CISTER

Research Centre in Real-Time & Embedded Computing Systems

Journal Paper

HTTU-Net: Hybrid Two Track U-Net for Automatic Brain Tumor Segmentation

Nagwa M. Aboelenein

Piao Songhao

Anis Koubaa*

Alam Noor

Ahmed Afifi

HTTU-Net: Hybrid Two Track U-Net for Automatic Brain Tumor Segmentation

Nagwa M. Aboelenein, Piao Songhao, Anis Koubaa*, Alam Noor, Ahmed Afifi

*CISTER Research Centre

Rua Dr. António Bernardino de Almeida, 431

4200-072 Porto

Portugal

Tel.: +351.22.8340509, Fax: +351.22.8321159

E-mail:

https://www.cister-labs.pt

Abstract

Brain cancer is one of the most dominant causes of cancer death; the best way to diagnose and treat brain tumors is to screen early. Magnetic Resonance Imaging (MRI) is commonly used for brain tumor diagnosis; however, it is a challenging problem to achieve higher accuracy and performance, which is a vital problem in most of the previously presented automated medical diagnosis. In this paper, we propose a Hybrid Two-Track U-Net(HTTU-Net) architecture for brain tumor segmentation. This architecture leverages the use of Leaky Relu activation and batch normalization. It includes two tracks; each one has a different number of layers and utilizes a different kernel size. Then, we merge these two tracks to generate the final segmentation. We use the focal loss, and generalized Dice (GDL), loss functions to address the problem of class imbalance. The proposed segmentation method was evaluated on the BraTS'2018 datasets and obtained a mean Dice similarity coefficient of 0.865 for the whole tumor region, 0.808 for the core region and 0.745 for the enhancement region and a median Dice similarity coefficient of 0.883, 0.895, and 0.815 for the whole tumor, core and enhancing region, respectively. The proposed HTTU-Net architecture is sufficient for the segmentation of brain tumors and achieves highly accurate results. Other quantitative and qualitative evaluations are discussed, along with the paper. It confirms that our results are very comparable expert human-level performance and could help experts to decrease the time of diagnostic

HTTU-Net: Hybrid Two Track U-Net for Automatic Brain Tumor Segmentation

NAGWA M. ABOELENEIN[1,2], **PIAO SONGHAO**[1], **ANIS KOUBAA**[3,4], **ALAM NOOR**[3],
AND AHMED AFIFI[2,5]

[1]School of Computer Science, Harbin Institute of Technology, Harbin 150001, China
[2]Faculty of Computers and Information, Menoufia University, Menoufia 32511, Egypt
[3]Robotics and Internet-of-Things Laboratory, Prince Sultan University, Riyadh 12435, Saudi Arabia
[4]CISTER, INESC-TEC, ISEP, Polytechnic Institute of Porto, 4200-465 Porto, Portugal
[5]College of Computer Science and Information Technology, King Faisal University, Hofuf 31982, Saudi Arabia

Corresponding author: Piao Songhao (piaosh@hit.edu.cn)

ABSTRACT Brain cancer is one of the most dominant causes of cancer death; the best way to diagnose and treat brain tumors is to screen early. Magnetic Resonance Imaging (MRI) is commonly used for brain tumor diagnosis; however, it is a challenging problem to achieve higher accuracy and performance, which is a vital problem in most of the previously presented automated medical diagnosis. In this paper, we propose a Hybrid Two-Track U-Net(HTTU-Net) architecture for brain tumor segmentation. This architecture leverages the use of Leaky Relu activation and batch normalization. It includes two tracks; each one has a different number of layers and utilizes a different kernel size. Then, we merge these two tracks to generate the final segmentation. We use the focal loss, and generalized Dice (GDL), loss functions to address the problem of class imbalance. The proposed segmentation method was evaluated on the BraTS'2018 datasets and obtained a mean Dice similarity coefficient of 0.865 for the whole tumor region, 0.808 for the core region and 0.745 for the enhancement region and a median Dice similarity coefficient of 0.883, 0.895, and 0.815 for the whole tumor, core and enhancing region, respectively. The proposed HTTU-Net architecture is sufficient for the segmentation of brain tumors and achieves highly accurate results. Other quantitative and qualitative evaluations are discussed, along with the paper. It confirms that our results are very comparable expert human-level performance and could help experts to decrease the time of diagnostic.

INDEX TERMS Brain tumor segmentation, deep neural networks, U-net, fully convolutional network, BraTS'2018 challenge.

I. INTRODUCTION

Brain tumor segmentation plays a crucial role in the diagnosis and planning of cancer treatment. Gliomas are the most general principle for brain tumors. It can be classified as Low-Grade (LGG) and High-Grade Gliomas (HGG). LGG tumors are less aggressive, while HGG tumors are malignant, growing, and rapidly invading surrounding tissues [1]. Patients with high-grade gliomas, even under treatment, do not live on average for more than 14 months after diagnosis [2]. Possible treatments include surgery, followed by or combined with chemotherapy and radiotherapy [3]. Specialists can use ultrasound, computed tomography (CT), or MRI for patient screening. MRI is especially helpful for Gliomas assessment

as extra information can be extracted. Common stages of MRI screening are fluid-attenuated reversal (FLAIR), T1-weighted (T1), T1-weighted contrast-enhanced (T1c), and T2-weighted (T2).

Manual segmentation requires a great deal of time and effort, and it is a subjective process. Automatic or semi-automatic methods are therefore necessary [4]. The main challenges of brain tumor segmentation are that it can occur in almost any form and size anywhere in the brain, and the tumor has a low contrast to the surrounding tissue. Deep learning-based techniques have outperformed traditional methods. One of the most common effective segmentation methods is the fully convolutional neural network (FCN). Zhao *et al.* proposed in [5] a technique for the segmentation of the brain tumors by integrating FCN and Conditional Random Fields(CRF's). Badrinarayanan *et al.* [6], presented

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

SegNet, a deep convolutional model for semantic segmentation presented SegNet, a deep convolutional model for semantic segmentation consists of an encoder network and a corresponding decoder network followed by a pixel-wise classification layer. Recently, in the field of image segmentation, convolutional neural networks (CNNs) based algorithms have accomplished excellent success [7]. In this area, in particular, U-Net-based models [8] are used extensively. The U-Net architecture is a fully convolutional model consisting of encoding and decoding parts. The U-Net model offers several advantages [9] for the segmentation process, such as its capacity to simultaneously use the context and global place needed to generate a good segmentation map and can use very few training samples and provide better segmentation results. The contributions of this paper can be summarized as follows.

