

Research Article

DRT-Unet: A Segmentation Network for Aiding Brain Tumor Diagnosis

Shujing Li and Linguo Li 

College of Information Engineering, Fuyang Normal University, Fuyang 236041, China

Correspondence should be addressed to Linguo Li; llg-1212@163.com

Received 27 May 2022; Revised 8 August 2022; Accepted 22 August 2022; Published 9 September 2022

Academic Editor: Beijing Chen

Copyright © 2022 Shujing Li and Linguo Li. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Using image segmentation techniques to assist physicians in brain tumor diagnosis is a hot issue in computer technology research. Although most brain tumor segmentation networks to date have been based on U-Net, the prediction results are depending on which are not well generalized and need to be further improved. As the depth of the network increases, the gradients of the network vanish together with the decrease of the accuracy; meanwhile, the large number of parameters in the network will cause data redundancy. Moreover, a single modality of MRI images cannot adequately segment tumor details. Therefore, a segmentation network with an improved U-Net model is proposed in this paper, which combines Dilated Convolution-Dense Block-Transformation Convolution-Unet (hereafter referred to as DRT-Unet). The network adopts the combination of dilated convolution, dense residual block, and transposed convolution. In the coding process, a dilated convolution block and a local feature residual for fusing dense block are adopted to replace the 3×3 convolution layers on each layer in U-Net, and a transition layer is used for down-sampling. In the decoding process, a local feature residual is adopted for fusing dense blocks; meanwhile, a deconvolution structure with up-pooling and transposed convolution cascade is used. By connecting the decoded output features with the encoded low-level visual features, the information on transition layer loss is obtained. The experiments in this paper are carried out on BraTs2018 and BraTs2019 datasets; as a result, the DRT-Unet network can effectively segment tumor lesion regions.

1. Introduction

Brain tumors are a general term for tumors of the nervous system that grow inside the skull, second only to tumors of the lung, stomach, uterus, and breast, accounting for approximately 5% of systemic tumors, 70% of childhood tumors [1], and more than 2.4% of deaths [2]. Magnetic resonance imaging (MRI) is one of the most commonly used diagnostic techniques in clinical care, which is particularly important in the diagnosis of brain tumors. It is noninvasive, accurately providing the shape, size, and location of the brain tumors without the patients receiving high ionizing radiation, as well as having good soft tissue contrast [3]. Accurate segmentation of brain tumors is of essential importance for disease diagnosis, pathological research, and later surgical plan determination.

Although accurate segmentation of brain tumors is required in clinical research, it is usually filled with challenges, mainly including image artifacts, noise, and low contrast, as well as considerable variations in tumor shape, size, and location from case to case. What is more, the segmentation of brain tumors is more challenging since the boundaries between the structures of brain tumors are fuzzy and the internal structures are similar. Manual segmentation of the brain tumors, which depends on the doctor's expertise and experience, is quite cumbersome. Therefore, the study of a method that can automatically, accurately, and effectively segment brain tumors is of great significance for clinical diagnosis and surgery.

In recent years, deep learning has achieved a series of successes in many fields such as image, audio, and natural language. Among them, convolutional neural networks

(CNNs) have important references in computer vision tasks [4–9], and significant progress has been made in semantic image segmentation [10, 11]. CNNs learn visual and semantic features in images during the training process, reducing the complexity of the network model and making it possible to train networks in depth. In summary, deep CNNs have a wide range of applications in medical image processing [12–14]. According to the different input and output methods, the image segmentation method based on deep learning can be divided into block segmentation and end-to-end segmentation, and the latter is mainly realized through the encoder-decoder structure. The complete image or image block is input, and the type probability of each pixel in the output image is decoded, so as to achieve the purpose of tumor region segmentation. The relevant model of this method is mainly based on the U-Net network, which proposes a symmetric structure with jump connections to retain image details, becoming the mainstream framework for most image segmentation tasks. Although the improved method based on U-Net improves the segmentation performance, there is still room for improvement in network depth and generalization. In recent years, the concept of identity mapping has been introduced to balance the depth and performance of the network. However, the use of residual blocks to adjust the number of channels makes the number of channels increase dramatically, resulting in data redundancy. Unimodal MRI images cannot complete the full segmentation of tumor-related areas and details, and the use of multimodal brain tumor images can make up for the above weaknesses.

In this paper, the characteristics and performance of each model are combined and a multimodal brain tumor segmentation method is proposed for the DRT-Unet network, which is similar to U-Net in the overall framework. The exact contributions of this paper are as follows:

- (1) The ordinary convolution with dilated convolution is combined to expand the sensory field and optimize the feature extraction capability. While introducing two mapping methods, 3×3 ordinary convolution and dilated convolution in parallel can obtain a sensory field larger than 9 frames, with a greater sensory field and better feature extraction capability. Each pixel in the output feature map can respond to a larger area in the image.
- (2) The dense block used in this paper consists of a dense layer and a residual fusion of local features. The “jump connection” of ResNets is introduced in the down-sampling process, which is combined with the dense block to preserve and propagate the rich low-level visual features [15], such as brain tumor brightness, color, texture, and other features that directly stimulate vision.
- (3) In the decoding process, dense blocks are fused by using local feature residuals to form a cascaded deconvolution structure, so that the output image has the same dimension as the input image, while in the decoding process, low-level visual features from

the encoding process are connected with the same dimension and channels; meanwhile, features are fused to obtain the missing information after the transition layer in the encoding path.

2. Related Work

Currently, most deep CNNs used for brain tumor segmentation networks are end-to-end. End-to-end brain tumor segmentation networks use an encoding-decoding approach where the input is the whole image or image block. The features are extracted by encoding in the convolutional layer, which is then decoded to obtain the class probability of each pixel point in the whole image or image block finally. Such a segmentation method is mainly based on FCN [16] and U-Net networks [17]. Raza et al. [18] proposed a hybrid model based on a deep residual network and U-Net, which takes the residual network as the encoder to deal with the problem of gradient disappearance, as well as uses low-level and high-level features to predict. Nevertheless, this method ignores the context information, resulting in high computational costs. Zhang et al. [19] proposed a multi-scale mesh aggregation network. By introducing an improved inception module to replace the standard convolution, effective information is extracted and aggregated from different receptive fields, and the network aggregation strategy is adopted to gradually refine shallow features. However, the number of network parameters is large, accompanied by low segmentation efficiency. Chen et al. [20] proposed a symmetric network based on a deep convolution neural network, which expanded the functional mapping between low-level and high-level features by adding symmetric masks in multiple layers, and combined the prior knowledge of symmetry with brain tumor segmentation; however, the effect of low-contrast tumor segmentation was poor. Wang et al. [21] proposed a segmentation network based on a segmented attention module, which extracted useful information in connected features through different attention mechanisms and discarded redundant information to realize selective aggregation of features. What is more, Wang et al. [22] proposed extracting multi-scale image features by using a spatial module composed of multiple parallel dilated convolution layers and deepening the network structure by using a residual module. Shen et al. [23] proposed a multi-task full convolutional network for the automatic segmentation of brain tumors. Based on the hierarchical relationship between tumor substructures, the network takes multimodal MRI images and their symmetric differential images as inputs to extract multi-level background information. Experiments showed that the proposed multi-task FCN outperformed single-task FCN for all subtasks. However, there were limitations in the FCN-based approach for predicting low-resolution images [24]. Based on U-Net network, Hao et al. [25] proposed a new network for brain tumor segmentation. They used a comprehensive data enhancement scheme to preprocess the data and conducted the experiments with BraTs2015 dataset. The DSC values obtained in the intact tumor region, the core tumor region, and the enhanced tumor region were 0.86, 0.86, and 0.65,

