0	1.1
Metric	p-value		H0	
Dice	0		Rejected	
PPV	0		Rejected	
Sen	0		Rejected	
HD	0		Rejected	

both datasets, especially achieving high-performance segmentation while maintaining low model complexity.

### G. STATISTICAL SIGNIFICANCE ANALYSIS

The statistical significance of experimental results was assessed by Friedman's test and Holm's post hoc test. The Friedman test calculates the mean rank of the compared means and determines whether the observed differences are statistically significant. The significance level is set to 0.05, i.e. if the p-value is below this threshold, the null hypothesis ( $H_0$ ) is rejected, indicating a significant difference. Subsequently, the statistical significance of the differences between the control method and the other methods was further analyzed using Holm's post hoc test procedure. This experiment chose the BraTS 2019 data set as an example. Table 9 shows the statistical significance analysis results of the Friedman test and Holm post hoc test on the four indicator data of each model on the BraTS 2019 data set.

Based on the analysis of the experimental data in TABLE 9, the proposed BTIS-Net model's performance on the BraTS 2019 dataset is as follows: For the Dice metric, the best-performing model is GoogleNet (ranked 1.9), while the worst is BTIS-Net (ranked 11.8), placing our model last in this metric. For the PPV metric, the best model is V-Net (ranked 1.3), and the worst is RU-Net2+ (ranked 9.0); BTIS-Net ranks second, showing excellent performance. For the Sen metric, the top model is VGGNet (ranked 1.5), and the lowest is TransAttUNet (ranked 11.5); BTIS-Net ranks sixth. For the HD metric, BTIS-Net is the best (ranked 1.1), while GoogleNet is the worst (ranked 11.2). Friedman and Holm post-hoc tests confirm that the differences across metrics are statistically significant. In summary, although BTIS-Net performs poorly in the Dice metric, it excels in PPV and HD metrics, particularly ranking first in HD, demonstrating strong capabilities in detail localization, thereby providing significant advantages in the overall evaluation.



## V. CONCLUSION

Within this investigation, an efficient and high-performance 3D MRI BTIS model named BTIS-Net is introduced. Specifically, we introduce the 3D-DSC, DDRB module, and CAS module into the classical 3D U-Net model to improve the model's efficiency and segmentation performance. As an improved 3D U-Net model, BTIS-Net has three main features: first, it reduces the number of parameters and thus saves computational resources. Second, the DDRB module expands the sensing field, capturing a wider range of image context information. This enhances the model's ability to understand the overall image context and structure, thereby improving segmentation performance. Third, the CAS module optimizes the segmentation performance of the confusion area. Experiments on the BraTS dataset show that the proposed BTIS-Net model improves in detail segmentation, overall segmentation, and efficiency compared with other benchmark models. However, there are some limitations in this study. To illustrate this point, the segmentation of MRI and brain tumor images is the sole focus of this research. Here are two areas where future research could be conducted: Firstly, the BTIS-Net model may be expanded to handle multimodal medical image segmentation problems, which involve merging data from many modalities to enhance the accuracy and performance of segmentation. In addition, the BTIS-Net model can be applied to different types of brain disease image segmentation tasks utilizing migration learning, which can further verify its generalization ability and applicability.

## ACKNOWLEDGMENT

Dataset used in the preparation of this article were retrieved from MICCAI's Dataset on Brain Tumor Segmentation.

