AUT Journal of Electrical Engineering

AUT J. Elec. Eng., 56(2) (2024) 203-212
DOI: 10.22060/eej.2023.22395.5538

RDCU-Net: A Multi-Scale Residual Dilated Convolution U-Net with Spatial Pyramid Pooling for Brain Tumor Segmentation

M. Soltani-Gol[1], A. Asgharzadeh-Bonab[2]*, H. Soltanian-Zadeh[1], J. Mazloum[3]

[1] Department of Electrical and Computer Engineering, College of Engineering, University of Tehran, Tehran, Iran
[2] Department of Electrical and Computer Engineering, College of Engineering, Urmia University, Urmia, Iran
[3] Department of Electrical Engineering, Shahid Beheshti Aeronautical University of Science and Technology, Tehran, Iran.

ABSTRACT: Tumors refer to abnormal growth of cells in the body. Early diagnosis of tumors plays a crucial role in improving treatment conditions , quality of life and patient survival. Deep learning methods are effective for medical image segmentation, but they struggle with tumors in magnetic resonance images (MRI) due to variations in intensity and appearance. Existing models like U-Net face challenges due to the integration of high-level and low-level features, leading to confusion. Our proposed model addresses the above issues by utilizing two techniques and fewer parameters compared to the existing methods, achieving higher accuracy. In the first technique, dilated convolution (DC) blocks with proportional rates are used to integrate high-level and low-level features. The second technique involves selecting dilated spatial pyramid (DSP) blocks, which increase the receptive field of features while maintaining their resolution, contributing to the network's generalization. The proposed model improves training, network depth, and feature extraction by incorporating a residual block. It outperforms the traditional U-Net model in terms of segmentation accuracy and network stability. We evaluated the model using the BraTS 2018 dataset, obtaining Dice coefficients of 0.906, 0.817, and 0.839 for the whole tumor (WT), the enhancing tumor (ET), and the tumor core (TC), respectively.

Review History:

Received: May, 08, 2023
Revised: Sep. 08, 2023
Accepted: Dec. 19, 2023
Available Online: Mar. 01, 2024

Keywords:

image segmentation
deep convolutional neural network
magnetic resonance imaging (MRI)
brain tumor

1- Introduction

Brain tumors are known to be one of the most dangerous types of tumors all around the world. Glioma, which is the most common primary brain tumor, arises from the cancerous growth of glial cells in the spinal cord and brain. Different degrees of tissue and malignancy exist in gliomas, and the average survival time for glioblastoma patients has been determined to be less than 14 months after diagnosis [1]. Magnetic Resonance Imaging (MRI) is a non-invasive and popular method for diagnosing brain tumors. This method helps medical professionals in the detection, sizing, and positioning of brain tumors by producing high-quality images and using various tissue contrasts [2]. Although MRI images are very useful as a diagnostic strategy for brain tumors, manual segmentation and structural analysis of brain tumors on MRI are difficult and time-consuming and can only be performed by experienced radiologists [3].

Despite its existence for several decades, artificial intelligence (AI) has recently gained significant attention, particularly in the field of medical imaging, due to the introduction of deep learning algorithms and a focus on multi-modality imaging [4]. Ronneberger et al. [5] presented a paper on medical image segmentation titled "U-Net" network in 2015. The paper introduced a new architecture for medical image segmentation, which could provide accurate segmentation of images using a small number of training images. The purpose of this architecture was to design a multi-layer neural network that could perform medical image segmentation with high accuracy and without the need for a large number of training images. In 2016, Pereira et al. [6] successfully increased the accuracy of brain tumor segmentation by designing a fully convolutional architecture using three stages: data normalization, data augmentation, and N4ITK filter for correcting MRI bias field.

Milletari et al. [7] introduced the V-Net architecture in 2016, based on a three-dimensional structure and U-Net for architecture design. In this article, a number of steps were used to reduce the dimensions using three-dimensional kernels to slide over the input tensors. In 2017, Mohammad Havaei et al. [8] presented a paper titled "Cascade Architecture with Two Pathways" based on a two-dimensional structure for feature extraction. In this method, local features are extracted using 7x7 kernels, and global features are extracted using 13x13 kernels. By combining all features, the desired output is obtained. In 2020, Yang and colleagues [9] introduced a new architecture for neural networks that combines U-Net and Residual block. This article addressed the problem of gradient disappearance in deep networks using the Residual block. It proposed a network with improved performance by extracting more features in the deeper layers. In 2019, Nouri et al. [10] combined the U-Net architecture with the Attention-Guide for brain tumor segmentation. This architecture, using the

*Corresponding author's email: a.asgharzadeh@urmia.ac.ir

attention mechanism block in the two-dimensional network, resolved the model's confusion for high-grade (HGG) and low-grade (LGG) data. In this article, the BraTS 2017-2018 dataset was used for the network's training.

Another architecture called DRAU-Net (Double Residual Attention U-Net) has also been proposed with a new structure in [11]. In this article, a combination of U-Net structures, Residual Double blocks, and attention mechanisms have been used to address the issues associated with the above networks. In convolutional networks, the use of Max-Pooling layers leads to a reduction in the resolution of extracted features. This can limit the accuracy of the model, and as the network depth increases for feature extraction, the learning process of the model becomes more complex. In 2015, the dilated convolutional structure was introduced by Yu et al. Their studies indicate that the dilated convolutional operator is suitable due to its ability to expand the receptive field while preserving the resolution of features [12]. Then a hypothesis was proposed that by changing the network structures to Dilated Residual Networks (DRNs), the effects of altering the receptive field could be reduced, leading to an improvement in model accuracy. These changes, including increasing the receptive field through the application of dilated convolutions along with preserving skip connections in deep networks, can lead to a remarkable improvement in generalization and the output of deep networks.

