IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received April 19, 2019, accepted June 17, 2019, date of publication June 24, 2019, date of current version July 12, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2924744

Dual U-Net for the Segmentation of Overlapping Glioma Nuclei

XIELI LI[1,2], YUANYUAN WANG[1,2,3], QISHENG TANG[4], ZHEN FAN[4], AND JINHUA YU[1,2,3]

[1]Department of Electronic Engineering, Fudan University, Shanghai 200433, China
[2]Institute of Functional and Molecular Medical Imaging, Fudan University, Shanghai 200040, China
[3]Key Laboratory of Medical Imaging Computing and Computer-Assisted Intervention, Shanghai 200433, China
[4]Department of Neurosurgery, Huashan Hospital, Fudan University, Shanghai 200040, China

Corresponding author: Yuanyuan Wang (yywang@fudan.edu.cn)

This work was supported by the National Basic Research Program of China (2015CB755500).

ABSTRACT The morphology and surroundings of cells have been routinely used by pathologists to diagnose the pathological subtypes of gliomas and to assess the malignancy of tumors. Thanks to the advent and development of digital pathology technology, it is possible to automatically analyze whole slides of tissue and focus on the nucleus in order to derive a quantitative assessment that can be used for grading, classification, and diagnosis. During the process of computer-assisted diagnosis, the accurate location and segmentation of nuclei from hematoxylin and eosin (H&E)-stained histopathological images is an important step. In this paper, we proposed a U-Net-based multi-task learning network in which the boundary and region information is utilized to improve the segmentation accuracy of glioma nuclei, especially overlapping ones. To refine the segmentation, a classification model is used to predict the boundary, a regression model is used to predict the distance map, and the final segmentation is obtained by using the fusion layers. The proposed approach was compared with other specially designed boundary-aware methods by using a pathological section dataset that consists of 320 glioma cases from the Huashan Hospital at Fudan University. Both the pixel-level and object-level evaluations showed that the structural modification is effective in segmentation with an F1-score of 0.82, a Hausdorff distance (HD) of 3.95, and an aggregated Jaccard index (AJI) of 0.66 (+0.46%, −3.75%, and +4.09% compared with the unimproved methods, respectively). In addition, comparative experiments on multi-organ nuclei segmentation (MoNuSeg) open dataset proved the advanced nature of the proposed method in the field of nuclei segmentation, especially separating touching objects. The proposed method obtains an AJI of 0.59 and an F1-score of 0.79.

INDEX TERMS Cancer research, deep learning, digital pathology, nuclei segmentation.

I. INTRODUCTION

The histological phenomena observed by pathologists when viewing tissue biopsy images are central for pathological grading, diagnosis and treatment of cancer [1]. Conventionally, a tumor tissue specimen processed by dehydration, sectioning, staining and fixation will be examined optically by pathologists via a microscope [2]. However, recently, with the application of whole slide imaging, this analysis has become digital. One necessary and important step within digitalized histological analyses is the nuclear segmentation of pathological images. Manual annotation and counting lack reproducibility, require considerable effort, and are time consuming. Therefore, automated segmentation methods are highly needed to improve the robustness of nuclear location and segmentation.

However, automating nuclei detection is not a trivial task and can be challenging for a number of reasons. First, there are various kinds of cells with different nuclear shapes, even in one typical tumor patch. Among the whole slide images (WSIs) of gliomas that we processed, there were triangular nuclei from nerve cells, tiny round nuclei from glial cells, oval nuclei from astrocytes, dark round nuclei with light rings from oligodendrocytes, slender nuclei from endothelial cells, and irregular nuclei from malignant tumor cells (see Fig. 1). The methods must be robust enough to detect all kinds of nuclei effectively. Second, one of the major challenges of nuclear segmentation is the separation of overlapping objects [1]. Overlapping occurs when the cells