## REFERENCES

- [1] T. A. Soomro, L. Zheng, A. J. Afifi, A. Ali, S. Soomro, M. Yin, and J. Gao, "Image segmentation for MR brain tumor detection using machine learning: A review," *IEEE Rev. Biomed. Eng.*, vol. 16, pp. 70–90, 2023, doi: [10.1109/RBME.2022.3185292](https://doi.org/10.1109/RBME.2022.3185292).
- [2] P. Jyothi and A. R. Singh, "Deep learning models and traditional automated techniques for brain tumor segmentation in MRI: A review," *Artif. Intell. Rev.*, vol. 56, no. 4, pp. 2923–2969, Apr. 2023.
- [3] M. Tamilarasi, "Performance analysis of glioma brain tumor segmentation using CNN deep learning approach," *IETE J. Res.*, vol. 69, no. 5, pp. 2400–2411, Jul. 2023, doi: [10.1080/03772063.2021.1892536](https://doi.org/10.1080/03772063.2021.1892536).
- [4] M. Mokhtar, H. Abdel-Galil, and G. Khoriba, "Brain tumor semantic segmentation using residual U-Net++ encoder-decoder architecture," *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. 6, pp. 1110–1117, 2023.
- [5] H. T. Halawani, "Salp swarm algorithm with multilevel thresholding based brain tumor segmentation model," *Comput., Mater. Continua*, vol. 74, no. 3, pp. 6775–6788, 2023, doi: [10.32604/cmc.2023.030814](https://doi.org/10.32604/cmc.2023.030814).
- [6] R. Kalam, C. Thomas, and M. A. Rahiman, "Brain tumor detection in MRI images using adaptive-ANFIS classifier with segmentation of tumor and edema," *Soft Comput.*, vol. 27, no. 5, pp. 2279–2297, Mar. 2023, doi: [10.1007/s00500-022-07687-4](https://doi.org/10.1007/s00500-022-07687-4).
- [7] P. S. Chaitanya and S. K. Satpathy, "A multilevel de-noising approach for precision edge-based fragmentation in MRI brain tumor segmentation," *Traitement du Signal*, vol. 40, no. 4, pp. 1715–1722, Aug. 2023, doi: [10.18280/ts.400440](https://doi.org/10.18280/ts.400440).
- [8] D. Kreuzberger, N. Kühl, and S. Hirschl, "Machine learning operations (MLOps): Overview, definition, and architecture," *IEEE Access*, vol. 11, pp. 31866–31879, 2023.
- [9] R. Pitchai, P. Supraja, A. R. Sulthana, T. Veeramakali, and C. M. Babu, "MRI image analysis for cerebrum tumor detection and feature extraction using 2D U-ConvNet and SVM classification," *Pers. Ubiquitous Comput.*, vol. 27, no. 3, pp. 931–940, Jun. 2023, doi: [10.1007/s00779-022-01676-y](https://doi.org/10.1007/s00779-022-01676-y).
- [10] U. G. Naidu, R. Thiruvengatanadhan, S. Narayana, and P. Dhanalakshmi, "Character level segmentation and recognition using CNN followed random forest classifier for NPR system," *Int. J. Adv. Comput. Sci. Appl.*, vol. 13, no. 11, pp. 12–18, 2022, doi: [10.14569/ijacsa.2022.0131102](https://doi.org/10.14569/ijacsa.2022.0131102).
- [11] T. Magadza and S. Viriri, "Efficient nnU-Net for brain tumor segmentation," *IEEE Access*, vol. 11, pp. 126386–126397, 2023, doi: [10.1109/ACCESS.2023.3329517](https://doi.org/10.1109/ACCESS.2023.3329517).
- [12] T. Tarasiewicz, M. Kawulok, and J. Nalepa, "Lightweight U-Nets for brain tumor segmentation," in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, vol. 12659, A. Crimi and S. Bakas, Eds., Cham, Switzerland: Springer, 2021, doi: [10.1007/978-3-030-72087-2\\_1](https://doi.org/10.1007/978-3-030-72087-2_1).
- [13] G. Du, X. Cao, J. Liang, X. Chen, and Y. Zhan, "Medical image segmentation based on U-Net: A review," *J. Imag. Sci. Technol.*, vol. 64, no. 2, Mar. 2020, Art. no. 020508, doi: [10.2352/j.imagingisci.technol.2020.64.2.020508](https://doi.org/10.2352/j.imagingisci.technol.2020.64.2.020508).
- [14] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: Learning dense volumetric segmentation from sparse annotation," in *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, Cham, Switzerland: Springer, 2016, doi: [10.1007/978-3-319-46723-8\\_49](https://doi.org/10.1007/978-3-319-46723-8_49).