This network has been able to prevent model confusion without reducing resolution in deeper layers, and in the task of image segmentation, it has shown better performance compared to regular convolutional networks [13]. By making minor changes to the structure of the DRNs network, Luper et al. were able to evaluate the network for the first time using the Brats 2017 dataset. They believed that increasing the Receptive Field through dilated convolutions, as well as preserving residual connections, could enhance the generalization of deep network outputs for volumetric medical images [14]. Chen et al. managed to reduce computational costs using 3D multi-fiber units with fewer parameters. They employed dilated 3D convolutions with a stride of 2 and achieved results by performing dimension reduction of extracted features without utilizing Max-Pooling layers. In the end, they utilized dilated convolutions to enlarge the relevant field and understand the multi-scale spatial correlations of brain tumor lesions in 3D [15]. In 2020, a new architecture named DCU-Net was introduced by Yang et al. In this study, the method of dilated convolution residual block (RD-Skip) was employed to expand the receptive field of extracted features in the down-sampling path. This approach had the capability to enhance the fusion of high-level and low-level features in both down-sampling and up-sampling paths. Additionally, in another approach, Dilated Spatial Pyramid Pooling (DSPP) blocks were used as substitutes for Max-Pooling to preserve image clarity and extracted features. These modifications led to increased accuracy in target detection, particularly in identifying smaller tumors [16]. We propose a new architecture called RDCU-Net (Residual Dilated Convolution U-Net) for brain tumor segmentation using expanded convolution structures and Residual blocks. DC (Dilated Convolution) and DSP (Dilated Spatial Pyramid) blocks are used to integrate features and increase the resolution of the extracted features. The main contribution of our study compared to existing methods based on U-Net networks and dilated convolutions is the use of residual blocks. The use of residual blocks is employed to increase the network depth and extract meaningful features relevant to tumor detection. This method facilitates network training and enhances model accuracy with increasing depth. This results in the transfer of meaningful features from the input blocks to deeper layers of the network, enabling better generalization from the network's output. With residual learning, these blocks experience an improvement in the gradient vanishing problem, which means a more optimal transfer of gradients and information from one layer to the next. This leads to improved conclusions during training and an increase in model accuracy. Additionally, increasing the network depth, due to the absence of the vanishing gradient problem, enables the extraction of more complex and abstract features from tumor images. In conclusion, focusing on increasing the generalizability of the model is very important for the accurate diagnosis of tumors and allows the model to segment tumor regions accurately.

2- Materials and Methods

2- 1- BraTS 2018 Dataset

Considering the importance of brain tumors and their potential unavoidable consequences, providing open-source datasets to facilitate research in this area is very important. In this regard, the BraTS 2018 dataset is known as a valuable dataset. This dataset includes 210 three-dimensional MRI of patients with high-grade gliomas (HGG) and 75 MRI of low-grade gliomas (LGG). Clinical MRI of patients with varying degrees of tissue heterogeneity are included in this database. These data include four types of imaging sequences, FLAIR, T1, T1-contrast with gadolinium, and T2. Each three-dimensional MRI has dimensions of 155×240×240 and can be viewed from three planes: axial, coronal, and sagittal. All labels in this segmentation, including background, necrotic tissue, non-enhancing tumor, edema around the tumor, and enhancing tumor, have been classified by experienced radiologists. In this study, the Brats 2018 database (https://www.med.upenn.edu/sbia/brats2018/data.html) has been used for network training and evaluation. Additionally, for quantitative evaluation of results (such as Dice score and Hausdorff), it is necessary to upload the proposed network results to the BraTS 2018 platform online and evaluate them, as the validation set labels in the BraTS dataset are not available.

2- 2- Pre-Processing

Intensity inhomogeneity in MRI has a negative impact on the results of image analysis. In this study, the N4ITK algorithm [17] has been used to correct image inhomogeneity, which uses a multi-scale optimization method. In the first layer of the network, the data is normalized and the outliers