- We propose HTTU-Net architecture to address challenges of brain tumor segmentation which occur in almost any form and size anywhere in the brain; also tumor has a low contrast to the surrounding tissue and address class imbalance problem. HTTU-Net not only extracts more semantic information but also gives more consideration to the information of small-scale brain tumors, which improves the segmentation of brain tumors.
- HTTU-Net is based on the excellent achievement of U-Net based architectures. Our technique is trained using multimodal image patches. It also updates the U-Net network by adding batch normalization at the end of each block to reduce the mean and variance problems and stable the layers. Our architecture, the first track, focuses on the tumor's form and size while the second track captures the contextual information. Each track consists of a different number of convolution blocks and uses a different kernel size to handle the different tumor sizes.
- We have introduced a new hybrid loss feature, combining Focal Loss and Generalized Dice Loss functions, to mitigate the class imbalance.
- We demonstrate that the proposed strategy improves the precision of the initial U-Net and also alleviates the issue of overfitting. We experiment with Brats 2018 dataset, and our architecture shows superior performance.

Section II presents a brief survey of the related work. In Section III, we describe the material and methods. Section IV shows the experimental results, and the discussion will show in Section V Finally, conclusion and future work are provided in Section VI.

II. RELATED WORK

Several CNN-based approaches for brain tumor segmentation have been suggested in the literature, and significant progress has been made [10]. Pereira *et al.* [11] proposed a 2D CNN network with a small kernel size (i.e., 3 × 3). They trained two distinct models, one for HGG and another one for LGG. They also used a max-pooling layer of stride two

and applied a dropout to the dense layers only. The model utilizes the activation function of Leaky rectified linear units (LeakyReLU) [12]. A two-stage cascade network was also implemented by Havaei *et al.* [13] that contains a local path and a global path. This architecture can encode contextual characteristics, both locally and globally. Zhao and Jia [14] presented a multiscale CNNs model, through which not only local and global features are learnt, but also complementary information from various MRI image modality is combined.

Le and Pham [15], suggested full-convolution U-Net networks acquire features from a multimodal MRI training dataset and then apply Extremely Randomized Trees (Extra-Trees) to segment the abnormal tumor cells. Wang *et al.* [16] used three binary classifiers, one for each task, to develop a triple-cascaded brain tumor segmentation architecture. Consequently, they can decrease over-fitting and make it simple to train. However, one limitation of the cascade architecture is that it is not end-to-end, and it takes a long time to train. Dong *et al.* [17], established a deep convolution network based on U-Net for automatic tumor segmentation, which was evaluated using data set from BRATS' 2015. This approach uses the Dice loss function. In [18], Chen proposed a separable 3D U-Net architecture using the prevalent U-Net structure, dividing each 3D convolution into three branches in a parallel. He and Fang [19] proposed three pathways U-Net structure to segment the brain tumor; each modality is processed in a single pathway. Table 1 summarizes the datasets, methods, and results of the most similar related works on brain tumor segmentation based on deep learning methods with our work where Enhancing Tumor (ET), Whole Tumor (WT) and Tumor Core (TC).

## III. MATERIAL AND METHODS
A. DATASET AND DATA PREPARATION
1) DATASET
The BraTS' 2018 [20] contest provides a large training set of 210 HGG MRI scans and 75 LGG MRI scans. The size of each MRI scan is 240 × 240 × 155, and each case has FLAIR, T1, T1-enhanced, and T2 volumes. The dataset is co-registered, re-sampled to 1 mm3, and skull-stripped. Fig 1 shows an example of the data as well as the ground truth.Segment brain tumors, including necrosis, edema, non-enhancing, and enhancing tumor

2) DATA PREPARATION
The N4ITK bias correction method [21] is applied to MRI volumes to alleviate non-homogeneity and intensity variations. Each slice is then normalized using the slice's mean and standard deviation. In this work, to reduce the impact of the class imbalance issue, we perform data augmentation that comprises rotation, translation, horizontal and vertical flipping [22]. During the training, 2D patches of 128 × 128 × 4 size are sampled randomly to decrease computational time, and all zero-intensity patches from the training set will be removed [13].

TABLE 1. Comparison of our paper with reviewed brain tumor segmentation methods.

Ref	Dataset	Methods	Performance (Dice Scores)		
			ET	WT	TC
[11]Pereira etal,2016	BRATS2013	CNN with small (3x3) filters for deeper architecture	0.72	0.84	0.62
[13]Havaei et el ,2016	BRATS2013	Cascaded CNNs for simultaneous local and global processing	0.79	0.88	0.73
[17]Dong et al,2017	BRATS2015	U-Net based deep convolution networks	0.86	0.86	0.65
[15]pham et al,2018	BRATS2013	U-Net-based fully convolutional ,Extremely randomized trees as classifier	0.81	0.85	0.72
[19]Fang et al ,2018	BRATS2018	Fully automatic three pathways U-Net from two modality MRI	0.726	0.856	0.722
[18]Chen etal,2019	BRATS2018	propose the S3D-UNet architecture using the U-Net structure	0.783	0.836	0.689
Our paper	BRATS2018	propose Hybrid two Track architecture using the U-Net structure	0.808	0.865	0.745
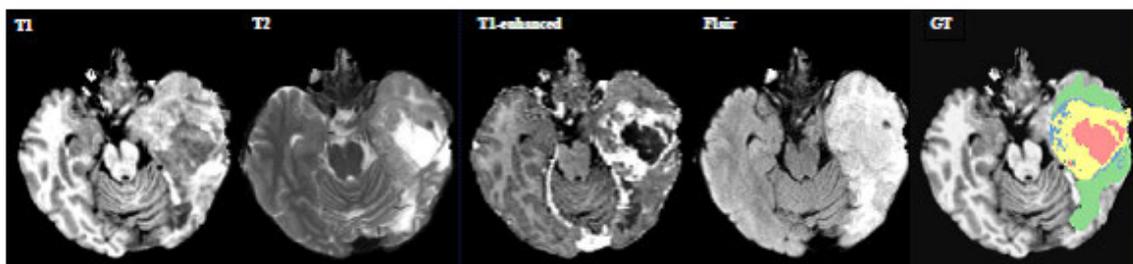
FIGURE 1. Sample MRI images and ground truth labels, from left to right, T1, T2, T1-enhanced, and the label images; Green: edema, yellow: enhancing tumor, red: necrosis and non-enhancing.

B. METHODS

1) U-NET BASED DEEP CONVOLUTIONAL NETWORKS

We will briefly explain the original U-Net framework in this subsection. The U-Net architecture can be regarded as an auto-encoder where there are several contraction blocks in the contraction part. Each block includes two 3×3 convolutional layers followed by a 2×2 pooling layer, and it uses the rectified linear unit (ReLU) activation function. It doubles the number of feature maps from one block to the next to be able to effectively learn the complex structures, increasing the number of feature maps from 1 to 1024. Each block in the expanding part consists of two 3×3 convolutional layers followed by a 2×2 up-sampling layer. In order to maintain symmetry, the number of feature maps is reduced by half from each block to the next. There is a collection of stacked convolution/pooling layers in the contracting or downsampling part, whereas the expansion part consists of up/transposed convolutional layers. Skip connections are used in U-Net to append the feature maps of the corresponding contraction layer at the end of each expansion block. These connections guarantee that the contracting part features will be used in image reconstruction, and the final layer is a 1×1 convolutional layer to produce the segmentation results.

