Efficient Brain Tumor Segmentation with Dilated Multi-fiber Network and Weighted Bi-directional Feature Pyramid Network

Thanh Hau Nguyen[1,3], Cong Hau Le[1], Dinh Viet Sang[1], Tingting Yao[2], Wei Li[3], Zhiyong Wang[3]

[1]School of Information and Communications Technology, Hanoi University of Science and Technology Hanoi, Vietnam
{*thanhhau097, leconghau.hit*}*@gmail.com, sangdv@soict.hust.edu.vn*
[2]College of Information Science and Technology, Dalian Maritime University, Liaoning, China
ytt1030@dlmu.edu.cn
[3]School of Computer Science, The University of Sydney, Sydney, Australia
{*weiwilson.li, zhiyong.wang*}*@sydney.edu.au*

***Abstract*—Brain tumor segmentation is critical for precise diagnosis and personalised treatment of brain cancer. Due to the recent success of deep learning, many deep learning based segmentation methods have been developed. However, most of them are computationally expensive due to complicated network architectures. Recently, multi-fiber networks were proposed to reduce the number of network parameters in U-Net based brain tumor segmentation through efficient graph convolution. However, the efficient use of multi-scale features has not been well explored between contracting and expanding paths except simple concatenation. In this paper, we propose a light-weight network where contracting and expanding paths are connected with fused multi-scale features through bi-directional feature pyramid network (BiFPN). The backbone of our proposed network has a dilated multi-fiber (DMF) structure based U-net architecture. First, conventional convolutional layers along the contracting and expanding paths are replaced with a DMF network and an MF network, respectively, to reduce the overall network size. In addition, a learnable weighted DMF network is utilized to take into account different receptive sizes effectively. Next, a weighted BiFPN is utilized to connect contracting and expanding paths, which enables more effective and efficient information flow between the two paths with multi-scale features. Note that the BiFPN block can be repeated as necessary. As a result, our proposed network is able to further reduce the network size without clearly compromising segmentation accuracy. Experimental results on the popular BraTS 2018 dataset demonstrate that our proposed light-weight architecture is able to achieve at least comparable results with the state-of-the-art methods with significantly reduced network complexity and computation time. The source code of this paper will be available at Github.**

I. Introduction

MRI (Magnetic Resonance Imaging) has been widely used for the diagnosis and treatment of brain cancer. Effective and efficient segmentation of brain tumors such as gliomas is critical for precision and personalized medicine [1]. However, it has been subjective and time-consuming for well-trained physicians to manually annotate various brain tumor regions from MRI scans.

Over the years, many automatic brain tumor segmentation techniques have been developed [2], [3]. For example, in [2], visual saliency features were proposed for segmenting brain tumors. Recently, due to the great success of deep learning techniques, many deep learning based brain tumor segmentation methods have been proposed [4], [5]. In [4], Havaei *et al.* proposed a two-pathway architecture for brain tumor segmentation by utilizing the popular 2D convolutional operator and predicting the label of each pixel. Based on the fully covolutional network [6] proposed for image segmentation, Ronnerberger *et al.* proposed U-Net [7] as an encoder-decoder architecture for medical image segmentation Since then, many U-Net variants (e.g., [8]) have also been proposed. In order to take 3D spatial context of MRI images, various 3D versions of U-Net including 3D U-Net [9] and V-Net [10] were proposed. Although these methods have achieved promising segmentation results, their computational complexity is quite high and more efficient networks are needed for practical applications.

There have been several attempts to explore light-weight network architectures for efficient brain tumor segmentation. For example, Nuechterlein *et al.* proposed 3D-ESPNet [11] for efficient volumetric brain tumor image segmentation by using efficient spatial pyramid (ESP) module with pyramidal refinement block. Chen *et al.* proposed S3D-UNet [12] to reduce the number of learnable network parameters by decomposing 3D convolutions into three parallel branches. However, as a result, segmentation accuracy has been compromised. In [13], a 3D dilated multi-fiber (DMF) network was proposed for real-time dense volumetric segmentation by leveraging the 3D multi-fiber unit aggregating a group of light-weight 3D convolution networks. However, like other U-Net based segmentation networks, the connection between contracting and expanding paths has not been well explored for devising effective and efficient new networks, except simple concatenation.