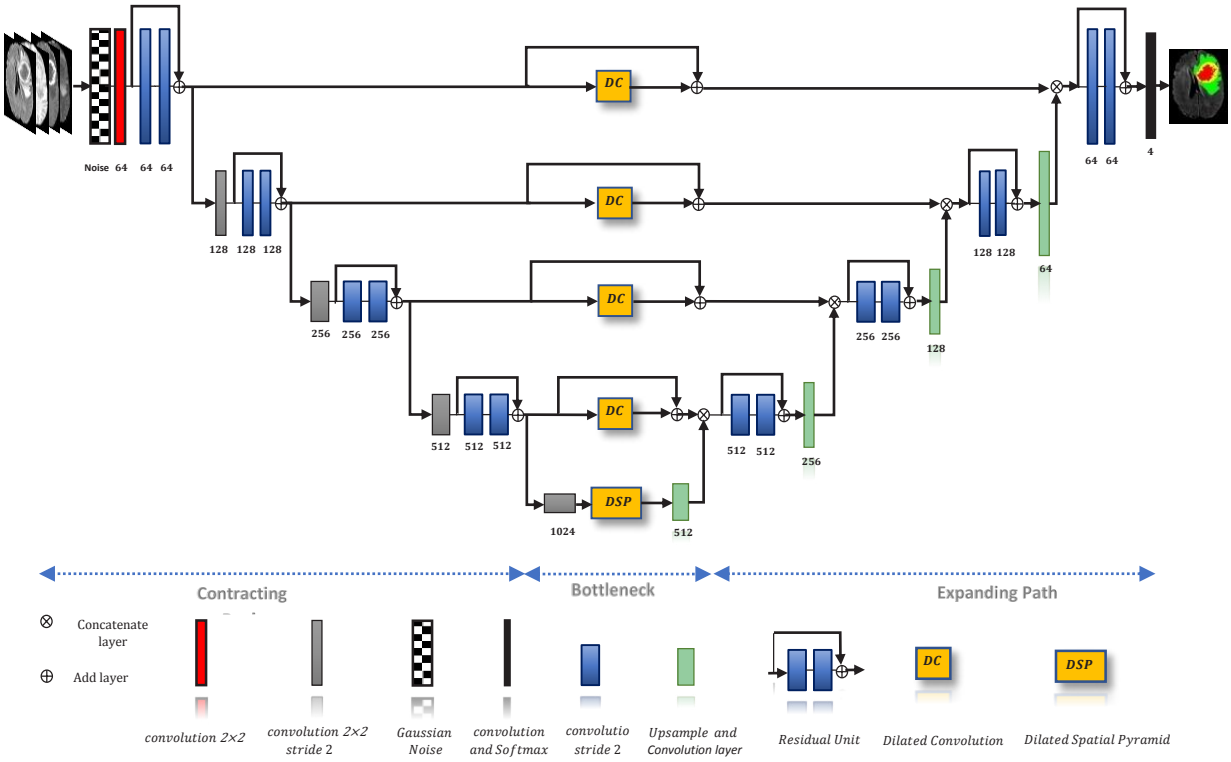

Fig. 1. The architecture of the proposed RDCU network.

are removed to improve the learning process. Since a three-dimensional human brain volume includes 155 × 240 × 240 voxels, to reduce the computational load of the network, the input data is divided into sections. Furthermore, the lack of valuable information in the margins of the images allows its dimensions to be reduced to 192 × 192 pixels. Ultimately, 70% of the 2D images are used for the training process, 20% for validation, and 10% for testing.

2- 3- Proposed Network

Our proposed network is based on the U-Net architecture [5], Residual blocks, and some changes in layer connections, as shown in Fig. 1. Like the original U-Net, the proposed network has two main parts: 1- Encoding part: input images are encoded using convolutional layers. 2- Decoding part: the output image is recovered using deconvolutional layers. In the encoding path, four convolutional layers with stride 2 are responsible for reducing the dimension of the extracted features. The original U-Net used Max-Pooling layers are used for this purpose. In addition, Residual blocks [18] are used after each convolutional layer in the encoding path to help accelerate the training and convergence of the network in deeper layers. Each Residual block consists of BN layers, a PReLU activation function, and a 3×3 convolution, as shown in Fig. 2. Deepening the network by adding repeated convolutional layers is not recommended, as this operation can cause the problem of gradient vanishing in the layers, leading to saturation of accuracy and its decrease. Therefore, it is necessary to use residual blocks to solve these problems. Additionally, after each block, a dropout layer is added to prevent overfitting. In the decoding path, we include four residual blocks with the same structure. This path also has four Up-Sampling layers, each of which doubles the dimensions of the extracted features. In the last layer of the network, the softmax non-linear activation function is used to convert the features to the probability value of predicting the tumor class. Each path connection in the original U-Net is formed by a skip connection .[19] In our proposed network, this is connected by DC blocks shown in Fig. 1. The DC block represents the use of dilated convolutions, which combine the extracted features in the previous layers and combine them with the Up-Sampling output. In the Bottleneck path, the DSP block is used to express more accurate features for the presence of extracted features in the previous layers. Finally, this block can identify more brain tumor features. DC and DSP blocks will be described in detail below. The advantages and disadvantages of these methods are described as follows:

- **Advantages:** Increasing detection accuracy, multi-scale receptive field enhancement, preservation of resolution clarity, parameter reduction, improved detail resolution, and capability to understand complex patterns and features.
- **Disadvantages:** Sensitivity to selected rates and significant time cost in network design for selecting rates.

In the forward path, a combination of binary cross-entropy loss and Dice loss functions are used to increase network convergence, while in the backward path, the Adam

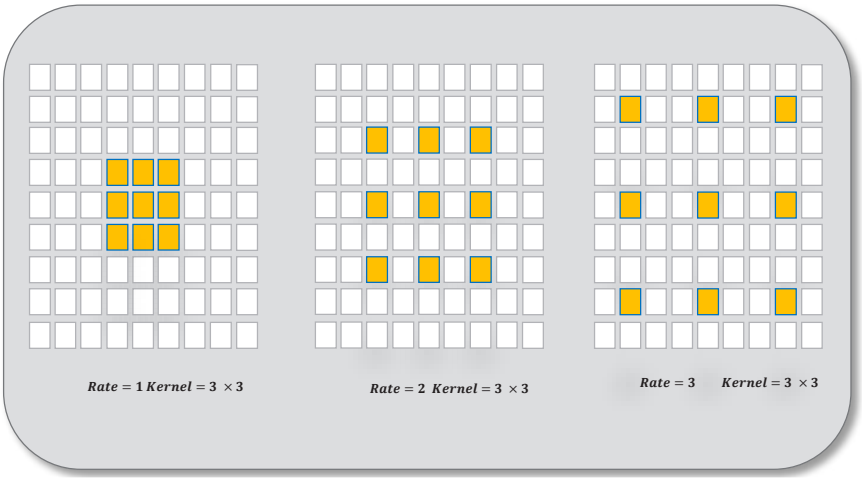

Fig. 2. Dilated convolutions with different rates.

optimizer with a learning rate of 1e-5 is adopted. Finally, the last output layer (Softmax) is responsible for transforming the features into the probability values for predicting the classes [11].

