Neurocomputing 421 (2021) 195–209

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Brain tumor segmentation of multi-modality MR images via triple intersecting U-Nets ☆

Jinjing Zhang [a,*], Jianchao Zeng [a,*], Pinle Qin [a], Lijun Zhao [b]

[a] North University of China, No. 3 Xueyuan Road, Jiancaoping District, Shanxi Province, Taiyuan 030051, China
[b] Taiyuan University of Science and Technology, No. 66 Waliu Road, Wanbailin District, Shanxi Province, Taiyuan 030051, China

ARTICLE INFO

Article history:
Received 28 March 2020
Revised 25 June 2020
Accepted 10 September 2020
Available online 28 September 2020

Keywords:
Multi-modality MR images
Glioma segmentation
U-Net
Cross-entropy loss

ABSTRACT

In this paper, we propose a triple intersecting U-Nets (TIU-Nets) for brain glioma segmentation. First, the proposed TIU-Nets is composed of binary-class segmentation U-Net (BU-Net) and multi-class segmentation U-Net (MU-Net), in which MU-Net reuses multi-resolution features from BU-Net. Second, we introduce a segmentation soft-mask predicted by BU-Net, that is, candidate glioma region is generated by removing most of non-glioma backgrounds, which guides multi-category segmentation of MU-Net in a weighted manner. Third, an edge branch in MU-Net is leveraged to enhance boundary information of glioma substructure, which facilitates to locate glioma true boundaries and improve segmentation accuracy. Finally, we propose a sigmoid-evolution based polarized cross-entropy loss (S-CE) to resolve class unbalance problem, and apply S-CE loss to soft-mask prediction loss in BU-Net, multi-class segmentation loss in MU-Net and edge prediction loss in edge branch. Experimental results have demonstrated that the proposed 2D/3D TIU-Nets achieves a higher segmentation accuracy than corresponding 2D/3D state-of-the-art segmentation methods including FCN, U-Net, SegNet, CRDN, IVD-Net, FCDenseNet, DeepMedic, DMFNet, etc, evaluating on publicly available brain tumor segmentation challenge 2015 (BRATS2015) datasets. To show the universality of the proposed method, we also give a comparison of segmentation performance on BrainWeb dataset.

1. Introduction

Glioma is a kind of common primary cell tumor, which is caused by glial cell changes in the brain and spinal cord. According to CBTRUS's professional investigation [1], glioma accounts for about 27% of all primary brain tumors. Generally, glioma is classified into two categories of malignant glioma and benign glioma based on the property of invasiveness. Since the malignant glioma accounts for about 80% of brain tumors, it has appealed people's attention to study the glioma treatment. Afterwards, WHO [2] finely classifies glioma into four categories from grade I to grade IV according to mitosis. Low grade gliomas of grade I and grade II have a longer survival than high grade gliomas of grade 3 and grade IV. Low grade gliomas may become higher grade gliomas over time. Hence, doctors need to predict the probability and point-in-time of grade transformation of glioma by quantitatively analyzing glioma substructures [3].

Magnetic resonance imaging (MRI) of human brain, which captures brain soft tissue clearly, is a class of noninvasive method. The glioma substructures include edema, necrosis, enhancing and non-enhancing. Multi-modality MR images offer a kind of complementary information to recognize glioma substructures. By analyzing glioma on MRI, doctors can diagnose brain glioma, predict glioma growth tendency [4] and make elaborate treatment plans [5]. As we all know, glioma substructures are identified by combining tumor pathological priors with four modalities of MRI: flair, t1, t1c and t2. Meanwhile, experts can qualitatively estimate the glioma substructures by meticulously observing multi-modality MRI. More importantly, quantitative analysis is essential to accurately diagnose glioma invasion. However, this analysis requires to manually annotate glioma substructures, which is time consuming. To save annotation time and manual labor, automatic glioma segmentation of multi-modality MRI is adopted to segment glioma substructures efficiently, which can avoid a lot of complicated and tedious works.

Compared with traditional segmentation methods that need to get features by manually-designed feature extractor, deep neural network (DNN) based methods are capable of extracting a variety of deep features and segmenting object automatically. Hence, on

☆ This is an example for title footnote coding.

* Corresponding authors.
E-mail address: 1595928799@qq.com (J. Zhang).

https://doi.org/10.1016/j.neucom.2020.09.016
0925-2312/

account of superiority of DNN, more and more researchers pay close attention to the studies on DNN based segmentation methods [6]. Among existing DNN structures, U-Net [7] is one of the most efficient architectures and achieves satisfactorybiomedical segmentation results, since encoding features with locating information and decoding features with semantic information are fused hierarchically in U-Net. However, there are still some problems when U-Net is applied to glioma segmentation. For example, annoying discrete points always appear in a form of high intensity in MRI. Meanwhile, the unbalanced class distribution between glioma regions and normal tissue regions makes it hard to learn features belonging to tumor region with small proportion, while blurred boundaries of glioma region may make DNN hard to accurately recognize each tissue class.

For the above problem of discrete segmentation points, some post-processing methods have been proposed to refine glioma segmentation results [8–10]. However, these methods are independent with glioma segmentation and make segmentation inefficient. To reduce discrete points efficiently and automatically, multi-stage cascaded DNN architectures [11–16] are widely used for multi-modality MRI glioma segmentation [17]. In these architectures, the whole glioma region is segmented in the first stage, and then substructure segmentation is constrained within this region in the second stage.

Although these cascaded DNN methods achieve an efficient automatic reduction of discrete segmentation points during glioma segmentation, there are still some deficiencies of network design. For example, [11–13] with more than two U-Nets are complicated, which obtain high performance at the cost of memories and time. Each of these U-Net achieves specific segmentation task independently neglecting the relationships among these tasks. Since multi-task learning can improve the generalization of DNNs, [14–16] use this learning way to increase segmentation accuracy and employ feature sharing to reduce the parameter number of DNNs. However, a better way of feature sharing is still need to be researched. Sharing too many features among multiple tasks always makes specific-task features become fewer than those without feature sharing, which finally leads to the reduction of discriminative ability [14,15]. Different from only using shared features for each task in [14,15], CU-Net [16] not only uses the shared decoding features in the first U-Net, but also uses specific-task features. However, it equally treats two U-Nets with the same number of features in each layer without considering the complexity of different tasks.

The large variation of proportion among multiple classes makes it difficult to correctly predict pixels belonging to small class, which is a class unbalance problem of segmentation task. In general, there are two widely adopted ways to solve this problem. In the first way, they over-sample pixels belonging to the class with small proportion, while they down-sample pixels coming from the class with large proportion. Patch based DNN segmentation methods [18,19,8] prefer to employ the first way. In the second way, different from cross-entropy loss [20] treating all pixels equally, weighted semantic segmentation loss assigns a weight to each pixel to make training insensitive to class unbalance [7,21–23]. Specifically, dice loss [7,23] assigns a weight to each pixel by calculating the proportion of each class. However, some difficultly-classified pixels belonging to large proportion are weakened with low weights. To accurately focus on the difficultly-classified pixel, focal loss [24] and difficulty-guided cross-entropy loss [21,22] are designed to pay more attention to these pixels. Nonetheless, there are also some limitations for these losses. Specifically, the probability of each pixel belonging to the true class in focal loss [24] is mapped as a weight by a power function. However, the weight in focal loss cannot efficiently highlight some pixels with low probability, since it cannot clearly discriminate easily-classified pixels and difficultly-classified pixels, although power function can enlarge weight difference between high probability pixel and low probability pixel. The difficulty-guided maps are generated according to specific prior that pixels in class boundaries are difficult to be classified correctly [21]. However, difficulty-guided map weakens some difficultly-classified pixels that do not belong to class boundaries. In [22], hard threshold in segmentation loss can efficiently select difficultly-classified pixels, but hard threshold cannot be updated according to image content and this hard threshold hinders gradient back-propagation during training.

As for blurred boundary problem in glioma, the boundary enhancement in the segmentation process can effectively improve the segmentation accuracy near the glioma boundaries. In general, adding edge information into DNN [25] is a direct way to alleviate segmentation difficulty around object boundaries. Edge information has been widely used as a kind of low-level features to improve the performance of computer vision tasks such as image super-resolution [26], image segmentation [21,27], image fusion [28] etc. Meanwhile, edge detection operators can be applied to obtain edge information from images. However, they are independent from image segmentation and heavily depend on tradition edge detectors, which cannot be accelerated by the hard ware such as graphic process unit (GPU). In [25], holistically nested edge detection [29] is used to assist image segmentation. However, this method is a complex architecture, which spends a lot of network parameters and computational memory. To simplify DNN-based edge generation for segmentation process, many segmentation methods [21,27,30–34] reuse previous shallow features from segmentation network to generate boundary information, and edge loss is used to regularize the segmentation network for boundary feature enhancement. In [21], boundary detection task and segmentation task share almost all the parameters in DNN, where it ignores the difference between these two tasks. Therefore, boundary enhancement method in [21] may hinder segmentation tasks. In the kidney and glioma segmentation methods [33,34], edge information is leveraged as the supplementary information of encoding features to improve accuracy of medical image segmentation. However, boundary information of encoding features is weakened by subsequent convolutions and de-convolutions in the decoding path.

To resolve these problems mentioned above, our contributions are four-folds in this paper:

1) Inspired by glioma hierarchical structure as shown in Fig. 1, we propose a triple intersecting U-Nets (TIU-Nets) including binary-class segmentation U-Net (BU-Net) and multi-class segmentation U-Net (MU-Net). In TIU-Nets, the candidate regions of valid glioma substructures are proposed to be predicted, which can greatly reduce the disturbance of isolated pixels in complex background. Meanwhile, the structure of TIU-Nets facilitates to obtain high precision of segmentation when the light-weight DNN with a limited model size is required by practical edge-equipment.
2) In TIU-Nets, the segmentation soft-mask from BU-Net is applied to guide multi-class segmentation of MU-Net. The multiplication of this soft-mask with input images and multi-resolution features from the decoding path of BU-Net are respectively fed into MU-Net, and then the multi-resolution features from MU-Net are fused with these features from BU-Net. As a result, triple feature intersection between BU-Net and MU-Net enriches discriminative features in MU-Net.
3) We propose a novel sigmoid-evolution based polarized cross-entropy loss function, namely S-CE loss, to alleviate class unbalance problem. This proposed loss function assigns high weight values approaching 1 to the difficultly-classified pixels,

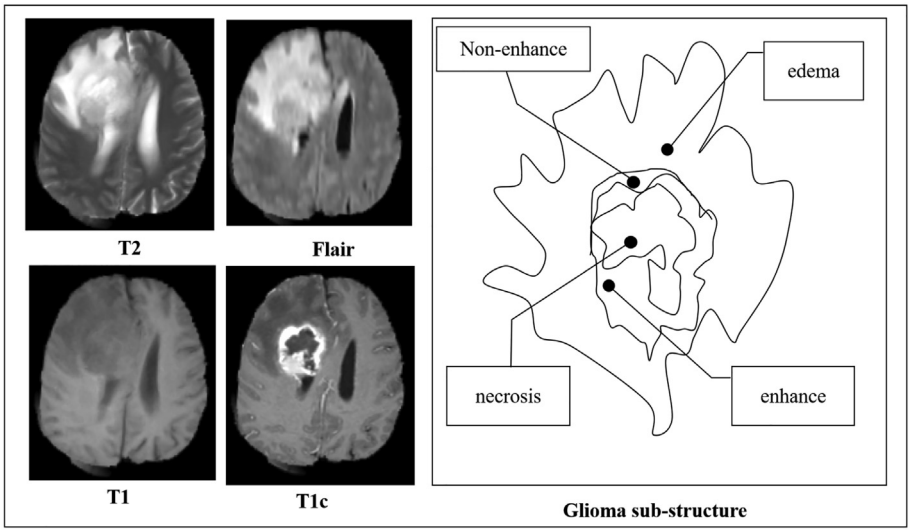


Fig. 1. The diagram of glioma substructure. Four modalities of MR images are shown in the left part of the figure, while the subregions of glioma are shown in the right part of the figure.

while weakening the influence of the easily-classified pixels on the average loss of all pixels by using small weights near 0.