With various path aggregation strategies, features at dif-

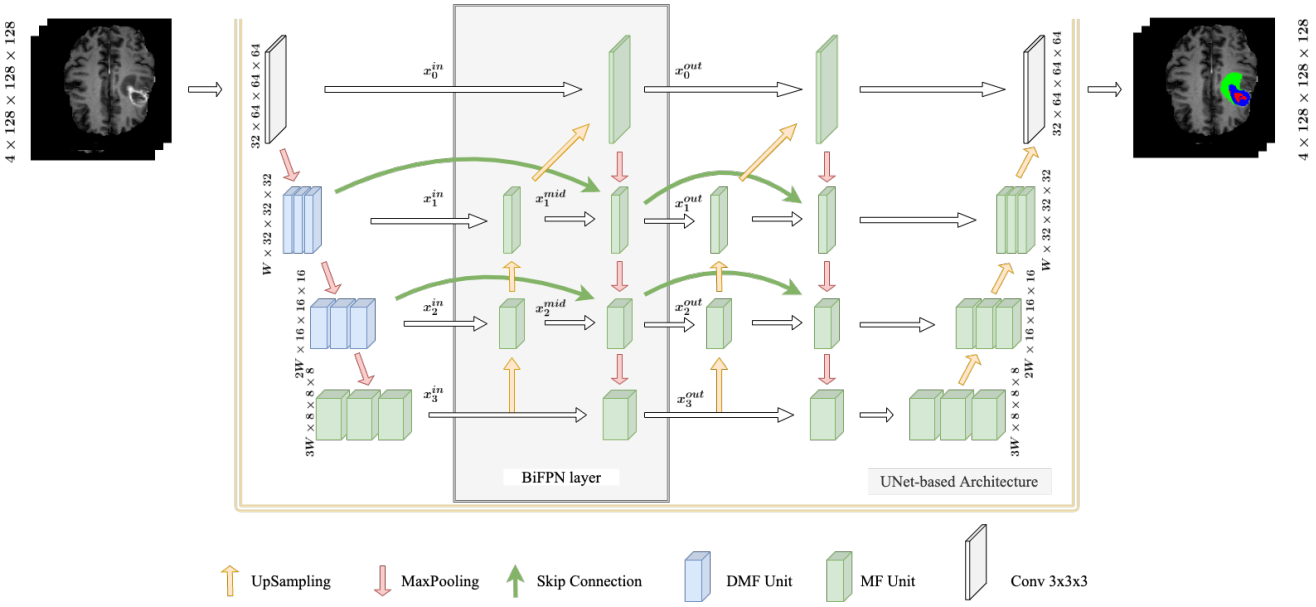


Fig. 1: Illustration of our proposed light-weight model, where W denotes the number of channels in a convolution block.

ferent levels or layers can be integrated for more effective feature learning. Skip connection [14] was first used to capture and propagate extra contextual information from low resolution layers and high resolution ones. Based on the idea of skip connection, several advanced techniques have been developed, such as Feature Pyramid Network (FPN) [15], Neural Architecture Search FPN (NAS-FPN) [16], Bi-directional FPN (BiFPN) [17], and Adaptively spatial feature fusion (ASFF) [18]. In particular, BiFPN is one of the most efficient and has been successfully used for scalable and efficient object detection [17].

Therefore, in this paper, based on the existing DMF based U-Net framework, we propose to connect the corresponding layers along contracting and expanding paths with weighted BiFPN, instead of simple concatenation. That is, a layer on the expanding path will take inputs from both its precedent expanding layer and the features fused from multiple layers along the contracting path. As a result, features learned at the encoder layers along the contracting path can be more effectively passed to the decoder layers along the expanding path. In order to preserve small network size, multi-fiber network is also utilized in BiFPN network.

In summary, the key contributions of our proposed network can be summarized as follows:

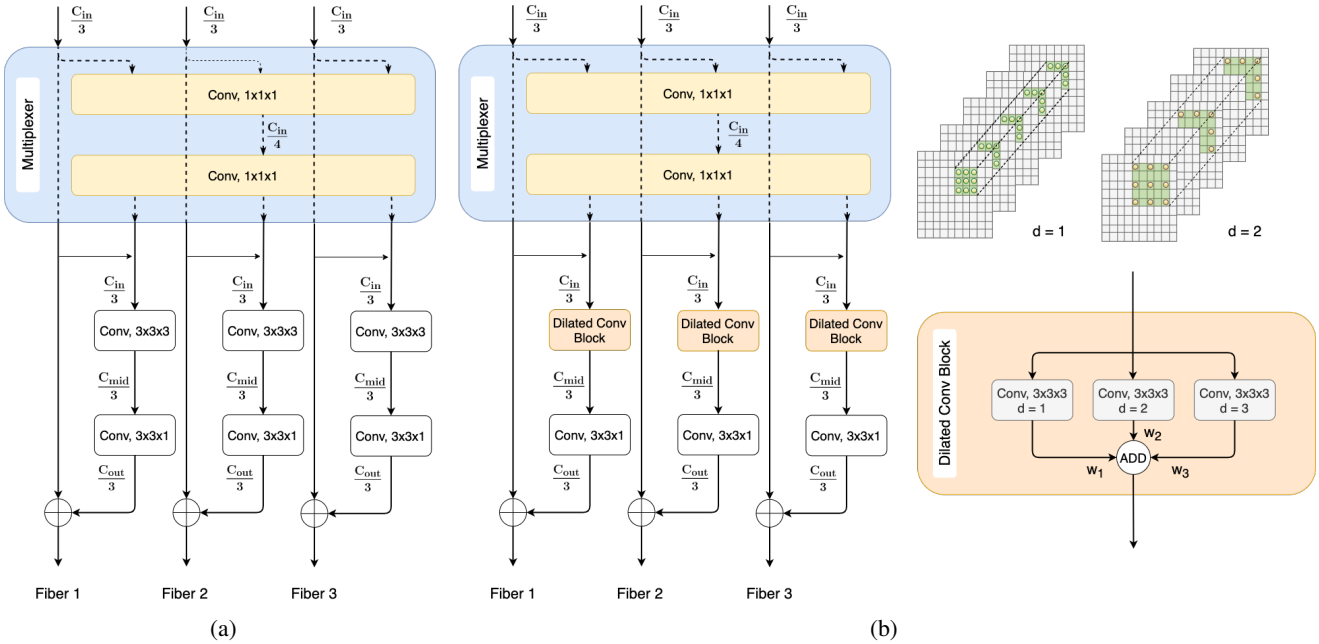
- We propose one of the first U-Net architectures where multi-scale features along the contracting path are fused as an input to the decoding layers along the expanding path, so that the learned multi-scale features can be more effectively passed to the decoding layers for improved segmentation performance.
- We develop a light-weight network by adopting the scalable and efficient BiFPN for connecting the contracting and expanding paths. In addition, we further reduce the network size of BiFPN by replacing conventional convolutional layers with MF structure.
- Our proposed model is able to achieve segmentation performance on the popular BraTS 2018 dataset comparable to that of the state-of-the-art with about 1/20-th of its network size.

II. Related work

Many deep learning based brain tumor segmentation methods have been proposed. For example, in [19], a unified attentional Generative Adversarial Network (GAN) was proposed without pair based supervision. As our work is based on U-Net architecture, in this section, we only review methods closely related to U-Net. As summarized in [20], existing methods can be categorized into three groups: data processing, model designing methods, and optimizing methods. Considering the symmetric architecture of U-Net consisting of a contracting path (i.e., encoding path) and an expanding path (i.e., decoding path), in this section, we divide the U-Net based methods into two categories: structure based methods which focus on designing new convolutional structures along the contracting and expanding paths and constraint based methods which focus on designing new constraints to enhance learning.

A. Structure based Methods

Structure based methods aim to increase the learning ability of convolution networks through various strategies such as enlarging receptive field and utilizing attention mechanism. In [21], a spatial pyramid pooling (SPP) module was proposed to produce a feature map with multi-level pooling. With SPP modules, multi-scale features can be extracted from input images with arbitrary sizes. Motivated by the idea of SPP [21], Atrous Spatial Pyramid Pooling (ASPP) [22] was proposed to derive multi-scale features of multiple receptive fields



(a) (b)

Fig. 2: Illustration of MF and DMF units. (a) MF Unit. Each fiber is a residual unit with two convolution layers. Information from all input branches is fused by an information routing block, called *multiplexer*, before being passed to fibers. (b) DMF Unit. One convolution layer in each fiber is replaced by a dilated convolution block, which has multiple dilated convolution layers and each dilated convolution layer has a different dilation rate.

by using multiple parallel dilated convolutional layers with different sampling rates. In order to address the limitation of ASPP where features have a uniform resolution from previous convolutional layers, receptive field block (RFP) [23] was proposed by using multiple dilated convolutional layers. As a result, the model focuses more on the center point of the kernel than the father ones.

Recently, several light-weight network architectures have been proposed for efficient brain tumor segmentation, such as 3D-ESPNet [11], S3D-UNet [12], and DMFNet [13]. A number of methods have also been proposed to reduce the memory footprint of the proposed segmentation model. For example, in [24], a partially reversible U-Net was proposed to leverage the benefit of the reversible architecture which allows to exactly recover each layer's outputs from those of the subsequent layers and eliminate the need to store activations for back-propagation.

B. Constraint based Methods

Constraint based methods aim to introduce new constraints by designing extra branches with extra tasks. In [25], a reconstruction branch was proposed to introduce an autoencoder-like reconstruction constraint. Similarly, various multi-task learning based methods were proposed to add new branches in parallel to the expanding path [26], [27]. For example, in [27], two extra tasks, reconstruction and classification, were added along the segmentation decoder by sharing the same encoder. As a result, a total loss can be derived from the three tasks.

Another line of the constraint based methods explore the hierarchical relationship among the tumor classes. In [28], a two-stage cascaded U-Net was proposed to segment different types of brain tumors from coarse to fine levels. similarly, in [29], Tu-Net was proposed with three cascaded U-Net to drive three types of tumor masks, respectively. In [30], a deep cascaded attention network was also proposed.