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

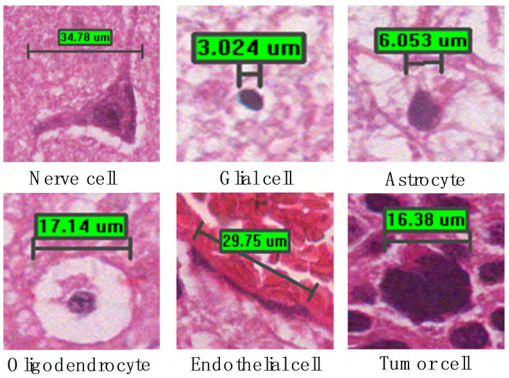

FIGURE 1. Six kinds of nuclei with various shapes and sizes in gliomas. From top to bottom and left to right are the pictures of a nerve cell, a glial cell, an astrocyte, an oligodendrocyte, an endothelial cell and a tumor cell. The approximate size is given in the green label.

proliferate or nuclei from different depths are projected in the same place. The undersegmentation of overlapping nuclei results in the incorrect estimation of the nuclear density, size and morphology. Third, the inevitable variation of the digital image quality is another problem to be solved in nuclear segmentation. Variations are introduced technically by staining, fixating and scanning the tissue samples and biologically by the development of the tumor. In conclusion, the diversity and complexity of the nuclear phenotypes make segmentation difficult.

Considering these challenges and the development of semantic segmentation, many automated algorithms have been proposed. Two comprehensive review papers introduced most of the nuclear segmentation techniques [1], [3]. We summarize these techniques as the conventional segmentation methods and the deep-learning based methods.

The conventional segmentation methods include thresholding following morphological operations [4], [5], watershed segmentation [6], deformable models [7], graph-based models [8], [9], and their variants along with multitudes of pre- and post-processing steps and are designed to achieve the aforementioned goals. These methods focus on the distribution, variation, gradient and other color features. However, due to inter- and intra-nuclear color variations in the wide spectrum of tissue morphologies, these methods fail to generalize the scene of a whole slide nuclear segmentation.

Techniques based on machine learning can obtain better results because they can be trained to recognize shape and color variations. In recent years, convolutional neural networks (CNNs) have dominated the field of image classification [10], [11] and image segmentation [12]. Based on the recognition ability of CNNs, fully convolutional networks improve the efficiency of semantic segmentation [13], [14]. Broadly speaking, previous studies of deep learning frameworks on nuclear segmentation can be categorized into two classes. (1) CNN-based nuclear detection with deformable models. For this kind of method, the CNN is applied to generate a probability map of a nuclear centroid or area and algorithms like the watershed transform, clustering, and active contour are utilized to post-process the boundary of nuclei. In [15], a typical simple model of a CNN is proposed to segment electron microscopy images. The deep neural network pixel-by-pixel predicts the probability that the center of the scanning patch is a neuronal membrane, and the segmentation result of the whole image is given by thresholding the spatially smoothed probability map. In [16], the probability map produced by the learned CNN model is regarded as the result of nuclei detection and the initial contour of the nuclei. Next, shape deformation based on the active contour model is performed to segment these nuclei. In conclusion, this kind of method is computationally expensive for large-scale images because nuclear segmentation requires pixel-wise classification via a sliding-window and a post-processing algorithm [3]. Therefore, previous studies are focused on the fast scanning strategy [17]–[20] and unsupervised segmentation models [21]–[26]. (2) FCN-based pixel wise segmentation. For this kind of method, the deep learning model is expected to give pixel-wise segmentation results end-to-end. Generally speaking, this kind of segmentation model consists of an encoder-decoder architecture that includes fully convolutional layers and is sometimes embedded with a refining algorithm such as watershed transform and conditional random field (CRF). In [14], a symmetric structure called U-Net is proposed to segment the neuronal membranes end-to-end and achieved state-of-art performance on the 2015 ISBI challenge. In [27], the FCN learning of the nucleus region is used to guide the graph construction and generate an in-region cost function. In other words, the latter method is more efficient and time-economical than the former.