respectively. Although the performance of the network can be improved as more layers are added to the network, degradation and gradient disappearance can occur with the deepening of the network. The ResNets proposed by He et al. [26] in 2015 cope with this problem by introducing the concept of residual blocks. ResNets perform well in a range of image recognition, localization, and detection tasks, such as ImageNet and COCO object detection. The literature [24] proposes RefineNet, a generalized multi-path optimization network whose components are connected by using residuals according to the idea of identity mapping, so as to achieve efficient end-to-end training. Experimentally, the method proved to improve the performance of the segmentation. However, the addition of a multi-path refinement network on ResNets increased the parameters of that network as well.

DenseNet [27], as the best paper of CVPR2017, does not take deepening and widening the network as the two ways to improve the performance of the network, but considers the feature perspective, greatly reducing the number of parameters as well as alleviating the problem of gradient disappearance through feature reuse and bypass settings. The authors in the literature [28] proposed a fully convolutional network for semantic segmentation, i.e., FC-DenseNet, by fusing dense blocks in DenseNet with jump connections of ResNets. Kaku et al. [29] proposed a brain tumor segmentation network named DenseUnet by incorporating the dense block structure into U-Net and conducted experiments in Mindboggle-101 and New York University (NYU) artificial correction dataset. The best Dice values of 0.819 ± 0.011 and 0.800 ± 0.012 were obtained, respectively, which were better than the segmentation performance of U-Net.

3. Network Model

On the basis of the advantages and disadvantages of FCN, U-Net, ResNets, and DenseNet and the computational principles of CNNs, this paper proposes a segmentation network of DRT-Unet, whose network structure is shown in Figure 1. The DRT-Unet network is similar to U-Net in the overall framework; meanwhile, a dense fusion of dilated convolution blocks, as well as local feature residuals, is used in the encoding process. In the coding process, the dilated convolution block and the local feature residual fusion dense block are used instead of two repetitive 3×3 convolution layers in U-Net: the dilated convolution block consists of dilated convolution and normal convolution, which can expand the perceptual field without losing local information. An $X \times X$ convolution layer can make the value of each pixel feel an area of X^2 size; for example, a 3×3 convolution layer can obtain the receptive field of 9 lattice size, but the parallel connection of ordinary convolution and dilated convolution can not only obtain receptive fields larger than 9 lattices, but also introduce two mapping methods at the same time. Therefore, the combined hole-convolutional block has larger receptive fields and better feature extraction ability. The dense block in this paper is composed of a dense layer and residual fusion of local features, and the identity mapping of

ResNet is connected to the dense block and the coding process, so as to retain and spread more low-level visual features. The transition layer is used for down-sampling in the coding process. The decoding process is implemented by using a deconvolution structure with local feature residual fusion dense blocks and up-pooling and transposed convolution cascades, while the decoding process connects with the lower-level visual characteristics in the encoding process with the same dimension and number of channels, and the features with high-level semantic information are then fused to generate new features to obtain the missing information after the transition layer in the encoding path. (Since then, the deconvolution in the following refers to the cascading operation of up-pooling and transposed convolution in the decoding network).

It is required that the feature maps remain the same size in the same dense block, so only a transition layer between different dense blocks is implemented for down-sampling. To further decrease the network parameters, a 1×1 convolution operation is inserted between every two dense blocks. The transition layer is the TD (transition down) module in Figure 1, whose specific structure is BN + Conv (1×1) + 2×2 max-pooling. The number of layers in the down-sampling part of the network is set according to the number of layers in the first four dense blocks in FC-DenseNet.

3.1. Dilated Convolution Block. Deep CNNs usually use down-sampling or convolutional layers to enhance the perceptual field of the network, which, however, will reduce the spatial resolution. In order to be able to balance both resolution and perceptual field, the literature [30] proposes dilated convolution, also known as dilated convolution. Therefore, in this paper, the coding process of the network uses the dilated convolution at each layer instead of the twice repeated 3×3 convolution used in the U-Net network.

The computational effort of the dilated convolution is comparable to that of the conventional convolution, except that the sampling density of data is changed. However, in the results of a layer obtained by using the dilated convolution, the neighboring pixels are obtained from the convolution of independent subsets. There is a lack of correlation between them as well as a problem of local information loss, while the network model in this paper adopts the dilated convolution and normal convolution in parallel to obtain this information and solve the problem of lack of correlation between the convolution results. Input image or image block, a 3×3 dilated convolution layer, and a 3×3 ordinary convolution layer are used to extract features under different receptive fields, and the feature extraction results under two different mapping methods are obtained. Two different feature extraction result matrices are spliced, and the combined hole convolution block output results are obtained through the activation function. Compared with 3×3 ordinary convolution, the result has a larger receptive field and richer contextual information. The structure of the dilated convolution block used in this paper is shown in Figure 2.