Note that, except the methods which aim for high efficiency, most of the abovementioned methods are very computationally expensive in both 2D and 3D convolutional neural networks, which makes it difficult to train and deploy them to real-time applications. In addition, the connection between contracting and expanding paths of U-Net has not been well explored to further reduce the network complexity for higher efficiency, except simple concatenation such as directly bridging the corresponding layers of the contracting and expanding paths.

III. Proposed Method

As shown in Fig. 1, our proposed method is based on the basic U-Net architecture and consists of two key components: 1) dilated multi-fiber (DMF) unit (a dilated version of the multi-fiber (MF) unit) along contracting and expanding paths, and 2) BiFPN layer connecting the corresponding layers between contracting and expanding paths. The network takes the four modalities of a 3D MRI scan as input and produces a segmentation mask of three types of tumor regions. In this section, we will explain the details of the two key components: DMF unit and BiFPN layer.

A. Dilated Multi-Fiber (DMF) Unit

We first explain the MF unit based on which DMF unit is built. An MF unit is actually a set of residual sub-blocks or

branches in parallel. That is, one conventional residual block is split into a set of sub-blocks which will then be combined together through addition or concatenation. Each sub-block or branch is called a fiber.

The output of such MF unit can be calculated as follows:

$$\mathbf{y} = \sum_{i=1}^{g} f_i(\mathbf{x}_i), \tag{1}$$

where g is the number of branches, $\mathbf{x}$ is the input vector of a conventional residual block with C channels, $\mathbf{x}_i$ is the input vector for the i-th branch, and each branch will have C/g channels. f_i can be an arbitrary function, which is convolution in our proposed architecture. As shown in Fig. 2, there are 3 fibers for illustration. That is, $g = 3$.

While the total number of channels is the same as the conventional residual architecture, the total number of parameters can be reduced effectively through such channel grouping. Let C_{in} denote the number of input channels, C_{mid} denote the number of middle channels, and C_{out} denote the number of output channels with two convolutional layers in each fiber. Let's assume that the size of the kernel is constant, denoted by k (e.g., $27 = 3 \times 3 \times 3$ for a 3D filter), the number of parameters of the conventional residual architecture, $P_{residual}$, can be calculated as follows:

$$P_{residual} = k * (C_{in} * C_{mid} + C_{mid} * C_{out}). \tag{2}$$

When the conventional residual block is split into g branches, the number of parameters comes to:

$$\begin{aligned} &P_{multi-branch} \\ &= g * k * (C_{in}/g * C_{mid}/g + C_{mid}/g * C_{out}/g) \\ &= k * (C_{in} * C_{mid} + C_{mid} * C_{out})/g = P_{residual}/g. \end{aligned} \tag{3}$$

That is, the total number of parameters is significantly reduced by a factor of g.

Clearly, such a grouping strategy breaks the coherence among the branch channels, which could compromise the model's capacity of learning representations. To rectify this issue and facilitate information flow among fibers, a light-weight convolution block called *multiplexer* is introduced at the beginning of the branches. The multiplexer works as a router that mixes and redirects features from all the branches. As shown in Fig. 2, a multiplexer block aggregates information from all the branches by using two $1 \times 1 \times 1$ convolution layers to mix the information without significantly increasing the network size. The number of output channels of the first convolution layer in multiplexer is reduced four times to further reduce the total number of parameters. As a result, the residual component in each fiber takes input from both individual split features and the ones produced by the multiplexer.

In order to capture multi-scale features through the receptive fields with different sizes, dilated convolution is introduced to replace the first conventional convolutional layer in a fiber of a DMF unit. As shown in Fig. 2 (b), three dilation rates are used and the dilated convolutional outputs are aggregated with learnable weights w_1, w_2 and w_3:

$$y = \frac{w_1 * f_1(x) + w_2 * f_2(x) + w_3 * f_3(x)}{w_1 + w_2 + w_3 + \epsilon}, \tag{4}$$

where f_i is a dilated convolution layer, these dilated convolution layers have different dilation rates, and ϵ is used to avoid the division by zero problem.

In summary, MF Unit and DMF Unit could help accelerate the learning process, reduce the model redundancy, as well as maintain good learning capacity.

B. Bi-directional Feature Pyramid Network (BiFPN) Layer

BiFPN [17] is an efficient version of the Path-aggregation FPN (PANet) which adds a bottom-up path on to the top-down path based multi-scale feature fusion of FPN. As shown in Fig. 1, in order to balance the efficiency and effectiveness of the network, the bottom-up pathway has been simplified to reduce the number of convolutional blocks.