In the research process of deep-learning based segmentation models, many new approaches have emerged to solve the challenge of overlapping targets. Hao Chen et al proposed the deep contour network (DCAN), which generates both the object and contour predictions in the segmentation of histopathological images [28]. DCAN took the boundary information into consideration in order to separate touching or overlapping targets. Haocheng shen et al improved the DCAN to the boundary-aware FCN by adding a combination stage at the end of the network [29]. The results proved that the convolutional deduction of the combination stage works better than the conditional judgment. Based on the idea of being boundary/contour-aware, a method that fuses two FCNs to segment nuclei was proposed in [30]. The two FCNs are independently trained to recognize boundaries and overlapping targets, while the third FCN is trained as a fusion network for the separation of targets. Similar methods of boundary recognition occurred in [31]. A tri-partition CNN with a simple architecture was used predict an image pixel by pixel and classified the background, nuclei and the boundary outside the nuclei. In addition, Peter Naylor et al formulated the problem as a regression task by replacing the binary annotation with a distance map to train the network [32]. By concentrating on the nuclear core area instead

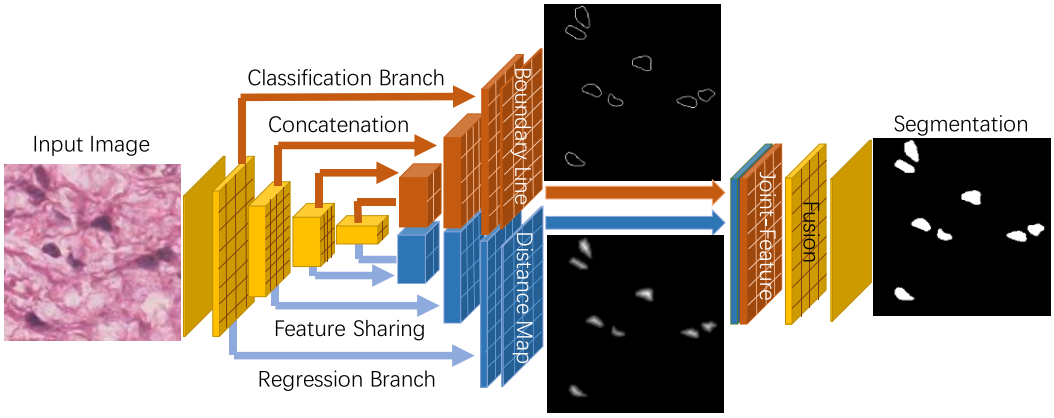

FIGURE 2. Flowchart of the proposed method. The top branch is a classification model for boundary segmentation, and the lower branch is a regression model for the distance map prediction. The intermediate results are combined for the final segmentation.

of the fuzzy boundary, the segmentation performance is better.

In this paper, we propose a novel dual U-Net architecture to solve the challenging problem of nucleus segmentation. The proposed network adopts the encoder-decoder structure from U-Net, and it shares the downsampling path yet equips a pair of upsampling paths: one classifies the nuclear region, and the other is sensitive to the nuclear boundary. To refine the segmentation, the nuclear boundaries are obtained by the classification model while the coarse regions are given by the regression of the distance map. By combining the information of the coarse region and boundary via some special fusion layers, the model is capable of segmenting overlapping nuclei in gliomas. The main advantages of the proposed model are that it (1) enables the simultaneous segmentation of multiple nuclei in glioma and (2) improves the separation of touching and overlapping nuclei. The remainder of this article is organized as follows. Section II gives the details of the proposed model and its theoretical basis. Section III introduces the dataset and the evaluation metrics we used in the experiments. Section IV presents the comparison results of the nuclear segmentation, including the experiments of the training optimization, the robustness testing, the effectiveness of the improvements, and the advanced nature of the model. Section V discusses these performances, and Section VI concludes the paper.

II. PROPOSED METHOD

In this section, we describe in detail the construction of our proposed dual U shape network for overlapping cell segmentation and the theoretical explanation of the network structure.

To make full use of the boundary information and the region information, the contour/boundary-aware [28], [29] structure is employed. An end-to-end segmentation network based on U-Net [14] that was modified according to the images of glioma cells is utilized as the backbone of the method. Moreover, to improve the recognition ability of overlapping targets, the deep regression of the distance map [32] is embedded. Finally, all the information is integrated by the fusion layers to produce the refined segmentation results. The flow chart of the proposed method is shown in Fig. 2.