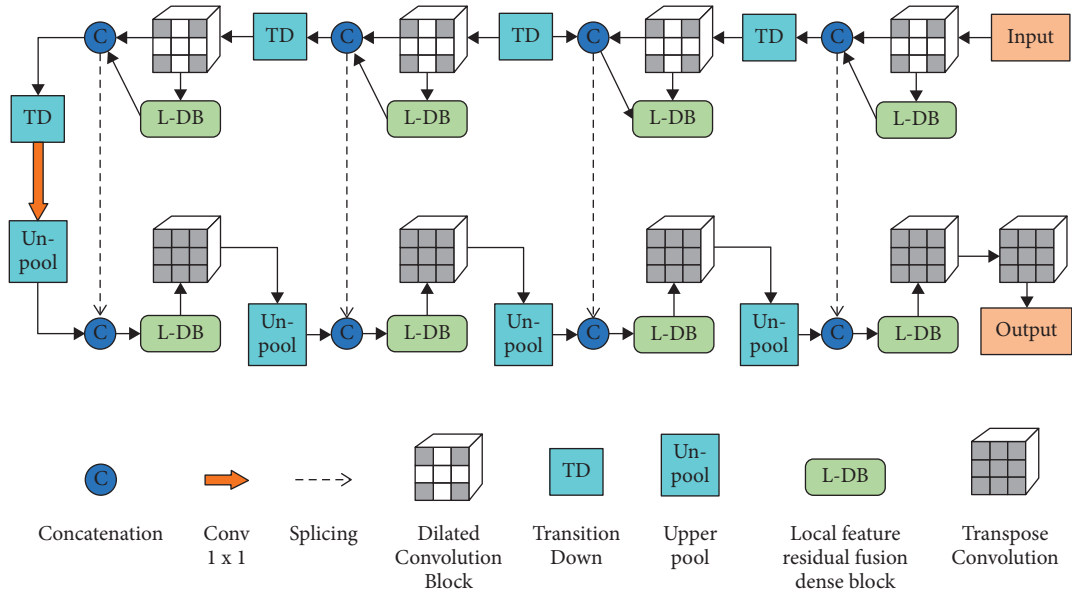


FIGURE 1: DRT-Unet structure diagram.

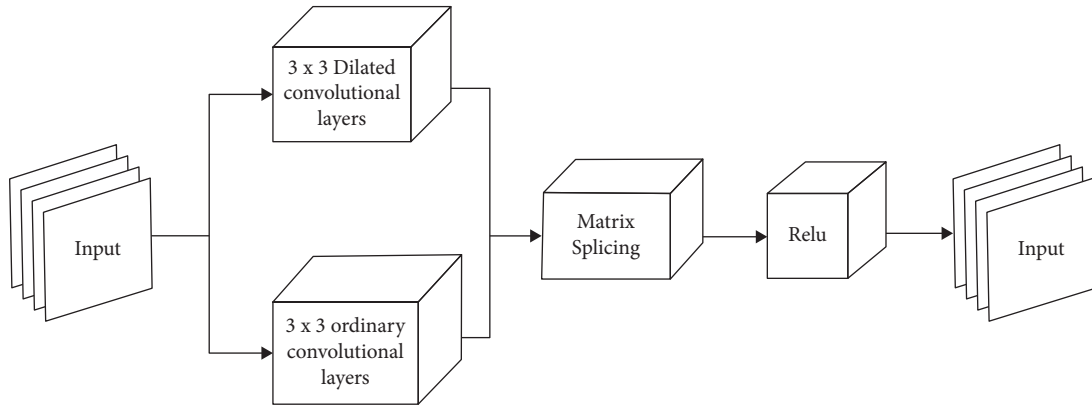


FIGURE 2: Structure of the dilated convolution block.

3.2. Dense Residual Block. Recent research has shown that CNNs can be trained more deeply, accurately, and efficiently if shorter connections are used between their input and output layers. Based on this conclusion, the literature [27] proposed DenseNets, which consist of dense blocks that connect each layer with all previous layers. The features on the input of each of these layers are the outputs of all previous layers, and the features on the output of each layer will be used as the input of all subsequent layers. This dense connection now makes a direct connection on the input and loss at each layer, so the dense network can mitigate the problem of gradient disappearance. Given the above advantages of dense blocks, the dense block of local feature residual fusion from the literature [15] is cited in this paper, which consists of two parts, i.e., a densely connected block and a local feature residual block fusion.

3.3. Deconvolutional Network. The DRT-Unet network uses local feature residuals to fuse dense blocks and up-pooling and transposed convolutional cascade with a deconvolution

structure to realize the decoding process of feature map size scaling. The information generated by the dense block during the encoding process is lost after the transition layer, but this lost information can be obtained in the decoding path by making a jump connection to the encoding path. Thus, the feature map after the up-pooling operation is jump-connected with the features of the same layer in the coding network to form the input of the next dense block. In order to reduce the spatial dimension, the input of the dense block in the deconvolution-decoding network is not cascaded with its output. As shown in Figure 1, the coding network in the upper half is a feature extractor that extracts feature descriptions from the input image, while the decoding network in the lower half is a shape generator used to generate segmentation targets from the extracted feature maps. It can be seen that the deconvolution-decoding network is almost a mirror result of the convolution-encoding network.

The feature map recovered by up-pooling becomes a sparse feature map due to the presence of a large number of 0 elements. The transposed convolution refers to the

transposition of the convolution kernel learned in the process of convolution. Compared with the sparse feature map obtained after up-pooling, this sparse feature map is used to form a dense feature map so as to correspond multiple feature maps to one feature map. In order to maintain the same size as the feature map obtained from the up-pooling, the obtained dense feature map needs to be cropped accordingly. Using the convolution kernel learned by transposed convolution and up-pooling, the shape of the target object based on the reconstruction is obtained in the deconvolution network. By applying the structure of up-pooling and transposed convolutional cascades in the decoding path and combining jump connections to compensate for the missing information, a deeper dense block network is constructed without generating a feature map explosion.

4. Experimental Results and Analysis

4.1. Dataset and Preprocessing. The datasets used in this paper were derived from BraTs2018 and BraTs2019, in which each case has four modalities, namely, T1-weighted imaging, T2-weighted imaging, contrast-enhanced T1ce, and liquid-attenuated inversion recovery column Flair [31], with different imaging modalities providing different information about brain tumors (each modality represents a different response to different tumor tissues) [32]. Although MRI images can quickly and effectively detect changes in water content in the sensing region and provide rich diagnostic information, a single MRI modality image cannot adequately subdivide the tumor in the region of interest and therefore cannot solve the problem of precise regional segmentation. Besides, using different MRI modalities can compensate for the above weaknesses. Hence, the slices of four modalities are used in this paper as the input of the segmentation network.

In this paper, BraTs2018 is selected as the training set, which contains 285 cases. Among them, 210 cases of HGG and 75 cases of LGG are included [33]. BraTs2019 has added 49 cases of HGG and 1 case of LGG based on BraTs2018, the new addition of which is used as the validation set. The dimension of each MRI image in the dataset is $155 \times 240 \times 240$. MRI images are represented as stereoscopic pixels in NIFTI format, and a series of preprocessing operations are required to fit the 2D network in this paper. A dichotomous segmentation was used to cut the brain tumor cases from the cross-sectional plane and obtain 2D images. A z-score approach is then used to normalize each modal image [34]. To alleviate the inter-category imbalance problem, slices without lesions in the image are discarded. To enhance the performance of model segmentation, the original 2D images are cropped from width and height dimensions of 240 mm to 160 mm in this paper, and the linear features and corresponding distribution relations of the image distribution are not changed during cropping. After the above steps, the images were divided into 4 channels and saved as an array of formats for data training and validation. Finally, the training set contains 17,925 slices and the validation set contains 7,750 slices.