In this architecture, multi-level features are passed to a BiFPN layer to produce more effective high-level features. The features are fused in both two directions: bottom-up and top-down, this technique gives our model the ability to learn cross-scale features via cross-scale connections. These features could be combined by using weighted feature fusion approach or concatenating them together. We can stack multiple BiFPN layers repeatedly to obtain more high-level features. As shown in Fig. 1, we can calculate the output of the BiFPN layer as follows:

$$\begin{aligned} x_2^{mid} &= Block(F(x_2^{in}, Resize(x_3))) \\ x_1^{mid} &= Block(F(x_1^{in}, Resize(x_2^{mid}))) \\ x_0^{out} &= Block(F(x_0^{in}, Resize(x_1^{mid}))) \\ x_1^{out} &= Block(F(x_1^{in}, x_1^{mid}, Resize(x_0^{out}))) \\ x_2^{out} &= Block(F(x_2^{in}, x_2^{mid}, Resize(x_1^{out}))) \\ x_3^{out} &= Block(F(x_3^{in}, Resize(x_2^{out}))), \end{aligned} \tag{5}$$

where *Block* is a convolution block, and F is a multi-scale feature fusion function.

To further improve its efficiency, we utilize MF unit as the implementation of the *Block*. The *Resize* function is used for matching the resolution to combine them. When resizing up, we can use UpSampling or DeConvolution layer; when resizing down, we use Max Pooling layer.

Similar to the techniques used in DMF unit, we use the linear weighted feature fusion to combine features so that the model can learn and differentiate the importance of input features, and also enable the model a fast fusion process. The outputs x_0^{out}, x_1^{out}, x_2^{out}, x_3^{out} that we obtained from BiFPN [17] layer have the same shapes with the input features x_0,

x_1, x_2, x_3, so we can cascade another BiFPN layer with these outputs.

With the BiFPN layer, our model can be scaled both in the number of channels inside the BiFPN layer as well as the number of layers. Therefore, a family of architectures could be derived in a scalable manner, depending on the balance between segmentation accuracy and model efficiency.

IV. Experimental Results and Discussions

A. Dataset

The BraTS 2018 dataset [32] which was provided by the Brain Tumor Segmentation 2018 challenge [33], was used in our experiments. This dataset contains a large number of multi-parametric magnetic resonance imaging (mpMRI) scans with associated delineations of the relevant tumor sub-regions. The training set consists of 285 cases, and the validation set consists of 66 cases. Each mpMRI case consists of four scanned images: (1) a native T1-weighted scan (T1), (2) a post-contrast T1-weighted scan (T1ce), (3) a native T2-weighted scan (T2), and (4) a T2 Fluid Attenuated Inversion Recovery (T2-FLAIR) scan.

The volume size of each component scan is $240 \times 240 \times 155$. The inputs for our model will be the concatenation of these four scans, which has size $4 \times 240 \times 240 \times 155$, together with its corresponding ground truth segmentation result as output. With this dataset, we focus on the segmentation of intrinsically heterogeneous brain tumors, called gliomas. There are three sub-regions that we need to predict: (1) Enhancing Tumor (ET), (2) Tumor Core (TC), also known as the gross tumor, (3) Whole Tumor (WT), also known as a complete tumor. A special characteristics of this dataset is that the tumor core entails ET, and WT entails both TC and ET.

B. Evaluation Metrics

Similar to [13], segmentation accuracy is measured with two metrics: dice score and Hausdorff95 distance, and model complexity are measured in terms of the number of network parameters.

1) Dice Score: Dice Score (also known as Dice Similarity Coefficient) measures the positive samples identified by an algorithm against the overall samples of both the ground truth samples and the false positive samples identified by the algorithm.

More specifically, in the semantic segmentation problem, the ground truth is represented as a mask A where value 1 denotes a positive sample, and value 0 otherwise. With a predicted segmentation map B produced by a segmentation algorithm, Dice Score can be calculated as follows:

$$DS = \frac{2 \cdot |A \cap B|}{2 \cdot |A \cap B| + |B \setminus A| + |A \setminus B|}, \tag{6}$$

where $|A|$ denotes the number of positive pixels in A, $A \cap B$ denotes the intersection of A and B, and $A \setminus B$ denotes the number of positive pixels appearing in A but not in B.

2) Hausdorff Distance: Hausdorff distance measures the distance between two subsets of a metric space. It has been frequently used in evaluating the performance of medical semantic segmentation methods because of being a robust measurement of the largest segmentation error.

Given two point sets $X = \{x_1, x_2, ..., x_m\}$ and $Y = \{y_1, y_2, ..., y_n\}$, one-sided Hausdorff distance can be calculated as follows:

$$hd(X, Y) = max_{x \in X}(min_{y \in Y}(d(x, y)) \tag{7}$$

$$hd(y, X) = max_{y \in Y}(min_{x \in X}(d(y, x)), \tag{8}$$

where $d(x, y)$ is the distance between x and y. Similarly, we use Euclidean distance in this paper.