4) Edge branch in MU-Net is proposed to estimate boundaries of glioma substructure. This branch reuses multi-resolution features from decoding path of MU-Net to obtain edge detection maps and is supervised by an S-CE edge loss during training to resolve class unbalance problem, that is, the amount of boundary pixels is far lower than that of backgrounds.

The rest of this paper is organized as follows. In Section 2, we look back on DNN based image segmentation and semantic segmentation loss. Then, we will introduce the proposed method in Section 3. After that, experimental results and analysis are given in Section 4. At last, we conclude this paper in Section 5.

2. Related works

2.1. DNN based image segmentation

DNN has been widely applied to image segmentation because of its superior fitting ability and automatic discrimination. Generally, DNN can be divided into two categories according to the size of input data: image-based and patch-based. Image-based DNN methods learn a mapping between input images and segmentation maps. Compared with image-based DNN, patch-based DNN methods use the whole patch to only classify the patch's center pixel. The advantages of patch-based DNN include data augmentation and 3D model training. Specifically, it disassembles image into many patches so that it augments amount of training data to avoid over fitting [18,19,8]. Moreover, patch-based DNN is usually applied to 3D medical image segmentation to save memory. However, as compared with 2D patch-based DNN, image-based DNN is more efficient, because it predicts all pixels of the whole image rather than pixels in partial region of image at once during segmentation. FCN is the first image-based segmentation DNN, which consists of encoding part for feature extraction and decoding part of mapping semantic features to segmentation results. Following the work of FCN [17], many image-based methods [7,35–37] are proposed to improve performance of FCN. Based on the structure of FCN network, U-Net [7] improves FCN [17] by using hierarchical feature fusion between encoding and decoding path of U-Net. Because of its simple structure with high segmentation accuracy, it is widely applied in medical image segmentation. Many researchers attempt to ameliorate the U-Net [21,11–16]. Moreover, LSTM is applied to improve U-Net by learning the relationships among slices [38,24].

2.2. Semantic segmentation loss

Cross-entropy [20] is widely applied to DNN based classification tasks [39,40], especially to DNN based image semantic segmentation tasks [17,36]. The cross-entropy loss can be written as follows:

$$L_{CE} = -\sum_{k=1}^{K} l_k * ln(p_k) \tag{1}$$

where K means number of classes, k is k-th class, $l_k \in \{0, 1\}$ indicates wether this pixel belongs to k-th class, and p_k is the predicted probability belonging to k-th class. However, the cross-entropy is unable to highlight pixels that are difficult to be classified into true class since it treats all pixels equally. Later, focal loss [41] is used to solve class unbalance class problem by focusing on optimizing difficultly-classified pixels and has been applied to semantic segmentation task [42]. This loss assigns weights to pixels, which can be written as Eq. (2), where high probability p means these pixels are easier to be classified correctly, vice versa. Different focal loss curves with different γ are shown in Fig. 2, where they use a power function to transform probabilities as weights.

$$L_{Focal} = -(1-p)^{\gamma} \sum_{k=1}^{K} l_k * ln(p_k) \tag{2}$$

Although weights of focal loss [41] are assigned according to difficulty that pixels are classified correctly, its weights cannot enlarge difference between the pixel with the high probability of correct classification and that with the low probability significantly. As shown in Fig. 2, focal loss with low value of γ can enhance pixels with low probability indeed, but this loss cannot suppress pixels with high probability well. With the increase of γ, focal loss obviously suppresses pixels with high probability, but it also assigns low weights to some difficultly-classified pixels (See Fig. 2). Recently, improved focal loss [43] and difficulty-guided loss [21] use a new way to identify difficultly-classified pixels and easily-classified pixels. However, improved focal loss [43] always focuses on the wrong classification probability of each pixel rather

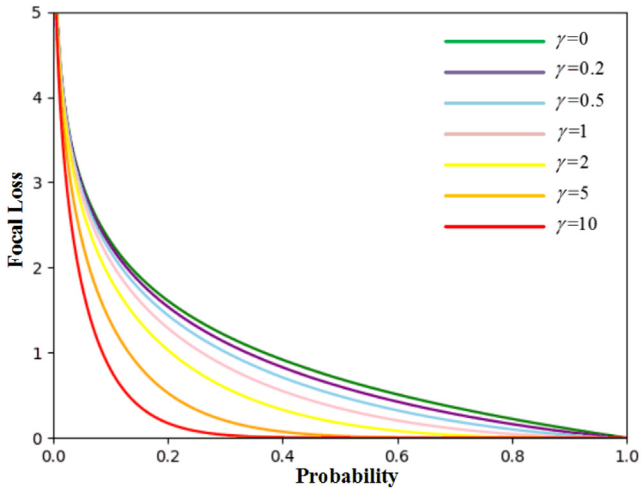


Fig. 2. The curve comparisons of focal loss with different γ.

than difficultly-classified pixels. Unlike foal loss, difficulty-guided loss [21] selects difficultly-classified pixels according to limited priors, which are unable to focus on difficultly-classified pixels that are not selected by this limited priors. As a consequence, it lacks some effective methods to find the difficultly-classified pixels as fully as possible and to concentrate on these difficultly-classified pixels effectively.

3. The proposed method

In this section, we will introduce the proposed method for brain tumor segmentation. First, we elaborate the architecture of the proposed TIU-Nets, as shown in Fig. 3. Second, we give a description of each component in the proposed TIU-Nets, that is, binary segmentation U-Net (BU-Net) and multi-class segmentation U-Net (MU-Net). Specifically, BU-Net is responsible to predict soft-mask of the whole glioma region (WGR) to guide glioma substructure segmentation. In consideration of blurring characteristic of glioma boundaries, we introduce an edge detection branch in MU-Net regularized by edge loss to extract glioma contour information. Third, we elaborate on the training loss of the proposed TIU-Nets for brain tumor semantic segmentation. The training loss of the proposed TIU-Nets contains three components: soft-mask prediction loss, glioma boundary prediction loss and multi-class segmentation loss. It should be emphasized that these three losses use the proposed S-CE loss function.

3.1. TIU-Nets

DNN is widely adopted to extract discriminative features for an automatic semantic segmentation way, whose network parameters are optimized in an end-to-end manner. In these methods, DNN works as a non-linear mapping from multi-modality MR images to a group of masks that label region of interesting in these images, which can be written as:

$$S = \mathscr{D}(I) \tag{3}$$

in which I is the input image, $\mathscr{D}$ denotes the DNN, and S is the segmentation result of I. We can denote $\mathscr{D}$ and S as the non-linear mapping of TIU-Nets and the glioma segmentation results respectively. TIU-Nets is composed of BU-Net and MU-Net. As shown in Fig. 3, both multi-resolution features and the soft-mask from BU-Net are reused by MU-Net for multi-class segmentation. The soft-mask from BU-Net is used as a guided map for MU-Net segmentation. Meanwhile, we reuse multi-resolution features from BU-Net to enrich discriminative features in MU-Net after multiplying these features with soft mask in a pointwise manner. In TIU-Nets, the soft-mask assigns high weights to the WGR, while very small weights are given to the background. As a result, it forces MU-Net to segment substructures of glioma within the candidate regions rather than the whole image. Since the design of the proposed TIU-Nets is a compact structure, the total parameters of this network can be greatly reduced by efficient soft-mask guided multi-resolution feature reutilization. Obviously, the TIU-Nets is a light-weight semantic segmentation model resulting in computational cost reduction.

3.2. BU-Net

As described above, BU-Net is utilized to generate a segmentation soft-mask. Since WGR is very clear in MR images which occupies a relatively large volume of the whole image and appears with high intensity on flair and t2 weighted MR images, soft-mask is relatively easy to be predicted. Based on this observation, we design the BU-Net as a shallow network using the compound operation twice in encoding path. The compound operation contains a sequence of two padding-3×3 convolutions with a stride of 1 and one 2×2 max-pooling as shown in Fig. 3(a). On account of the urgent requirements of low complexity model, we choose to reduce feature resolution, and then extract semantic features from these low resolution features. As shown in Fig. 3, before a sequence of the compound operations in the encoding path, we extract features from input images with a 3×3 padding convolution and then down-sample these features by a 3×3 padding convolution with a stride of 2. Because of symmetrical structure of BU-Net, it also uses a compound operation twice in decoding path containing a 2×2 de-convolution with a stride of 2, skip connection from encoding path and two 3×3 padding convolutions. After compound operation, we use a 2×2 de-convolution to up-sample features to obtain the same resolution as input image. Afterwards, a 1×1 convolution followed by a sigmoid activation function is utilized to generate segmentation soft mask. Additionally, for each 3×3 convolution in BU-Net, it is followed by a sequence of operations including batch normalization (BN) and ReLU activation to strengthen its nonlinear capability. It is noteworthy that there is no activation as well as BN after de-convolution.

In BU-Net, skip connections are used to finely fuse encoding features of location information and decoding features of semantic information at each resolution. Since the max-pooling operation may give rise to a lot of location information missing, these skip connections can compensate these missing information. The compensation information from encoding path is used to be fused with high-level semantic features in decoding path to accurately identify fairly small regions and even pixels. Different from the pixel-wise addition manner in FCN [17], BU-Net employs the convolutional operation to fuse features from encoding and decoding path, in which this kind of operation works as fusion rules, whose weights can be learned by optimization. Apparently, it is more flexible and more efficient than the pixel-wise addition operation in FCN [17]. The nonlinear mapping of BU-Net can be written as follows:

$$M = f_U(f_M(f_D(X))) \tag{4}$$

where X is multi-modality MR images, and M is probability map predicted by BU-Net. f_D denotes a transformation from these images to low-resolution features, while f_U denotes a transformation from low-resolution feature maps to the segmentation maps. Between these two operations, f_M is a nonlinear function, which maps low-level features to high level-features.