2) TWO-TRACK U-NET ARCHITECTURE

Severe class imbalance is one of the main problems of brain tumor segmentation, with healthy vowels accounting for 98% of the total vowels, necrosis, edema, and non-enhanced

represent 0.18% 1.1% and 0.38% respectively. We designed and improved U-Net architecture in this article to solve the problem, consisting of two distinct tracks. These two tracks use small and large convolutional kernels and follow the same encoder/decoder structure. These tracks are defined as the first track and the second track. The receptive field size of the first route is 3×3, and it has five convolutional blocks, while the second track uses a 5×5 convolutional kernels and four convolutional blocks. The proposed two-tracks architecture improves the original U-Net architecture. We apply a batch normalization after each convolution block to preserve regulated gradient levels, accelerate convergence, and minimize the effect of inner shift covariates so that the network parameters do not alter rapidly during back-propagation. The Leaky Rectified Linear Unit (LeakyReLU) activation function with 0.1 leakage factor is used in this architecture. Fig 2 shows the suggested architecture. In the following subsections, the details of the two tracks will be described, and their parameters will be shown in Table 1. The motivation behind the design of this model is to improve the efficiency of the original U-Net model, process different tumor sizes, and reduce the issue of class imbalance. The details of each track are explained in the remainder part of this section.

a: THE FIRST TRACK

The first track's contracting part consists of 5 convolutional blocks. Every block has two convolutional layers. We use the
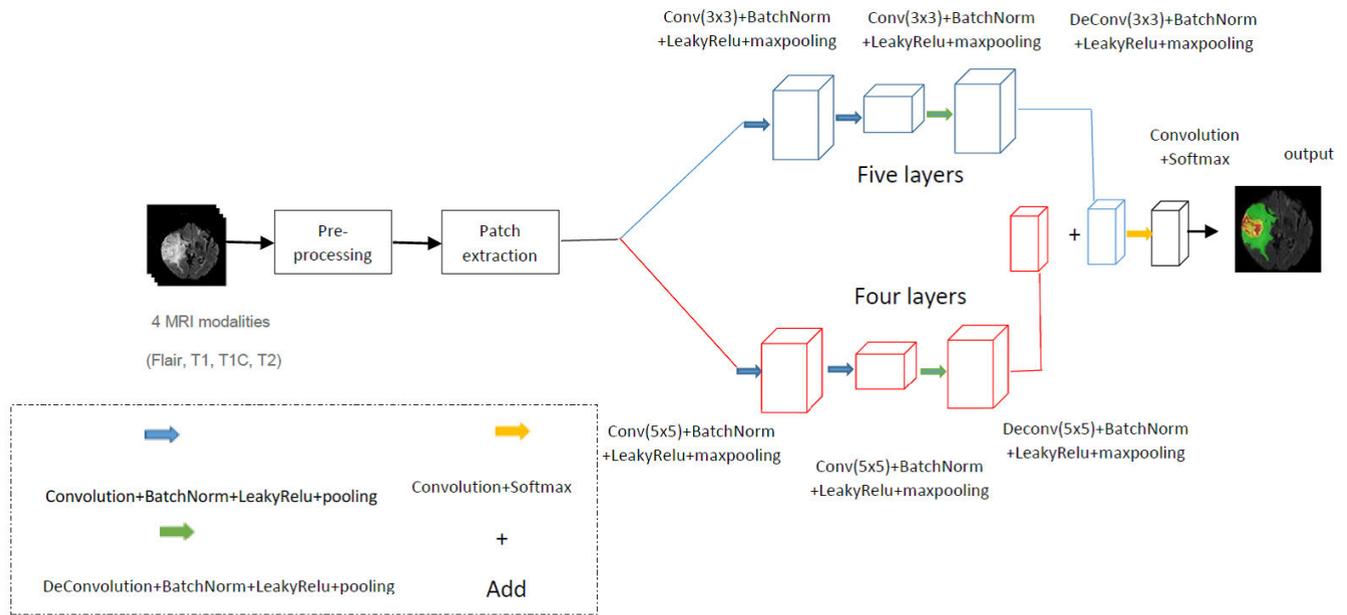
FIGURE 2. The proposed HTTU-Net architecture. The first and second tracks are shown respectively in the colors blue and red.

TABLE 2. Two-Track U-Net parameters setup.

Name	type	output size
First track		128x128x4
Conv1(repeat2 times)	conv+BatchNorm+ leaky Relu	128x128x64
Pooling1	pooling	64x64x64
Conv2(repeat2 times)	conv+BatchNorm+ leakyRelu	64x64x128
Pooling2	pooling	32x32x128
Conv3(repeat2 times)	conv+BatchNorm+leakyRelu	32x32x256
Pooling3	pooling	16x16x256
Conv4(repeat2 times)	conv+BatchNorm+leakyRelu	16x16x512
Pooling4	pooling	8x8x512
Conv5(repeat2 times)	conv+BatchNorm+leakyRelu	8x8x1024
Deconv4	Deconvolution + BatchNorm + leakyReLu	16x16x512
Concate4 (Deconv4 + Conv4)	Concatenate	16x16x1024
Deconv4_1	Convolution + BatchNorm +leaky ReLu	16x16x512
Deconv3	Deconvolution + BatchNorm + leakyReLu	32x32x256
Concate3(Deconv3 + Conv3)	Concatenate	32x32x512
Deconv3_1	Convolution + BatchNorm +leaky ReLu	32x32x256
Deconv2	Deconvolution + BatchNorm + leakyReLu	64x64x128
Concate2(Deconv2 + Conv2)	Concatenate	64x64x256
Deconv2_1	Convolution + BatchNorm +leaky ReLu	64x64x128
Deconv1	Deconvolution + BatchNorm + leakyReLu	128x128x64
Concate1 (Deconv1 + Conv1)	Concatenate	128x128x128
Deconv1_1	Convolution + BatchNorm + leakyReLu	128x128x64
First track_last	Convolution + BatchNorm + leakyReLu	128x128x4
Second track use the parameter but with less layers and different filter size		
Add (First track+ second track)	Add	128x128x8
Conv6	Convolution + softmax	128x128x4

LeakyReLU activation function in the proposed architecture and carry out batch normalization at the end of each block. For all convolutional layers, this track utilizes a 3×3 kernel. The amount of filters for the first, second, third, fourth, and fifth blocks is 64, 128, 256, 512, and 1024. At the end of each block, a max-pooling layer is used to reduce the size by half. Similarly, the expanding part consists of 5 blocks each block starts with a deconvolutional layer with a kernel size of 3×3 and a stride of 2, doubling the size of feature maps, thus increasing the size of feature maps from 8×8 to 128×128. Table 2 shows the parameters for each layer.