Since our method is a multitasking segmentation in a sense, we hereby make some symbolic explanations. In this section, let $\mathcal{I}$ be the space of RGB images (in particular, hematoxylin and eosin (H&E) stained histopathological images) where $\mathcal{I} = \mathbb{R}^{r\times c\times 3}$. Let $\mathcal{A}$ be the space of annotated images (the inner part of the nuclei) where $\mathcal{A} = \{0, 1\}^{r\times c}$. The target of our model is to find a prediction function f that establishes a map from $\mathcal{I}$ to $\mathcal{A}$. Assuming an unseen image $I \in \mathcal{I}$, f helps us to predict its annotated image $A \in \mathcal{A}$. By the means of some transformations, it is easy to get the region annotation $R \in \{0, 1\}^{r\times c}$ (the whole part of nuclei) and the boundary annotation $B \in \{0, 1\}^{r\times c}$ (the boundary line separating the nuclei), which provide more useful information in the search of f.

A. DUAL U SHAPE NETWORK

After the proposition of the FCN, several novel pixel-wise networks came into being for various kinds of segmentation tasks. What these networks have in common is the downsampling-upsampling structure (or the encoder-decoder architecture). For example, SegNet [12], which consists of a VGG-like encoder, an inverse decoder and special unpooling layers, was trained for scene segmentation. U-Net [14] is an elaborate deformation of the FCN-VGG16 [13] and likewise consists of an encoder-decoder pair. Compared to a natural image, the content of a biological image is simpler. Therefore, although the architecture of U-Net is not so deep, it performs well in biological tasks. For the sake of the model's size and performance, we take the structure of U-Net as the foundation of our network. The structure of the proposed network is presented in Table 1 and Table 2.

In our study, the digital pathological images of glioma are cropped into small image blocks, and thus, the input

TABLE 1. Dual U-Net Architecture.

Name	Input	Type	Kernel Size+Stride	Output size
Section1: Downsampling branch				
ConvBlock1.1	RGB image	Initial	–	$64 \times 128 \times 128$
Pool1	ConvBlock1.1	Max pooling	$2 \times 2 + 2$	$64 \times 64 \times 64$
ConvBlock1.2	Pool1	Downsampling	–	$128 \times 64 \times 64$
Pool2	ConvBlock1.2	Max pooling	$2 \times 2 + 2$	$128 \times 32 \times 32$
ConvBlock1.3	Pool2	Downsampling	–	$256 \times 32 \times 32$
Pool3	ConvBlock1.3	Max pooling	$2 \times 2 + 2$	$256 \times 16 \times 16$
ConvBlock1.4	Pool3	Downsampling	–	$512 \times 16 \times 16$
Pool4	ConvBlock1.4	Max pooling	$2 \times 2 + 2$	$512 \times 8 \times 8$
ConvLayer1.5	Pool4	Convolution+BN+ReLu	$3 \times 3 + 1$	$1024 \times 8 \times 8$
DropOutLayer	ConvLayer1.5	Dropout	$drop\_rate = 0.5$	$1024 \times 8 \times 8$
ConvLayer1.6	DropOutLayer	Convolution+BN+ReLu	$3 \times 3 + 1$	$512 \times 8 \times 8$
Section2: Upsampling Branch for Boundary Segmentation				
Deconv2.1	ConvLayer1.6	ConvTransposition	$4 \times 4 + 2$	$512 \times 16 \times 16$
Concate2.1	Deconv2.1+ ConvBlock1.4	Concatenation	–	$1024 \times 16 \times 16$
ConvBlock2.1	Concate2.1	Upsampling	–	$256 \times 16 \times 16$
Deconv2.2	ConvBlock2.1	ConvTransposition	$4 \times 4 + 2$	$256 \times 32 \times 32$
Concate2.2	Deconv2.2+ ConvBlock1.3	Concatenation	–	$512 \times 32 \times 32$
ConvBlock2.2	Concate2.2	Upsampling	–	$128 \times 32 \times 32$
Deconv2.3	ConvBlock2.2	ConvTransposition	$4 \times 4 + 2$	$128 \times 64 \times 64$
Concate2.3	Deconv2.3+ ConvBlock1.2	Concatenation	–	$256 \times 64 \times 64$
ConvBlock2.3	Concate2.3	Upsampling	–	$64 \times 64 \times 64$
Deconv2.4	ConvBlock2.3	ConvTransposition	$4 \times 4 + 2$	$64 \times 128 \times 128$
Concate2.4	Deconv2.4+ ConvBlock1.1	Concatenation	–	$128 \times 128 \times 128$
ConvBlock2.4	Concate2.4	Classification	–	$2 \times 128 \times 128$
Section3: Upsampling Branch for Distance Map Regression				
		Repeat Section2, without ConvBlock2.4		
ConvBlock3.4	Concate3.4	Regression	–	$1 \times 128 \times 128$
Section4: Fusion Layers for Final Segmentation				
Concate4.1	ConvBlock2.4+ ConvBlock3.4	Concatenation	–	$3 \times 128 \times 128$
ConvLayer4.1	Concate4.1	Convolution+BN+ReLU	$3 \times 3 + 1$	$64 \times 128 \times 128$
ConvLayer4.2	ConvLayer4.1	Convolution+BN+ReLU	$3 \times 3 + 1$	$64 \times 128 \times 128$
Segmentation	ConvLayer4.2	Convolution	$1 \times 1 + 1$	$2 \times 128 \times 128$
Total Loss	ConvBlock2.4+ ConvBlock3.4+ Segmentation	Cross Entropy+ Mean Square Error+ Cross Entropy	–	$1 \times 1 \times 1$

