IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 9 September 2023, accepted 23 October 2023, date of publication 6 November 2023,
date of current version 15 November 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3329517

RESEARCH ARTICLE

Efficient nnU-Net for Brain Tumor Segmentation

TIRIVANGANI MAGADZA AND SERESTINA VIRIRI, (Senior Member, IEEE)
School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal, Durban 4000, South Africa
Corresponding author: Serestina Viriri (viriris@ukzn.ac.za)

ABSTRACT Brain tumors are one of the leading causes of death in adults. They come in various shapes and sizes from one patient to another. Sometimes, they infiltrate surrounding normal tissues, making it challenging to delineate tumor boundaries. Despite extensive research, the prognosis is still low. Accurate and timely brain tumor segmentation is critical for treatment planning and disease progression monitoring. Automatic segmentation of brain tumors using deep learning methods has produced high-quality and reproducible segmentation results. Specifically, the encoder-decoder networks, like the U-Nets, have dominated the previous BraTS Challenges because of their superior performance. Due to the importance of high-quality segmentation, most state-of-the-art models focus more on pushing the boundaries of the current methods at the expense of computational complexity. The computational budget for practical applications is minimal, requiring technological solutions that balance accuracy and available computational resources. In this study, we extended the U-Net model in the nnU-Net by replacing the basic 3D convolution blocks with bottleneck units utilizing depthwise-separable convolutions. Furthermore, we introduced the shuffle attention mechanism in the skip connections to compensate for the slight loss in segmentation accuracy due to a reduction in the number of parameters. On the brain tumor dataset BraTS 2020, our network achieves dice scores of 79.2%, 91.2%, and 84.8% for enhancing tumor (ET), whole tumor (WT), and tumor core (TC), respectively, with only 2.51M parameters and 55.26G FLOPS. Extensive experimental results of the BraTS 2020 dataset reviewed that the proposed modifications achieved competitive performance at a lower computational cost. The code for this project is available at https://github.com/tmagadza/EfficientNNUNET.git.

INDEX TERMS Brain tumor segmentation, depthwise-separable convolutions, group convolution, shuffle attention, U-Net.

I. INTRODUCTION

A brain tumor is the abnormal growth of cells in any part of the brain. Their exact causes are not yet known [1]. However, the risk factors include a family history of brain tumors, metastases, and exposure to ionizing radiation. There are about 120 types of tumors, with gliomas being the most common and one of the leading causes of death among adults [2]. The World Health Organization broadly classifies gliomas into low-grade (Grade I and II) and high-grade (Grade III and IV) tumors. The low-grade tumors are less aggressive, with a life expectancy that spans many years. On the other hand, high-grade tumors are much more aggressive, with a median survival rate of fewer than two years, and require immediate treatment [3].

Timely, accurate, and reproducible segmentation of brain tumors is critical for diagnosis, treatment planning, and monitoring of disease progression. In clinical practice, segmentation is done manually by a high-trained radiologist. This process is tedious and time-consuming and suffers from intra and inter-rater variability [3], [4]. Consequently, manual segmentation is only used for qualitative assessment or visual inspection.

Meanwhile, in recent years, automatic brain tumor segmentation has been slowly becoming a viable solution to manual segmentation. It requires minimal human involvement if not none at all. However, it also presented its unique challenges. Brain tumors come in different shapes, sizes, and locations from one patient to another, limiting the use of prior

The associate editor coordinating the review of this manuscript and approving it for publication was Orazio Gambino.

knowledge of the shape and location of anatomic tissues. The most aggressive tumors often diffuse into surrounding tissues, making delineating tumor boundaries difficult. Furthermore, segmentation only depends on comparing pixel intensities between normal brain parts and lesions. Despite these challenges, automatic brain tumor segmentation is still a promising solution for quantitatively assessing brain tumors.

More recently, deep learning methods for automatic brain tumor segmentation have attracted much attention among the research community owing to their success in various computer vision applications. Applying deep learning techniques to medical image analysis requires expertise in choosing the appropriate network for the task at hand and making numerous decisions regarding hyper-parameters, preprocessing and post-processing techniques, training schemes, data augmentation, etc. [5]. A slight mistake in the configuration of these methods will lead to a significant drop in performance. For example, methods based on U-Net [6] like structure have been dominating the BraTS challenge [7]. Still, the performance of these methods varies significantly, signifying the importance of expected knowledge for the task at hand [5].

In 2020, Isensee et al. [8] proposed an open-source self-configuring deep learning framework for biomedical image segmentation, which they dubbed nnU-Net.[1] Their framework automates the entire segmentation pipeline, including configuring any medical dataset, preprocessing, network architecture, training, and post-processing without human input. nnU-Net has set a new state of the art in various semantic segmentation challenges [8]. In the context of Brain Tumor segmentation, Isensee et al. [7] investigated the suitability of nnU-Net for brain tumor segmentation while applying BraTS-specific modification, and their method came first in BraTS 2020 Challange. Again, in BraTS 2021 Challenge, Luu and Park [9] proposed several modifications to the nnU-Net, including using a larger network, swapping batch normalization with group normalization, and adopting axial attention in the decoder. Their method also came first.

Despite several benefits that nnU-Net brings to medical image segmentation, it needs more computational costs. At its core, nnU-Net is an instance of basic U-Net architecture. It makes use of standard convolution, which is computationally expensive. By using 3D convolutions, which have been shown to perform better than 2D counterparts, the number of parameters increases substantially, making it practically impossible to train the model reasonably for a given computational budget.

This work investigated the effects of reducing the nnU-Net framework's computational complexity on the model's segmentation performance on brain tumor segmentation tasks.

[1]https://github.com/MIC-DKFZ/nnUNet

Our main contributions can be summarized as follows:

1) We propose swapping all standard convolutions with depthwise separate convolutions to reduce the number of network parameters and improve the efficiency of the network.
2) We introduce bottleneck units to reduce the number of parameters further.
3) We adopt the 3D shuffle attention mechanism in skip connections to improve the segmentation performance of the network. Moreover, we introduced residual connections to avoid network degradation.
4) We extensively evaluate the proposed modifications using BraTS 2020 dataset.

The rest of the paper is organized as follows: Section II reviews related work. Section III describes the dataset used and the proposed modifications to the nnU-Net framework. Section IV presents the experimental results, which are discussed in Section V. Lastly, Section VI provides concluding remarks.

II. RELATED WORK

A. U-NET LIKE ARCHITECTURE

Since the introduction U-Net [6] in 2015, the encoder-decoder-like structure became the de-facto standard for biomedical segmentation. The U-Net architecture uses an encoder pathway to extract rich semantic and global information by successively reducing the spatial resolution by half and doubling the number of feature maps. The decoder gradually doubles the spatial resolution to recover the spatial resolution while reducing the feature maps by half. Skip connections combine the encoder's finer features and the decoder's course features. Dong et al. [10] proposed a 2D U-Net that was optimized using soft dice loss to mitigate the unbalanced nature of the BraTS 2015 dataset. Their methods applied extensive data augmentation techniques to improve segmentation performance. Myronenko [11], the winner of BraTS 2018, proposed an encoder-decoder network with an asymmetrically larger encoder to extract more deep features. Their method uses a variational autoencoder branch to regularize the shared encoder. The author observed that increasing the width of the network improved performance. Their approach is computationally expensive due to standard convolutions and large input patch sizes. Insensee et al. [12] developed a U-Net Like 3D architecture, which was trained using large patch size, dice loss, and extensive data augmentation. Deep supervision was used to improve gradient propagation to lower layers further. Li et al. [13] proposed an up-skip connection between the encoder and decoder to improve the information flow. Their network incorporated an inception module and used cascading training strategy to segment tumor regions sequentially. Zhao, Y. et al. [14] investigated the usefulness of various schemes in data processing, model designing, and optimization as applied to general DCNN design and training for the 3d brain tumor segmentation. Their method won second place in the BraTS 19 Challenge.