2- 3- 1- Dilated Convolution (DC) Block

In the past, dilated convolution was referred to as "convolution with dilated filter". This operator played a crucial role in the a'trous algorithm for wavelet decomposition and analysis [20]. In deep neural networks, Max-Pooling layers and stride-in filters are commonly used for subsampling the input images and reducing the dimensionality of the extracted features. However, this operation is not suitable for accurately expressing the features, and it leads to a decrease in the resolution of the features in each layer. Ultimately, by subsampling the same features, the dimensions of the original image are restored, which results in reduced accuracy in output segmentation. For this reason, in 2015, this problem was resolved using dilated convolution [12]. The fundamental difference between dilated convolution and regular convolution is the use of an additional parameter for determining the spacing of the kernel in dilated convolution, as shown in Fig. 2. Unlike Max-Pooling layers and stride in regular convolution that reduce the dimensions of the features, dilated convolution expands over the input image by selecting a dilation rate. In this operation, empty positions are always filled with zeros. Finally, a sliding operation is performed, and the extracted features are obtained. The size of the features is determined based on the input tensors using:

$$L_{l+1} = \frac{L_1 + 2p - k - (k-1)(r-1)}{S^*} + 1 \qquad (1)$$

where L_{l+1} represents the output of the extracted features, L_1 is the input tensor, p is the padding, k is the kernel size, r is the dilation rate, and S^* is the stride in dilated convolution.

In the original U-Net network, all convolutions are performed in the usual way. The high-level and low-level features are combined in a classic way, which confuses the model during the training process. This also leads to neglecting details in the extracted features, which can reduce the accuracy of tumor segmentation. In this paper, instead of using the direct method in the U-Net network, DC blocks with Skip Connections are used to eliminate the problem of direct integration of high-level and low-level features. These blocks preserve the meaningful feature information to a very large extent and help to generalize the network more accurately for segmentation output. The details of the convolution layers, along with the expansion rate and the number of filters in the DC blocks, are presented in Table 1.

After applying the first filter with dimensions of 64, the extracted features are generated with the largest dimensions. However, in dilated convolutions, we will use a Dilated Rate with a maximum value equal to 16. To explain further, the higher the number of dimensionality reduction layers in the encoder path, the higher the rate naturally is, and this increase is determined by the ratio of 2 to the power of *n*. For example, if we reduce the dimension three times in the encoder path, its maximum value is equal to 8. Another reason is to get meaningful information and wider features from the input image after the first filter. In the second step, as the number of filters increases from 64 to 128, there is another dimension reduction step. For this reason, a dilation rate of 8 is also used for this stage. When we reach a lower level of extracted features, the selection of the rate will become 2 [12]. Previous work [12]-[16] has shown that the dilation rate is adaptable

Table 1. DC block specifications.

Name of Block	Operation	Kernel size	Number of Kernel	Dilation rate
DC 1	Conv + PReLU	1 × 1	64	16
DC 2	Conv + PReLU	1 × 1	128	8
DC 3	Conv + PReLU	1 × 1	256	4
DC 4	Conv + PReLU	1 × 1	512	2

Table 2. Specifications of the layers in a DSP block.

Layer Name	Operation	Kernel size	Number of Kernel	Dilation rate
BN	Batch Normalization	--	--	--
Conv1	Conv 2D	3 × 3	1024	2
Conv2	Conv 2D	3 × 3	1024	4
Conv3	Conv 2D	3 × 3	1024	8
Conv4	Conv 2D	3 × 3	1024	16

and compatible. Moreover, repeated experiments during the training phase also validate this choice.

2- 3- 2- Dilated Spatial Pyramid (DSP) Block

In the classic U-Net model, Max-pooling layers serve two main purposes. In the first stage, by reducing the computations, the increase in the complexity of the network is prevented, and in the second stage, by enlarging the receptive field, the ability to detect features is increased. However, after each use of these

layers, the resolution of the extracted features decreases, resulting in the loss of Surface features. In this paper, we use the DSP block to simultaneously preserve the resolution of the extracted features and increase the feature receptive field. Fig. 3 shows the DSP block. Firstly, the previously extracted features are processed using the BN operation. Then, we slide over the previously extracted features using four convolutional blocks with expanded rates of 2, 4, 8, and 16, and finally, the information on the features is combined. By using the DSP block in the bottleneck, we can evaluate the significant feature information with higher accuracy than the classic U-Net and increase the accuracy of brain tumor segmentation. The specifications of the DSP block layers are listed in Table 2.

As explained in section 2-3-1, the use of DC blocks in the network is one of its strengths**.** In 2015, the U-Net network was introduced for the purpose of segmenting medical images. One of the weaknesses of the classic U-Net network is its direct connections for fusing high-level and low-level features, which can confuse the model. To address this issue, we can employ two techniques: 1. Attention Mechanism; and 2. DC blocks.