The output sizes are given for an example input of 128×128; the details of the ConvBlock are given in Table II.

TABLE 2. Detailed settings of convblock with five different types.

Layer	Initial	Downsampling	Upsampling	Classification	Regression
Input	$3 \times 128 \times 128$	$D \times H \times W$	$D \times H \times W$	$128 \times 128 \times 128$	$128 \times 128 \times 128$
Conv1	$3 \times 3 \times 3 \times 64$	$3 \times 3 \times D \times 2D$	$3 \times 3 \times D \times \frac{D}{2}$	$3 \times 3 \times 128 \times 64$	$3 \times 3 \times 128 \times 64$
BN1+ReLU1	✓	✓	✓	✓	✓
Conv2	$3 \times 3 \times 64 \times 64$	$3 \times 3 \times 2D \times 2D$	$3 \times 3 \times \frac{D}{2} \times \frac{D}{4}$	$1 \times 1 \times 64 \times 2$	$1 \times 1 \times 64 \times 1$
BN2+ReLU2	✓	✓	✓	×	×
Output	$64 \times 128 \times 128$	$2D \times H \times W$	$\frac{D}{4} \times H \times W$	$2 \times 128 \times 128$	$1 \times 128 \times 128$

The size of the input and output is given by $depth \times height \times width$, and the size of the convolutional layer is given by $kernal\_h \times kernal\_w \times channel \times feature$.

size is modified as 128×128 pixels. The downsampling path in our network, which is known as the encoding part, includes four downsampling blocks, each containing two 3×3 convolutional layers and one 2×2 max pooling layer. Symmetrically, the upsampling path or the decoding part has four upsampling blocks, each of which is embedded with one

2×2 deconvolutional (convolution transposition) layer and two 3×3 convolutional layers. In order to keep the size consistent and avoid discarding image fringe information, all the feature maps are padded with 0 before the 3×3 convolution operations. Moreover, we placed Batch Normalization [33] and a Rectified Linear Unit [34] between all convolutions. Inspired by the idea of a contour/boundary-aware network, we copied the upsampling path for multitask segmentation. After four downsamplings, the data flow is split into two lines. The one for the boundary classification defined as f_B is trained using the boundary annotation B. The other for the coarse region segmentation defined as f_R is trained using the region information. Each upsampling block in both branches concatenates the feature map from the last layer and the same-scale intermediate results in the downsampling block into one. In this way, each encoder-decoder branch could be regarded as one complete U-Net for a specific task. Meanwhile, the training errors that are back-propagated from the two decoders synchronously tune the parameters in the encoding part. In our hypothesis, there is a high correlation between region and boundary segmentation. Therefore, it is reasonable that the two segmentation tasks partly share the same features.

B. REGRESSION OF THE DISTANCE MAP

In the task of nuclear segmentation, the most annoying problem is the undersegmentation of overlapping or touching nuclei. For most learning based approaches, loss obtained by calculating the belonging of pixels is usually applied for object segmentation task. This practice works well when we want to get the position and outline of the object. However, in the optimization process, the contribution of some misclassified pixels to the global loss can be neglected, which will lead to the improvement of the pixel-level accuracy, yet also lead to undersegmentation of individual objects. Therefore, the close or overlapping nuclei tend to be segmented as one in histopathology images. Thus, we want to refine the coarse segmentation before the fusion stage.