Then the bidirectional Hausdorff distance between X and Y could be formulated as follows:

$$HD(X, Y) = max(hd(X, Y), hd(Y, X)). \tag{9}$$

The final bidirectional Hausdorff distance $HD(X, Y)$ indicates the largest distance from a point in one of the two sets to its closest point in the other set. In brain tumor segmentation, the Hausdorff distance calculates the distance between prediction boundaries and ground-truth boundaries.

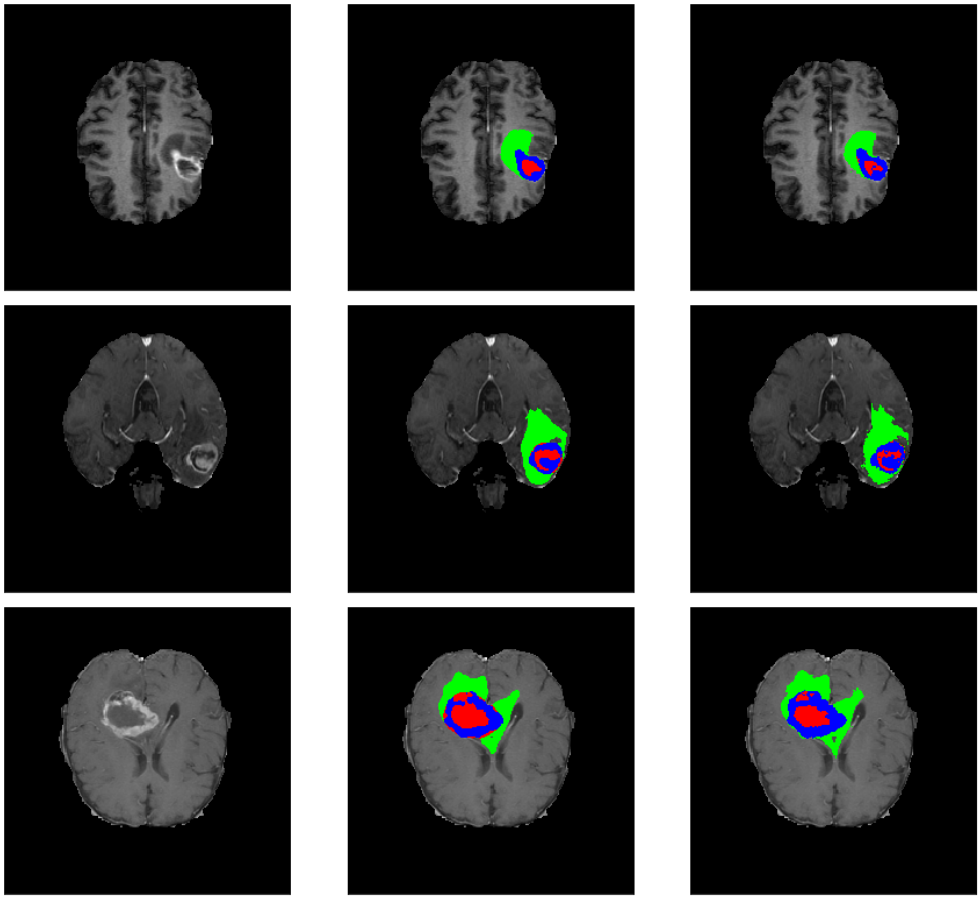
C. Performance Evaluation

We trained our models on an NVIDIA V100 GPU with a batch size 4, Adam optimizer, learning rate 0.001, and weight decay 1e-5. Similar to [13], we cropped the original scans to size $128 \times 128 \times 128$ due to memory restriction of our GPU, then applied several augmentation techniques to these crops to increase training data, such as random rotation, random intensity change, and random flip. The results may be improved by using larger size of crop area. The state-of-the-art model NVDLMED trained with random crop of size $160 \times 190 \times 128$, it makes the model learn the image content better by remaining more information in crop area. We trained with these settings in the first 500 epochs and then reduced the learning rate to 0.0001 and trained another 200 epochs. The labels in this dataset include the background, non-enhancing and necrotic tumor, peritumoral edema and GD-enhancing tumor. The loss function we used in the training phase is Generalized Dice Loss. At the test phase, we used the full-size image to predict and applied test time augmentation (TTA) technique.