Attention Mechanism: The attention mechanism allows neural networks to focus on specific features of the data at each processing stage. This focus can be achieved by assigning different weights to features, reflecting their significance. In other words, by assigning larger weights to more important features, attention to detail becomes evident. For this reason, by integrating these features at different levels and using the focus of weights on each feature at each level, the problem of direct integration and confusion in the model is solved.

DC Blocks: These blocks not only prevent the direct integration of features between different levels but also enable the detection and understanding of broader patterns and features within the data. Here, dilation increases the size of the receptive field. By increasing the size of the receptive field, the network becomes capable of understanding larger and more complex patterns within the data. This feature is especially useful in tasks such as recognizing complex patterns in medical images that require understanding of the information at different levels. Finally, in addition to the superiority of this block over the attention mechanism, the use of residual learning can be added to it, which is done by skip connection.

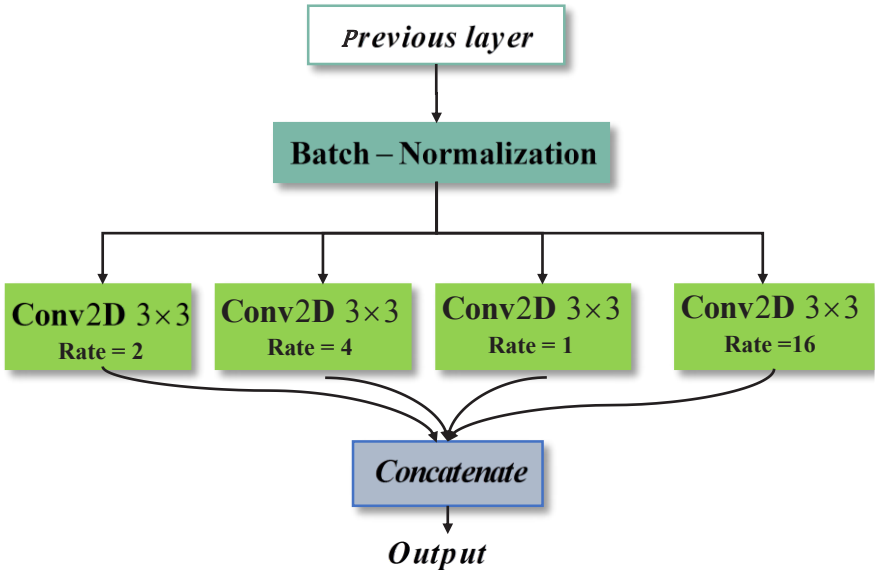

Fig. 3. Illustration of a DSP block.

Dilated Spatial Pyramid (DSP) Blocks: The two main features of these blocks are increasing the receptive field and reducing the calculations in the designed network. These blocks with different rates are good alternatives to Max-Pooling layers, which reduce the resolution of the extracted features and are not suitable for representing detailed features. By utilizing DSP blocks, it is possible to simultaneously increase the receptive field and preserve the clarity and resolution of the extracted features.

2- 3- 3- Residual Block

The main task of residual blocks is to solve the problem of the vanishing gradient and the reduction of depth in deep neural networks during the training process. Usually, in convolutional neural networks, increasing the number of layers leads to the vanishing gradient problem, where the gradient values become very small towards the initial layers, which results in less effect in the training process. Residual blocks, with their specific architecture, solve this problem and allow the network to achieve greater depth. To improve the performance of the network, Batch Normalization and PReLU layers are placed inside the block. First, the Batch Normalization layer is applied to the extracted features, and then the PReLU activation function with its trainable parameters transfers the extracted features to the next layer. Finally, a convolution layer with pre-defined initial values performs the processing. This operation is repeated once again inside each Residual block to improve the extracted features. Residual blocks can increase the network's generalization power and stability in the training process, and help with the network's learning. These blocks allow networks to extract complex and abstract features from the input data. These blocks extract details and complex patterns that play important roles in the diagnosis of brain tumors. This is particularly evident in complex tasks such as the segmentation of tumor regions. By deepening the network and enabling the extraction of more meaningful features, the model gains a better ability to differentiate between various regions and objects in the input image. This leads to a significant improvement in detection accuracy and enables better recognition of different objects and regions. These blocks have skip connections that act as direct paths for transferring gradients through network layers. This feature helps to reduce the vanishing gradient problem, which may cause instability in the training process. By effectively transferring gradient information, these blocks generally contribute to the robustness and stability of the network during the learning stages.

2- 3- 4- Loss Function

In order to perform pixel-wise segmentation in the proposed network, a combined loss function has been used. The use of an optimal loss function is crucial, as an incorrect choice can have negative effects on the output segmentation. In this regard, the combination of the Dice and binary cross-entropy loss functions has been employed to obtain accurate results. The Dice loss function [21] has the adaptive ability to balance classes, and the binary cross entropy loss function [22] improves convergence speed.

$$Loss_{Dice} = 1 - \frac{1}{N}\sum_{n=1}^{N}\sum_{m=1}^{M}\frac{2y_{t(n,m)}y_{p(n,m)}}{y_{t(n,m)} + y_{p(n,m)} + \varepsilon} \quad (2)$$

Table 3. Comparison of the proposed model with the U-net Model using training data.