4.2. Evaluation Indicators. In order to measure the effect of DRT-Unet network on brain tumor segmentation in a comprehensive and multi-faceted way, this paper adopts Dice [35], positive predictive value (PPV), sensitivity, intersection over union (IoU) ratio [36], and Hausdorff distance (95% HD) [37] as evaluation indexes, and the prediction results are compared with the real labeled data to show the segmentation effect from a visual perspective.

Dice is used to measure the resemblance between the segmentation result and the true value. The value of Dice is 1 when the segmentation result is best, and 0 when it is worst, which is defined by the following formula:

$$\text{Dice}\langle A, B \rangle = 2 \times \frac{|A \cap B|}{(|A| + |B|)} = \frac{2TP}{FP + 2TP + FN}. \quad (1)$$

PPV indicates the proportion of samples with positive predictions that are correctly predicted, and it is defined by the following formula:

$$\text{PPV} = \frac{TP}{TP + FP}. \quad (2)$$

Sensitivity indicates the proportion of samples predicted to be positive to the total positive samples (true-positive rate), which is defined by the following formula:

$$\text{Sensitivity} = \frac{TP}{TP + FN}. \quad (3)$$

IoU is a measure of the accuracy of the detection object, which is defined by the following formula:

$$\text{IoU} = \frac{TP}{FP + TP + FN}. \quad (4)$$

The Hausdorff distance (95% HD) is sensitive to contour information. The more this value tends to 0, the more accurate the predicted value is. In order to exclude the instability and unreasonableness of the segmented data caused by a few outliers, the parameter 95% was chosen as the maximum distance quartile, which is defined as follows:

$$\begin{aligned} 95\%HD &= \max\{d_{XY}, d_{YX}\} \\ &= \max \left\{ \max_{x \in X} \min_{y \in Y} \left\{ \max_{y \in Y} \min_{x \in X} \right\} \right\}, \end{aligned} \quad (5)$$

where A and B are the actual expert data values and model prediction values, respectively. TP indicates a positive sample with a positive model prediction, TN indicates a negative sample with a negative model prediction, FP represents a negative sample with a positive model prediction, and FN denotes a positive sample with a negative model prediction.

4.3. Experimental Parameter Settings. In this paper, PyTorch library is adopted to build DRT-Unet network, while Adam optimizer is used to train the method, with the training batch

being 20 and the training round being 400. The initial learning rate is set to 2×10^{-4} , with the weight decay coefficient set to 0.0002. k in the dense block is an indicator of the number of feature map output per layer in each dense block, which is set differently in this paper. Since Dice is more commonly used than the other three evaluation metrics, only the average value of Dice over the three segmentation regions is used as the reference standard in the stage of determining the parameter k . The results of model segmentation under different k are shown in Table 1, where the optimal data are in bold.

The best segmentation results are obtained when the value of k is 16. The experimental results show that the smaller the k is, the better the segmentation effect is; meanwhile, the network can be avoided to become too wide.

4.4. Experimental Results and Discussion. In order to verify the impact of each module in DRT-Unet network on the segmentation performance, this paper takes the U-Net network model as the basis, then adds the local feature residual fusion dense block and the dilated convolution in turn, respectively, to improve it, and finally, compares the obtained experimental results with DRT-Unet. As shown in Table 2, the experimental results in the table are mean values, where the optimal data are in bold.

U-Net, as the reference basic framework in this paper, obtained Dice coefficient of 0.810, precision of 0.822, recall of 0.919, and 95% HD of 1.157. After replacing the two 3×3 convolutions on each layer in the U-Net coding process with a dense block of local feature residual fusion, all four metrics are improved, with precision improved by 2.8% over the U-Net network, which indicates that the dense block of local feature residual fusion can effectively propagate and retain low-level visual features, and can reduce the information loss in deep network training through the fusion of local features. When the dilated convolution block is added to the encoding process, precision and recall are improved by 1.7% and 1.6%, respectively, based on the previous step, indicating that the fusion of normal convolution and dilated convolution can expand the perceptual field, obtain richer features, and provide more detailed information. As can be seen from the data in Table 2, the values of Dice and 95% HD have significantly changed, increasing by 2.2% and decreasing by 2.9%, respectively, compared with the previous method, which indicates that the combination of up-pooling and transposed convolution can effectively capture the global features and detailed features, and recover the extracted features well to the original pixels. Finally, the Dice value of DRT-Unet is 0.861, the precision value is 0.881, the recall value is 0.948, and the 95% HD value is 1.112. Compared with the U-Net network, these four metrics are improved by 5.1%, 5.9%, 2.9%, and 5.3%, respectively, which fully demonstrates the effectiveness of the proposed method in this paper.

In order to further prove the segmentation performance of this method, the classical deep learning segmentation networks FCN8s, U-Net, and the methods in literature [29] (DenseUnet), literature [38] (DeepResUnet), literature [39]

TABLE 1: Dice coefficient values at different k .

k	Dice coefficient value
16	0.861
32	0.796
48	0.789

TABLE 2: Comparison of different segmentation network models in each index.

Network model	Dice	Precision	Recall	95% HD
U-net	0.810	0.822	0.919	1.157
L-DB + U-Net	0.823	0.850	0.929	1.146
Dilated conv + L-DB + U-Net	0.839	0.867	0.945	1.133
DRT-Unet	0.861	0.881	0.948	1.104

(H 2 NF-Net), as well as literature [40] (MCA-ResUNet) are compared with DRT-Unet network, and all experiments use multimodal images of the same dataset as the input to the network. The goal of this paper is to segment three regions, WT, TC, and ET. WT is the intact tumor, which represents a blue region in the figure. Preoperative MRI images showing the extent and volume of the edema of the intact tumor can achieve high-precision localization of the tumor. TC is the core tumor, which corresponds to the white region in the figure and is a malignant tumor evolving from glial cells in the brain. The red region belongs to ET, which is the tumor-enhancing necrotic region, and is composed of necrotic cells. In this paper, the Flair sequence with the most obvious bright contrast is selected as the original contrast image from four modalities, T1, T2, Flair, and T1ce. Three cases with different characteristics are selected for doing visual effect comparison, the results of which are shown in Figures 3–5. In the figures, (a) is the original Flair sequence image, (b) is the manual segmentation label, and (c), (d), (e), (f), (g), (h), and (i) are the segmentation results of FCN8s, U-Net, DenseUnet, DeepResUnet, H 2 NF-Net, MCA-ResUNet, and DRT-Unet, respectively.