- [15] J. Zhang, Y. Jin, J. Xu, X. Xu, and Y. Zhang, "MDU-Net: Multi-scale densely connected U-Net for biomedical image segmentation," in *Proc. Comput. Vis. Pattern Recognit.*, Sep. 2018. [Online]. Available: <https://doi.org/10.48550/arXiv.1812.00352>
- [16] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: Redesigning skip connections to exploit multiscale features in image segmentation," *IEEE Trans. Med. Imag.*, vol. 39, no. 6, pp. 1856–1867, Jun. 2020, doi: [10.1109/TMI.2019.2959609](https://doi.org/10.1109/TMI.2019.2959609).
- [17] N. Ibtihaz and M. S. Rahman, "MultiResUNet: Rethinking the U-Net architecture for multimodal biomedical image segmentation," *Neural Netw.*, vol. 121, pp. 74–87, Jan. 2020, doi: [10.1016/j.neunet.2019.08.025](https://doi.org/10.1016/j.neunet.2019.08.025).
- [18] Z. Zhang, C. Wu, S. Coleman, and D. Kerr, "DENSE-INception U-Net for medical image segmentation," *Comput. Methods Programs Biomed.*, vol. 192, Aug. 2020, Art. no. 105395, doi: [10.1016/j.cmpb.2020.105395](https://doi.org/10.1016/j.cmpb.2020.105395).
- [19] O. Oktay, J. Schlemper, L. Le Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, B. Glocker, and D. Rueckert, "Attention U-Net: Learning where to look for the pancreas," in *Proc. Comput. Vis. Pattern Recognit.*, Sep. 2018. [Online]. Available: <https://doi.org/10.48550/arXiv.1804.03999>
- [20] W. Zhu, Y. Huang, L. Zeng, X. Chen, Y. Liu, Z. Qian, N. Du, W. Fan, and X. Xie, "AnatomyNet: Deep learning for fast and fully automated whole-volume segmentation of head and neck anatomy," *Med. Phys.*, vol. 46, no. 2, pp. 576–589, Feb. 2019, doi: [10.1002/mp.13300](https://doi.org/10.1002/mp.13300).
- [21] Q. Jin, Z. Meng, C. Sun, H. Cui, and R. Su, "RA-UNet: A hybrid deep attention-aware network to extract liver and tumor in CT scans," *Frontiers Bioeng. Biotechnol.*, vol. 8, Dec. 2020, Art. no. 605132, doi: [10.3389/fbioe.2020.605132](https://doi.org/10.3389/fbioe.2020.605132).
- [22] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-Net: Fully convolutional neural networks for volumetric medical image segmentation," in *Proc. 4th Int. Conf. 3D Vis. (3DV)*, Stanford, CA, USA, Oct. 2016, pp. 565–571, doi: [10.1109/3DV.2016.79](https://doi.org/10.1109/3DV.2016.79).
- [23] J. Dolz, I. Ben Ayed, and C. Desrosiers, "Dense multi-path U-Net for ischemic stroke lesion segmentation in multiple image modalities," in *Proc. Int. MICCAI Brainlesion Workshop*, vol. 11383, 2018, pp. 271–282, doi: [10.1007/978-3-030-11723-8\\_27](https://doi.org/10.1007/978-3-030-11723-8_27).
- [24] F. Isensee, J. Petersen, A. Klein, D. Zimmerer, P. F. Jaeger, S. Kohl, J. Wasserthal, G. Koehler, T. Norajitra, S. Wirkert, and K. H. Maier-Hein, "nnU-Net: Self-adapting framework for U-Net-based medical image segmentation," 2018, *arXiv:1809.10486*. [Online]. Available: <https://doi.org/10.48550/arXiv.1809.10486>
- [25] Q. Jin, Z. Meng, T. D. Pham, Q. Chen, L. Wei, and R. Su, "DUNet: A deformable network for retinal vessel segmentation," in *Proc. Comput. Vis. Pattern Recognit.*, 2018.
- [26] J. Li, Z. L. Yu, Z. Gu, H. Liu, and Y. Li, "MMAN: Multi-modality aggregation network for brain segmentation from MR images," *Neurocomputing*, vol. 358, pp. 10–19, Sep. 2019, doi: [10.1016/j.neucom.2019.05.025](https://doi.org/10.1016/j.neucom.2019.05.025).
- [27] H. Dong, G. Yang, F. Liu, Y. Mo, and Y. Guo, "Automatic brain tumor detection and segmentation using U-Net based fully convolutional networks," in *Proc. Annu. Conf. Med. Image Understand. Anal. Switzerland*: Springer, Jul. 2017, pp. 506–517, doi: [10.1007/978-3-319-60964-5\\_44](https://doi.org/10.1007/978-3-319-60964-5_44).