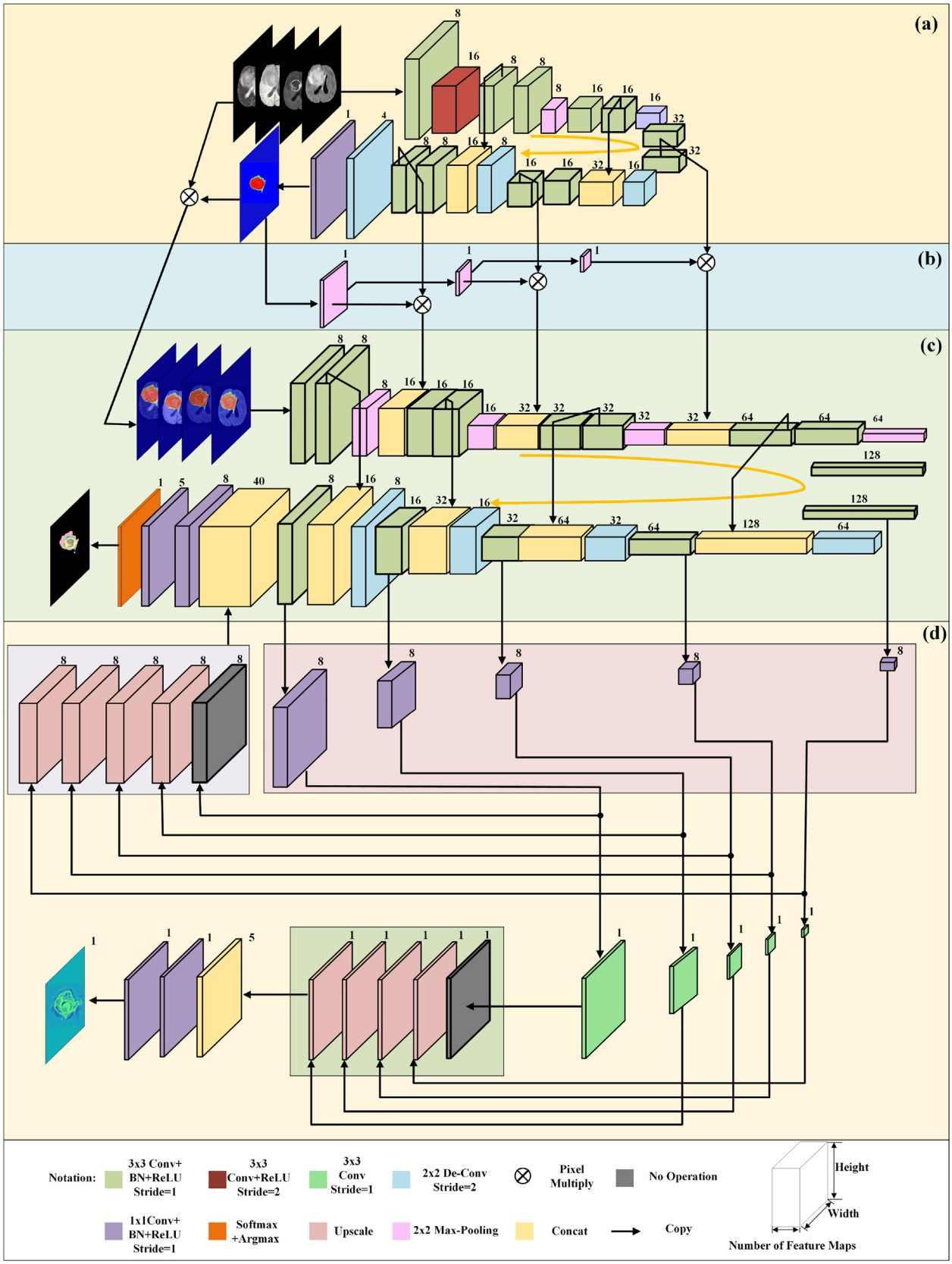


Fig. 3. The structure of the proposed TIU-Nets. (a) BU-Net, (b) Multi-resolution guidance generation, (c) MU-Net, (d) Edge prediction branch. The rectangle cuboid on bottom right corner denotes feature maps, whose width and height are the size of feature maps, while the colors of rectangle denote different operations. The colored feature maps are obtained by using corresponding-color operation. The cuboid with dark gray indicates that these feature maps are obtained by copying previous feature maps. The operations of different color are listed at the bottom.

3.3. MU-Net

In MU-Net, segmentation soft-mask from BU-Net is utilized to guide the MU-Net by multiplying the inputs of MU-Net with the soft-mask. Then, the MU-Net can predict substructures mainly from WGR. Due to fairly complex details in glioma substructures, the encoding path in MU-Net should contain more convolutional layers than that in BU-Net to extract more high level information. As shown Fig. 3(c), a compound operation as done in BU-Net is performed four times in encoding path. Meanwhile, we employ the skip connection between encoding features and decoding features to transmit encoding features to decoding path and fuse these features at different resolutions. To enrich features in encoding path, encoding features in MU-Net are fused with decoding features from BU-Net at different resolutions, as shown in Fig. 3(b). To remove most of non-glioma regions, feature maps from BU-Net are multiplied pixel-wisely with soft-mask before fusion.

Because glioma subregions with irregular shapes are small and adjacent, it needs more detail information to discriminate these substructures. Therefore, MU-Net fuses encoding low level features and decoding discriminative features by one convolution. For multi-class segmentation, we use soft-max activation after the 1×1 convolution in the last layer. The function of MU-Net can be written as follows:

$$Y = f_{DE}(f_E(X \otimes M_0, F_1 \otimes M_1, F_2 \otimes M_2, F_3 \otimes M_3)) \tag{5}$$

where X is the multi-modality MR images, Y is the segmentation result of MU-Net. M_0, M_1, M_2 and M_3 are obtained by rescaling soft segmentation mask M from BU-Net. $\otimes$ denotes pixel-wise multiplication. F_1, F_2 and F_3 are the feature maps in different resolutions from decoding path of BU-Net. From Eq. (5), it can be seen that image segmentation process is divided into encoding part f_E and decoding part f_{DE}.

3.4. Edge branch in MU-Net

To resolve the problem that the blurred glioma boundaries are difficult to be segmented accurately, we affiliate edge detection branch to MU-Net. The edge branch reuses decoding features of MU-Net to generate boundary features, as shown in Fig. 3(d). To keep parameters number of edge branch as small as possible, we use a few number of convolutions to obtain legible edge features. The edge branch, as shown in Fig. 3, reuses multi-resolution encoding features of MU-Net, and use a convolution with spatial size of 1×1 to recombine features at each resolution. Beside, edge loss regularizes these multi-resolution features to maintain glioma boundary features. After that, we up-sample these multi-resolution features to the same resolution with the input images. In edge branch, intermediate features are fed into the last layers of decoding path to strengthen boundary information of features in decoding path of the MU-Net. In order to extract edge features from multi-resolution features, we use 3×3 padding convolution with a stride of 1 to obtain an edge map at each resolution before up-sampling. After up-sampling multi-resolution edge maps, we fuse these features by two 1×1 convolutions and generate edge maps. Finally, the sigmoid activation function is employed in the last 1×1 convolution layers. It is noteworthy that we apply canny edge detector on glioma mask label to obtain glioma boundary maps as the edge label for glioma boundary prediction loss. The process of edge branch can be written as following:

$$Y_1 = f_M(f_{UP}(f_{E_2}(f_{E_1}(F_1, F_2, F_3, F_4, F_5)))) \tag{6}$$

$$Y_2 = f_{UP}(f_{E_1}(F_1, F_2, F_3, F_4, F_5)) \tag{7}$$

where F_1, F_2, F_3, F_4 and F_5 are multi-resolution features in encoder part of BU-Net. Y_1 and Y_2 are two outputs. Y_1 is the predicted edge. Meanwhile, Y_2 is the intermediate output from edge branch, which is added to the last layers in decoding path of MU-Net. f_{E_1} denotes feature extraction in the first step using a 1×1 convolution, while f_{E_2} is feature extraction in the second step using a 3×3 convolution. f_M denotes mapping from convolutional feature map to an edge map. f_{UP} means up-sampling low-resolution features to high-resolution features. As shown in Eq. (7), it up-samples intermediate outputs of Eq. (6) as Y_2.

3.5. Loss in TIU-Nets

In TIU-Nets, we use multi-task loss to optimize the framework, because it can highlight the relationship of these tasks compared with single task loss. On account of feature reusing, features in decoding path are learned according to more than one loss during training so that it improves the generalization ability of these features. In TIU-Nets, reused features from BU-Net are learned via the supervision of three losses, and reused features from MU-Net are learned according to the supervision of substructure segmentation loss and edge loss. In summary, the proposed multi-task loss can be written as follows:

$$L_{total} = \lambda_1 L_{seg} + \lambda_2 L_{mask} + \lambda_3 L_{edge} \tag{8}$$

in which L_{seg} is the loss of glioma substructure segmentation in MU-Net, L_{mask} is the loss of soft-mask prediction in BU-Net, and L_{edge} is the loss of glioma boundary prediction in edge branch. λ_1, λ_2, λ_3 are three weights of the L_{seg}, L_{mask} and L_{edge} respectively. Because of class unbalance problem in three tasks, all these losses use the proposed S-CE loss to optimize the proposed TIU-Nets.

3.5.1. Sigmoid-evolution based polarized cross-entropy loss (S-CE)

On account of class unbalance problem in these tasks, we propose S-CE loss that focuses on pixels with low probability of correct classification. As we all know, the sigmoid function is a soft threshold, which is differentiable. In S-CE loss, we use a threshold to distinguish pixels that are hard to be classified correctly from pixels that are easy to be classified correctly. In general, if the probability that pixel is classified correctly is higher than 0.5, the corresponding pixels will doubtlessly be classified correctly, so we regard these pixels as easily-classified pixels. On the contrary, when the probability of pixels is equal to or lower than 0.5, these pixels have a risk of wrong classification. What's more, in the case that the probability of pixels is equal to or lower than 0.5, the lower probability these pixels have, the harder they are to be classified correctly and optimized to the probability higher than 0.5. Thus, we treat these pixels as difficultly-classified pixels.

Based on the above analysis, we construct evolution sigmoid function with threshold of 0.5, where it maps probabilities that higher than 0.5 to the values closed to 0. On the contrary, probabilities lower than 0.5 are mapped to values closed to 1. This function is a monotone decreasing function and polarizes probability to values approaching zero or one. Besides, we use γ to adjust polarization degree. The S-CE loss can be written as Eq. (9):

$$f_{sigmoid} = \frac{1}{1 + e^{\frac{\gamma}{2} - \gamma(1-p)}} \tag{9}$$

where p is the probability that the pixel is classified correctly, γ is a factor that determines the curve steepness of the sigmoid-evolution function.