From the experimental results of the above three cases, the segmentation results of FCN8s are relatively rough. Figures 3 and 4 clearly reveal that the segmentation at the edge of the tumor is unclear, and the TC and ET regions cannot be finely segmented, with the poor overall effect. DenseUnet uses dense blocks to compensate for the detailed information of U-Net, which substantially improves the segmentation effect. However, by observing Figure 3, we find that the segmentation of WT edge region appears hollow. In Figure 5, we find that the segmented edge branching region is broken, and the contour is incoherently connected. DeepResUnet uses residual blocks to fuse multidimensional features. Although the segmentation results are generally better, the generalization ability is poor and there are many fragmented points at the boundary of WT region. Compared with U-Net segmentation results, DRT-Unet segmentation in WT area is more accurate, and the false segmentation region is smaller. It is obviously shown in Figure 4 that U-Net is unable to perform fine segmentation for areas with irregular edges, while the outline of DRT-Unet is closer to manual labels, showing that the dilated convolution block

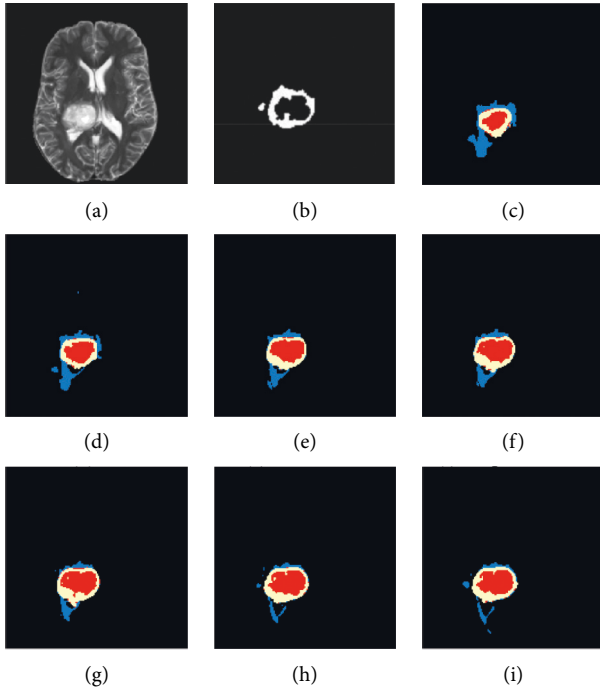


FIGURE 3: Case 1: weak contrast between light and dark inside the image, and difficulty in further fine segmentation of heterogeneous regions. (a) Original image. (b) Handmade labels. (c) FCN8s. (d) U-Net. (e) DenseNet. (f) DeepResUNet. (g) H 2 NF-Net. (h) MCA-ResUNet. (i) DRT-UNet.

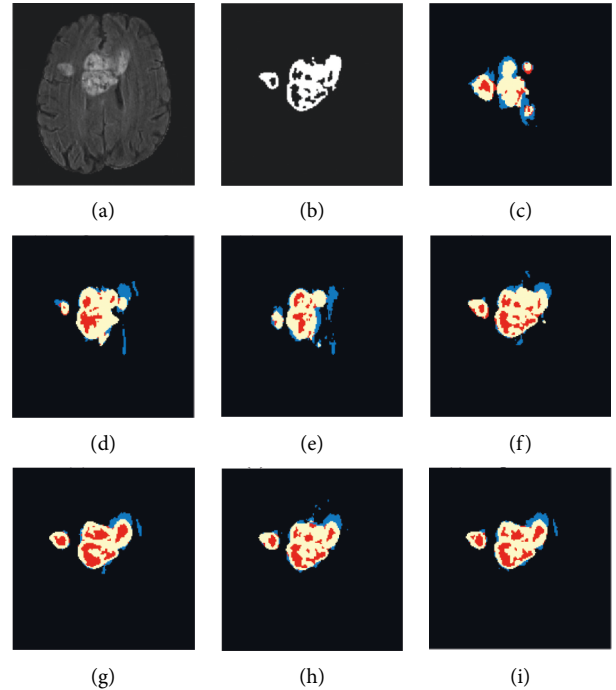


FIGURE 5: Case 3: the image is characterized by complex edges and many contour bifurcations, which can reflect the segmentation of contour details. (a) Original image. (b) Handmade labels. (c) FCN8s. (d) U-Net. (e) DenseNet. (f) DeepResUNet. (g) H 2 NF-Net. (h) MCA-ResUNet. (i) DRT-UNet.

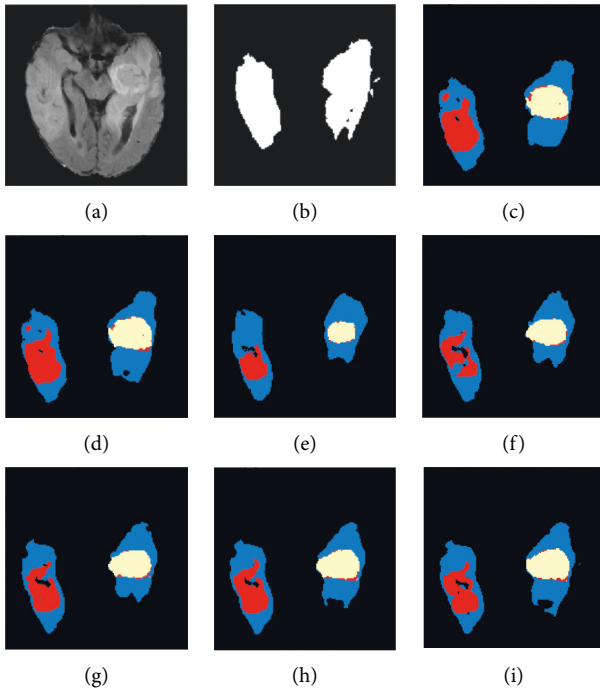


FIGURE 4: Case 2: tumor spread and metastasis can reflect segmentation effects on multiple regions. (a) Original image. (b) Handmade labels. (c) FCN8s. (d) U-Net. (e) DenseNet. (f) DeepResUNet. (g) H 2 NF-Net. (h) MCA-ResUNet. (i) DRT-UNet.

can obtain more abundant features; in addition, the combination of up-pooling and transposed convolution can effectively capture detailed features. DRT-UNet segmentation in each region is more complete compared with the above methods, the contrast between the core tumor region and the intact tumor region is clearer, and the contour line segmentation also performs better in detail than the existing algorithms.