B. REGION-BASED TRAINING

Wang et al. [15] developed a method that exploits brain tumors' hierarchical nature by segmenting partially overlapping regions one after the other in a cascading fashion. Their method uses anisotropic convolution to balance between accuracy and model complexity. Multi-scale feature fusion was exploited for robust segmentation. The shallow layers learn to represent local and low-level features while deep layers learn to represent more global and high-level features. Their method was not end-to-end. Each network was trained separately, increasing the time required for both training and testing [16]. Wang et al. [17] extended their previous work [15] to incorporate uncertainty estimation gathered from test time augmentation. The paper showed that uncertainty estimation could identify false positives and improve segmentation performance. Unfortunately, their method required a longer time to train. In [18], Zhoul et al. adopted the multi-task learning approach instead of training three networks separately, combining the three tasks in a single model. The paper adopted the curriculum learning scheme, gradually introducing each task as the learning proceeded.

C. LIGHT-WEIGHT NETWORKS

Chen et al. [16] used anisotropic convolutions to split the standard 3D convolution into three parallel branches, each extracting features from different orthogonal views. The use of separable convolution has the benefit of reducing the number of parameters. Their model replaces all the standard convolution operations in the U-Net structure with separable convolutions. Chen et al. [19] exploited group convolution to reduce model complexity. Each group is split into two three branches using weighted 3D dilated convolution for multi-scale learning. A multiplexer unit facilitates information sharing between each group or fiber. Zhou et al. [20] utilized the shufflenetV2 units in the encoder to reduce the number of parameters, while in the decoder, residual units are used to address network degradation. Luo, Z et al. [21] proposed hierarchical decoupled convolution to reduce the number of parameters in an encoder-decoder structure. Peng et al. [22] proposed a U-Net variant that utilizes weighted dilated convolutions to learn multi-scale features. The authors used group convolutions to reduce the number of parameters in the network. Furthermore, the authors used dense residual blocks to improve segmentation performance. In [23], the authors used a 3D inverted residual module to reduce the computational complexity of 3D models. Their methods achieved competitive results on BraTS 2018 while using few computational resources. Zhang et al. [24] exploited shuffle units and depthwise separable convolutions to reduce the number of network parameters and operations.

D. ATTENTION MECHANISM

Noori et al. [25] proposed a 2d encoder-decoder networks structure that utilizes residual units to improve network training and apply channel attention after concatenating low-level and high-level features. The authors argue that it is improper to concatenate features from low-level and high-level features without weighing them. Empirical results demonstrate the effectiveness of channel attention in improving segmentation performance. Zhang et al. [26] proposed a 2d encoder-decoder network structure that incorporates residual units and attention gates in the skip connection. Experiment results showed the effectiveness of attention gates in improving network performance. Cao et al. [27] proposed a UNet-like network structure that utilizes 3D Shuffle Attention in the encoder and skip connections. The authors adopted an optimized shuffle unit as a basic building block. The authors did not report on the complexity analysis of their method.

III. MATERIALS AND METHODS

A. DATA

We used the BraTS 2020 dataset [3], [28], [29] that contains 369 training and 125 validation subjects for training and validation of our model. As illustrated in Figure 1, all subjects have native (T1), post-contrast T1-weighted (T1Gd), T2-weighted (T2), and T2 Fluid Attenuated Inversion Recovery (T2-FLAIR) volumes that were acquired using varying clinical protocols and scanners from nineteen (19) institutions. The training set was also comprised of manually annotated ground truth by one to four raters applying the same annotation protocol, and experienced radiologists approved their annotations. In contrast, the ground truth for the validation set was not made public. Instead, the researchers can use the online evaluation platform[2] to evaluate models. All scans were co-registered to the same anatomical template, interpolated to the same resolution (1mm3), skull-stripped, and had an original image size of $240 \times 240 \times 155$. We also used the BraTS 2021 dataset [3], [29], [30] which includes 1251 training cases and 219 validaton cases. The structure and format of BraTS 2021 is consistent with the BraTS 2020 dataset.

B. NNU-NET BASELINE

Our baseline model was an instantiation of nnU-Net [7], a winning model for BraTS 2020. The model follows an encoder-decoder structure with skip connections linking the two pathways, as presented in Figure 2. As a 3D U-Net [31], the model takes in a large input patch of $128 \times 128 \times 128$, with four 3D MRI image modalities concatenated in the channel dimension. The network comprises five (5) resolution levels. In the encoding pathway, each level reduces the spatial resolution by half using strided convolution and doubles the feature maps starting from base feature maps of 32 up to a maximum of 320. Two consecutive convolution blocks were applied in each layer, each performing $3 \times 3 \times 3$ convolution followed by instance normalization [32] and then Leaky Relu non-linearity. In the

[2]https://ipp.cbica.upenn.edu/

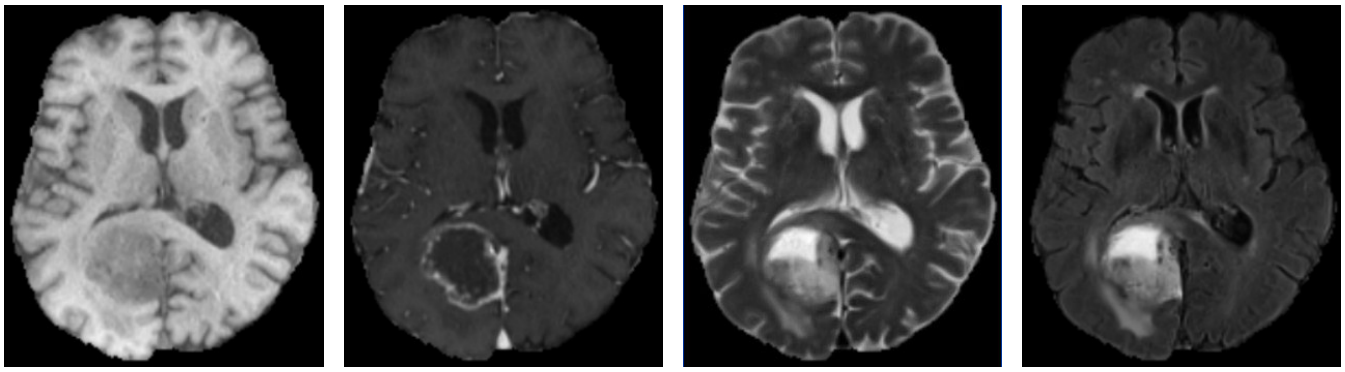


FIGURE 1. Examples of different MRI imaging modalities. From left to right: T1, T1ce, T2, and FLAIR.

decoding path, each layer gradually reduces the number of feature maps by half while doubling the spatial resolution with transpose convolutions. Convolution blocks in the decoding path follow the same structure as the encoding path. $1 \times 1 \times 1$ convolution followed by sigmoid non-linearity is performed after the last layer to reduce the number of feature maps to 3. Deep supervision was also used to improve network training in all layers along the decoding path except the two lowest resolutions. To improve the segmentation performance, we directly optimize the three partially overlapping regions: whole tumor, tumor core, and enhancing tumor, instead of providing labels that include: edema, non-enhancing tumor, and necrosis and enhancing tumor. Aggressive data augmentation techniques were applied on the fly using the batchgenerators framework.[3] Specifically, we applied rotation, scaling, elastic deformation, additive brightness augmentation, and gamma augmentation as described in [7]. The loss function was a summation of dice and binary cross-entropy losses, which has been shown to improve segmentation performance [33].