Networks	Dice		
	ET	WT	TC
U-Net	0.741	0.860	0.783
Propose Network	0.824	0.913	0.846

$$Loss_{BCE} = -\frac{1}{N}\sum_{n=1}^{N}\sum_{m=1}^{M} y_{t(n,m)} \log\left(y_{p(n,m)}\right) \quad (3)$$

N and M respectively correspond to the number of samples (pixels) and the number of classes in each image. $y_{t(n,m)}$ represents the actual label value, while $y_{p(n,m)}$ indicates the predicted value by the model for sample n and class m. To prevent issues of undefined in the denominator of the equation, a small positive value is considered with a variable named ε. Basically, the formula of the Dice Loss cost function measures the model's error by comparing the actual and predicted values for each class in each sample. This cost function aims to measure the similarity between the actual image and the predicted one and encourages the model to correctly identify different regions of the image. The goal is to minimize the cost function so that the model improves. To this end, using formulas (2)-(3), the final loss function is obtained. To prevent the denominator from becoming zero, ε is used, and the constant value α is determined to be 0.5 through empirical calculation and multiple attempts. Finally, the total loss is calculated as:

$$Loss_{T} = Loss_{BCE} + \alpha\, Loss_{Dice} \quad (4)$$

2- 3- 5- Evaluation Metrics

According to brain tumor segmentation papers, two standard criteria, namely Dice Score and Harsdorf95, have been used. The results related to the Dice Score, based on the model evaluation process using the training data with available labels, are presented in Table 3 (Note: these data were not used during the training phase). Also, the results of both Dice Score and Harsdorf 95 criteria are calculated in the online platform using validation data whose labels were not available to us. Then, after comparing with the original labels, the results were announced to us and we reported them in Table 4.

Dice score: as shown in formula (5), it evaluates the degree of similarity between the model's outputs and the true labels. The Dice score indicates how well the brain tumor segmentation results match the true tumor mask.

$$DSC = \frac{2|X \cap Y|}{|X| + |Y|} \quad (5)$$

where |X| and |Y| are the cardinalities of X and Y, respectively.

Hausdorff Distance: Formula (6) describes how close the boundary of a segmented region is to the boundary of the ground truth region.

$$\mathrm{D}(\mathrm{X},\mathrm{Y}) = \max\left\{\sup_{\mathrm{x}\in\mathrm{X}}\mathrm{D}(\mathrm{x},\mathrm{Y}), \sup_{\mathrm{y}\in\mathrm{Y}}\mathrm{D}(\mathrm{X},\mathrm{y})\right\} \quad (6)$$

where $D(x,Y) = inf_{y\in Y} D(x,y)$, sup is the supremum, and inf is the infimum. Also, it is notable that D is the Euclidian distance between a point x ∈ X and a subset Y.

3- Results

In this study, all experiments and results were implemented on Google Colaboratory. In the proposed network, 147 HGG and 52 LGG data (which are equivalent to 70% of the Brats 2018 dataset) were assigned to the training phase, and the remaining data were used in the validation phase. It should be noted that during the training phase, the model evaluation process was achieved based on 5-fold cross-validation. The results of our proposed model evaluation are presented in the tables and images below. As seen in Table 3, DC blocks, DSP blocks, and residual connections have significantly improved the network's generalization compared to the classic U-Net. The segmentation results in Table 3 have been calculated using 10% of the training data, which has been separated for evaluation.

To evaluate the quality and visual performance of the network, we show an axial image with all four modalities (Flair, T1, T1ce, and T2). These images are presented in the first to fourth columns. The fifth column in Fig. 4 is dedicated to the masks, and the sixth column displays the predicted images by the network.

We evaluate our model using the BraTS 2018 evaluation dataset, which consists of 66 volumetric brain images (without access to labels). After performing label predictions, we consider them as a complete brain volume. This complete volume includes all three types of labels. The dimensions

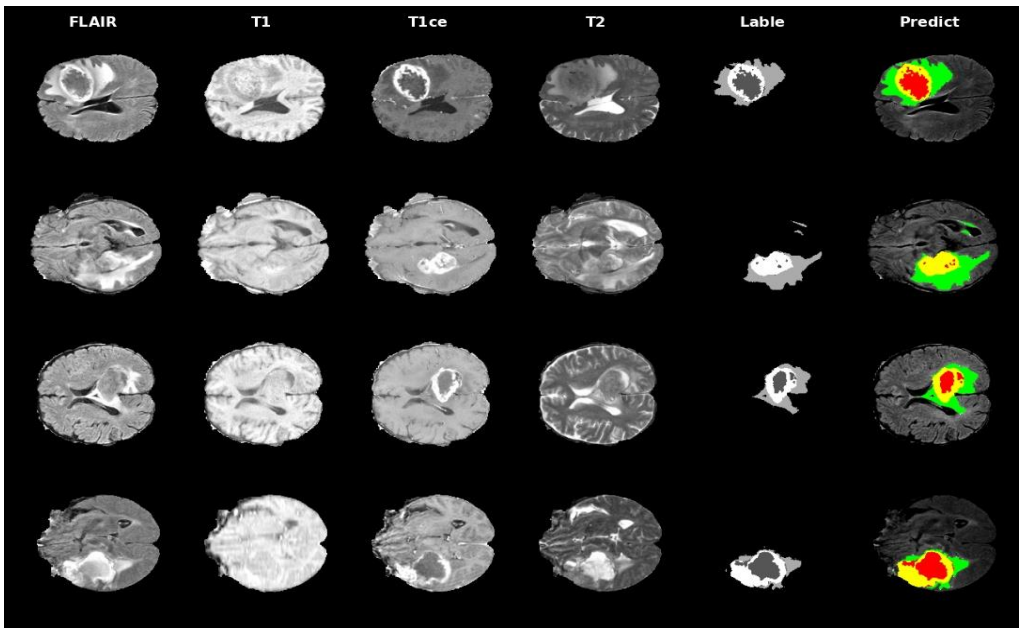

Fig. 4. The 5-fold cross-validation results for axial views on a multi-model MRI dataset.