Meanwhile, the IoU curves and loss function curves of the proposed methods in this paper are compared with those of FCN8s, FCN16s, FCN32s, U-Net, DenseUNet, and DeepResUNet, as shown in Figures 6 and 7. From these two figures, it can be seen that the IoU value of DRT-UNet is significantly higher than that of other methods, while the final loss function value is the lowest.

To better reflect the segmentation effect, the segmentation results of WT, TC, and ET were further and quantitatively analyzed by four assessment metrics, namely, Dice, precision, recall, and 95% HD, respectively. The whole tumor (WT) category includes all visible labels (a union of blue, yellow, and red labels), while the tumor core (TC) category is a union of red and yellow. Different from the two mentioned above, the enhancing tumor (ET) core category is only yellow (a hyperactive tumor part). The comparison results between DRT-UNet and other networks are shown in Tables 3–6, where the optimal data are shown in bold. As can

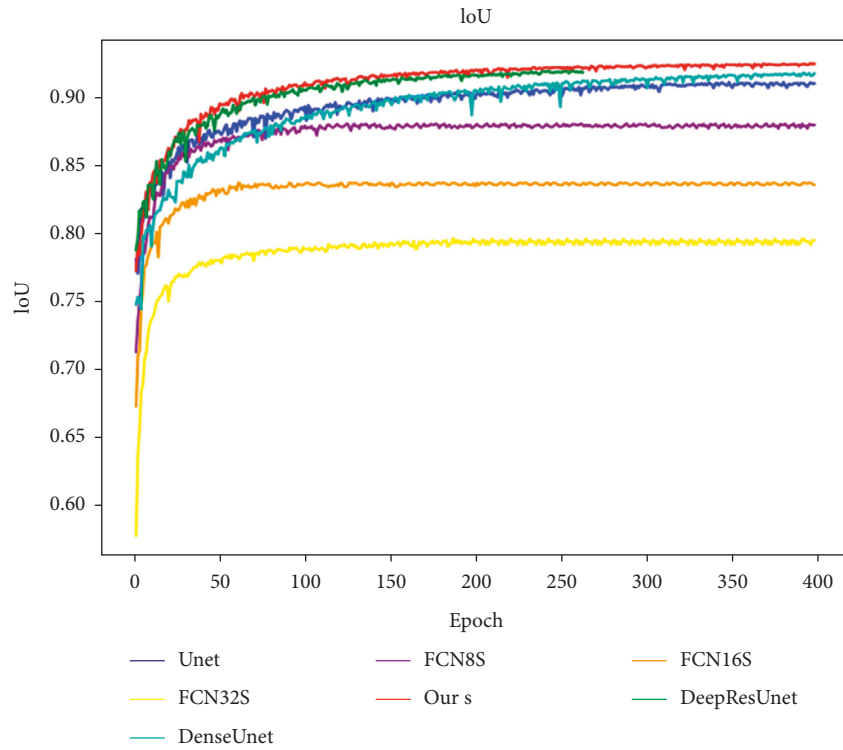


FIGURE 6: IoU curve comparison.

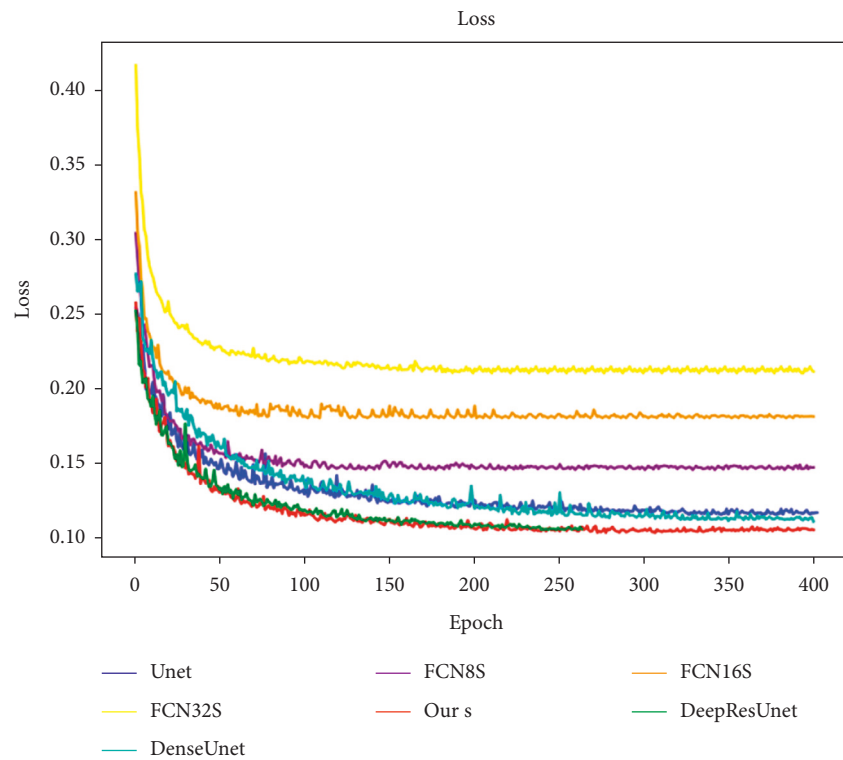


FIGURE 7: Comparison of loss function curves.

be seen from the table, the best data of DRT-Unet on these four evaluation indicators can reach 0.918, 0.93, 0.968, and 0.748, respectively. Besides, the values of each index on TC

are higher than those on WT and ET regions, which indicates that the TC region is relatively well segmented. Although the precision and Dice of DRT-Unet are slightly worse than

TABLE 3: Dice coefficient values under different networks.

Dice	Network								
	FCN8s	FCN16s	FCN32s	U-Net	DenseUnet	DeepResUnet	H2 NF-net	MCA-ResUNet	DRT-Unet
WT	0.564	0.539	0.418	0.769	0.792	0.799	0.913	0.849	0.842
TC	0.897	0.898	0.872	0.917	0.921	0.92	0.849	0.865	0.918
ET	0.531	0.495	0.351	0.743	0.765	0.765	0.79	0.784	0.823
Average	0.664	0.644	0.547	0.81	0.826	0.828	0.85	0.833	0.861

TABLE 4: Precision values under different networks.

Precision	Network								
	FCN8s	FCN16s	FCN32s	U-Net	DenseUnet	DeepResUnet	H2 NF-Net	MCA-ResUNet	DRT-Unet
WT	0.564	0.536	0.414	0.783	0.809	0.812	0.83	0.868	0.842
TC	0.911	0.905	0.897	0.935	0.932	0.926	0.912	0.89	0.918
ET	0.526	0.485	0.357	0.76	0.782	0.791	0.788	0.805	0.823
Average	0.667	0.642	0.556	0.826	0.841	0.843	0.854	0.854	0.861

TABLE 5: Recall values under different networks.