b: THE SECOND TRACK

The second track's contracting part consists of 4 convolutional blocks. Each block has two convolutional layers and followed by batch normalization (BN). We use the Leaky ReLU activation function and 5×5 kernel for all layers in this track. The amount of filters for the four blocks is 64, 128, 256, and 512. Similarly, the expanding part includes 4 convolutional blocks of convolutions, and each block begins with a deconvolutional layer of 5×5 kernel size with a stride of 2 and doubling the size of feature maps. The size of feature maps increases from 16×16 to 128×128.
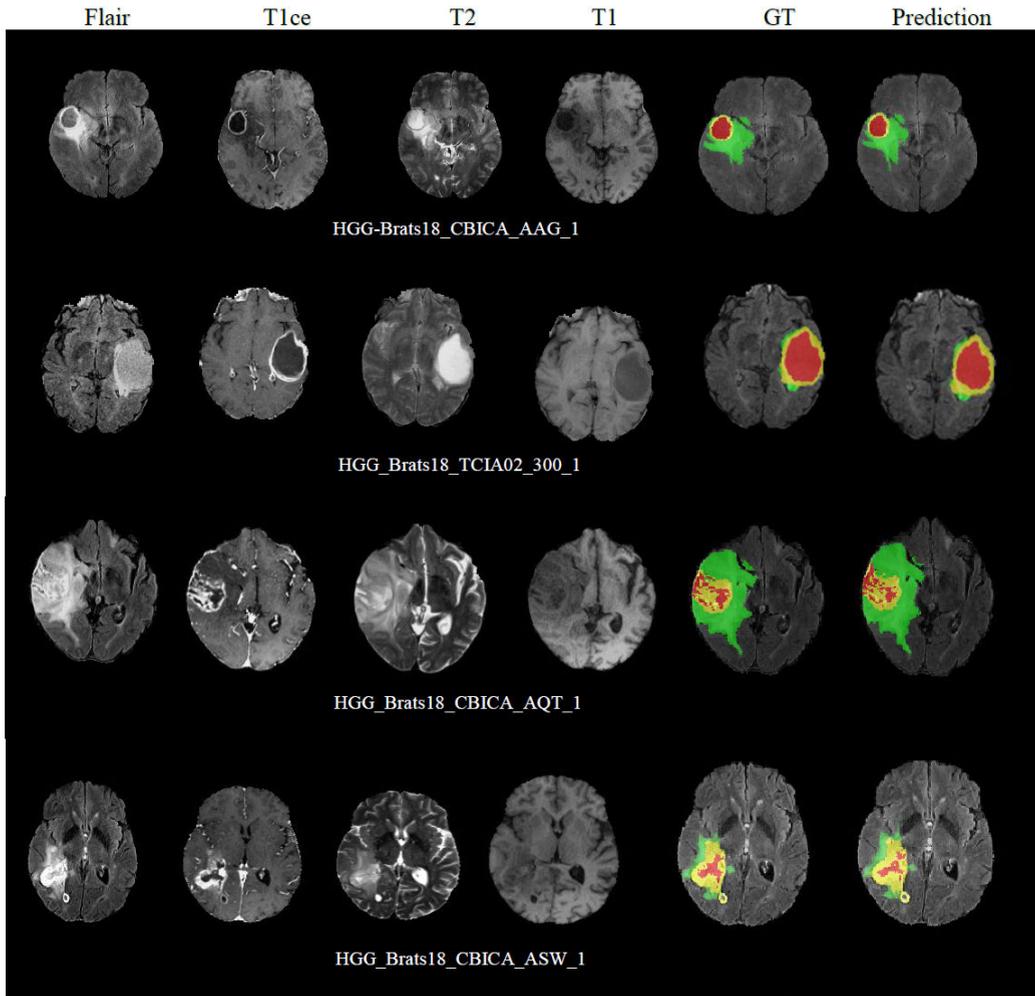
FIGURE 3. Sample segmentation results of four HGG cases from the BraTs'2018 training dataset. Labels are shown in different colors; Green for edema, yellow for enhancing tumor, red for necrosis and non-enhancing.

Finally, we concatenate the output of the first track with the output of the second track, and this concatenation is convolved and then followed by a soft-max function.

C. MODEL TRAINING

Multimodal image batches of size $128 \times 128 \times 4$ are used to train the proposed model. A batch of size 4 is used, and the stochastic gradient-based (SGD) [23] optimization algorithm is used for parameter optimization.

We set the momentum = 0.8. The initial learning rate = 0.0001 is decayed exponentially with decay factor 0.1. We trained the model for 60 epochs because the validation loss has not improved after that. Because the loss functions play an essential role in improving the model accuracy, we implemented a hybrid function as described in the following section.

1) HYBRID LOSS

Ultimately this work illustrates how important the choice of the loss function can be in a deep learning system, mainly when dealing with highly unbalanced problems. The selection of loss functions is also improved model accuracy. Accordingly, we use a hybrid loss function that combines the focal loss function and the Generalized Dice Loss (GDL). The most robust reliability across setups was observed when using GDL. The focal loss function is a modified version of binary cross-entropy and is aimed toward low-confidence labels also achieves state-of-the-art accuracy and speed.

a: THE GENERALIZED DICE SCORE (GDS)

Reference [24] proposed as a multi-class segmentation estimation method. As shown in equation (1);

$$GDL = 1 - 2\frac{(\sum\limits_{i}^{L} W_i \sum\limits_{i} g_{ik}\, p_{ik})}{(\sum\limits_{i}^{L} W_i \sum\limits_{i} g_{ik} + p_{ik})} \qquad (1)$$