To further analysis the S-CE loss, we give the visual curve comparisons of the sigmoid-evolution functions with different value of γ, as shown in Fig. 4(b). Compared with power function curves varied with probability as shown in Fig. 4(a), sigmoid-evolution functions varied with the probability tend to significantly enlarge weight difference between difficultly-classified pixels and those

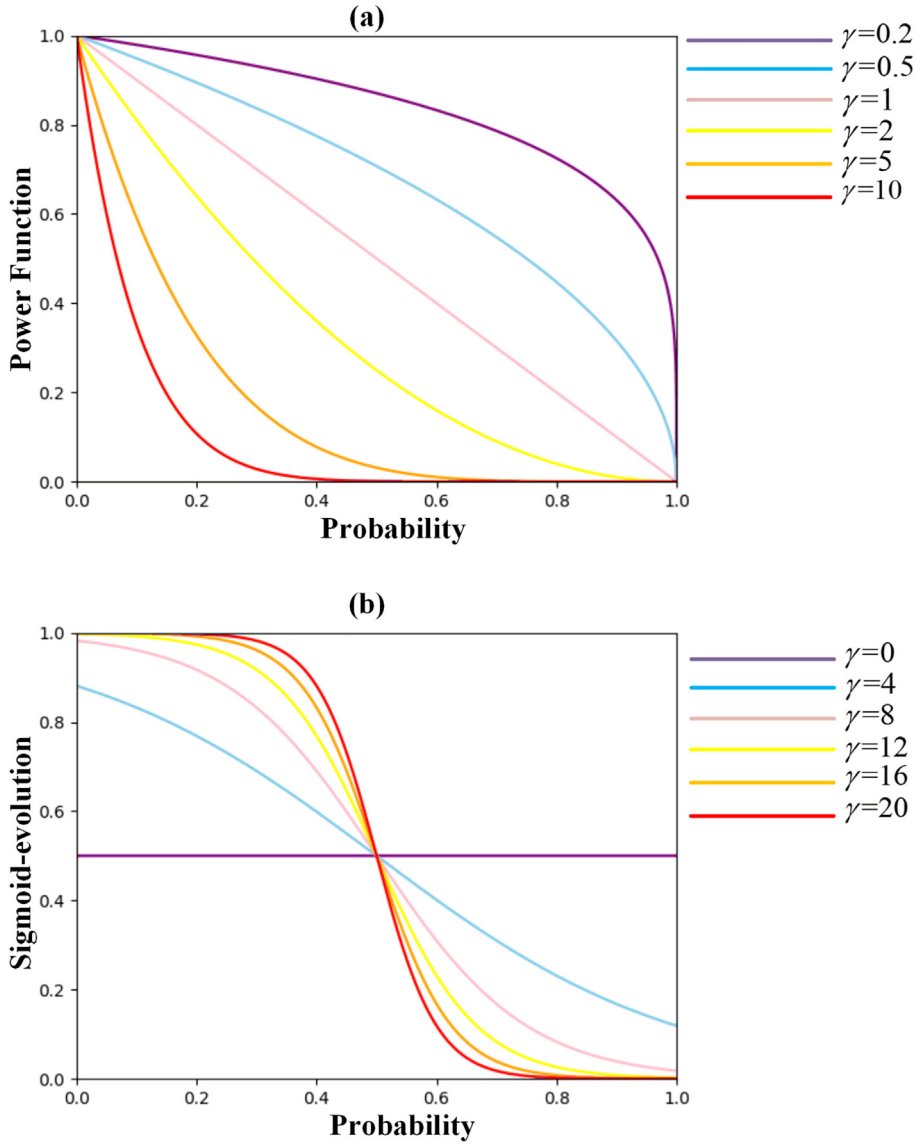


Fig. 4. Comparisons of power function curves in (a) applied in focal loss and sigmoid-evolution curves in (b) for S-CE loss.

of easily-classified pixels, so it can highlight difficultly-classified training pixels more efficiently by assigning a weight to each pixel. Accordingly, we use sigmoid-evolution function to assign a weight for the cross-entropy loss of each pixel. The proposed S-CE loss of one pixel can be written as following:

$$f_{S-CE} = \frac{1}{1+e^{\frac{\gamma}{2}-\gamma(1-p)}}\left(-\sum_{k=1}^{K} l_k ln(p_k)\right) \tag{10}$$

where $p \in [0,1]$ is the probability that the pixel is classified into true class, γ is a factor that determines the degree of polarization between low probability and high probability. $l_k \in \{0,1\}$ is a label of the pixel in k-th class, and p_k is a probability of the pixel belonging to k-th class.

We also show the curves of S-CE loss that varies with probability, when it is assigned with different γ in Fig. 5. Compared with focal loss, the proposed loss focuses on difficultly-classified pixels more efficiently. In S-CE loss, each pixel loss varies with its probability, as shown in Fig. 5. It is noteworthy that the proposed S-CE loss suppresses loss of pixels with probability higher than 0.5 by

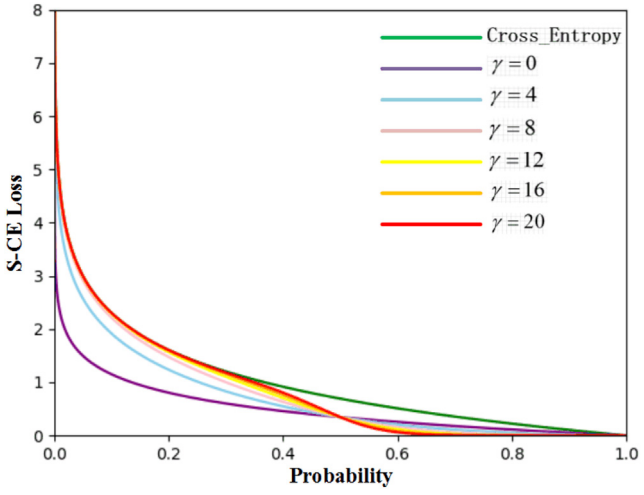


Fig. 5. The S-CE loss curves assigned with different γ.

assigning low weights near 0 to these pixels. Conversely, it enhances pixels with probability lower than 0.5 by assigning high weights near 1 to these pixels. As we can see in Fig. 5, green line represents standard cross-entropy loss varies with probability. The part of cross-entropy loss curves whose probability is lower than 0.5 is almost overlapped with that part of S-CE loss curves with different γ. It can be found that the proposed S-CE loss well retain the cross-entropy loss for pixels with low probability and suppress the loss of pixels with high probability. According to the above analysis of S-CE loss, it shows its powerful ability to select difficultly-classified pixels.

3.5.2. S-CE loss for TIU-Nets

Since class unbalance problem exists in these three tasks of TIU-Nets, we apply S-CE loss to these three tasks, which can be written as:

$$L = \frac{1}{\sum_{i\in O}\frac{1}{1+e^{\frac{\gamma}{2}-\gamma(1-p_i)}}} * \sum_{i\in O}\frac{1}{1+e^{\frac{\gamma}{2}-\gamma(1-p_i)}}\sum_{k=1}^{K} - l_{i,k} ln(p_{i,k}) \tag{11}$$

where $p_i \in [0,1]$ is the probability of pixel that belongs to true class, $p_{i,k}$ is the probability of pixel i that belongs to k-th class, $l_{i,k} \in \{0,1\}$ is the label indicator meaning whether pixel i belongs to k-th class, and O represents all pixels in MR image. K represents number of classes, L_{EU-Net} contains two categories: non-edge and edge regions, so $K = 2$ in EU-Net. L_{BU-Net} also contains two categories: WGR and non-glioma region, so $K = 2$ in BU-Net. L_{MU-Net} contains five categories: necrosis, enhance, non-enhance, edema and non-glioma, so $K = 5$. In MU-Net, γ is a factor of sigmoid function determining the degree of polarization between low probability and high probability, as shown in Fig. 5.

4. Experimental results and analysis

In this section, we will first introduce the implementation details of the proposed method. Then, we give a series of ablation studies on edge branch, BU-Net and S-CE loss. After that, our method is compared with several state-of-the-art semantic segmentation methods including FCN [17], SegNet [36], U-Net [7], the methods with VGGNet as backbone, the methods with ResNet50 as backbone, the methods with U-Net as backbone, IVD-Net [44], FCDenseNet [45] and CRDN [38], which are evaluated on brain tumor segmentation challenge 2015 (BRATS2015) dataset. To further show the effectiveness of the structure of TIU-Nets, we convert TIU-Nets to 3D segmentation model by replacing 2D operations such as 2D convolution, 2D batch normalization and 2D deconvolution with corresponding 3D operations. We compare 3D TIU-Nets with some state-of-the-art semantic segmentation methods containing DeepMedic [46], LSTM-UNet [47] and DMFNet [48], while our method is also compared with 3D FCN, 3D SegNet, 3D U-Net, and 3D CRDN. Because the data in BrainWeb consists of 2D images, we compare our 2D TIU-Nets with some 2D methods to further show the universality of the proposed method on BrainWeb dataset.

4.1. Implementation details

As mentioned above, we evaluate our method on BRATS2015 and BrainWeb dataset. BRATS2015 dataset contains 274 training patients (220 high grade, 54 low grade) and 110 testing patients. For each patient of BRATS2015 dataset, four modalities of each axial slice are imaging in the same anatomy position. Therefore, for 2D TIU-Nets training, we obtain 155 axial slice groups, among which each group contains four 2D axial slices from four modalities respectively. To remove some invalid slices, we select the slices containing glioma regions larger than 2% of the whole brain region for training. Meanwhile, for 3D TIU-Nets training, we augment training data on BRATS2015 by cropping 3D data with a size of $155 \times 240 \times 240$ into the 3D patches with a size of $32 \times 96 \times 96$. To remove some invalid patches, we also select the 3D patches containing glioma regions larger than 2% of the whole brain region for training. For TIU-Nets training and testing, we normalize four modalities (t1, t1c, t2, flair) of each patient from BRATS2015 dataset as the input and make intensity distribution of each modality follow the Gaussian distribution with mean value of 0 and variance of 1. The normalization can be written as:

$$\hat{X} = \frac{X - \mu}{\sigma} \tag{12}$$

in which X is a 3D MR image of each patient. $\hat{X}$ is the processed 3D MR image. μ and σ are the intensity mean and variance for X respectively.

For the BrainWeb dataset, it contains 399 training slices, where we randomly extract 239 slices for training and 160 slices for testing. We extract slices from three views of sagittal, axial and coronal. There are multiple MR modalities (t1, t2, pd) imaging in the same anatomy position for each anatomic position. Accordingly, we can get multi-modality 2D images from three views of sagittal, axial and coronal. Because of different size for three views, we crop slice to the same size as 180×180. After that, we concatenate three modalities as RGB image. Different from the normalization for BRATS2015 dataset, the intensity range of BrainWeb images is normalized to $[0, 255]$, which can be written as:

$$\tilde{X} = \frac{X - min(X)}{max(X) - min(X)} \times 255 \tag{13}$$

where X is an input 2D slice, $\tilde{X}$ is 2D slice after intensity normalization. When the proposed method is evaluated on the BrainWeb dataset, the preprocessed RGB image with three modalities (t1, t2, pd) are fed into TIU-Nets.