We compared our proposed method with two categories of segmentation methods: efficiency driven ones and accuracy driven ones. The former category includes our baseline algorithm DMFNet [13], S3D-UNet [12], and 3D-ESPNet [11], the latter category includes No New-Net [31], 3D U-Net [9], and NVDLEMED [8]. As shown in Table I, our proposed methods clearly outperform those efficiency-driven methods with small network size. For example, our proposed D3-W64 outperforms DMFNet W128 with nearly half of its network size. In particular, its performance is comparable to that of the state-of-the-art NVDLMED [8] with about 1/20-th of its size. This indicates that utilizing multi-fiber structure leads to small

Model	Params (M)	FLOPs (G)	Dice Score			Hausdorff95		
			ET	WT	TC	ET	WT	TC
Our D1-W64	**1.38**	**14.99**	79.64	90.63	84.92	2.69	4.30	**5.74**
Our D2-W64	<u>1.76</u>	<u>20.37</u>	80.08	<u>90.68</u>	<u>85.04</u>	3.05	4.36	<u>5.81</u>
Our D3-W64	2.14	25.75	<u>81.20</u>	89.79	84.42	<u>2.59</u>	**3.61**	7.04
MFNet W128 [13]	3.19	20.61	79.91	90.43	84.61	2.68	4.68	6.31
DMFNet W128 [13]	3.88	27.04	80.12	90.62	84.54	3.06	4.66	6.44
S3D-UNet [12]	3.32	75.20	74.93	89.35	83.09	4.43	4.71	7.74
3D-ESPNet [11]	3.80	76.51	73.70	88.30	81.40	5.30	5.46	7.85
No New-Net [31]	10.36	202.25	81.01	**90.83**	85.44	**2.41**	<u>4.27</u>	6.52
3D U-Net [9]	16.21	1669.53	75.96	88.53	71.77	6.04	17.10	11.62
NVDLMED [8]	40.06	1495.53	**81.73**	90.68	**86.02**	3.82	4.52	6.85

TABLE I: Quantitative comparison of segmentation performance on the BraTS 2018 dataset. For our proposed method, D denotes the number of BiFPN layers and W denotes the number of channels in each convolutional unit. The item in bold font in each column indicates the best performance, and the underlined one indicates the second-best performance.



(a) (b) (c)

Fig. 3: An example comparison of segmentation results, where each row is one slice of the scan, the peritumoral edema regions are represented in green, the non-enhancing and necrotic tumor regions are represented in red, and the GD-enhancing tumor regions (ET) are represented in blue. Whole tumor (WT) is the region including all the above regions; Tumor core includes the non-enhancing, necrotic tumor region, and enhancing tumor region. (a) Input MRI images (b) Our predictions (c) Ground-truth labels

network size and utilizing BiFPN to connect the contracting and expanding paths with multi-scale features leads to more effective segmentation without increasing model complexity.

An example of segmentation results is shown in Fig. 3. The results demonstrate that our model is able to produce a segmentation very close to the ground-truth. It is also

noticed that the boundary of the predicted segmentation looks smoother than that of the ground-truth, which may be due to the upsampling effect. It may be worthwhile to investigate edge based constraints in future research.

V. Conclusions

In this paper, we present a novel light-weight yet effective model for brain tumor segmentation. Our proposed network is one of the first U-Net architectures where enhanced connections other than simple concatenations are established between the contracting and expanding paths. Specifically, a dilated multi-filber (DMF) network is utilized in the encoder layers on the contracting path so that features are efficiently learned at different receptive field sizes and aggregated with learned weights. On the expanding path, multi-fiber (MF) network is utilized to reduce the network size. In order to establish better information flow between contracting and expanding paths, a weighted BiFPN is utilized to fuse the multi-scale features learned from the encoding layers in a scalable and efficient manner. The experimental results on the BraTS 2018 dataset show that our proposed method could achieve comparable results against the state-of-the-art methods with much smaller network size.

References

[1] P. Kickingereder, F. Isensee, I. Tursunova, J. Petersen, U. Neuberger, D. Bonekamp, and et al., "Automated quantitative tumour response assessment of MRI in neuro-oncology with artificial neural networks: a multicentre, retrospective study," *Lancet Oncology*, vol. 20, no. 5, pp. 728–740, May 2019.

[2] S. Banerjee, S. Mitra, B. U. Shankar, and Y. Hayashi, "A novel GBM saliency detection model using multi-channel MRI," *PloS one*, vol. 11, no. 1, p. e0146388, 2016.

[3] S. Banerjee, S. Mitra, and B. U. Shankar, "Automated 3D segmentation of brain tumor using visual saliency," *Information Sciences*, vol. 424, pp. 337–353, 2018.

[4] M. Havaei, A. Davy, D. Warde-Farley, A. Biard, A. Courville, Y. Bengio, C. Pal, P.-M. Jodoin, and H. Larochelle, "Brain tumor segmentation with deep neural networks," *Medical image analysis*, vol. 35, pp. 18–31, 2017.