C. NNU-NET MODIFICATIONS

1) REDUCED COMPUTATIONAL COMPLEXITY

A standard convolution operation is computationally expensive since it simultaneously performs spatial and channel-wise correlation in one go. An excessive amount of computation is required when using 3D MRI volumes with large patch sizes, which were shown to perform well as compared to 2D counterparts, making it difficult to train the resulting models. To reduce the number of parameters as well as computational complexity, we replaced all the standard convolution operations with depthwise separable convolutions, which apply $3 \times 3 \times 3$ convolution on each channel separately followed by $1 \times 1 \times 1$ convolution to project the output channels from previous operation to another channel space as illustrated in Figure 3(b). A depthwise separable convolution can be generalized as a group convolution with a group size equal to the number of input channels. We adopted the bottleneck unit as our basic building block with depthwise separable convolution in the middle, as shown in Figure 3(c). The module introduced an addition hyper-parameter, reduction ratio r, to reduce the number of input channels for the middle layer. We have fixed the value of r to 4.

[3]https://github.com/MIC-DKFZ/batchgenerators

2) SHUFFLE ATTENTION MECHANISM

The use of depthwise separable convolutions will significantly reduce network parameters, which may slightly reduce the segmentation accuracy. To compensate for the loss in performance, we introduced the shuffle attention (SA) [34] mechanism, which simultaneously applies spatial and channel attention. The attention will help the network focus more on all salient features of the task. The network can learn to capture the pixel-level correlations and channel dependency by combining spatial and channel attention. Numerous studies [35], [36], [37], [38] have shown that attention mechanisms can considerably enhance network performance.

Given an input feature map $I \in \mathbb{R}^{C\times H\times W\times D}$, where H, W, D, and C are the height, width, depth, and number of channels of the input feature map, respectively, SA first divides I into G groups along the channel dimension, i.e., $I = [I_1, \cdots, I_G], I_k \in \mathbb{R}^{C/G\times H\times W\times D}$. Then each sub group I_k is further split into two branches, denoted by $I_{k1}, I_{k2} \in \mathbb{R}^{C/2G\times H\times W\times D}$. As shown in Figure 4, the first branch is used to generate the channel attention map by applying global average pooling (GAP), which generates channel-wise statistics $s \in \mathbb{R}^{C/2G\times 1\times 1\times 1}$, to the input feature map, which can be calculated by shrinking I_{k1} through spatial dimension $H \times W \times D$:

$$s = \mathcal{F}_{gp}(I_{k1}) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} \sum_{t=1}^{D} I_{k1}(i, j, t) \qquad (1)$$

Furthermore, a linear transformation $\mathcal{F}_c(\cdot)$ is perform, followed by a simple gating mechanism with sigmoid activation σ to produce the channel attention:

$$I'_{k1} = \sigma(\mathcal{F}_c(s)) \cdot I_{k1} = \sigma(W_1 s + b_1) \cdot I_{k1} \qquad (2)$$

where $W_1 \in \mathbb{R}^{C/2G\times 1\times 1\times 1}$ and $b_1 \in \mathbb{R}^{C/2G\times 1\times 1\times 1}$ are parameters used to scale and shift s.

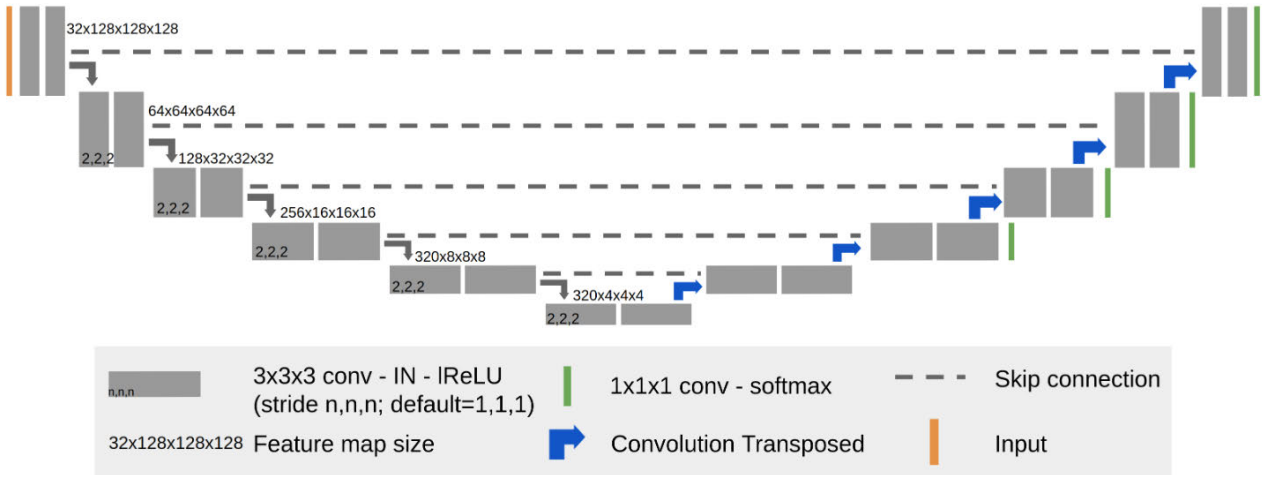


FIGURE 2. Baseline model as generated by the nnU-Net framework. (adapted from [7]).

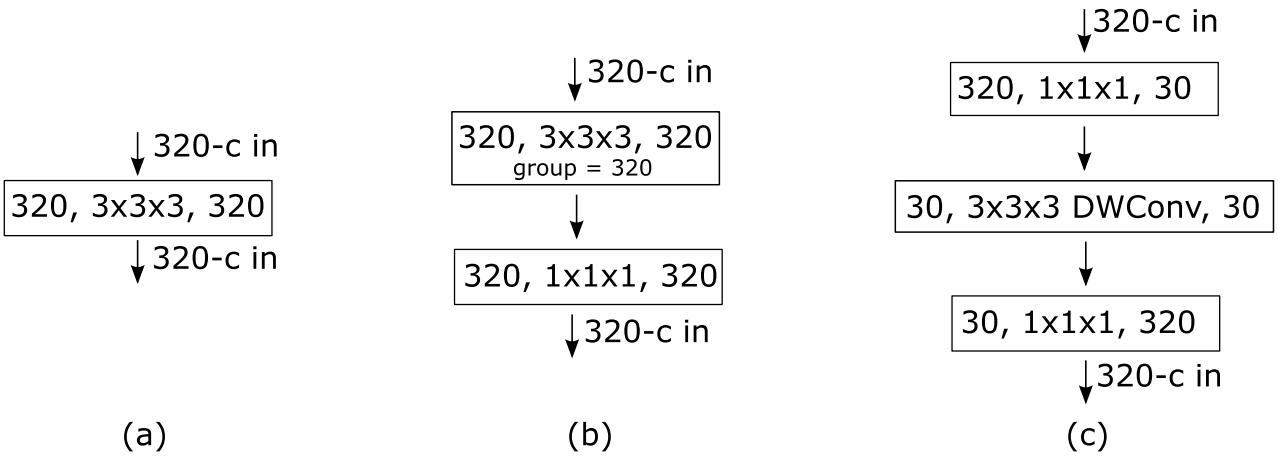


FIGURE 3. Basic building blocks. (a) Standard convolution block. (b) DWConv: Depthwise separable convolution block. (c) Bottlenet unit with depthwise separable convolution block in the middle.

The second branch generates the spatial attention by firstly obtaining spatial-wise statistics through Group Norm (GN) over I_{k2} followed by a linear transformation $\mathcal{F}_c(\cdot)$. The final output of spatial attention is given by:

$$I'_{k2} = \sigma(W_2 \cdot GN(I_{k2}) + b_2) \cdot I_{k2} \tag{3}$$

where W_2 and b_2 are parameters with shape $\mathbb{R}^{C/2G\times1\times1\times1}$.

Then, a concatenation operation is applied to the two branches to make the number of channels as the same as the number of input, i.e., $I'_k = [I'_{k1}, I'_{k2}] \in \mathbb{R}^{C/2G\times H\times W\times D}$. All the sub-groups are then aggregated, followed by the "channel shuffle" operation to enable information communication between different sub-groups. The final output of the SA module is the same size as I.