where L is the total number of labels, k denotes the batch size, W_i represent the weight assigned to the ith label. As proposed
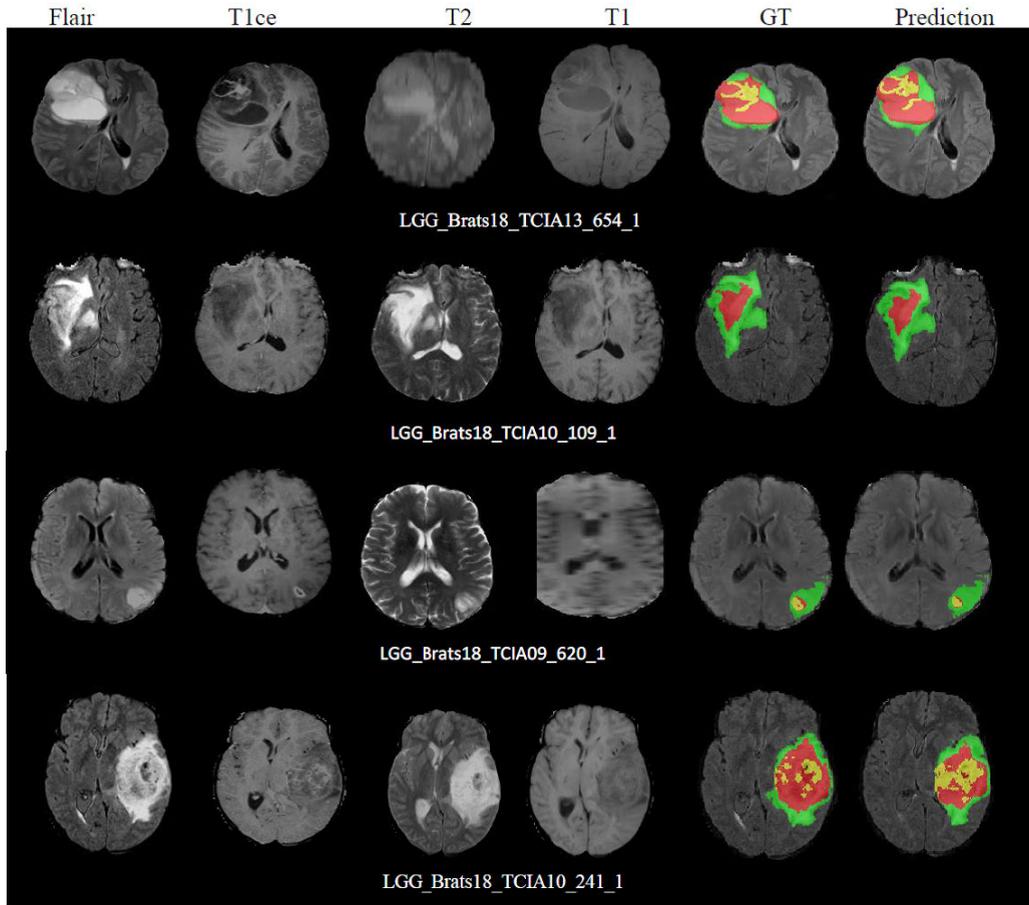
FIGURE 4. Sample segmentation results of four LGG cases from the BraTS'2018 dataset. Labels are shown in different colors; Green for edema, yellow for enhancing tumor, red for necrosis, and non-enhancing.

in [21], we set $W_i = \frac{1}{\sum_k g_{ik}}$. p_{ik} and g_{ik} representing the value of the (i^{th}, k^{th}) pixel of the segmented binary image and image of binary ground truth.

b: THE FOCAL LOSS
For the segmentation task, we select a multiclass focal loss [25] as shown in equation (2) :

$$Focal(p, q) = \frac{\sum\limits_{x,y,z}\sum\limits_{k} p_{x,y,z}^k (1 - q_{x,y,z}^k)^\gamma \log(q_{x,y,z}^k)}{\sum\limits_{k} p_{x,y,z}^k} \quad (2)$$

where $(1 - q_{x,y,z}^k)^\gamma$ is a modulating factor, The focusing parameter γ smoothly adjusts the rate at which easy examples are down-weighted. The setting of $\gamma > 0$ can reduce the relative loss for well-classified samples and putting a focus on hard and misclassified samples. In contrast, the focal loss is equal to the original cross-entropy loss when $\gamma = 0$ (we found $\gamma = 2$ to work best in our experiments). And p is the model's estimated probability for the class. The proposed hybrid loss HL function is shown in equation (3):

$$HL = GDL + FL \quad (3)$$

IV. EXPERIMENTAL RESULTS
A. EVALUATION METRICS
We evaluate the segmentation results using four metrics; Dice similarity coefficient (DSC), Hausdorff distance (HD), Sensitivity, and Specificity. The Dice similarity score measures the similarity between the ground truth and the resulted segmentation mask. It is defined in Equation (4):

$$DSC = \frac{2TP}{(FP + 2TP + FN)} \quad (4)$$

where TP, FP, and FN are the number of true positive, false positive, and false negatives, respectively. Sensitivity is also defined as a true positive rate as in equation (5). Specificity is also used to calculate the true negative rate as in equation (6).

$$Sensitivity = \frac{TP}{TP + FN} \quad (5)$$

$$Specificity = \frac{TP}{TP + FP} \quad (6)$$

Hausdorff Distance indicates the tumor border segmentation quality. It calculates the maximum distance between any two surfaces, A and B, as in equation (7).

$$h(A, B) = max_{a \in A} min_{b \in B} d(a, b) \quad (7)$$

TABLE 3. Quantitative result of segmentation of BraTS'2018 training dataset using Dice and Sensitivity metrics.

	Dice			Sensitivity		
	ET	WT	TC	ET	WT	TC
Mean	0.741	0.852	0.812	0.768	0.885	0.820
Std.Dev.	0.282	0.110	0.252	0.269	0.150	0.269
Median	0.801	0.879	0.878	0.859	0.908	0.874
25 quantile	0.743	0.840	0.784	0.752	0.868	0.803
75quantile	0.846	0.905	0.897	0.951	0.954	0.948

TABLE 4. Quantitative result of segmentation of BraTS'2018 training dataset using Specificity and Hausdorff distance metrics.

	Specificity			Hausdorff95		
	ET	WT	TC	ET	WT	TC
Mean	0.999	0.998	0.997	3.304	8.250	3.301
Std.Dev.	0.042	0.023	0.042	1.824	3.315	1.872
Median	0.999	0.999	1	3.211	8.401	3.251
25 quantile	0.999	0.998	1	2.503	6.172	4.103
75quantile	0.999	0.999	1	4.924	10.67	4.254

where a and b are the set of points in A and B, respectively. And, d(a,b) is Euclidean metric between these points [26].

B. PERFORMANCE ON BraTs'2018 TRAINING DATASET

In our experiments, 160 subjects from Brats training dataset is used for training and 40 subjects for validation purposes. We extract 25,000 multimodal patches from each case to form the 4,000,000 patches training set. We carry out our experiments using Keras framework TensorFlow backend [27]. The training was carried out on an Intel Corei7 3.5GHz machine using NVIDIA GeForce GTX 1070. The segmentation results for eight cases are shown in Figures 3 and 4.

Fig 3 shows the results of four HGG tumor samples, while Fig 4 shows the results of four LGG tumor samples. In these figures, columns one to four display one axial slide from Flair, T1c, T2, and T1 modalities, respectively. The fifth column shows the ground truth (GT) labels, and sixth columns display the segmentation results, where intratumor areas can be distinguished by color code: yellow for enhancing tumor, green for edema and necrotic, and red for non-enhancing. Tables 3 and 4 presents the quantitative evaluation of segmentation results where Tumor core (TC) is the union of necrosis & non-enhancing tumor and enhancing tumor (ET). A whole tumor (WT) is the union of necrosis & non-enhancing tumor, edema, and enhancing tumor. Mean, standard deviation, median, and 25th and 75th percentile are given for Dice and Sensitivity metrics in Table 3 and Specificity, and Hausdorff distance in Table 4.The quantitative results presented in this section indicate that the proposed HTTU-Net model produces accurate segmentation results. It achieved a Dice similarity coefficient of 0.852, 0.812, and 0.741 for the segmentation of the whole tumor, core tumor, and enhancing tumor, respectively.