In the region segmentation branch, we borrowed the idea of deep regression from [32] to improve the coarse segmentation. In the binary ground truth, values inside the object are the same and value jumps near the boundary. But in the distance map, the value assigned to the background gently evolves to the value assigned to the object along the gradient, which is more in line with the description of the nuclei boundary in histopathology images. Nuclei boundaries are often fuzzy and ill defined, especially between touching nuclei where the local color characteristics on both sides are consistent. For these local patches, binary annotation assigns a same value to two nuclei and leaves a pixel-wide gap at the boundary which stands a good chance to be ignored during the optimization. The well-trained model has no ability to distinguish between interior and edge pixels, so high values of the score map on both sides submerge the boundary line. However, the distance map assigns a numerical valley between touching nuclei and assigns a buffer area with moderate numerical change along the fuzzy boundary. Although the incorrectly predicted value of the boundary line contributes little to the global loss, the loss caused by the numerical valley and slop prediction errors cannot be negligible. Thus, the well-trained model is able to acquire the patterns of different nuclei regions and produce regression map with continuously varying values from which touching objects can easily be separated. The rationale of distance map is that it can be regard as a series ground truth versions resulting from successive erosions. Hence, the distance map can provide more useful information and models can acquire the difference between background, boundary, interior and core area. Moreover, since two branches of the proposed model share part of the features, the distance map interacts with boundary segmentation. Because the distance map partly contains boundary information and have more in common with boundary annotation than binary ground truth, two branches reinforce each other. With the increasing of the coarse segmentation accuracy, the fine result is improved.

Distance maps are used for this part of the network training and take the place of binary maps. Similar to [32], we define the space of the distance map as $\mathcal{D} = \{D; D = Dist\,(R)\,, R \in \mathcal{R}\}$, where $Dist$ is the distance transform for each pixel from the Chebyshev distance to the closest background. For each individual nucleus, the nucleus that overlaps with it is regarded as the background. The Chebyshev distance between pixels $x = (i_x, j_x)$ and $y = (i_y, j_y)$ is $D_{ch} = max(\left|i_x - i_y\right|, \left|j_x - j_y\right|)$. Meanwhile, the Softmax layers in the region segmentation branch were removed, and the mean square error (MSE) took the place of cross entropy in the loss function. For the regression task, it can be regarded as a well-constructed U-Net whose output ought to be consecutive positive integers rather than one-hot vectors. Thus, MSE is used as the loss function for distance map regression. We define the prediction of the distance map from input image I as $\hat{D}$, i.e., $\hat{D} = f_R(I)$. The loss in this branch is defined as follows:

$$L_R\left(D, \hat{D}\right) = \frac{1}{np} \sum\nolimits_{i,j} \left(D(i,j) - \hat{D}(i,j)\right)^2. \tag{1}$$

In the testing case, we use a threshold value between 0 and 1 to get the coarse region segmentation.

C. FUSION LAYERS

After obtaining the boundary and the coarse region segmentation, we applied additional convolutional layers as the fusion method. Different from [29], we concatenate the predicted distance map with the probability map of the boundary. Therefore, the input depth of the fusion layer is 3. Two 3×3 convolutional layers retransform the integrated information into feature maps, and one 1×1 convolutional layer predicts the refined segmentation result. We define the fusion function as f_F. The total loss of our method consists of three parts: the loss from boundary segmentation, the loss from distance map regression and the loss from fusion segmentation. The loss

function is given as follows:

$$
\begin{aligned}
&L_{tatal}\left(I; D, B, A\right)\\
&= L_R\left(D, f_R\left(I\right)\right) + L_B\left(B, f_B\left(I\right)\right) + L_F\left(A, f_F\left(f_R\left(I\right), f_B\left(I\right)\right)\right)\\
&= \frac{1}{np}\sum\nolimits_{i,j}\Big[\alpha(D(i,j) - \hat{D}(i,j))^2 - \beta logP\left(B\left[i,j\right]; \hat{B}\left[i,j\right]\right)\\
&\quad - \gamma logP\left(A\left[i,j\right]; \hat{A}\left[i,j\right]\right)\Big], \qquad (2)
\end{aligned}
$$

where $\hat{B}$ and $\hat{A}$ have similar definitions as $\hat{D}$. α, β, and γ control the weight of each loss due to the different calculations of the MSE and cross entropy. By minimizing the function above, we get the well-convergent dual U-Net, in which f_R allows us to get the distance map, f_B helps us to recognize the boundary and the output of f_F is the precise nuclear segmentation result.