D. TRAINING

Our model was implemented in Pytorch[4] using opensource framework for biomedical segmentation[5] [8]. Each network takes an input patch of 128 × 128 × 128, with four 3D MRI image modalities concatenated in the channel dimension. We normalize each input channel independently by subtracting the mean and dividing it by the standard deviation. Data augmentation, which comprised random rotation and scaling, elastic deformation, additive brightness augmentation, and gamma scaling, was applied on the fly. The loss function was a summation of batched dice and cross-entropy loss. We optimize all the networks with stochastic gradient descent with an initial learning rate of 0.01 and Nesterov momentum of 0.99. The learning rate was decayed with a polynomial schedule:

$$lr = 0.01 \times (1 - \frac{epoch}{400})^{0.9} \tag{4}$$

Each network was trained for a total of 400 epochs, with each epoch defined as 250 iterations, on an NVIDIA Tesla V100 16GB GPU. During inference, we post-processed each subject by replacing the enhancing tumor with the tumor core when the predicted volume was less than some threshold. The configuration of our models was as follows:

[4]https://pytorch.org/
[5]https://github.com/MIC-DKFZ/nnUNet

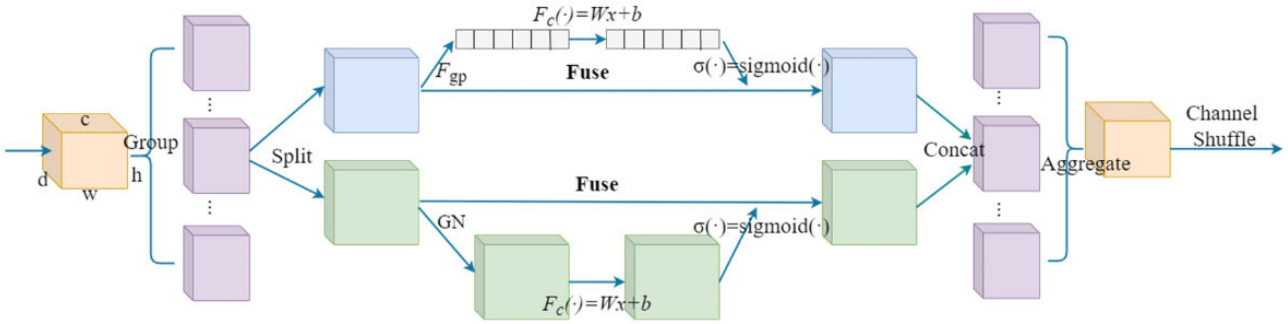


FIGURE 4. The 3D Shuffle Attention Module. The input feature map is first divided into sub-groups along the channel dimension. Then, each sub-group is further divided into two branches, the channel, and spatial attention branches. A concatenation operation is used to join features from the two branches. Afterward, all sub-groups are aggregated, followed by a channel shuffle operation to enable information communication between different sub-groups. (adapted from [34]).

- **BL**: baseline nnUnet-Net without modifications (see Section III-B).
- **BL + DS**: replaced all standard convolution with depthwise separable convolutions
- **BL + BU**: baseline with bottleneck unit as a basic building block
- **BL** $+DS+$ **BU**: baseline with bottleneck unit with depthwise separable convolution in the middle as described in Section III-C1
- **BL** $+DS+BU+$ **SA**: baseline with depthwise separable convolutions, bottleneck unit, shuffle attention in both the encoder and skip connections.
- **BL + R**: baseline with residua connection
- **BL** $+DS+$ **R / BL** $+DS+$ **R***: baseline with depthwise separable convolutions and residual connection. * indicates that the ReLu is applied after the addition of residual units.
- **BL** $+DS+$ **R*+ AA**: baseline with depthwise separable convolutions, residual connection, and shuffle attention skip connections.

IV. RESULTS

A. PERFORMANCE COMPARISON OF THE PROPOSED METHOD

Due to a 12-hour limitation on CHPC,[6] we trained all model configurations for a maximum of 400 epochs. Each configuration was trained with all 369 training cases and evaluated with the 125 validation cases. The validation results of each configuration, as computed by the online evaluation platform, are presented in Table 1. The results show that the baseline configuration (BL) performance for both the dice score and Hausdauf distance is relatively high. Introducing the bottleneck unit (BU) to the baseline showed a decrease in dice score for enhancing tumor and tumor core by 1.9% and 0.4%, respectively. From the results, we can also see a slight increase in Hausdorff distance in enhancing tumor, whole tumor, and tumor core by 9.80 mm, 0.12 mm, and 0.74 mm, respectively.

On the other hand, replacing all the convolution blocks in the baseline with the depthwise separable convolutions (DS) produced similar if not better, results. For Example, the dice score in the tumor core improved by 0.5% while remaining the same for the whole tumor and marginally decreased by 0.4% in enhancing the tumor. As for the Haursdoff distance, the results in Table 1 show an increase of 5.80 mm in enhancing tumor and an improvement of 2.14 mm in the tumor core. The BL + DS + BU model achieved slightly less performance as compared to the BL + DS model. At the same time, the BL + DS + BU + SA model shows a slight improvement in performance in dice score and the Haursdoff distance for enhancing tumor compared to the BL + DS + BU model.

Residual units can help reduce degradation in deep networks like U-Net structure [39]. From Table 1, we did not observe any benefits of residual connections with ReLu before addition (BL + R and BL + DS + R) except for Hausdourff distance, where we observed a decrease of 2.80 mm in the whole tumor and an increase of 3.00 mm in the tumor core. Interestingly, applying ReLu after the addition further decreases the performance. Introducing Shuffle Attention to the skip connection of BL + DS + R* substantially improved performance in the Hausdorff distance of the tumor core.

B. MODEL COMPLEXITY

Table 1 also reports on the complexity of the different model configurations in terms of floating-point operations (FLOPS) and a number of parameters (Params) as computed by the THOP[7] python library. The table shows that the BL + DS model balanced model complexity and segmentation performance well. Specifically, it achieved 82% and 90% reduction in floating-point operations and several parameters, respectively, without affecting the segmentation performance. A combination of depthwise separable convolutions and bottleneck units (BL + DS + BU) further reduced the model complexity at a slight reduction in segmentation performance. The results show that the Shuffle

[6]https://www.chpc.ac.za/

[7]https://github.com/Lyken17/pytorch-OpCounter

TABLE 1. Performance comparison on the BraTS 2020 validation set (125 cases). Metrics are computed by the online evaluation platform. See Section III-D for decoding the abbreviations. ET - Enhancing tumor, WT - Whole tumor, TC - Tumor core.

Model	Dice			HD95				
	ET	WT	TC	ET	WT	TC	FLOPS	Params.
BL	**0.796**	**0.912**	0.843	23.515	**4.337**	8.340	308.711G	25.708M
BL + BU	0.777	0.911	0.839	32.313	4.453	9.081	65.608G	3.804M
BL + DS	0.792	**0.912**	**0.848**	29.312	4.410	6.200	55.263G	2.513M
BL + DS + BU	0.782	0.910	0.847	26.676	4.758	8.881	**52.333G**	**2.510M**
BL + DS + BU + SA	0.788	0.905	0.837	26.525	5.626	9.101	52.501G	2.510M
BL + R	**0.796**	0.910	0.844	**23.490**	4.591	5.923	328.842G	31.088M
BL + DS + R	0.792	0.910	0.841	26.421	5.203	9.192	75.394G	7.894M
BL + DS + R*	0.784	0.909	0.841	32.331	4.580	9.368	75.394G	7.894M
BL + DS + R* + AA	0.788	0.908	0.845	29.612	5.108	**5.919**	75.461G	7.894M

Best values are shown in bold, and second best are underlined.

Attention barely increases the computation cost. Because of strided convolutions for downsampling and upsampling, introducing residual units resulted in a slight increase in floating point operation due to the $1 \times 1 \times 1$ convolution to match the dimensions in both branches before addition.