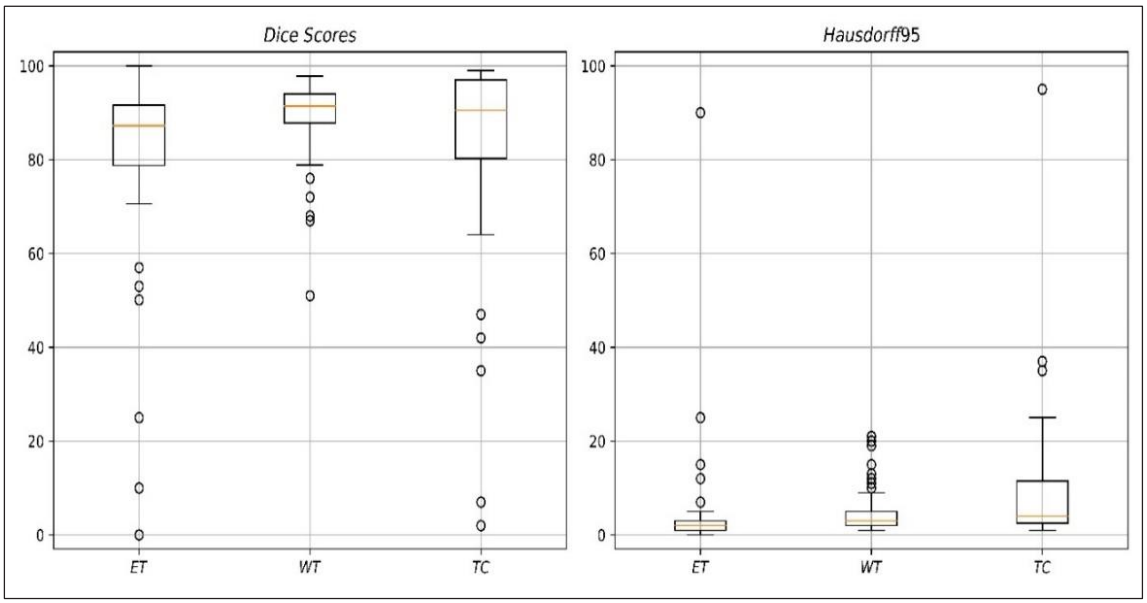

Fig. 5. Box plots of the 5-fold cross-validation results on axial views. ET is the enhancing tumor, WT is the whole tumor, and TC is the tumor core.

of this volume should match the original label dimensions. Finally, the 66 predicted labels are uploaded to the online platform of the University of Pennsylvania at the address (https://ipp.cbica.upenn.edu/). After performing a comparison with the ground truth labels, the final results are announced. The evaluation results are shown in the form of a box plot in Fig. 5. Furthermore, these results demonstrate that the network performs successfully in segmenting input MRI with high accuracy.

3- 1- Comparison with Existing Methods

In this section, our model is evaluated using the test dataset of Brats 2018 and is compared with metrics such as Dice and Hausdorff distance. In Table 4, a comparison between our

Table 4. Comparison of our method with other methods using BRAST 2018 validation set.

Methods	Type	Dice			Harsdorf95		
		ET	WT	TC	ET	WT	TC
Zhang et al. [23]	2D	0.772	0.872	0.808	6.04	5.62	8.36
Noori et al. [10]	2D	0.813	0.895	0.823	2.93	4.05	6.34
Havaei et al. [8]	2D	0.730	0.880	0.790	-	-	-
Vijay et al. [24]	3D	0.837	0.895	0.872	7.13	34.1	7.97
Jiang et al. [25]	2D	0.833	0.888	0.837	2.65	4.61	4.13
Propose Network	2D	0.817	0.906	0.839	3.31	3.98	6.11

study and existing research and advanced methods is provided. This table illustrates the evaluation of the predicted labels which have been computed by the online platform. In this comparison, our study exhibits its maximum and minimum values according to the Dice Score and Hausdorff95 metrics for the WT (Whole Tumor) category. The minimum value is based on the maximum distance between the predicted label and the ground truth label, while the maximum value indicates their highest similarity.

4- Conclusion

In this article, a deep learning-based method called RDCU-Net is proposed for segmenting three regions of brain tumors. We believe that this method can overcome issues such as direct feature fusion, low accuracy of segmentation at region boundaries, and reduced resolution compared to classic U-Net. The high-level and low-level feature fusion in this method is done by the DC block. Additionally, due to the presence of dilated convolutions in each pathway, the receptive field is expanded, and the network's confusion problem is resolved. The DSP block is a suitable replacement for Max-pooling, despite having dilated convolutions with different rates. The DSP block has two important features. First, it can expand the feature receptive fields, and second, it can preserve the resolution (clarity) of extracted features to a great extent. This is why

it leads to more accurate detection of smaller targets Finally, to deepen the network for extracting meaningful features and transferring them to the network's depth for ease of training and network stability, Residual blocks have also been used.

References

[1] E. G. Van Meir, C. G. Hadjipanayis, A. D. Norden, H. K. Shu, P. Y. Wen, and J. J. Olson, "Exciting new advances in neuro-oncology: the avenue to a cure for malignant glioma," CA: a cancer journal for clinicians, vol. 60, no. 3, pp. 166-193, 2010.