III. DATASETS AND EVALUATION METRICS

A. GLIOMA NUCLEI DATASET

The method proposed in this paper was evaluated by using an H&E stained glioma sections dataset. We generated a dataset at Huashan Hospital of Fudan University consisting of annotated H&E stained histology images at 40x magnification (see Table 3). All slides are taken from various glioma patients (grade, II to IV; type, astrocytoma, oligodendroglioma, or glioblastoma). They were scanned using an Aperio Scanscope XT and saved as WSIs.

TABLE 3. Composition of the glioma nuclei dataset.

Data subset	WSI	Image	Patch
Training	224	1120	17920
Validation	32	160	2560
Independent testing	64	320	5120
Total	320	1600	25600

For the 320 WSIs, we extracted five typical images with an average size of 245 × 240 from different areas of the tissue per WSI. Among all of these images, the longest length is 412 pixels, and the shortest is 143 pixels. The number, the morphology and the surroundings of the cells in these images have great variations because they were selected in order to present the heterogeneity of the data.

We invited experts to annotate a large number of nuclei. The overall dataset consists of 1600 images with a total of 32255 annotated nuclei. The maximum number of nuclei is 116 while the minimum is 2 in one image. Each image has an associated ground truth annotation depicting the inside and the outer edge of the nucleus and background area. Here, the inner and edge labels are used to generate A and B in Section II, respectively, while R is the union of A and B. Samples of the image with annotations are shown in Fig. 3. The annotation was performed by experienced pathologists. One annotates the image, another checks the first, and the third arbitrates if there is a disagreement.

Due to the limitation of the nuclear number, we augment the dataset by random rotation, color disturbances and elastic deformation. Since each image has a different size, we crop them into patches of 128×128 by random translation. Finally, 25600 patches are involved in the experiments, with 70% for training, 10% for validation and 20% for independent testing.

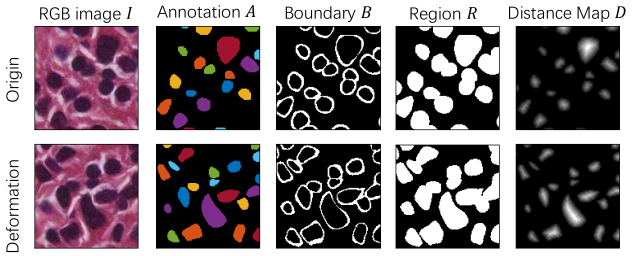

FIGURE 3. Sample images with annotations and symbolic explanations from our dataset. The first row is the original image extracted from the WSI, and the second row is an augmented image deformed by elastic deformation and rotation in our training dataset. *A* and *B* are generated from manual annotations, and *R* is the union of *A* and *B*. *D* is derived from *R* by using the distance transformation and region disconnection.

We used the glioma nuclei dataset to find the optimal training conditions and verify the effectiveness of the structure improvement.

B. MULTI-ORGAN NUCLEI DATASET

The second dataset is multi-organ nuclei segmentation (MoNuSeg) dataset. In total, this dataset involves 30 images from 7 different organs with 21623 annotated nuclei. Compared with the first dataset, MoNuSeg is more comprehensive and professional, and many methods were trained on it. Therefore, it is more suitable for verifying the universality and the advancement of the model. We used the same subset of images used in [31], [32], [35] for training and testing, so the results can be directly compared. We cut 16 training images into overlapping pieces, augmented them, and finally obtained 25600 patches.

C. EVALUATION MEASURES

Because the aim of the proposed method is accurate segment overlapping glioma nuclei, we employed four types of metrics to assess the performance in terms of the pixel-level and the object-level.