C. PERFORMANCE COMPARISON WITH THE STATE-OF-THE-ART

1) PERFORMANCE COMPARISON WITH THE STATE-OF-THE-ART METHODS WITHOUT MODEL ENSEMBLE

For a fair comparison, Table 2 list the results without a model ensemble of the top performances in the BraTS 2020 validation dataset except for the result for Isensee et al. [7], since they did not present the results for a single model. Yuan [40] and Wang et al. [41] are the top participants of the BraTS 2020 challenge, and only results without model ensemble are listed here. We also included single model results of our previous work [42], Raza et al. [43], and Daza et al. [44]. From the results, it is evident that our proposed method achieves superior performance with minimum computation complexity.

Y. Yuan [40] won third place in the BraTS 2020 challenge by aggregating the output feature maps from all the encoding layers with high-level feature maps of each decoding layer using skip connections. Yuan's method achieves superior performance against state-the-art in Hausdorff distance for enhancing tour. However, our lightweight method outperforms Yuan's method in the other metrics.

Wang et al. [41] won second place in the BraTS 2020 challenge. Their methods utilize two interconnected pathways, which take a pair of modalities each. From the results, it is clear that our method demonstrated superior performance in all metrics.

In our previous work [42], we partially utilized depthwise separable convolutions in both the encoder and the decoder. Although our previous work shows competitive performance, it has many floating point operations. In contrast, our proposed work is superior in all metrics.

On the other hand, Raza et al. [43] adopted residual units in the encoding pathway resulting in superior performance in dice score for the enhancing tumor, the worst performance in dice score for the whole tumor, and a comparable performance in the remaining metrics. The computational complexity of their method is relatively high. Similarly, Daza et al. [44] proposed a lightweight method with superior performance in the Hausdorff distance for enhancing tumors. Our approach remains superior in other metrics.

2) PERFORMANCE COMPARISON WITH THE STATE-OF-THE-ART METHODS WITH MODEL ENSEMBLE

Table 3 reports on the aggregate performance of the state-of-the-art methods with the model ensemble on the BraTS 2020 validation dataset. Isensee et al. [7] won the first price, followed by Jia et al. [45] and Wang et al. [41] for the second price, and then Yuan [40] for the third place in the BraTS 2020 challenge. We have also included model ensemble results for Wang et al. [46]. From the table, it is clear that the state-of-the-art methods achieved the best performance at the cost of computational complexity.

Isensee et al. [7] applied nnU-Net [8] with BraTS specific modifications and extensive data augmentation to the brain tumor segmentation problem. Their winning method, which is an ensemble of 25 models, uses basic U-Net structures, with each model trained for 1000 epochs. Results show that Isensee et al.'s method is superior in dice for enhancing tumor and Hausdorff distance for the whole tumor as compared to the state-of-the-art methods. Additionally, their 5 model ensemble also shows similar performance. However, our method achieves the same results as Isensee et al.'s method in dice for the whole tumor and a slight improvement in Hausdorff distance for the tumor core and comparable performance for the other metrics while using significantly fewer computational resources.

Jia et al. [45] proposed a two-stage cascaded model that maintains high-resolution feature representation and uses a Non-local attention mechanism to aggregate contextual information from all layers. Again, their best method was an ensemble of 27 models, each trained for 450 epochs. Despite high performance in dice for the whole tumor and Hausdorff distance for the tumor core, their method is

TABLE 2. Mean performance metrics on BraTS 2020 Validation dataset as compared to the state-of-the-art without model ensemble. See Section III-D for decoding the abbreviations. ET - Enhancing tumor, WT - Whole tumor, TC - Tumor core.

Model	Dice			HD95					
	ET	WT	TC	ET	WT	TC	Epochs	FLOPS	Params.
Y. Yuan [40]	0.785	0.904	0.842	**20.35**	5.49	8.34	300	-	16.50M
Wang et al. [41]	0.785	0.907	0.837	32.25	4.39	8.34	1000	-	-
Magadza et al. [42]	0.774	0.898	0.824	29.82	6.78	7.36	100	616.00G	6.90M
Raza et al. [43]	**0.800**	0.866	0.836	29.82	6.78	7.36	100	374.04G	30.47M
Daza et al. [44]	0.794	0.897	0.845	29.82	**3.59**	6.47	-	**49.82G**	4.02M
BL + DS (ours)	0.792	**0.912**	**0.848**	29.31	4.41	6.20	400	55.26G	**2.51M**
BL + DS + R (ours)	0.792	0.910	0.841	26.42	5.20	9.19	400	75.39G	7.89M
BL + DS + R* + AA (ours)	0.788	0.908	0.845	29.61	5.11	**5.92**	400	75.46G	7.89M

Best values are shown in bold, and second best are underlined.

still computationally expensive. Compared to our method, as shown in Table 3, our best single model achieves comparable results at low computation costs.

Wang et al. [41] and Yuan [40] ensemble 9 models and 11 models to secure second and third places receptively. Their ensemble improved performance in all metrics except for Wang et al. [41], which marginally increased Hausdorff distance for the enhancing tumor and tumor core. In comparison, our model achieved competitive results.

Lastly, Wang et al. [46] introduced Transformer to the encoder-decoder structure for brain tumor segmentation to model long-range dependencies. Their 5 model ensemble achieved superior performance in the Hausdorff distance for the enhancing tumor with relatively huge computational costs. In contrast, our lightweight method demonstrated superior performance in other metrics at significantly low computation costs.

3) PERFORMANCE COMPARISON ON BRATS 2021 DATASET.

Table 4 compares our best-performing model with the state-of-the-art models on the BraTS 2021 dataset 5-fold cross-validation results. By the time of writing this paper, the online evaluation platform for the BraTS 2021 dataset[8] was no longer available. The table shows that our model performed well on the enhancing tumor region and performed slightly poorly on both the whole tumor and tumor core. On the other hand, our model uses very light resources as compared to the state-of-the-art.

D. QUALITATIVE ANALYSIS

In Figure 5, we show qualitative overview of the segmentation performance of BL + DS model on the validation set. To avoid cherry picking [7], we systematically selected cases by first computing an average over the three validation regions and then picked the best, worst, median, and 75th and 25th percentile. The clearly show that the segmentation quality of our model is quite high overall. However, in the worst scenarios, it completely fail to segment small enhancing tumor lesion.

[8]https://www.synapse.org/#!Synapse:syn25829067/wiki/610863

V. DISCUSSION

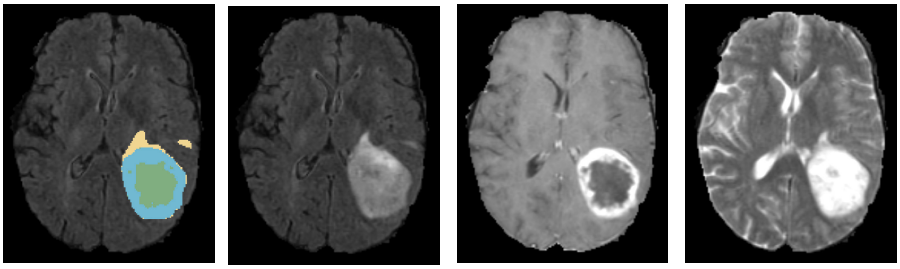
Automatic brain tumor segmentation is paramount for the timely, reproducible, and accurate delineation of tumor substructures. Although deep learning methods have demonstrated superior performance than traditional methods in the past few years, automatic brain tumor segmentation is still an open challenge. Brain liaisons appear in different shapes, sizes, and locations from one patient to another, rendering prior knowledge useless. Moreover, deep learning methods require massive training datasets and computational resources [20]. One would need a model with competitive performance for a limited computational budget for practical application.

Unfortunately, as shown in Table 3, most state-of-the-art methods focus more on improving segmentation performance at the cost of high computation resources. These methods are usually an ensemble of multiple models. For Example, Isensee et al.'s method [7], which won the first prize in the BraTS 2020 challenge, is an ensemble of 25 models. Each model needed to be trained separately for 1000 epochs before their results could be aggregated. Furthermore, their method is computationally expensive when applied to 3D MRI scans due to standard convolutions.