[5] J. Kleesiek, G. Urban, A. Hubert, D. Schwarz, K. Maier-Hein, M. Bendszus, and A. Biller, "Deep MRI brain extraction: A 3D convolutional neural network for skull stripping," *NeuroImage*, vol. 129, pp. 460–469, 2016.

[6] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 3431–3440.

[7] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2015, pp. 234–241.

[8] A. Myronenko, "3D MRI brain tumor segmentation using autoencoder regularization," in *International MICCAI Brainlesion Workshop*, 2018, pp. 311–320.

[9] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: learning dense volumetric segmentation from sparse annotation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2016, pp. 424–432.

[10] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-Net: Fully convolutional neural networks for volumetric medical image segmentation," in *International Conference on 3D vision (3DV)*, 2016, pp. 565–571.

[11] N. Nuechterlein and S. Mehta, "3D-ESPNet with pyramidal refinement for volumetric brain tumor image segmentation," in *International MICCAI Brainlesion Workshop*, 2018, pp. 245–253.

[12] W. Chen, B. Liu, S. Peng, J. Sun, and X. Qiao, "S3D-UNet: separable 3d u-net for brain tumor segmentation," in *International MICCAI Brainlesion Workshop*, 2018, pp. 358–368.

[13] 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 *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2019, pp. 184–192.

[14] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.

[15] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *IEEE conference on computer vision and pattern recognition (CVPR)*, 2017, pp. 2117–2125.

[16] G. Ghiasi, T.-Y. Lin, and Q. V. Le, "NAS-FPN: Learning scalable feature pyramid architecture for object detection," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 7036–7045.

[17] M. Tan, R. Pang, and Q. V. Le, "EfficientDet: Scalable and efficient object detection," *arXiv preprint arXiv:1911.09070*, 2019.

[18] S. Liu, D. Huang, and Y. Wang, "Learning spatial fusion for single-shot object detection," *arXiv preprint arXiv:1911.09516*, 2019.

[19] W. Yuan, J. Wei, J. Wang, Q. Ma, and T. Tasdizen, "Unified attentional generative adversarial network for brain tumor segmentation from multimodal unpaired images," in *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2019, pp. 229–237.

[20] Y.-X. Zhao, Y.-M. Zhang, and C.-L. Liu, "Bag of tricks for 3D MRI brain tumor segmentation," in *International MICCAI Brainlesion Workshop*, 2019, pp. 210–220.

[21] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 9, pp. 1904–1916, 2015.

[22] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 4, pp. 834–848, 2017.

[23] S. Liu, D. Huang *et al.*, "Receptive field block net for accurate and fast object detection," in *European Conference on Computer Vision (ECCV)*, 2018, pp. 385–400.

[24] R. Brügger, C. F. Baumgartner, and E. Konukoglu, "A partially reversible u-net for memory-efficient volumetric image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2019, pp. 429–437.

[25] M. Frey and M. Nau, "Memory efficient brain tumor segmentation using an autoencoder-regularized u-net," in *International MICCAI Brainlesion Workshop*, 2019, pp. 388–396.

[26] X. Guo *et al.*, "Improving brain tumor segmentation with multi-direction fusion and fine class prediction," in *International MICCAI Brainlesion Workshop*, 2019, pp. 349–358.

[27] L. Weninger, Q. Liu, and D. Merhof, "Multi-task learning for brain tumor segmentation," in *International MICCAI Brainlesion Workshop*, 2019, pp. 327–337.

[28] Z. Jiang, C. Ding, M. Liu, and D. Tao, "Two-stage cascaded u-net: 1st place solution to brats challenge 2019 segmentation task," in *International MICCAI Brainlesion Workshop*, 2019, pp. 231–241.

[29] M. H. Vu, T. Nyholm, and T. Löfstedt, "Tunet: End-to-end hierarchical brain tumor segmentation using cascaded networks," in *International MICCAI Brainlesion Workshop*, 2019, pp. 174–186.

[30] H. Xu, H. Xie, Y. Liu, C. Cheng, C. Niu, and Y. Zhang, "Deep cascaded attention network for multi-task brain tumor segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2019, pp. 420–428.

[31] F. Isensee, P. Kickingereder, W. Wick, M. Bendszus, and K. H. Maier-Hein, "No New-Net," in *International MICCAI Brainlesion Workshop*, 2018, pp. 234–244.

[32] B. H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, Y. Burren, N. Porz, J. Slotboom, R. Wiest *et al.*, "The multimodal brain tumor image segmentation benchmark (BRATS)," *IEEE Transactions on Medical Imaging*, vol. 34, no. 10, pp. 1993–2024, 2014.

[33] S. Bakas, M. Reyes, A. Jakab, S. Bauer, M. Rempfler, A. Crimi, R. T. Shinohara, C. Berger, S. M. Ha, M. Rozycki *et al.*, "Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge," *arXiv preprint arXiv:1811.02629*, 2018.