[2] S. Bakas et al., "Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features," Scientific data, vol. 4, no. 1, pp. 1-13, 2017.

[3] A. Khosravanian, M. Rahmanimanesh, P. Keshavarzi, and S. Mozaffari, "Fast level set method for glioma brain tumor segmentation based on Superpixel fuzzy clustering and lattice Boltzmann method," Computer Methods and Programs in Biomedicine, vol. 198, p. 105809, 2021.

[4] Y. Sim et al., "Deep convolutional neural network–based software improves radiologist detection of malignant lung nodules on chest radiographs," Radiology, vol. 294, no. 1, pp. 199-209, 2020.

[5] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18, 2015: Springer, pp. 234-241.

[6] S. Pereira, A. Pinto, V. Alves, and C. A. Silva, "Brain tumor segmentation using convolutional neural networks in MRI images," IEEE transactions on medical imaging, vol. 35, no. 5, pp. 1240-1251, 2016.

[7] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in 2016 fourth international conference on 3D vision (3DV), 2016: Ieee, pp. 565-571.

[8] M. Havaei et al., "Brain tumor segmentation with deep neural networks," Medical image analysis, vol. 35, pp. 18-31, 2017.

[9] T. Yang, J. Song, L. Li, and Q. Tang, "Improving brain tumor segmentation on MRI based on the deep U-net and residual units," Journal of X-ray Science and Technology, vol. 28, no. 1, pp. 95-110, 2020.

[10] M. Noori, A. Bahri, and K. Mohammadi, "Attention-guided version of 2D UNet for automatic brain tumor

segmentation," in 2019 9th international conference on computer and knowledge engineering (ICCKE), 2019: IEEE, pp. 269-275.

[11] M. Soltani-Gol, M. Fattahi, H. Soltanian-Zadeh, and S. Sheikhaei, "DRAU-Net: Double Residual Attention Mechanism for automatic MRI brain tumor segmentation," in 2022 30th International Conference on Electrical Engineering (ICEE), 2022: IEEE, pp. 587-591.

[12] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," arXiv preprint arXiv:1511.07122, 2015.

[13] F. Yu, V. Koltun, and T. Funkhouser, "Dilated residual networks," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 472-480.

[14] M. Moreno Lopez and J. Ventura, "Dilated convolutions for brain tumor segmentation in MRI scans," in Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: Third International Workshop, BrainLes 2017, Held in Conjunction with MICCAI 2017, Quebec City, QC, Canada, September 14, 2017, Revised Selected Papers 3, 2018: Springer, pp. 253-262.

[15] C. Chen, X. Liu, M. Ding, J. Zheng, and J. Li, "3D dilated multi-fiber network for real-time brain tumor segmentation in MRI," in Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part III 22, 2019: Springer, pp. 184-192.

[16] T. Yang, Y. Zhou, L. Li, and C. Zhu, "DCU-Net: Multi-scale U-Net for brain tumor segmentation," Journal of X-ray Science and Technology, vol. 28, no. 4, pp. 709-726, 2020.

[17] N. Tustison and J. Gee, "N4ITK: Nick's N3 ITK implementation for MRI bias field correction," Insight Journal, vol. 9, 2009.

[18] S. Wu, S. Zhong, and Y. Liu, "Deep residual learning for image steganalysis," Multimedia tools and applications, vol. 77, pp. 10437-10453, 2018.

[19] O. Oktay et al., "Attention u-net: Learning where to look for the pancreas," arXiv preprint arXiv:1804.03999, 2018.

[20] M. Holschneider, R. Kronland-Martinet, J. Morlet, and P. Tchamitchian, "A real-time algorithm for signal analysis with the help of the wavelet transform," in Wavelets: Time-Frequency Methods and Phase Space Proceedings of the International Conference, Marseille, France, December 14–18, 1987, 1990: Springer, pp. 286-297.

[21] C. H. Sudre, W. Li, T. Vercauteren, S. Ourselin, and M. Jorge Cardoso, "Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations," in Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: Third International Workshop, DLMIA 2017, and 7th International Workshop, ML-CDS 2017, Held in Conjunction with MICCAI 2017, Québec City, QC, Canada, September 14, Proceedings 3, 2017: Springer, pp. 240-248.

[22] U. Ruby and V. Yendapalli, "Binary cross entropy with deep learning technique for image classification," Int. J. Adv. Trends Comput. Sci. Eng, vol. 9, no. 10, 2020.

[23] 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.

[24] S. Vijay, T. Guhan, K. Srinivasan, P. Vincent, and C.-Y. Chang, "MRI brain tumor segmentation using residual Spatial Pyramid Pooling-powered 3D U-Net," Frontiers in public health, vol. 11, p. 1091850, 2023.

[25] Z. Jiang, C. Ding, M. Liu, and D. Tao, "Two-stage cascaded u-net: 1st place solution to brats challenge 2019 segmentation task," in Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 5th International Workshop, BrainLes 2019, Held in Conjunction with MICCAI 2019, Shenzhen, China, October 17, 2019, Revised Selected Papers, Part I 5, 2020: Springer, pp. 231-241.

HOW TO CITE THIS ARTICLE

M. Soltani-Gol, A. Asgharzadeh-Bonab, H. Soltanian-Zadeh, J. Mazloum. RDCU-Net: A Multi-Scale Residual Dilated Convolution U-Net with Spatial Pyramid Pooling for Brain Tumor Segmentation. AUT J. Elec. Eng., 56(2) (2024) 203-212.

DOI: 10.22060/eej.2023.22395.5538