Our TIU-Nets is implemented on the platform of Pytorch [49] with NVIDIA TITAN RTX GPU. During training, the proposed TIU-Nets is initialized with the Kaiming initialization method and is optimized by stochastic gradient descent (SGD) optimizer. We use the initial learning rate of 0.01 with a momentum of 0.9 and a weight-decay of 0.0005. The batch size is set as 16 in TIU-Nets. Our model is trained with 100,000 iterations. For testing assessment, we upload the glioma segmentation results on the online evaluation platform in the publicly available website [50].

We evaluate the segmentation accuracy according to dice in BRATS2015. Dice indicates the degree of overlap between ground truth region and prediction region. We evaluate the substructures segmentation of glioma in three regions by dice index: complete, core and enhance. Complete region consists of edema, necrosis, enhancing and non-enhancing regions, while core region is composed of necrosis, enhancing and non-enhancing regions. We divide each image into four parts according to overlap between ground truth region and predict region, as shown in Fig. 6. The dice is calculated using these four parts according to Eq. (14).

$$dice = \frac{2 \times TP}{2 \times TP + FP + FN} \tag{14}$$

For the BrainWeb, both pixel accuracy (PA) and dice measurement are used to evaluate the performance of different semantic segmentation approaches. The index of PA denotes the proportion of pixels that are classified correctly to total pixels, which can be written as:

$$PA = \frac{TP + TN}{TP + TN + FP + FN}. \tag{15}$$

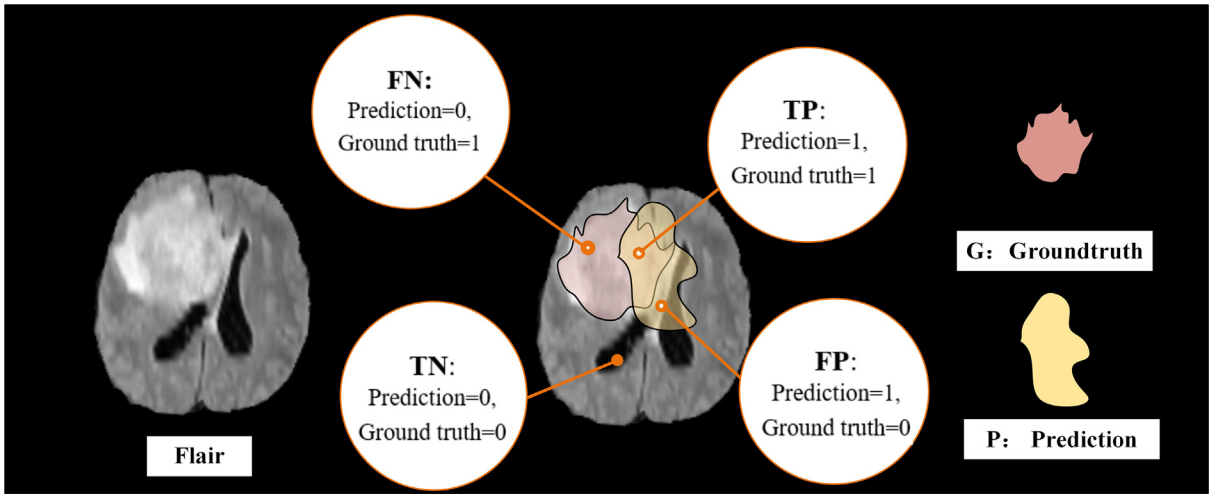


Fig. 6. The diagram of the overlapping between the ground-truth region and prediction region.

4.2. Ablation study

Since the blurred boundaries of glioma region are annoying and are hard to be discriminated for semantic segmentation, we propose an edge branch in TIU-Nets to enhance boundary information for accuracy improvement. To verify significance of edge branch, an ablation study about TIU-Nets without/with edge branch is conducted to compare their performance. As shown in Table 1, TIU-Nets with edge branch maintains performance in core and enhance region, but improves 1% in complete region.

As mentioned above, BU-Net highlights important regions in MU-Net, which plays a similar role to the attention model. Therefore, to validate efficiency of BU-Net, we compare TIU-Nets with attention module based TIU-Nets. In attention based TIU-Nets, we take place of BU-Net by attention module [51] in TIU-Nets. The attention module including spatial attention and channel attention is applied to MU-Net of TIU-Nets to select important regions. The segmentation results are shown in Table 2. From Table 2, we can find that dice score of MU-Net guided by BU-Net is 2%, 2% and 1% higher than that guided by attention model respectively in complete, core and enhance regions. From the above analysis, it is obvious that BU-Net is more efficient than attention model to enhance the significant regions.

Since class unbalance problem in these tasks, we propose S-CE loss that pays more attention to difficultly-classified pixels. To validate the efficiency of the proposed S-CE loss, we conduct an ablation study on different semantic segmentation loss such as cross-entropy loss, weighted cross-entropy loss, focal loss and S-CE loss for TIU-Nets with/without edge branch as shown in Table 3 and 4.

From Table 3, it can be observed that dice score of S-CE loss is equally matched with that of focal loss in complete region with the highest segmentation accuracy because both focal loss and S-CE loss tend to pay more attention to difficultly-classified pixels in blurred edema boundaries. However, the model trained with weighted cross-entropy loss has the low segmentation accuracy, whose dice score in complete region has the worst among these losses. Because it assigns weights according to the proportion for each class rather than the difficulty degree of classification, which makes some difficultly-classified pixels in edema region with large proportion to be weakened by assigning lower weights. For core region, S-CE loss shows the highest dice score with the best segmentation accuracy in core region. Nevertheless, the cross-entropy loss has the worst performance with the lowest dice score in enhance region, because it treats all pixels equally during training making pixels in enhance region with small proportion suppressed by pixels in large proportion region. Since core region occupies a small proportion and has clearer boundaries compared with complete region, weighted cross-entropy loss, which assigns weights according proportion of class, can well strengthen core region and achieve a better segmentation performance than cross-entropy loss. For enhance region, they almost have the same performance on dice for these losses. From Table 4, we also find that the proposed S-CE loss achieves the best dice score in complete, core and enhance regions. The model trained with weighted cross-entropy obtains the worst dice score in complete region, while the cross-entropy achieves the best score in enhance region. However, the dice score of focal loss is inferior than that of cross-entropy loss in complete region.

We also do an ablation study on γ in S-CE loss. The comparative results of TIU-Nets with edge branch are shown in Table 5 and the results of TIU-Nets without branch are shown in Table 6. From Table 5 and Table 6, it can be found that dice score of TIU-Nets with S-CE loss is higher than that with other losses when they are tested on three regions (complete, core and enhance).

To further study the effects of the loss proportion on segmentation performance, we change proportion of three losses in the total loss by adjusting corresponding weights λ_1, λ_2 and λ_3 for TIU-Nets training. We fix two of weights as value of 1 and change another one weight of three weights from low value to high value. From Table 7, we find that when the proportion of L_{seg} is larger than that

Table 1
Comparisons of TIU-Nets with edge branch and without edge branch. (Bold number is the best score in each column.)

Method	Dice		
	Complete	Core	Enhance
TIU-Nets(Without Edge)	0.81	**0.68**	**0.61**
TIU-Nets(With Edge)	**0.82**	**0.68**	**0.61**

Table 2
Comparisons of TIU-Nets with BU-Net and the one with the attention model. (Bold numbers is the best score at each column.)

Method	Dice		
	Complete	Core	Enhance
Attention Based TIU-Nets	0.8	0.66	0.6
TIU-Nets(with BU-Net)	**0.82**	**0.68**	**0.61**

Table 3
The performance comparisons of TIU-Nets without edge branch trained by different loss. (Bold number is the best score in each row.)

Loss		Cross-Entropy	Weighted Cross-Entropy	Focal	S-CE
Dice	Complete	0.80	0.76	**0.81**	**0.81**
	Core	0.65	0.67	0.67	**0.68**
	Enhance	**0.61**	0.59	0.60	**0.61**

Table 4
The performance comparison of TIU-Nets with edge branch using different loss. (Bold number is the best score in each row.)

Loss		Cross-Entropy	Weighted Cross-Entropy	Focal	S-CE
Dice	Complete	0.81	0.77	0.8	**0.82**
	Core	0.67	**0.68**	0.67	**0.68**
	Enhance	**0.61**	0.60	0.60	**0.61**

Table 5
The performance comparisons of TIU-Nets without edge branch when it is trained by using S-CE loss with different γ. (Bold number is the best score in each row.)

γ		4	8	12	16	20
Dice	Complete	0.81	0.81	0.81	**0.82**	0.81
	Core	0.67	0.67	**0.68**	0.64	**0.68**
	Enhance	0.60	0.60	**0.61**	0.60	**0.61**

Table 6
The performance comparison of TIU-Nets with edge branch when it is trained by using S-CE loss with different γ. (Bold number is the best score in each row.)

γ		4	8	12	16	20
Dice	Complete	0.81	0.81	**0.82**	0.81	0.81
	Core	0.66	0.66	**0.68**	0.67	0.67
	Enhance	0.60	0.61	0.61	0.61	**0.62**

Table 7
The performance comparison of TIU-Nets with different loss proportion, λ_1 , λ_2 and λ_3 are weights of L_{seg}, L_{mask} and L_{edge} respectively (Bold number is the best score in each column)

Loss Proportion			Dice		
λ_1	λ_2	λ_3	Complete	Core	Enhance
1	**1**	**0.1**	0.8	0.68	0.6
1	**1**	**0.5**	0.81	0.68	0.59
1	**1**	2	0.8	0.66	0.6
1	**1**	**10**	0.8	0.64	0.6
1	**0.1**	**1**	0.79	0.66	0.6
1	**0.5**	**1**	0.81	0.67	0.6
1	**2**	**1**	0.81	0.67	0.6
1	**10**	**1**	0.81	0.67	0.61
0.1	**1**	**1**	**0.82**	0.67	0.6
0.5	**1**	**1**	0.81	0.67	0.6
2	**1**	**1**	**0.82**	0.67	**0.62**
10	**1**	**1**	**0.82**	**0.7**	0.61
1	**1**	**1**	**0.82**	0.68	0.61

of L_{mask} and L_{edge}, it has better segmentation performance than other cases. However, it shows lower segmentation accuracy when L_{edge} occupies the larger proportion. From above observations, relatively speaking, it can be known that large proportion of L_{seg} and low proportion of L_{edge} are beneficial to improve performance of the proposed method.