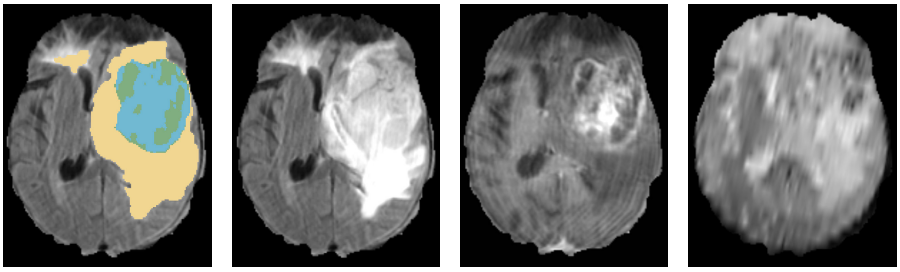
Similarly, Jia et al. [45] ensemble 27 models to win second place in the same challenge. These models chew a significant amount of computation resources to train them. With an increase in the training dataset set, as in BraTS 2021, computation resources are needed even more. The computational requirements may be prohibitive for clinical applications or out of reach for many researchers resulting in poor adoption rates.

Motivated by the above observations, we extended Isensee et al.'s work [7] by introducing depthwise separable convolutions to reduce the computational costs significantly. We also experimented with bottleneck units to further reduce the number of parameters at the expense of a slight loss in segmentation performance. As shown in Table 1, our model configuration with depthwise separable convolutions demonstrated a good balance between computation cost and segmentation performance compared to other configurations. The results are consistent with other previous studies [50], [51], [52]. Although residual units [53] may assist in

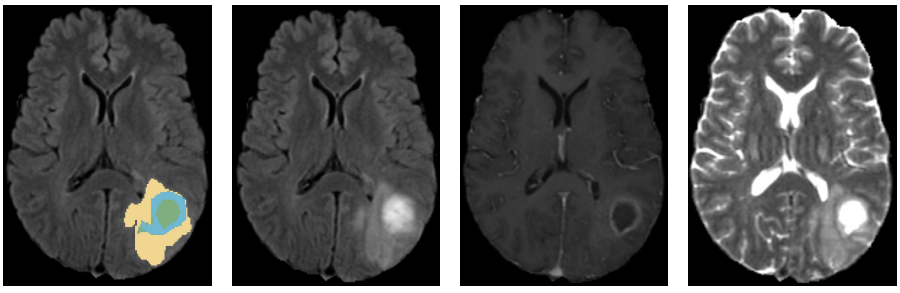
Best: BraTS20_Validation_040, whole: 0.97, core: 0.98, enh: 0.96



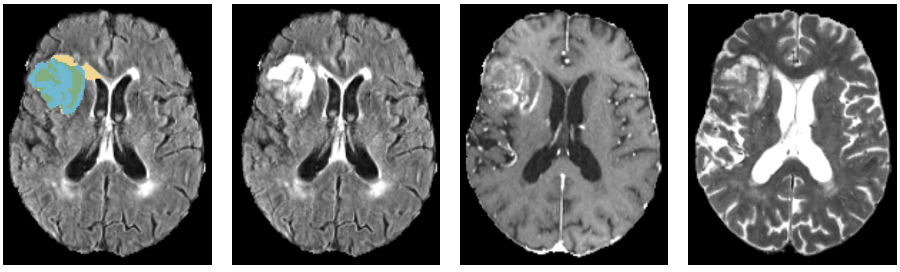
75th percentile: BraTS20_Validation_105, whole: 0.94, core: 0.92, enh: 0.91



Median: BraTS20_Validation_052, whole: 0.95, core:0.91, enh: 0.82



25th percentile: BraTS20_Validation_052, whole: 0.89, core:0.94, enh: 0.66



Worst: BraTS20_Validation_090, whole: 0.91, core:0.25, enh: 0.00

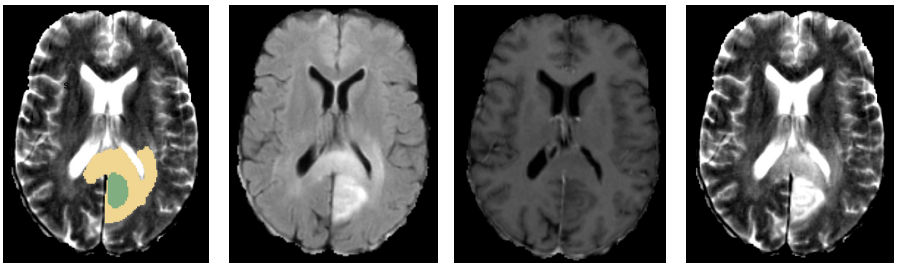


FIGURE 5. Qualitative validation set result. Selection criteria for cases were based on best,worst, median, and 75th and 25th percentile. From left to right:FLAIR image with overlay of generated segmentation, FLAIR image, T1ce image, and T2 image. Edema is shown in yellow, necrosis in green and enhancing tumor in blue.

TABLE 3. Mean performance metrics on BraTS 2020 Validation dataset as compared to the state-of-the-art with the model ensemble. See Section III-D for decoding the abbreviations. ET - Enhancing tumor, WT - Whole tumor, TC - Tumor core. † - Ensemble of models trained with 5-fold cross-validation.

Model	Dice			HD95					
	ET	WT	TC	ET	WT	TC	Epochs	FLOPS	Params.
Isensee et al. [7]	**0.799**	0.912	0.851	23.50	**3.69**	7.82	1000	539.56G	31.20M
Isensee et al. [7] †	**0.799**	0.912	**0.857**	26.41	3.73	5.64	1000	539.56G	31.20M
Jia et al. [45]	0.788	**0.913**	0.855	26.58	4.18	**4.97**	450	621.09G	26.07M
Jia et al. [45]†	0.784	**0.913**	0.835	26.50	4.18	5.52	450	621.09G	26.07M
Y. Yuan [40]	0.793	0.911	0.853	18.20	4.10	5.99	300	-	16.50M
Wang et al. [41]	0.787	0.908	0.856	35.01	4.71	5.70	1000	-	-
Wang et al. [46]†	0.787	0.901	0.817	**17.95**	4.97	9.77	8000	333.00G	32.99M
BL + DS (ours)	0.792	0.912	0.848	29.31	4.41	6.20	400	**55.26G**	**2.51M**
BL + DS + R (ours)	0.792	0.910	0.841	26.42	5.20	9.19	400	75.39G	7.89M
BL + DS + R* + AA (ours)	0.788	0.908	0.845	29.61	5.11	5.92	400	75.46G	7.89M

Best values are shown in bold, and second best are underlined.

TABLE 4. Cross-validation results on BraTS 2021 dataset. ET - Enhancing tumor, WT - Whole tumor, TC - Tumor core.

Model	Dice						
	ET	WT	TC	Avg.	Epochs	FLOPS	Params.
Jia et al. [47]	0.877	0.934	0.910	0.907	250	436.59G	16.85M
Siddiquee et al. [48]	**0.888**	0.935	0.921	0.915	300	-	-
Luu et al. [9]	0.882	**0.938**	**0.924**	0.915	1000	-	-
Liang et al. [49]	0.883	0.908	0.904	**0.926**	600	68.60G	20.40M
BL + DS (ours)	0.885	0.923	0.913	0.907	200	**55.25G**	**2.51M**

Best values are shown in bold, and second best are underlined.

mitigating network degradation [20], we did not observe any meaningful benefits. He et al. [54] observed that applying 1×1 convolution in residual skip connection will result in poor performance, especially when the number of residual units is high. In the future, we will experiment with residual units in the decoding path as in [20].