C. PERFORMANCE ON BraTs'2018 TESTING

60 subjects from BraTs 2018 dataset are used for testing purposes. We present the Dice similarity and Sensitivity metrics for these cases in Table 5 and the Specificity and Hausdorff distance in Table 6. In Table 5, the method obtained a mean
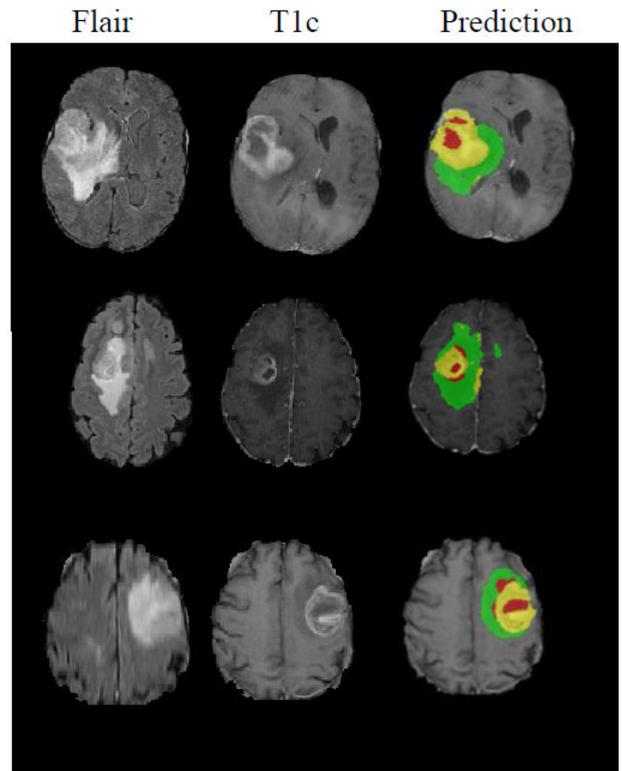

FIGURE 5. Sample segmentation of cases from the BraTS 2018 validation: the 74*th*, 10*th*, and 111*th*slices from the subject *Brats*18_*TCIA*02_230_1, *Brats*18_*CBICA_BLK*_1, *Brats*18_*TCIA*07_601_1. Labels are shown in different colors; Green for edema, yellow for enhancing tumor, red for necrosis and non-enhancing.

TABLE 5. Quantitative segmentation results for testing on BraTS 2018 training Set using Dice and Sensitivity metrics.

	Dice			Sensitivity		
	ET	WT	TC	ET	WT	TC
Mean	0.745	0.865	0.808	0.78	0.883	0.80
Std.Dev.	0.211	0.103	0.223	0.286	0.151	0.263
Median	0.815	0.883	0.895	0.868	0.894	0.882
25 quantile	0.76	0.858	0.77	0.703	0.87	0.77
75quantile	0.887	0.915	0.923	0.943	0.941	0.972

TABLE 6. Quantitative segmentation results for testing on BraTS 2018 training Set using Specificity and Hausdorf distance metrics.

	Specificity			Hausdorff95		
	ET	WT	TC	ET	WT	TC
Mean	0.999	0.999	0.998	4.43	7.53	8.811
Std.Dev.	0.053	0.031	0.032	2.441	3.461	2.961
Median	0.999	0.999	0.999	4.29	5.871	7.12
25 quantile	0.999	0.998	0.998	3.20	3.76	5.55
75quantile	1	0.999	0.999	5.09	5.09	9.05

ET, WT, and TC dice score of 0. 745, 0.865, and 0.808, respectively.

In Table 6, average HD scores of 4.43, 7.53and 8.81 for ET, WT, and TC, respectively, were obtained. The evaluation of our algorithm's performance on the Brat's 2018 validation is presented in Fig 5. We can observe that performance is consistent in both training and testing cases, indicating that this model is well generalized to invisible examples.
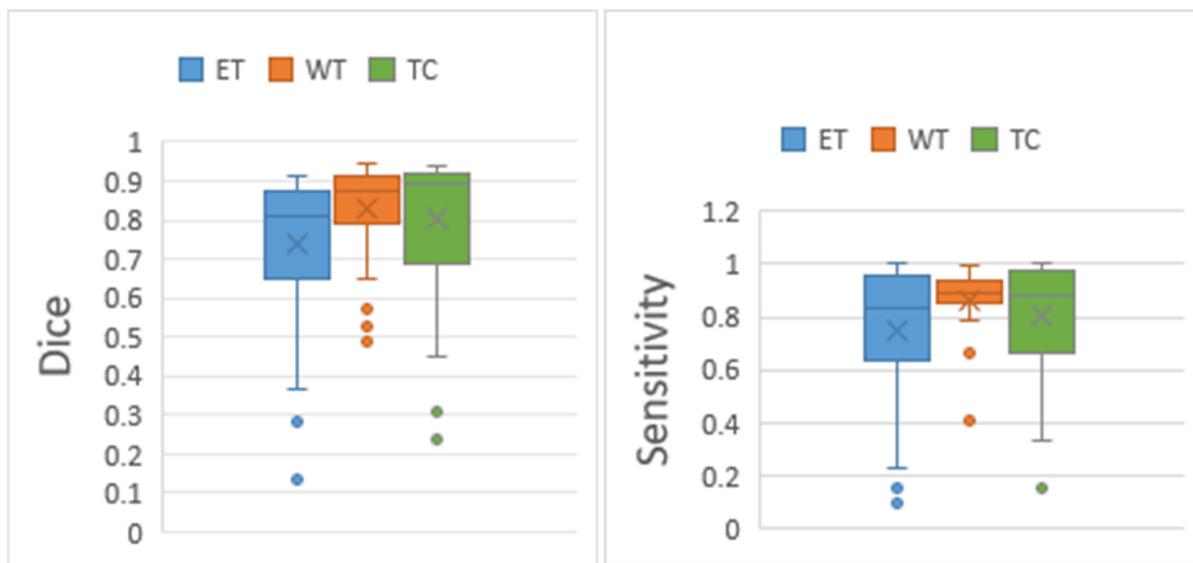
FIGURE 6. Boxplots of DSC and Sensitivity obtained from BraTS'2018. The 'x' marks the mean score, " marks outliers.
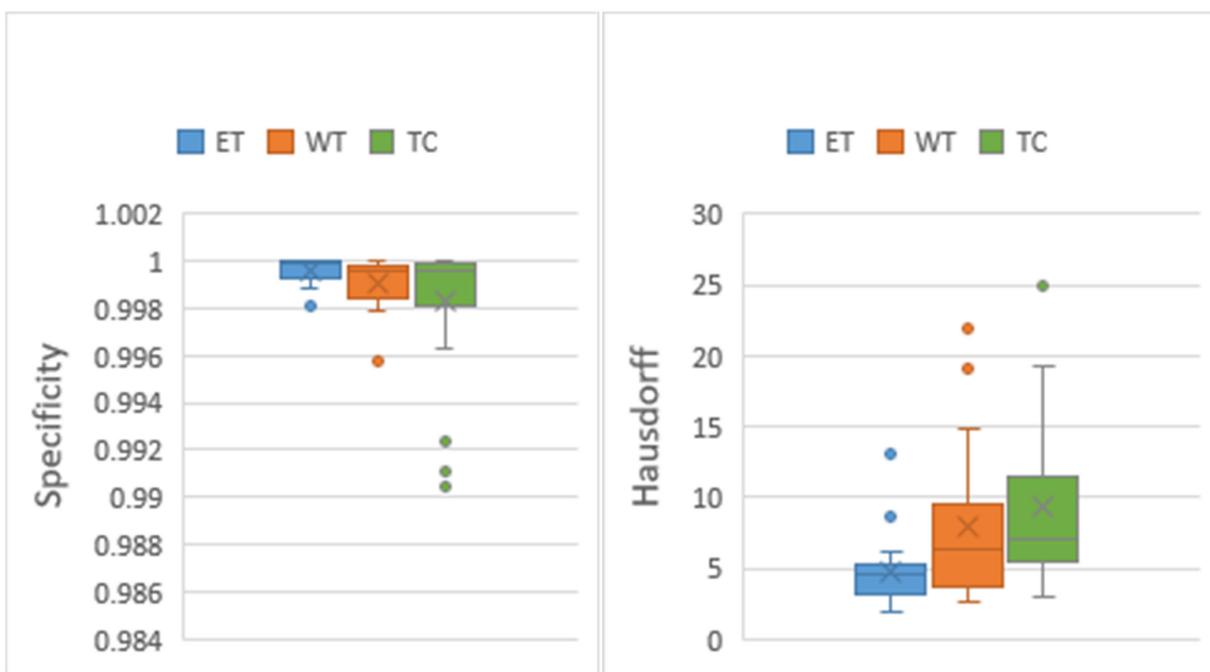