As we know from Eq. (9), Figs. 4(b) and 5, γ in S-CE loss determines the degree of polarization. To observe the influence of different γ on segmentation accuracy, we set different γ in S-CE loss whose reusult can be shown in Table 5 and Table 6. From Table 5 and Table 6, we find that the best performance in each evaluation index is almost distributed on the S-CE loss with $\gamma > 8$, that is to say, higher γ is always helpful for glioma segmentation. However, as shown in Table 5 and Table 6, it is not always better when we set γ with higher value in S-CE loss. Moreover, we find that when γ of S-CE loss is equal to 12, we can get segmentation results.

4.3. Performance evaluation

4.3.1. Evaluations on BRATS2015

To validate the efficiency of the proposed method, we compare the proposed methods in 2D and 3D versions respectively. Specifically, we compare 2D TIU-Nets with the state-of-the-art 2D methods including FCN [17], SegNet [36], U-Net [7], CRDN [38],

IVD-Net [44] and FCDenseNet [45], and also compare the composed methods by combining FCN [17], SegNet [36], U-Net [7], CRDN [38] with VGGNet, ResNet50 and U-Net, when these methods are evaluated on the BRATS2015 testing dataset. When VGGNet is used as backbone, we can get FCN, SegNet, UNetVGG, CRDNVGG. Similarly, FCNResNet50, UNetsResNet50, and CRDNResNet50 can be obtained if the backbone of each network uses ResNet50. We have FCNUNet, SegNetUNet, and CRDN when we use U-Net as backbone. Meanwhile, we compare 3D TIU-Nets with the state-of-the-art 3D methods including DMFNet [48], LSTM-UNet [47] and DeepMedic [46], and also compare the proposed method with FCN [17], U-Net[7], SegNet [36] and CRDN [38] by extending them to 3D version.

To assess the performance of these methods, dice is considered to evaluate these methods on three regions (complete, core, enhance), as shown in Table 8 and 9. We rank these methods of each evaluation index and calculate the mean rank for each method in Table 8.

From Table 8, it can be seen that TIU-Nets with edge branch and S-CE loss ranks number one, when it is compared with several state-of-the-art methods. The dice of our method is higher at least 2% than that of methods with VGGNet as backbone in complete region, while the parameter number of TIU-Nets is at most 3% of four VGGNet backbone methods. Meanwhile, in term of dice of complete region, our method outperforms FCNVGG. Although the dice score of our method is a little inferior to methods with UNetResNet50 as backbone in core region, it has 2%, 1% higher dice than methods with UNetResNet50 as backbone in complete region and in enhance region, and the parameter number of our TIU-Nets is 2% of these methods. Moreover, the dice of the proposed method is at least 1% better than methods with U-Net as backbone in complete region, but the parameter number of our TIU-Nets is at most 21% of methods with U-Net as backbone. Our method is equally matched CRDN in aspect of dice score, but the parameter number in our method are nearly one third of those in CRDN. To further verify the efficiency of TIU-Nets, we compare TIU-Nets with IVD-Net and FCDenseNet. It has at least 2% higher dice than IVD-Net in complete and core regions. Meanwhile, the its dice is also higher 2%, 1% than FCDenseNet respectively in complete and core regions.

From Table 9, 3D TIU-Nets ranks the number one, when it is compared with several state-of-the-art methods. Compared with 2D TIU-Nets in Table 8, 3D TIU-Nets has 3%, 1% and 2% higher dice in the complete, core and enhance regions. Our method is higher at least 1% than U-Net in aspect of dice but the parameter number of our TIU-Nets is 25% of that in U-Net. Meanwhile, dice of 3D TIU-Nets is 1%, 3% and 4% higher than CRDN in complete, core and enhance regions respectively. However, the parameter number in our method is nearly one third of that in CRDN. Meanwhile, it also has 3%, 2% and 5% higher dice than LSTM-UNet in complete, core and enhance regions respectively. Although 3D TIU-Nets has more parameters than DeepMedic method, its dice is at least 4% higher than DeepMedic in all regions. Furthermore, its dice is higher at least 1%, 2% and 5% than DMFNet, LSTM-UNet and 3D FCN in complete, core and enhance regions respectively. Because of serious over-fitting of SegNet, we reduce the parameter number of SegNet by a quarter. However, as shown in Table 9, the performance of SegNet is inferior to the proposed method.

To visually compare performance of these methods, we select two representative slice of patients and compare our segmentation results with those state-of-the-art methods in Fig. 7. From Fig. 7, we find that 3D methods including 3D FCN, 3D U-Net, 3D CRDN, 3D SegNet, DeepMedic, DMFNet, LSTM-UNet and 3D TIU-Nets show higher segmentation accuracy than 2D models including methods with U-Net as backbone, methods with VGGNet as backbone, methods with ResNet50 as backbone, IVD-Net, FCDenseNet and TIU-Nets. Compared with 2D methods, segmentation results from 3D methods have better continuity in and clearer boundaries. In 2D methods, TIU-Nets has better results along glioma boundaries and contains less segmentation noises than other segmentation methods. Meanwhile, in segmentation results of 3D TIU-Nets, subregions are more continuous, and the contours of subregions are clearer and more precise compared with those obtained from other 3D methods. From the analysis above, it can be known that the parameter number of the proposed method is far less than other methods, while our method has the best performance in term of dice score for brain tumor segmentation.

4.3.2. Evaluations on the BrainWeb

To further show the generalization of our method, the BrainWeb dataset [38] is used for training and testing. BrainWeb dataset consists of simulated MR images with three kinds of modalities: t1, t2 and pd. In this dataset, 239 slices with three modalities are chosen for training and validation, and we use 160 slices for testing. As described above, we preprocess these MR images according to Eq. (14), before they are fed into the proposed TIU-Nets. To demonstrate the efficiency of the proposed method, we compare it with several state-of-the-art methods of FCN [17], SegNet [36], U-Net

Table 8
Performance evaluation on segmentation of different state-of-the-art methods in BRATS2015 (Bold number is the best score in each column, and the number in the bracket is the performance ranking of specific measurement.)

Method	Index				
	Dice			Parameters	Mean Rank
	Complete	Core	Enhance		
FCN	0.79(4)	0.66(4)	0.60(2)	50.42 M(12)	5.50
SegNet	0.78(5)	0.63(6)	0.54(5)	29.45 M(11)	6.75
UNetVGG	0.78(5)	**0.69(1)**	**0.61(1)**	25.86 M(10)	4.50
CRDNVGG	0.78(5)	0.67(3)	0.59(3)	14.91 M(8)	4.75
FCNResNet50	0.79(4)	0.67(3)	0.59(3)	115.84 M(13)	5.75
UNetResNet50	0.80(3)	**0.69(1)**	0.60(2)	71.87 M(12)	4.50
CRDNResNet50	0.78(5)	0.67(3)	0.59(3)	23.66 M(9)	5.00
FCNUNet	0.79(4)	0.66(4)	0.60(2)	1.24 M(2)	3.00
SegNetUNet	0.78(5)	0.63(6)	0.54(5)	2.36 M(5)	5.25
U-Net	0.80(3)	0.65(5)	**0.61(1)**	1.94 M(4)	3.25
CRDN	0.81(2)	**0.69(1)**	**0.61(1)**	1.25 M(3)	1.75
IVD-Net	0.77(6)	0.66(4)	**0.61(1)**	13.61 M(7)	4.50
FCDenseNet	0.8(3)	0.67(3)	**0.61(1)**	3.46 M(6)	3.25
TIU-Nets	0.80(3)	0.65(5)	**0.61(1)**	**0.49 M(1)**	2.50
TIU-Nets + S-CE	0.81(2)	0.68(2)	**0.61(1)**	**0.49 M(1)**	1.50
TIU-Nets + S-CE + edge	**0.82(1)**	0.68(2)	**0.61(1)**	**0.49 M(1)**	**1.25**

Table 9
Dice evaluation on segmentation of different state-of-the-art 3D methods in BRATS2015 (bold numbers are the best score in each column, and the number in the bracket is the performance ranking of specific measurement.)

Method	Index				
	Dice			Parameters	Mean Rank
	Complete	Core	Enhance		
3D FCN	0.84(3)	0.67(2)	0.58(5)	117.59 M(8)	4.50
3D SegNet	0.79(5)	0.67(2)	0.62(2)	5.53 M(6)	3.75
3D U-Net	0.84(2)	0.67(2)	0.62(2)	5.65 M(7)	3.25
3D CRDN	0.84(2)	0.66(3)	0.59(3)	3.86 M(4)	3.00
DeepMedic	0.79(5)	0.63(4)	0.59(3)	**0.66 M(1)**	3.25
DMFNet	0.79(5)	0.62 (5)	0.57(5)	3.88 M(5)	5.00
LSTM-UNet	0.82(4)	0.66(3)	0.58(4)	85.99 M(14)	6.25
3D TIU-Nets	**0.85(1)**	**0.69(1)**	**0.63(1)**	1.44 M(2)	**1.25**

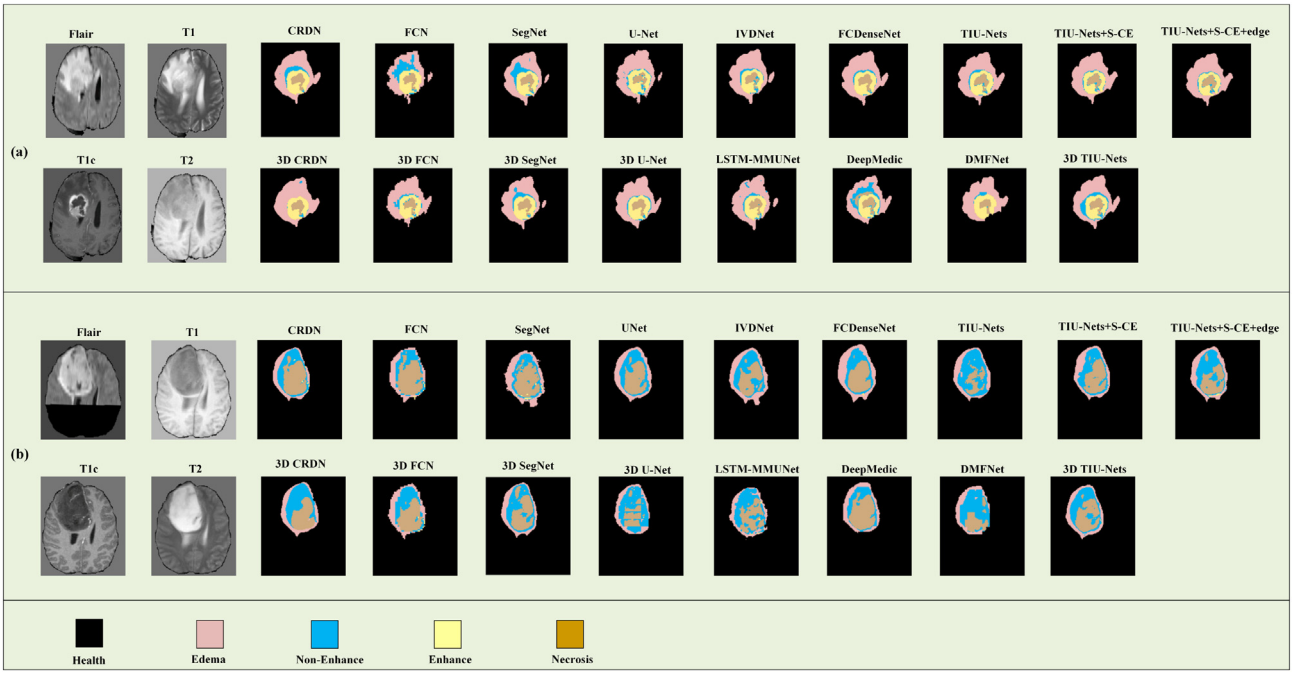


Fig. 7. The visual comparison of brain tumor segmentation results generated by different DNN methods on two samples (a and b) of glioma images from BRATS2015 testing dataset. The sample in (a) is an MR slice of patient with high grade glioma, and the sample in (b) is an MR slice of patient with low grade glioma. For each part, the methods in the first line are 2D methods, while 3D methods are in the second line.