The benefits of attention mechanism have been studied extensively in natural language processing [55], computer vision [34], [56], [57], [58] as well in medical image segmentation [23], [27], [59]. In this work, we adopted a lightweight Shuffle Attention mechanism [34] to squeeze in extra segmentation performance without introducing noticeable computational costs. Table 1 shows that the attention mechanism in the skip connections significantly improved the Hausdorff distance for the tumor core.

It is evident in Table 2 and Table 3 that our proposed method is both competitive and efficient regarding computation resources. Table 2 shows that our single model nearly outperforms state-of-the-art methods without model ensemble in all metrics. However, as shown in Table 3, the model ensemble is essential to garner extra segmentation performance. Nevertheless, our single model achieved comparable performance using significantly few computational resources. Table 4, our model performed slight poor for both the whole tumor and tumor core regions. However, performance can be boosted by increasing the number of training iterations. Visual inspection in Figure 5 demonstrated that the segmentation quality of our method is high overall. Sometimes, our method fails to segment small regions of enhancing tumors. In the future, we will experiment with an ensemble of lightweight models to improve segmentation performance.

VI. CONCLUSION

This paper proposed some modifications to the nnU-Net framework to reduce computational complexity while maintaining competitive segmentation performance. Specifically, we replaced all convolution blocks with depthwise separable convolutions. We adopted the bottleneck units to minimize the trainable network parameters further. We applied the Shuffle Attention mechanism to the skip connections to improve performance without introducing additional computational costs. Moreover, we utilized residual units to prevent network degradation. Experimental results on BraTS 2020 validate the effectiveness of the proposed method. Our method achieves competitive results while consuming significantly few computational resources.

ACKNOWLEDGMENT

The authors would like to thank CHPC for the provision of state-of-the-art computing infrastructure.

REFERENCES

[1] S. Turcan and D. Cahill, "Origin of gliomas," *Seminars Neurol.*, vol. 38, no. 1, pp. 5–10, Feb. 2018.

[2] *Brain Tumors and Brain Cancer*. Accessed: Mar. 27, 2023. [Online]. Available: https://www.hopkinsmedicine.org/health/conditions-and-diseases/brain-tumor

[3] B. H. Menze et al., "The multimodal brain tumor image segmentation benchmark (BRATS)," *IEEE Trans. Med. Imag.*, vol. 34, no. 10, pp. 1993–2024, Oct. 2015.

[4] A. Işın, C. Direkoğlu, and M. Şah, "Review of MRI-based brain tumor image segmentation using deep learning methods," *Proc. Comput. Sci.*, vol. 102, pp. 317–324, Dec. 2016.
[5] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. W. M. van der Laak, B. van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, Dec. 2017.
[6] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," 2015, *arXiv:1505.04597*.
[7] F. Isensee, P. F. Jäger, P. M. Full, P. Vollmuth, and K. H. Maier-Hein, "nnU-Net for brain tumor segmentation," in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries* (Lecture Notes in Computer Science), A. Crimi and S. Bakas, Eds. Cham, Switzerland: Springer, 2021, pp. 118–132.
[8] F. Isensee, P. F. Jaeger, S. A. A. Kohl, J. Petersen, and K. H. Maier-Hein, "nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation," *Nature Methods*, vol. 18, no. 2, pp. 203–211, Feb. 2021.
[9] H. M. Luu and S.-H. Park, "Extending nn-UNet for brain tumor segmentation," 2021, *arXiv:2112.04653*.
[10] 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 *Medical Image Understanding and Analysis* (Communications in Computer and Information Science), M. V. Hernández and V. González-Castro, Eds. Cham, Switzerland: Springer, 2017, pp. 506–517.
[11] A. Myronenko, "3D MRI brain tumor segmentation using autoencoder regularization," 2018, *arXiv:1810.11654*.
[12] F. Isensee, P. Kickingereder, W. Wick, M. Bendszus, and K. H. Maier-Hein, "Brain tumor segmentation and radiomics survival prediction: Contribution to the BRATS 2017 challenge," 2018, *arXiv:1802.10508*.
[13] H. Li, A. Li, and M. Wang, "A novel end-to-end brain tumor segmentation method using improved fully convolutional networks," *Comput. Biol. Med.*, vol. 108, pp. 150–160, May 2019.
[14] Y.-X. Zhao, Y.-M. Zhang, and C.-L. Liu, "Bag of tricks for 3D MRI brain tumor segmentation," in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries* (Lecture Notes in Computer Science), A. Crimi and S. Bakas, Eds. Cham, Switzerland: Springer, 2020, pp. 210–220.
[15] G. Wang, W. Li, S. Ourselin, and T. Vercauteren, "Automatic brain tumor segmentation using cascaded anisotropic convolutional neural networks," 2017, *arXiv:1709.00382*.
[16] W. Chen, B. Liu, S. Peng, J. Sun, and X. Qiao, "S3D-UNet: Separable 3D U-Net for brain tumor segmentation," in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, vol. 11384, A. Crimi, S. Bakas, H. Kuijf, F. Keyvan, M. Reyes, and T. van Walsum, Eds. Cham, Switzerland: Springer, 2019, pp. 358–368.
[17] G. Wang, W. Li, S. Ourselin, and T. Vercauteren, "Automatic brain tumor segmentation based on cascaded convolutional neural networks with uncertainty estimation," *Frontiers Comput. Neurosci.*, vol. 13, p. 56, Aug. 2019.
[18] C. Zhou, C. Ding, X. Wang, Z. Lu, and D. Tao, "One-pass multi-task networks with cross-task guided attention for brain tumor segmentation," *IEEE Trans. Image Process.*, vol. 29, pp. 4516–4529, 2020.
[19] C. Chen, X. Liu, M. Ding, J. Zheng, and J. Li, "3D dilated multi-fiber network for real-time brain tumor segmentation in MRI," 2019, *arXiv:1904.03355*.
[20] X. Zhou, X. Li, K. Hu, Y. Zhang, Z. Chen, and X. Gao, "ERV-Net: An efficient 3D residual neural network for brain tumor segmentation," *Expert Syst. Appl.*, vol. 170, May 2021, Art. no. 114566.
[21] Z. Luo, Z. Jia, Z. Yuan, and J. Peng, "HDC-Net: Hierarchical decoupled convolution network for brain tumor segmentation," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 3, pp. 737–745, Mar. 2021.
[22] Y. Peng and J. Sun, "The multimodal MRI brain tumor segmentation based on AD-Net," *Biomed. Signal Process. Control*, vol. 80, Feb. 2023, Art. no. 104336.
[23] Y. Liu, X. Du, D.-H. Wang, and S. Zhu, "A lightweight brain tumor segmentation network based on 3D inverted residual modules," in *Proc. 11th Int. Conf. Comput. Pattern Recognit.* New York, NY, USA: Association for Computing Machinery, May 2023, pp. 149–155.
[24] R. Zhang, S. Jia, M. J. Adamu, W. Nie, Q. Li, and T. Wu, "HMNet: Hierarchical multi-scale brain tumor segmentation network," *J. Clin. Med.*, vol. 12, no. 2, p. 538, Jan. 2023.
[25] M. Noori, A. Bahri, and K. Mohammadi, "Attention-guided version of 2D UNet for automatic brain tumor segmentation," in *Proc. 9th Int. Conf. Comput. Knowl. Eng. (ICCKE)*, Oct. 2019, pp. 269–275.
[26] J. Zhang, Z. Jiang, J. Dong, Y. Hou, and B. Liu, "Attention gate ResU-Net for automatic MRI brain tumor segmentation," *IEEE Access*, vol. 8, pp. 58533–58545, 2020.
[27] Y. Cao, W. Zhou, M. Zang, D. An, Y. Feng, and B. Yu, "MBANet: A 3D convolutional neural network with multi-branch attention for brain tumor segmentation from MRI images," *Biomed. Signal Process. Control*, vol. 80, Feb. 2023, Art. no. 104296.
[28] S. Bakas et al., "Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge," 2018, *arXiv:1811.02629*.
[29] S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J. S. Kirby, J. B. Freymann, K. Farahani, and C. Davatzikos, "Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features," *Sci. Data*, vol. 4, no. 1, Sep. 2017, Art. no. 170117.
[30] U. Baid et al., "The RSNA-ASNR-MICCAI BraTS 2021 benchmark on brain tumor segmentation and radiogenomic classification," 2021, *arXiv:2107.02314*.
[31] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: Learning dense volumetric segmentation from sparse annotation," 2016, *arXiv:1606.06650*.
[32] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Instance normalization: The missing ingredient for fast stylization," 2016, *arXiv:1607.08022*.
[33] F. Isensee, P. Kickingereder, W. Wick, M. Bendszus, and K. H. Maier-Hein, "No new-Net," 2018, *arXiv:1809.10483*.
[34] Q.-L. Zhang and Y.-B. Yang, "SA-Net: Shuffle attention for deep convolutional neural networks," 2021, *arXiv:2102.00240*.
[35] F. Wang, M. Jiang, C. Qian, S. Yang, C. Li, H. Zhang, X. Wang, and X. Tang, "Residual attention network for image classification," 2017, *arXiv:1704.06904*.
[36] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional Block Attention Module," in *Computer Vision—ECCV*, vol. 11211, V. Ferrari, M. Hebert, C. Sminchisescu, and Y. Weiss, Eds. Cham, Switzerland: Springer, 2018, pp. 3–19.
[37] Y. Yuan, "Evaluating scale attention network for automatic brain tumor segmentation with large multi-parametric MRI database," in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, vol. 12963, A. Crimi and S. Bakas, Eds. Cham, Switzerland: Springer, 2022, pp. 42–53.
[38] O. Oktay, J. Schlemper, L. L. 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. 1st Conf. Med. Imag. Deep Learn.*, May 2018, pp. 1–10.
[39] L. H. Shehab, O. M. Fahmy, S. M. Gasser, and M. S. El-Mahallawy, "An efficient brain tumor image segmentation based on deep residual networks (ResNets)," *J. King Saud Univ.-Eng. Sci.*, vol. 33, no. 6, pp. 404–412, Sep. 2021.
[40] Y. Yuan, "Automatic brain tumor segmentation with scale attention network," in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries* (Lecture Notes in Computer Science), A. Crimi and S. Bakas, Eds. Cham, Switzerland: Springer, 2021, pp. 285–294.
[41] Y. Wang, Y. Zhang, F. Hou, Y. Liu, J. Tian, C. Zhong, Y. Zhang, and Z. He, "Modality-pairing learning for brain tumor segmentation," in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries* (Lecture Notes in Computer Science), A. Crimi and S. Bakas, Eds. Cham, Switzerland: Springer, 2021, pp. 230–240.
[42] T. Magadza and S. Viriri, "Brain tumor segmentation using partial depthwise separable convolutions," *IEEE Access*, vol. 10, pp. 124206–124216, 2022.
[43] R. Raza, U. I. Bajwa, Y. Mehmood, M. W. Anwar, and M. H. Jamal, "DResU-Net: 3D deep residual U-Net based brain tumor segmentation from multimodal MRI," *Biomed. Signal Process. Control*, vol. 79, Jan. 2023, Art. no. 103861.
[44] L. Daza, C. Gómez, and P. Arbeláez, "Cerberus: A multi-headed network 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, pp. 342–351.