- [28] D. Liu, H. Zhang, M. M. Zhao, X. J. Yu, S. W. Yao, and W. Zhou, "Brain tumor segmentation based on dilated convolution refine networks," in *Proc. 16th IEEE Int. Conf. Softw. Eng. Res., Manag. Appl.*, Kunming, China, Jun. 2018, pp. 113–120, doi: [10.1109/SERA.2018.8477213](https://doi.org/10.1109/SERA.2018.8477213).
- [29] X. M. Kong, G. X. Sun, Q. Wu, J. Liu, and F. M. Lin, "Hybrid pyramid U-Net model for brain tumor segmentation," in *Proc. 10th Int. Conf. Intell. Inf. Process. IX*, Nanning, China, 2018, pp. 346–355, doi: [10.1007/978-3-030-00828-4\\_35](https://doi.org/10.1007/978-3-030-00828-4_35).
- [30] A. Beers, K. Chang, J. Brown, E. Sartor, C. Mammen, E. Gerstner, B. Rosen, and J. Kalpathy-Cramer, "Sequential 3D U-Nets for biologically-informed brain tumor segmentation," 2017, *arXiv:1709.02967*.
- [31] K. Kamnitsas, C. Ledig, V. F. J. Newcombe, J. P. Simpson, A. D. Kane, D. K. Menon, D. Rueckert, and B. Glocker, "Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation," *Med. Image Anal.*, vol. 36, pp. 61–78, Feb. 2017, doi: [10.1016/j.media.2016.10.004](https://doi.org/10.1016/j.media.2016.10.004).
- [32] X. Feng, N. J. Tustison, S. H. Patel, and C. H. Meyer, "Brain tumor segmentation using an ensemble of 3D U-Nets and overall survival prediction using radiomic features," *Frontiers Comput. Neurosci.*, vol. 14, Apr. 2020, doi: [10.3389/fncom.2020.00025](https://doi.org/10.3389/fncom.2020.00025).
- [33] T. Magadza and S. Viriri, "Brain tumor segmentation using partial depthwise separable convolutions," *IEEE Access*, vol. 10, pp. 124206–124216, 2022, doi: [10.1109/ACCESS.2022.3223654](https://doi.org/10.1109/ACCESS.2022.3223654).
- [34] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2261–2269.
- [35] Q. Huang, C. Wu, C. Xia, Y. Wang, and C.-C.-J. Kuo, "Semantic segmentation with reverse attention," in *Proc. Brit. Mach. Vis. Conf.*, 2017, doi: [10.48550/arXiv.1707.06426](https://doi.org/10.48550/arXiv.1707.06426).
- [36] J. Wang, X. Zhang, L. Guo, C. Shi, and S. Tamura, "Multi-scale attention and deep supervision-based 3D UNet for automatic liver segmentation from CT," *Math. Biosci. Eng.*, vol. 20, no. 1, pp. 1297–1316, 2022.
- [37] M. S. Aslam, M. Younas, M. U. Sarwar, M. A. Shah, A. Khan, M. I. Uddin, S. Ahmad, M. Firdausi, and M. Zaindin, "Liver-tumor detection using CNN ResUNet," *Comput., Mater. Continua*, vol. 67, no. 2, pp. 1899–1914, 2021, doi: [10.32604/cmc.2021.015151](https://doi.org/10.32604/cmc.2021.015151).
- [38] R. Zaitoon and H. Syed, "RU-Net2+: A deep learning algorithm for accurate brain tumor segmentation and survival rate prediction," *IEEE Access*, vol. 11, pp. 118105–118123, 2023, doi: [10.1109/ACCESS.2023.3325294](https://doi.org/10.1109/ACCESS.2023.3325294).
- [39] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: A nested U-Net architecture for medical image segmentation," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support* (Lecture Notes in Computer Science), vol. 11045, Cham, Switzerland: Springer, 2018, pp. 3–11, doi: [10.1007/978-3-030-00889-5\\_1](https://doi.org/10.1007/978-3-030-00889-5_1).