Table 10
The comparisons of different segmentation methods in terms of PA and dice evaluation indexes on BrainWeb. (Bold number is the best score in each column, and the number in the bracket is the performance ranking of specific measurement.)

Method	Index			
	PA	Dice	Parameters	Mean Rank
FCN	0.9575(13)	0.9142(12)	50.42 M(13)	12.67
SegNet	0.9834(10)	0.9679(9)	29.45 M (12)	10.33
UNetVGG	0.9962(6)	0.9923(5)	25.86 M(11)	7.33
CRDNVGG	0.9964(5)	0.9927(4)	14.91 M(9)	6.00
FCNResNet50	0.9554 (14)	0.9115(13)	115.84 M(15)	14.00
UNetsResNet50	0.9954 (8)	0.9909(7)	71.87 M (14)	9.67
CRDNResNet50	0.9960(7)	0.9920(6)	23.66 M(10)	7.67
FCNUNet	0.9579(12)	0.9176(11)	1.24 M(3)	6.00
SegNetUNet	0.9715(11)	0.9455(10)	2.36 M(5)	8.67
U-Net	0.9945(9)	0.9892(8)	1.94 M(4)	7.00
CRDN	0.9967 (4)	0.9934(3)	1.23 M(2)	3.00
IVD-Net	0.9973(3)	**0.9950(1)**	13.61 M(8)	4.00
FCDenseNet	0.9975(2)	0.9946(2)	3.46 M(7)	3.33
TIU-Nets(small)	0.9975(2)	0.9946(2)	**0.49 M(1)**	**1.67**
TIU-Nets(large)	**0.9976(1)**	**0.9950(1)**	2.45 M(6)	2.67

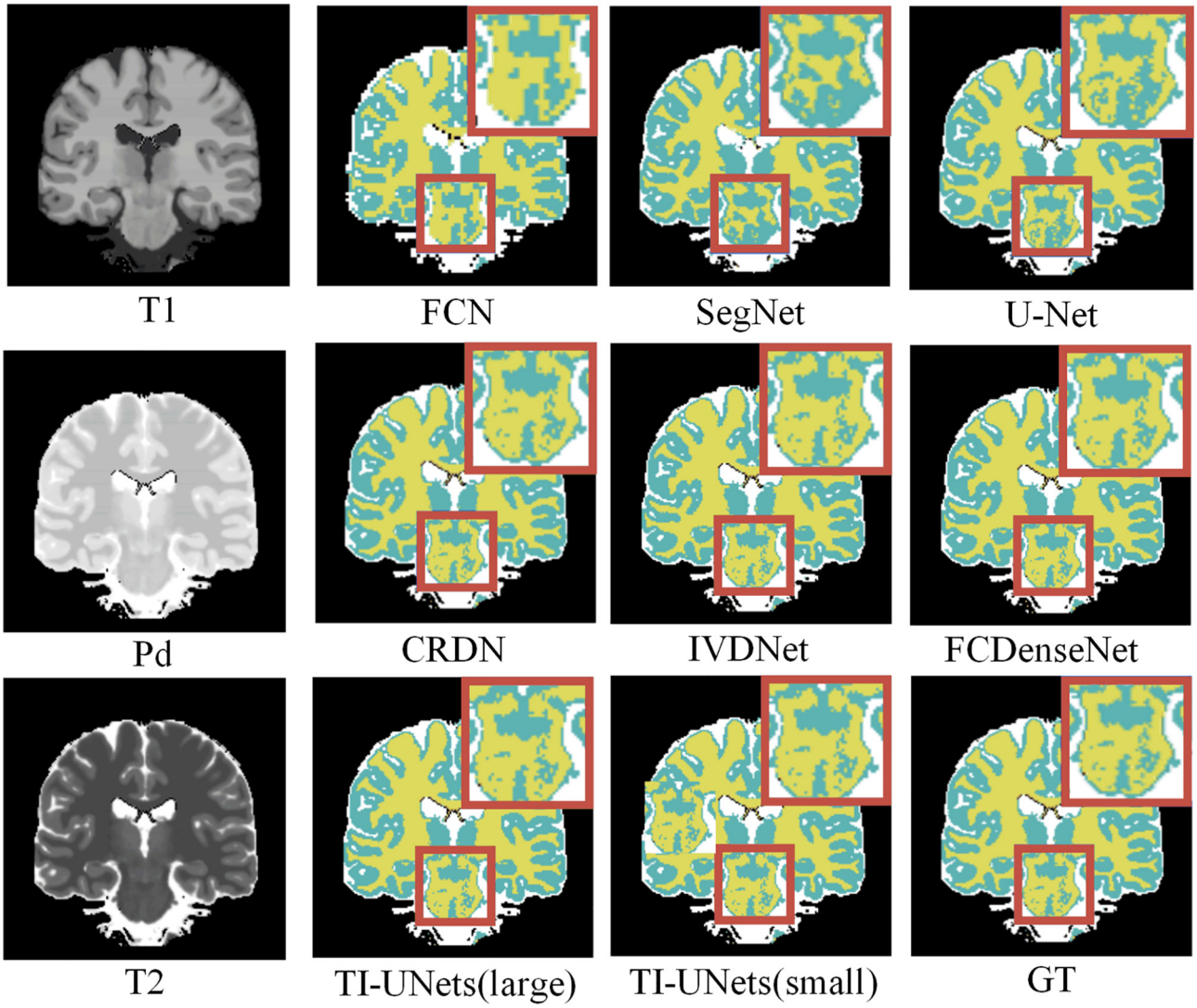


Fig. 8. The visual comparisons of different brain tissue segmentation results from BrainWeb dataset.

[7], CRDN [38], IVD-Net [44] and FCDenseNet [45] as shown in Table 10. Because the tissue boundaries are very clear and each tissue occupies a relative large proportion, the S-CE loss and edge branch have no significant improvements in segmentation. Accordingly, we train TIU-Nets without edge branch on BrainWeb by using cross-entropy loss.

From Table 10, it can be observed that the PA and dice improvement of our TIU-Nets is at least 0.02% and 0.06% respectively, when our TIU-Nets is compared with U-Net backbone methods. At the same time, our TIU-Nets has the PA and dice improvements with at least 0.07% and 0.13%, as compared with the VGGNet backbone methods. As for the ResNet backbone methods, the PA and dice improvement of our TIU-Nets is at least 0.11% and 0.2% respectively. TIUNets with the fewest parameter number shows similar performance, when it is compared with FCDenseNet and IVD-Net. Besides, our TIU-Nets outperforms these methods when it is extend to a larger verison.

To show the performance intuitively, we compare the segmentation images by different methods, as shown in Fig. 8. We select an axial slice of patient to show the visual segmentation comparison of these methods. From Fig. 8, it can be found that the close-up from slice segmentation predicted by our method contains fewer noises and finely details than those of other methods.

5. Conclusion

In this paper, we propose a multi-task driven triple intersecting U-Nets for brain tumor segmentation. To reduce noise disturbance in MR images, a segmentation soft-mask from BU-Net is estimated to guide multi-category segmentation in MU-Net. Meanwhile, we propose an S-CE loss to efficiently strengthen difficultly-classified pixels for training, which can alleviate class imbalance problem. To improve segmentation accuracy of glioma boundaries, we introduce an edge branch to enhance edge features in glioma boundaries. Compared with the state-of-the-art 2D methods, the proposed TIU-Nets has the best performance, but contains the fewest parameters, when testing on BRTAS2015 and BrainWeb dataset. We also compare the proposed 3D TIU-Nets with several 3D methods on BRATS2015, it shows that the proposed method has the best segmentation results. In the future, we will research on simultaneously multi-modality MR image segmentation and fusion based on deep neural networks.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

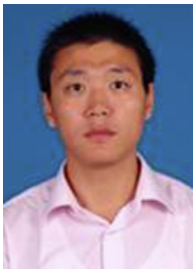
This work was supported by Doctoral Scientific Research Starting Foundation of Taiyuan University of Science and Technology (No. 20192023), Funding Awards for Outstanding Doctors Volunteering to Work in Shanxi Province (No. 20192055).