[45] H. Jia, W. Cai, H. Huang, and Y. Xia, "H^2NF-Net for brain tumor segmentation using multimodal MR imaging: 2nd place solution to BraTS challenge 2020 segmentation task," in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries* (Lecture Notes in Computer Science), A. Crimi and S. Bakas, Eds. Cham, Switzerland: Springer, 2021, pp. 58–68.
[46] W. Wang, C. Chen, M. Ding, H. Yu, S. Zha, and J. Li, "TransBTS: Multimodal brain tumor segmentation using transformer," in *Medical Image Computing and Computer Assisted Intervention—MICCAI* (Lecture Notes in Computer Science), M. de Bruijne, P. C. Cattin, S. Cotin, N. Padoy, S. Speidel, Y. Zheng, and C. Essert, Eds. Cham, Switzerland: Springer, 2021, pp. 109–119.
[47] H. Jia, C. Bai, W. Cai, H. Huang, and Y. Xia, "HNF-Netv2 for brain tumor segmentation using multi-modal MR imaging," 2022, *arXiv:2202.05268*.
[48] M. M. R. Siddiquee and A. Myronenko, "Redundancy reduction in semantic segmentation of 3D brain tumor MRIs," 2021, *arXiv:2111.00742*.
[49] J. Liang, C. Yang, and L. Zeng, "3D PSwinBTS: An efficient transformer-based Unet using 3D parallel shifted windows for brain tumor segmentation," *Digit. Signal Process.*, vol. 131, Nov. 2022, Art. no. 103784.
[50] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1800–1807.
[51] N. Ma, X. Zhang, H.-T. Zheng, and J. Sun, "ShuffleNet V2: Practical guidelines for efficient CNN architecture design," in *Computer Vision—ECCV*, vol. 11218, V. Ferrari, M. Hebert, C. Sminchisescu, and Y. Weiss, Eds. Cham, Switzerland: Springer, 2018, pp. 122–138.
[52] D. Zhang, Y. Song, D. Liu, C. Zhang, Y. Wu, H. Wang, F. Zhang, Y. Xia, L. J. O'Donnell, and W. Cai, "Efficient 3D depthwise and separable convolutions with dilation for brain tumor segmentation," in *AI 2019: Advances in Artificial Intelligence* (Lecture Notes in Computer Science), J. Liu and J. Bailey, Eds. Cham, Switzerland: Springer, 2019, pp. 563–573.
[53] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 2015, *arXiv:1512.03385*.
[54] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," 2016, *arXiv:1603.05027*.
[55] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in Neural Information Processing Systems*, vol. 30. Red Hook, NY, USA: Curran Associates, 2017.
[56] J. Fu, J. Liu, H. Tian, Y. Li, Y. Bao, Z. Fang, and H. Lu, "Dual attention network for scene segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 3141–3149.
[57] J. Ho, N. Kalchbrenner, D. Weissenborn, and T. Salimans, "Axial attention in multidimensional transformers," 2019, *arXiv:1912.12180*.
[58] L. Chen, H. Zhang, J. Xiao, L. Nie, J. Shao, W. Liu, and T.-S. Chua, "SCA-CNN: Spatial and channel-wise attention in convolutional networks for image captioning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6298–6306.
[59] 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," 2018, *arXiv:1804.03999*.

TIRIVANGANI MAGADZA received the B.Tech. degree in computer science from the Harare Institute of Technology, Zimbabwe, in 2012, and the M.Tech. degree in computer science from Jawaharlal Nehru Technological University, Hyderabad, India, in 2016. He is currently pursuing the Ph.D. degree in computer science with the School of Mathematics, Statistics, and Computer Science, University of KwaZulu-Natal, Durban, South Africa. His research interests include medical image analysis, computer vision, high-performance computing, wireless sensor networks, and natural language processing.

SERESTINA VIRIRI (Senior Member, IEEE) received the B.Sc. degree in mathematics and computer science and the M.Sc. and Ph.D. degrees in computer science. He is currently a Full Professor of computer science with the School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal, South Africa. He is also a rated Researcher by the National Research Foundation (NRF) of South Africa. Since 1998, he has been with the academia. He has published extensively in several artificial intelligence, and computer vision-related accredited journals and international and national conference proceedings. His main research interests include artificial intelligence, computer vision, image processing, machine learning, medical image analysis, pattern recognition, and other image processing-related fields, such as biometrics, medical imaging, and nuclear medicine. He is a reviewer of several machine learning and computer vision-related journals. He has also served on program committees for numerous international and national conferences.