Recall	Network								
	FCN8s	FCN16s	FCN32s	U-Net	DenseUnet	DeepResUnet	H2 NF-Net	MCA-ResUNet	DRT-Unet
WT	0.94	0.913	0.91	0.925	0.928	0.94	0.923	0.86	0.95
TC	0.918	0.926	0.922	0.935	0.966	0.966	0.9	0.845	0.968
ET	0.908	0.879	0.853	0.909	0.914	0.915	0.894	0.917	0.923
Average	0.922	0.906	0.895	0.923	0.936	0.942	0.915	0.874	0.947

TABLE 6: 95% HD values under different networks.

95% HD	Network								
	FCN8s	FCN16s	FCN32s	U-Net	DenseUnet	DeepResUnet	H2 NF-Net	MCA-ResUNet	DRT-Unet
WT	1.809	2.211	2.598	1.334	1.42	1.345	1.1	2.592	1.252
TC	0.788	0.829	0.913	0.711	0.698	0.749	1.14	1.595	0.748
ET	2.114	2.27	2.738	1.417	1.476	1.407	2.61	2.739	1.315
Average	1.579	1.77	2.083	1.154	1.198	1.167	1.617	2.31	1.105

DenseUnet in the TC region, all the metrics of WT and ET are improved to various degrees compared with other methods, and the mean values of all the metrics of the proposed method are the highest in all three regions, indicating that DRT-Unet can segment WT, TC, and ET better and achieve more satisfactory segmentation results.

5. Conclusion

At present, the majority of brain tumor segmentation methods are based on two networks, FCN and U-Net, but the network connection based on FCN is not fine-grained and ignores the relationship between different pixel points. The U-Net model is experimentally proven to be slightly improved compared with FCN, but the overall generalization of prediction results is not strong and needs to be improved to a certain depth. To address these problems, this paper proposes a DRT-Unet network for the accurate segmentation of brain tumors, where four MRI modality images are used as input, and a dilated convolution block is used to expand the perceptual field in the coding process, so that the network can obtain richer and more detailed

feature information. Meanwhile, a dense block of local feature residual fusion is used in the coding process to propagate and preserve low-level visual features, reducing the information loss in deep network training through the fusion of local features. The DRT-Unet network adopts a dense block of local feature residual fusion and a deconvolution structure of up-pooling and transposed convolution cascade to achieve the decoding process of feature map size enlargement. The up-pooling and transposed convolution play a key role in recovering the global features and detailed features of the image. It can be seen from the experiments in this paper and the comparison with other methods that the DRT-Unet method can achieve effective segmentation of brain tumor lesions. Moreover, compared with the other four segmentation methods, the proposed network in this paper has better performance in visual effects and objective indexes.

Data Availability

The datasets used in this paper were derived from BraTs2018 and BraTs2019, and these are open-access data.

Conflicts of Interest

The authors declare that they have no conflicts of interest with any organization or entity in this manuscript.

Acknowledgments

This work was supported in part by the National Youth Natural Science Foundation of China (61802208), the Natural Science Foundation of Anhui (1908085MF207, KJ2020A1215, KJ2020A1216, and KJ2021A1251), the Excellent Youth Talent Foundation of Anhui (gxyqZD2019097 and gxyqZD2021142), the Postdoctoral Foundation of Jiangsu (2018K009B), the Quality Engineering Project of Anhui (2021jyxm1117, 2021kcszsfkc307, 2018mooc059, 2018kfk009, 2018sxzx38, 2018FXJT02, and 2018sxzx38), the Fuyang Normal University Doctoral Startup Foundation (2017KYQD0008), the Industry-University Cooperation Collaborative Education Project (201901258002).