FIGURE 7. Boxplots of Specificity and Hausdorff obtained from BraTS'2018. The 'x' marks the mean score, " marks outliers.

Table.7 shows a comparison between the performance of the proposed HTTU-Net model, the original U-net architecture, the first track, and second track models. Boxplots of Dice dispersion and Sensitivity are shown in Fig 6 and Specificity, and HD scores dispersion boxplots are shown in Fig 7. The boxplots plots report the minimum, lower quartile, median, upper quartile and maximum. Points that fall outside the interquartile range are considered outliers. It is evident from the boxplots that in most cases our algorithm produces significantly good segmentation accuracy.

V. DISCUSSION

The precise segmentation of gliomas has drawn considerable interest from medical doctors and researchers as a critical component of tumor detection, treatment preparation. Since manual segmentation of tumor regions is exhausting, and time ravage, it is important to develop effective computational methods for the segmentation of automatic brain tumor. So, most deep learning methods are proposed for brain tumor segmentation to solve the class imbalance problem. The suggested method provides more reliable output

TABLE 7. Comparison of our proposed model with one –track model.

Methods	ET	WT	TC
Original U-Net	0.69	0.852	0.794
First track	0.739	0.850	0.80
Second track	0.732	0.859	0.792
HTTU-Net	0.745	0.865	0.808

in segmentation than most previous methods. Accordingly, the results Fig. 3, and Fig. 4 indicate that the size, shape, location, and intensity of tumors in these eight samples are different, and also enhance the segmentation performance for small tumor regions. Generally, the results of the proposed segmentation architecture are comparable to those acquired by the experts (GT). In Table 5, Table 6. It is also observed that the small enhancements on the testing based on Brats 2018 training set are due to the reality that it contains more HGG tumor samples than LGG tumor samples. In Table 7. we use two tracks with small and large convolutional kernels sizes to investigate the impact of different parameters; the result reveals that the HTTU-Net model has achieved the best performance among all compared models. It increases 0.055 for ET, 0.013 for WT, and 0.014 for TC values over the original U-Net, it also increases 0.006for ET, 0.015 for WT, and 0.008 for TC values over the first track model and increases 0.013 for ET, 0.006 for WT, and 0.016 for TC values over the second track model. Due to the multi-modality nature of MRI and memory limitations of the current GPU,it is worth mentioning that in the proposed method, the training stage is time-consuming, which could be considered as a limitation. The training time for one epoch is around 10 min, but also the prediction phase rapidly processes the testing dataset.

VI. CONCLUSION AND FUTURE WORK
In this paper, we introduced an automatic approach for brain tumor segmentation using 2D HTTU-Net architecture. The proposed technique has been quantitatively evaluated using the BraTS'2018 dataset. It contains two tracks; each one consists of a different number of convolution blocks and uses a different kernel size to handle the different tumor sizes. We also developed a new hybrid loss function to alleviate the class imbalance problem by combining the focal loss and Generalized Dice Loss functions. Higher performance is achieved through HTTU-Net architecture, which solves brain tumors segmentation problems that can happen anywhere in the brain, in almost any type and size. The evaluation of the proposed approach verifies that our results are very comparable to those obtained manually by experts. In the end, a future work possibility may concentrate on 3D HTTU-Net.

REFERENCES
[1] D. N. Louis, A. Perry, G. Reifenberger, A. von Deimling, D. Figarella-Branger, W. K. Cavenee, H. Ohgaki, O. D. Wiestler, P. Kleihues, and D. W. Ellison, "The 2016 world health organization classification of tumors of the central nervous system: A summary," *Acta Neuropathologica*, vol. 131, no. 6, pp. 803–820, Jun. 2016.

[2] 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, Cancer J. Clinicians*, vol. 60, no. 3, pp. 166–193, 2010.

[3] G. Tabatabai, R. Stupp, M. J. van den Bent, M. E. Hegi, J. C. Tonn, W. Wick, and M. Weller, "Molecular diagnostics of gliomas: The clinical perspective," *Acta Neuropathologica*, vol. 120, no. 5, pp. 585–592, Nov. 2010.

[4] 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.

[5] X. Zhao, Y. Wu, G. Song, Z. Li, Y. Zhang, and Y. Fan, "A deep learning model integrating FCNNs and CRFs for brain tumor segmentation," *Med. Image Anal.*, vol. 43, pp. 98–111, Jan. 2018.

[6] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.

[7] B. Hariharan, P. Arbeláez, R. Girshick, and J. Malik, "Simultaneous detection and segmentation," in *Proc. Eur. Conf. Comput. Vis.*, in Lecture Notes in Computer Science, vol. 8695. Springer, 2014, pp. 297–312.

[8] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, in Lecture Notes in Computer Science, vol. 9351. Springer, 2015, pp. 234–241.