References

[1] Q. Ostrom, H. Gittleman, J. Fulop, M. Liu, R. Blanda, C. Kromer, J. Barnholtz, CBTRUS statistical report: primary brain and central nervous system tumors diagnosed in the united states in 2008–2012, Neuro-oncology 17 (2015) 1–62.
[2] P. Wesseling, D. Capper, WHO 2016 classification of gliomas, Neuropathol. Appl. Neurobiol. 44 (2018) 139–150.
[3] A. Claes, A. Idema, P. Wesseling, Diffuse glioma growth: a guerilla war, Acta Neuropathol. 114 (2007) 443–458.
[4] E. Neill, T. Luks, M. Dayal, J. Phillips, A. Perry, L. Jalbert, S. Nelson, Quantitative multi-modal mr imaging as a non-invasive prognostic tool for patients with recurrent low-grade glioma, J. Neuro-oncol. 132 (2017) 171–179.
[5] J. Henson, P. Gaviani, R. Gonzalez, MRI in treatment of adult gliomas, Lancet Oncol. 6 (2005) 167–175.
[6] A. Garcia, S. Orts, S. Oprea, V. Villena, J. Garcia, A review on deep learning techniques applied to semantic segmentation, in: arXiv:1704.06857, 2017.
[7] O. Ronneberger, P. Fischer, T. Brox, U-Net: convolutional networks for biomedical image segmentation (2015) 234–241.
[8] M. Havaei, A. Davy, D. Warde, A. Biard, A. Courville, Y. Bengio, H. Larochelle, Brain tumor segmentation with deep neural networks, Med. Image Anal. 35 (2017) 18–31.
[9] K. Kamnitsas, C. Ledig, V. Newcombe, J. Simpson, A. Kane, D. Menon, B. Glocker, Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation, Med. Image Anal. 36 (2017) 61–78.
[10] S. Pereira, A. Pinto, V. Alves, C. Silva, Brain tumor segmentation using convolutional neural networks in mri images, IEEE Trans. Med. Imag. 35 (2016) 1240–1251.
[11] G. Wang, W. Li, S. Ourselin, T. Vercauteren, Automatic brain tumor segmentation using cascaded anisotropic convolutional neural networks, in: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries – Third International Workshop, BrainLes 2017, 2017, pp. 178–190.
[12] A. Beers, K. Chang, J. Brown, E. Sartor, C. Mammen, E. Gerstner, J. Kalpathy, Sequential 3D U-Nets for biologically-informed brain tumor segmentation.
[13] X. Li, G. Luo, K. Wang, Multi-step cascaded networks for brain tumor segmentation, in: arXiv:1908.05887, 2019.
[14] C. Zhou, C. Ding, Z. Lu, X. Wang, D. Tao, One-pass multi-task convolutional neural networks for efficient brain tumor segmentation, in: Medical Image Computing and Computer Assisted Intervention-MICCAI 2018-21st International Conference, 2018, pp. 637–645.
[15] H. Shen, R. Wang, J. Zhang, S. McKenna, Multi-task fully convolutional network for brain tumour segmentation, in: Medical Image Understanding and Analysis, 2017, pp. 239–248.
[16] H. Liu, X. Shen, F. Shang, F. Ge, F. Wang, CU-Net: cascaded U-Net with loss weighted sampling for brain tumor segmentation, in: Multimodal Brain Image Analysis and Mathematical Foundations of Computational Anatomy – 4th International Workshop, MBIA 2019, 2019, pp. 102–111.
[17] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, IEEE Trans. Pattern Anal. Mach. Intell. 39 (2014) 640–651.
[18] P. Moeskops, M. Viergever, A. Mendrik, L. DeVries, M. Benders, I. Isgum, Automatic segmentation of MR brain images with a convolutional neural network, IEEE Trans. Med. Imag. 35 (2016) 1252–1261.
[19] P. Qin, J. Zhang, J. Zeng, H. Liu, Y. Cui, A framework combining DNN and level-set method to segment brain tumor in multi-modalities mr image, Soft. Comput. 23 (2019) 9237–9251.
[20] Z. Zhang, M.R. Sabuncu, Generalized cross entropy loss for training deep neural networks with noisy labels (2018) 8792–8802.
[21] S. Zhou, D. Nie, E. Adeli, J. Yin, J. Lian, D. Shen, High-resolution encoder-decoder networks for low-contrast medical image segmentation, IEEE Trans. Image Process. 29 (2019) 461–475.
[22] Z. Wu, C. Shen, A. Hengel, High-performance semantic segmentation using very deep fully convolutional networks, in: arXiv:1604.04339, 2016.
[23] P. Ferdin, Christ, M.E.A. Elshaer, F. Ettlinger, S. Tatavarty, M. Bickel, P. Bilic, M. Rempfler, M. Armbruster, F. Hofmann, M. D'Anastasi, W.H. Sommer, S. Ahmadi, B.H. Menze, Automatic liver and lesion segmentation in CT using cascaded fully convolutional neural networks and 3D conditional random fields, in: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016–19th International Conference, 2016.
[24] R. Azad, M. Asadi-Aghbolaghi, M. Fathy, S. Escalera, Bi-directional ConvLSTM U-Net with densley connected convolutions, in: 2019 IEEE/CVF International Conference on Computer Vision Workshops, ICCV Workshops 2019, 2019, pp. 406–415.
[25] D. Marmanis, K. Schindler, J. Wegner, S. Galliani, M. Datcu, U. Stilla, Classification with an edge: Improving semantic image segmentation with boundary detection, ISPRS J. Photogrammetry Remote Sens. 135 (2018) 158–172.
[26] M.S. Rad, B. Bozorgtabar, U. Marti, M. Basler, H.K. Ekenel, J. Thiran, SROBB: Targeted perceptual loss for single image super-resolution, in: 2019 IEEE/CVF International Conference on Computer Vision, ICCV 2019, 2019, pp. 2710–2719.
[27] S. Zhou, J. Wang, F. Wang, D. Huang, SE2Net: Siamese edge-enhancement network for salient object detection, in: arXiv:1904.00048, 2019.
[28] H. Xu, P. Liang, W. Yu, J. Jiang, J. Ma, Learning a generative model for fusing infrared and visible images via conditional generative adversarial network with dual discriminators, in: Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI 2019, 2019, pp. 3954–3960.
[29] S. Xie, Z. Tu, Holistically-nested edge detection, Int. J. Comput. Vis. 125 (1–3) (2017) 3–18.
[30] T. Ruan, T. Liu, Z. Huang, Y. Wei, S. Wei, Y. Zhao, Devil in the details: Towards accurate single and multiple human parsing, in: The Conference on Artificial Intelligence, AAAI2019, 2019, pp. 4814–4821.
[31] T. Liu, Y. Wei, Y. Zhao, S. Liu, S. Wei, Magic-Wall: Visualizing room decoration by enhanced wall segmentation, IEEE Trans. Image Process. 28 (2019) 4219–4232.
[32] X. Wang, H. Ma, X. Chen, S. You, Edge preserving and multi-scale contextual neural network for salient object detection, IEEE Trans. Image Process. 27 (2017) 121–134.
[33] A. Myronenko, A. Hatamizadeh, 3D kidneys and kidney tumor semantic segmentation using boundary-aware networks, in: arXiv:1909.06684, 2019.
[34] A. Hatamizadeh, D. Terzopoulos, A. Myronenko, End-to-end boundary aware networks for medical image segmentation, in: Machine Learning in Medical Imaging - 10th International Workshop, MLMI 2019, 2019, pp. 187–194.
[35] H. Noh, S. Hong, B. Han, Learning deconvolution network for semantic segmentation, in: 2015 IEEE International Conference on Computer Vision, ICCV, 2015, pp. 1520–1528.
[36] V. Badrinarayanan, A. Kendall, R. Cipolla, SegNet: a deep convolutional encoder-decoder architecture for image segmentation, IEEE Trans. Pattern Anal. Mach. Intell. 39 (2017) 2481–2495.
[37] L.C.L., G. Papandreou, I. Kokkinos, K. Murphy, A. Yuille, Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs, IEEE Trans. Pattern Anal. Mach. Intell. 40 (2017) 834–848.
[38] Y. Wen, K. Xie, L. He, Segmenting medical MRI via recurrent decoding cell, in: The Thirty-Fourth Conference on Artificial Intelligence, AAAI2020, 2020, pp. 12452–12459.
[39] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S.E. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, in: IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2015, 2015, pp. 1–9.
[40] M.D. Zeiler, R. Fergus, Visualizing and understanding convolutional networks, in: Computer Vision – ECCV 2014–13th European Conference, 2014, pp. 818–833.
[41] T.Y. Lin, P. Goyal, R. Girshick, K. He, P. Dollár, Focal loss for dense object detection, IEEE Trans. Pattern Anal. Mach. Intell. 42 (2017) 2999–3007.
[42] K. Doi, A. Iwasaki, The effect of focal loss in semantic segmentation of high resolution aerial image, in: 2018 IEEE International Geoscience and Remote Sensing Symposium, IGARSS 2018, 2018, pp. 6919–6922.
[43] N. Abraham, N.M. Khan, A novel focal tversky loss function with improved attention U-Net for lesion segmentation, in: 16th IEEE International Symposium on Biomedical Imaging, ISBI 2019, 2019, pp. 683–687.
[44] J. Dolz, C. Desrosiers, I.B. Ayed, IVD-Net: Intervertebral Disc localization and segmentation in MRI with a multi-modal unet, in: Computational Methods and Clinical Applications for Spine Imaging-5th, International Workshop and Challenge, CSI 2018, 2018, pp. 130–143.
[45] A. Kori, M. Soni, B. Pranjal, M. Khened, A. Varghese, G. Krishnamurthi, Ensemble of fully convolutional neural network for brain tumor segmentation from magnetic resonance images, in: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries-4th International Workshop, BrainLes 2018, 2018, pp. 485–496.
[46] K. Kamnitsas, E. Ferrante, S. Parisot, C. Ledig, A.V. Nori, A. Criminisi, D. Rueckert, B. Glocker, DeepMedic for brain tumor segmentation, in: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries – Second International Workshop, BrainLes 2016, with the Challenges on BRATS, ISLES and mTOP 2016, 2016, pp. 138–149.
[47] F. Xu, H. Ma, J. Sun, R. Wu, Y. Kong, LSTM multi-modal UNet for brain tumor segmentation, in: 2019 IEEE 4th International Conference on Image, Vision and Computing (ICIVC), 2019.
[48] J. Yuan, W. Zhou, T. Luo, DMFNet: deep multi-modal fusion network for RGB-D indoor scene segmentation, IEEE Access 7 (2019) 169350–169358.
[49] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Köpf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, S. Chintala, Pytorch: an imperative style, high-performance deep learning library, in: Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, 2019, pp. 8024–8035.
[50] B. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, L. Lanczi, The multimodal brain tumor image segmentation benchmark, BRATS, IEEE Trans. Med. Imag. 34 (2014) 1993–2024.
[51] S. Woo, J. Park, J. Lee, I.S. Kweon, CBAM: convolutional block attention module, in: Computer Vision – ECCV 2018–15th European Conference, 2018, pp. 3–19.

Jinjing Zhang received her B.S. from North University of China in Taiyuan, China, in 2014. She is studying for her doctorate at North University of China. She is mainly majoring medical image segmentation.

Jianchao Zeng received the B.Sc. degree from the Taiyuan Heavy Machinery Institute, Taiyuan, China, in 1982, and the M.Sc. and Ph.D. degrees from Xi'an Jiaotong University, Xi'an, China, in 1985 and 1990, respectively. He is currently a Professor with the School of Data Science and Technology, North University of China, Taiyuan. He has published over 200 international journals and conference papers. His current research interests include modeling and control of complex systems, intelligent computation, swarm intelligence, and swarm robotics.,Dr. Zeng is the Vice President of the North University of China and the Technical Committee on System Simulation of the Chinese Association of Automation and the Director of the China Simulation Federation.



Pinle Qin was born in 1978, and received his PhD degree from Dalian University of Technology. His research interests include big data,machine vision, three-dimensional reconstruction.

Lijun Zhao received his M.S. degree from Taiyuan University of Science and Technology in 2015 and PhD degree from Beijing Jiaotong University in 2019. He is currently an assistant lecturer in the Institute of Digital Media and Communication, Taiyuan University of Science and Technology. His research interests include image coding, multiple description coding, 3D video processing, pattern recognition and computer vision, etc.