References

- [1] T. X. Gao, Z. Lv, H. Y. Ding, and X. Y. Tang, "Research advances in magnetic resonance temperature imaging," *Chinese Journal of Medical Imaging*, vol. 22, no. 7, pp. 547–550, 2014.
- [2] R. Siegel, D. Naishadham, and A. Jemal, "Cancer statistics," *CA: A Cancer Journal for Clinicians*, vol. 62, no. 1, pp. 10–29, 2012.
- [3] P. L. Zhi and P. C. Lauterbur, "Principles of magnetic resonance imaging: a signal processing perspective," *Spie Optical Engineering*, Wiley-IEEE Press, Hoboken, NJ, USA, 2000.
- [4] W. Li, S. Cheng, K. Qian, K. Yue, and H. Liu, "Automatic recognition and classification system of thyroid nodules in CT images based on CNN," *Computational Intelligence and Neuroscience*, vol. 2021, pp. 1–11, Article ID 5540186, 2021.
- [5] Z. Qin, Y. Wang, H. Cheng, Y. Zhou, Z. Sheng, and V. C. M. Leung, "Demographic information prediction: a portrait of smartphone application users," *IEEE Transactions on Emerging Topics in Computing*, vol. 6, no. 3, pp. 432–444, 2018.
- [6] Z. Qin, Y. Zhang, S. Meng, Z. Qin, and K. K. R. Choo, "Imaging and fusing time series for wearable sensor-based human activity recognition," *Information Fusion*, vol. 53, pp. 80–87, 2020.
- [7] B. Xiao, K. Wang, X. Bi, W. Li, and J. Han, "2D-LBP: an enhanced local binary feature for texture image classification," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 29, p. 1, 2018.
- [8] H. Tang, B. Xiao, W. Li, and G. Wang, "Pixel convolutional neural network for multi-focus image fusion," *Information Sciences*, vol. 433–434, pp. 125–141, 2018.
- [9] B. Xiao, Y. Xu, X. Bi et al., "Follow the sound of children's heart: a deep learning-based computer-aided pediatric CHDs diagnosis system," *IEEE Internet of Things Journal*, vol. 7, no. 3, pp. 1994–2004, 2020.
- [10] Y. Ding, F. Chen, Y. Zhao, Z. Wu, C. Zhang, and D. Wu, "A stacked multi-connection simple reducing net for brain tumor segmentation," *IEEE Access*, vol. 7, pp. 104011–104024, 2019.
- [11] Y. Ding, C. Luo, C. Li, T. Lan, and Z. G. Qin, "High-order correlation detecting in features for diagnosis of Alzheimer's disease and mild cognitive impairment," *Biomedical Signal Processing and Control*, vol. 53, pp. 101564–101571, Article ID 101564, 2019.
- [12] M. Havaei, A. Davy, D. Warde-Farley et al., "Brain tumor segmentation with deep neural networks," *Medical Image Analysis*, vol. 35, pp. 18–31, 2017.
- [13] K. Kamnitsas, C. Ledig, V. F. J. Newcombe et al., "Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation," *Medical Image Analysis*, vol. 36, pp. 61–78, 2017.
- [14] B. Kayalibay, G. Jensen, and V. D. S. Patrick, "CNN-Based segmentation of medical imaging data," 2017, <https://arxiv.org/abs/1701.03056>.
- [15] Y. Ding, L. Gong, M. Zhan, C. Li, and Z. Qin, "A multi-path adaptive fusion network for multimodal brain tumor segmentation," *Neurocomputing*, vol. 412, no. 1, pp. 19–30, 2020.
- [16] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 4, pp. 640–651, 2017.
- [17] O. Ronneberger, F. Philipp, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proceedings of the 2015 International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 234–241, Springer, Munich, Germany, October 2015.
- [18] R. Raza, U. I. Bajwa, Y. Mehmood, M. W. Anwar, and M. H. Jamal, "dResU-Net: 3D Deep Residual U-Net Based Brain Tumor Segmentation from Multimodal MRI," *Biomedical Signal Processing and Control*, Article ID 103861, 2022.
- [19] Y. Zhang, Y. Lu, W. K. Chen, Y. K. Chang, H. M. Gu, and B. Yu, "MSMANet: a multi-scale mesh aggregation network for brain tumor segmentation," *Applied Soft Computing*, vol. 110, Article ID 107733, 2021.
- [20] H. Chen, Z. G. Qin, Y. Ding, L. Tian, and Z. Qin, "Brain tumor segmentation with deep convolutional symmetric neural network," *Neurocomputing*, vol. 392, pp. 305–313, 2020.
- [21] Y. L. Wang, Z. J. Zhao, S. Y. Hu, and F. L. Chang, "CLCU-Net: cross-level connected U-shaped network with selective feature aggregation attention module for brain tumor segmentation," *Computer Methods and Programs in Biomedicine*, vol. 207, no. 106154, 2021.
- [22] J. J. Wang, J. Gao, J. W. Ren et al., "DFP-ResUNet: Convolutional neural network with a dilated convolutional feature pyramid for multimodal brain tumor segmentation," *Computer Methods and Programs in Biomedicine*, vol. 208, Article ID 106208, 2021.
- [23] H. Shen, R. Wang, J. Zhang, and S. Mckenna, "Multi-task fully convolutional network for brain tumour segmentation," in *proceedings of the Annual Conference on Medical Image Understanding and Analysis*, pp. 239–248, Edinburgh, UK, July 2017.
- [24] G. Lin, M. Anton, C. Shen, and I. Reid, "RefineNet: multi-path refinement networks for high-resolution semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1925–1934, San Juan, PR, USA, June 2017.
- [25] D. Hao, Y. Guang, F. Liu, Y. Mo, and Y. Guo, "Automatic brain tumor detection and segmentation using U-net based fully convolutional networks," in *proceedings of the Annual Conference on Medical Image Understanding and Analysis*, pp. 506–517, Edinburgh, UK, July 2017.
- [26] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference*

- on *Computer Vision and Pattern Recognition*, pp. 770–778, San Juan, PR, USA, July 2016.
- [27] G. Huang, L. Zhuang, V. D. M. Laurens, and K. Q. Weinberger, “Densely connected convolutional networks,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4700–4708, Honolulu, HI, USA, July 2017.
- [28] S. Jégou, D. Michal, D. Vazquez, A. Romero, and Y. Bengio, “The one hundred layers tiramisu: fully convolutional DenseNets for semantic segmentation,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 11–19, Honolulu, HI, USA, July 2017.
- [29] A. Kaku, C. Hegde, J. Huang et al., “DARTS: DenseUnet-based automatic rapid tool for brain segmentation,” 2019, <https://arxiv.org/abs/1911.05567>.
- [30] JS. Rhee, “Certain radially dilated convolution and its application,” *Honam Mathematical Journal*, vol. 32, no. 1, pp. 101–112, 2010.
- [31] M. Brown, R Semelka, and T. K. Nishino, “MRI: basic principles and applications, 3rd edition,” *Medical Physics*, vol. 31, no. 1, p. 170, 2004.
- [32] L. Jin, M. Liu, J. Wang, F. Wu, T. Liu, and Y. Pan, “A survey of MRI-based brain tumor segmentation methods,” *Tsinghua Science and Technology*, vol. 19, no. 6, pp. 578–595, 2014.
- [33] A. Kermi, I. Mahmoudi, and M. T. Khadir, “Deep convolutional neural networks using U-net for automatic brain tumor segmentation in multimodal MRI volumes,” in *Proceedings of the International MICCAI Brainlesion Workshop*, pp. 37–48, Granada, Spain, September 2018.
- [34] T. Upadhaya, M. Yannick, S. Eric, P. J. L. Reste, and M. Hatt, “Prognostic value of multimodal MRI tumor features in glioblastoma multiforme using textural features analysis,” in *Proceedings of the IEEE 12th International Symposium on Biomedical Imaging (ISBI)*, Brooklyn, NY, USA, April 2015.
- [35] V. Thada and V. Jaglan, “Comparison of jaccard, dice, cosine similarity coefficient to find best fitness value for web retrieved documents using genetic algorithm,” *International Journal of Innovations in Engineering and Technology*, vol. 2, no. 4, pp. 202–205, 2013.
- [36] H. Rezaatofghi, T. Nathan, J. Y. Gwak, S. Amir, and S. Savarese, “Generalized intersection over union: a metric and a loss for bounding box regression,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 658–666, Long Beach, CA, USA, June 2019.
- [37] A. Garcia-Garcia, S. Orts-Escolano, S. Oprea, V. Villena-Martinez, and J. Garcia-Rodriguez, “A review on deep learning techniques applied to semantic segmentation,” 2017.
- [38] Z. Zhang, Q. Liu, and Y. Wang, “Road extraction by deep residual U-net,” *IEEE Geoscience and Remote Sensing Letters*, vol. 15, pp. 749–753, 2018.
- [39] H. Jia, W. Cai, H. Huang, and Y. Xia, “H 2 NF-net for brain tumor segmentation using multimodal MR imaging: 2nd place solution to BraTS challenge 2020 segmentation task,” *Brainlesion: glioma, multiple sclerosis, Stroke and Traumatic Brain Injuries*, pp. 58–68, 2020.
- [40] T. Cao, G. Wang, L. Ren, Y. Li, and H. Wang, “Brain tumor magnetic resonance image segmentation by a multiscale contextual attention module combined with a deep residual UNet (MCA-ResUNet),” *Physics in Medicine and Biology*, vol. 67, no. 9, Article ID 095007, 2022.