- [40] H. Ramamoorthy, M. Ramasundaram, R. S. P. Raj, and K. Randive, "TransAttU-Net deep neural network for brain tumor segmentation in magnetic resonance imaging," *IEEE Can. J. Electr. Comput. Eng.*, vol. 46, no. 4, pp. 298–309, Fall 2023, doi: [10.1109/ICJECE.2023.3289609](https://doi.org/10.1109/ICJECE.2023.3289609).
- [41] J. Sun, Y. Peng, Y. Guo, and D. Li, "Segmentation of the multimodal brain tumor image used the multi-pathway architecture method based on 3D FCN," *Neurocomputing*, vol. 423, pp. 34–45, Jan. 2021, doi: [10.1016/j.neucom.2020.10.031](https://doi.org/10.1016/j.neucom.2020.10.031).
- [42] Y. Chen and Y. Liu, "Automatic segmentation of hippocampal subfields MRI based on FPN-DenseVoxNet," in *Proc. Asia-Pacific Conf. Commun. Technol. Comput. Sci. (ACCTCS)*, Jan. 2021, pp. 58–62, doi: [10.1109/ACCTCS52002.2021.00020](https://doi.org/10.1109/ACCTCS52002.2021.00020).
- [43] H. Mzoughi, I. Njeh, A. Wali, M. B. Slima, A. BenHamida, C. Mhiri, and K. B. Mahfoudhe, "Deep multi-scale 3D convolutional neural network (CNN) for MRI gliomas brain tumor classification," *J. Digit. Imag.*, vol. 33, no. 4, pp. 903–915, Aug. 2020, doi: [10.1007/s10278-020-00347-9](https://doi.org/10.1007/s10278-020-00347-9).
- [44] T. Banzato, M. Bernardini, G. B. Cherubini, and A. Zotti, "A methodological approach for deep learning to distinguish between meningiomas and gliomas on canine MR-images," *BMC Vet. Res.*, vol. 14, no. 1, pp. 1–6, Dec. 2018, doi: [10.1186/s12917-018-1638-2](https://doi.org/10.1186/s12917-018-1638-2).
- [45] A. Younis, L. Qiang, C. O. Nyatega, M. J. Adamu, and H. B. Kawuwa, "Brain tumor analysis using deep learning and VGG-16 ensembling learning approaches," *Appl. Sci.*, vol. 12, no. 14, p. 7282, Jul. 2022, doi: [10.3390/app12147282](https://doi.org/10.3390/app12147282).



**LI LIU** received the master's degree in computer application technology from Jiangnan University, in 2013. She is currently pursuing the Ph.D. degree with Universiti Malaya. Her research interests include artificial intelligence, pattern recognition, image processing, medical imaging, and signal processing.



**KAUJIAN XIA** (Member, IEEE) received the B.S. degree from the School of Computer Science and Engineering, Soochow University, Suzhou, Jiangsu, China, the M.S. degree from the School of Information and Engineering, Jiangnan University, Wuxi, Jiangsu, China, and the Ph.D. degree from the School of Information and Control Engineering, China University of Mining and Technology, Xuzhou, Jiangsu, China. Currently, he is a Senior Engineer in computer science with Changshu Hospital, Soochow University, Changshu, Jiangsu, China; a Professor in medical information with the University of Malaya, Malaya; and a Graduate Supervisor with Jiangnan University. He is the author/co-author of more than 90 research papers in international/national journals, including IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS, IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEM, IEEE/ACM TRANSACTIONS ON COMPUTATIONAL BIOLOGY AND BIOINFORMATICS, *ACM Transactions on Internet Technology*, *ACM Transactions on Multimedia Computing, Communications, and Information Sciences—Applications*. His research interests include intelligent medical information, image processing, and machine learning.

• • •