[9] O. Cicek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: Learning dense volumetric," *Medical Image Computing and Computer-Assisted Intervention MICCAI*, vol. 2016, pp. 424–432, 2016.

[10] P. Dvorak, "Structured prediction with convolutional neural networks for multimodal brain tumor segmentation structured prediction with convolutional neural networks for multimodal brain tumor segmentation," in *Proc. Int. MICCAI Workshop Med. Comput. Vis.*, Spring/Nov. 2015, pp. 59–71.

[11] S. Pereira, A. Pinto, V. Alves, and C. A. Silva, "Brain tumor segmentation using convolutional neural networks in MRI images," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1240–1251, May 2016.

[12] A. L. Maas, A. Y. Hannun, and A. Y. Ng, "Rectifier nonlinearities improve neural network acoustic models," in *Proc. ICML Work. Deep Learn. Audio, Speech Lang. Process.*, vol. 28, p. 3, Jun. 2013.

[13] 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," *Med. Image Anal.*, vol. 35, pp. 18–31, Jan. 2017.

[14] L. Zhao and K. Jia, "Multiscale CNNs for brain tumor segmentation and diagnosis," *Comput. Math. Methods Med.*, vol. 2016, pp. 1–7, Mar. 2016.

[15] H. T. Le and H. T.-T. Pham, "Brain tumour segmentation using U-Net based fully convolutional networks and extremely randomized trees," *Vietnam J. Sci., Technol. Eng.*, vol. 60, no. 3, pp. 19–25, Sep. 2018.

[16] G. Wang, W. Li, S. Ourselin, and T. Vercauteren, "Automatic brain tumor segmentation using cascaded anisotropic convolutional neural networks," in *Proc. Int. MICCAI Brainlesion Workshop*, in Lecture Notes in Computer Science, vol. 10670. Springer, 2017, pp. 178–190.

[17] H. Dong, G. Yang, F. Liu, Y. Mo, and Y. Guo, "Automatic brain tumor detection and segmentation using U-net based fully convolutional networks," in *Proc. Annu. Conf. Med. Image Understand. Anal.*, in Communications in Computer and Information Science, vol. 723. Springer, 2017, pp. 506–517.

[18] "W. Chen, B. Liu, S. Peng, J. Sun, and X. Qiao, "S3D-UNET: Separable 3D U-Net for brain tumor segmentation," in *Proc. 4th Int. Workshop, BrainLes, Held Conjunct. (MICCAI)*, in Lecture Notes in Computer Science), vol. 11384. Granada, Spain: Springer, 2019.

[19] H. He and L. Fang, "Three pathways U-Net for brain tumor segmentation," in *Proc. Pre-Conf. Int. MICCAI BraTS Challenge*, 2018, pp. 119–126.

[20] (2018). *Multimodal Brain Tumor Segmentation Challenge*. Accessed: Nov. 1, 2018. [Online]. Available: https://www.med.upenn.edu/sbia/brats2018.html

[21] N. J. Tustison, B. B. Avants, P. A. Cook, Y. Zheng, A. Egan, P. A. Yushkevich, and J. C. Gee, "N4ITK: Improved n3 bias correction," *IEEE Trans. Med. Imag.*, vol. 29, no. 6, pp. 1310–1320, Jun. 2010.

[22] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017.

[23] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. 3rd Int. Conf. Learn. Represent.*, San Diego, CA, USA, 2014, pp. 1–15.

[24] C. H. Sudre, W. Li, T. Vercauteren, S. Ourselin, and M. J. 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* (Lecture Notes in Computer Science), vol. 10553. Springer, 2017, pp. 240–248.

[25] T.-Y. Cl:Aire Lin, P. Goyal, R. Girshick, and K. P. He & Dollar, "Focal loss for dense object detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 2, pp. 318–327, Feb. 2020, doi: 10.1109/TPAMI.2018.2858826.

[26] A. Kumarasiri, F. Siddiqui, C. Liu, R. Yechieli, M. Shah, D. Pradhan, H. Zhong, I. J. Chetty, and J. Kim, "Deformable image registration based automatic CT-to-CT contour propagation for head and neck adaptive radiotherapy in the routine clinical setting," *Med. Phys.*, vol. 41, no. 12, Nov. 2014, Art. no. 121712.

[27] M. Abadi *et al.*, "Tensorflow: A system for large-scale machine learning," in *Proc. 12th USENIX Symp. Oper. Syst. Design Implement. (OSDI)*, 2016, pp. 265–283.

NAGWA M. ABOELENEIN received the M.Sc. degree from the Information Technology Department, Faculty of Computers and Information, Menoufia University, Egypt, in 2016. She is currently pursuing the Ph.D. degree with the School of Computer Science and Technology, Harbin Institute of Technology, Harbin, China. Since October 2009, she has been working as a Demonstrator and an Assistant Lecturer with the Faculty of Computers and Information, Menoufia University. Her research interests include computer vision, image segmentation, feature extraction, artificial intelligence, and machine learning.

PIAO SONGHAO received the Ph.D. degree in computer science from the Harbin Institute of Technology (HIT), Harbin, China, in 2004. He is currently a Professor with the Department of Computer Science, Harbin Institute of Technology. His research interests include intelligent robot, computer vision, and robot navigation.

ANIS KOUBAA is currently a Professor of computer science, an Advisor to the Rector, and the Leader of the Robotics and Internet of Things Research Laboratory, Prince Sultan University. He is also Research and Development Consultant with Gaitech Robotics, China, and a Senior Researcher with CISTER/INESC TEC and ISEP-IPP, Porto, Portugal.

ALAM NOOR was born in Khyber Pakhtoon Khwa, Pakistan, in 1992. He received the B.S. degree in electronics from the University of Peshawar, Pakistan, and the M.S. degree in information and communication engineering from the Harbin Institute of Technology (HIT). He is currently an Artificial Intelligence Researcher with the Robotics and Internet of Things Laboratory, Prince Sultan University, Riyadh, Saudi Arabia. His research interests include digital signal and image recognition, facial expression, deep learning, machine learning, and data processing for real-time applications.

AHMED AFIFI received the master's degree from Menoufia University, Egypt, in 2005, and the Ph.D. degree from Chiba University, Japan, in 2011. He is currently an Assistant Professor with the College of Computer Science and Information Technology, King Faisal University (KFU), Saudi Arabia. Before joining KFU, in 2018, he was an Assistant Professor with Menoufia University, Egypt, from 2011 to 2018. He got two postdoctoral fellowships at the Center of Frontier in Medical Engineering, Chiba University, Japan, funded by the Egyptian Government and the Japan Society for the Promotion of Science (JSPS). He has published many conferences and journal papers on topics related to medical image analysis, computer vision, and machine learning. He has participated in many international scientific competitions in medical imaging. His research interests include image/video analysis and understanding, computer vision, virtual reality, surgery planning, artificial intelligence, and machine learning.

· · ·