At pixel-level, we evaluated models by pixel accuracy (PA) and the F1 score (equivalent to the Dice index and Zijdenbos similarity index). PA is the most basic segmentation metric, calculating the proportion of the correctly classified pixel. We used PA to measure the quality of the dual U-Nets under different training conditions. F1 score is defined as the harmonic mean between the recall and precision of the pixel level segmentation. The calculation of F1 is as follows:

$$F_1 = \frac{2 * Precision * Recall}{Precision + Recall}. \qquad (3)$$

We use F1 score to measure the segmentation performance of different models at the pixel level.

At object-level, we evaluated models by the Hausdorff distance (HD) [36] and the aggregated Jaccard index

(AJI) [31]. HD is used for assessing the similarity between the predicted object and the ground truth, even if it is less popular due to the computational complexity. Assuming that P_i is the ith predicted nucleus and G_i is the matched ground truth, HD is defined as the Euclidean distance of the furthest pixels on the contours of these two shapes, which means that the lower the HD, the higher the similarity is:

$$HD = \max\{\max_{p\in P_i}[Dist\,(p, G_i)]\,, \max_{g\in G_i}[Dist(g, P_i)]\}, \quad (4)$$

where G_i is the nucleus that has the maximum overlap area with P_i in ground truth map and will not match the remaining predictions. If no nucleus matches P_i, then HD will be assigned the distance from the furthest pixels inside P_i as a punishment. In this way, both the undersegmentation and oversegmentation get a low similarity, even if they have high pixel accuracy. The AJI is the extension of the global Jaccard index and is defined as follows:

$$AJI = \frac{\sum_{i=1}^{L} |G_i \cap P_i|}{\sum_{i=1}^{L} |G_i \cap P_i| + \sum_{i\in rest} |P_i|}, \quad (5)$$

where P_i is the predicted nucleus that maximizes the Jaccard Index with the ground truth nucleus G_i and *rest* refers to the collection of P_i with no match. AJI reflects the proportion between the common region of matched elements and the segmented results. Any imprecise segmentation, whether under- or oversegmentation, will lead to a decrease in AJI.

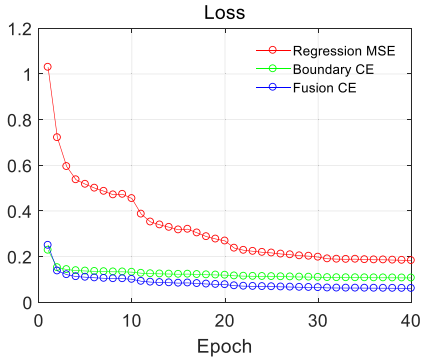

FIGURE 4. MSE loss curve and cross entropy loss curves during the pre-training.

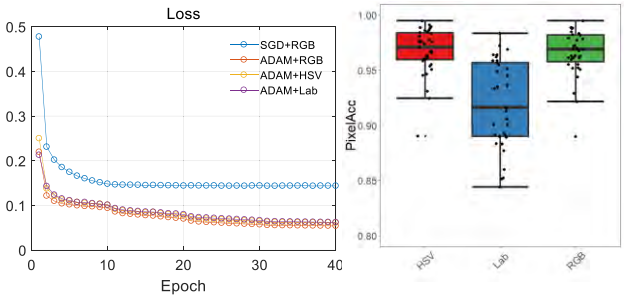

FIGURE 5. Loss curve of the models trained by different conditions (left) and pixel accuracy boxplot of models trained by three color spaces (right).

IV. EXPERIMENTS AND RESULTS

A. EXPERIMENTAL ENVIRONMENT

We performed the experiments using a NVIDIA TITAN Xp with 12GB memory and Windows 10 installed on an Intel(R) Xeon(R) CPU with 64 GB RAM. The deep learning algorithm is programmed in MATLAB R2017a with the MatCovNet toolkit [36], [37] .

B. TRAINING OPTIMIZATION

Before comparing with the other nuclei segmentation methods, the optimal training settings for dual U-Net should be studied. In this experiment, we took the weight of each part of the total loss function, the optimizer, and the color space of the training samples into consideration. The glioma nuclei dataset is used for training and testing.

Since the calculations of MSE and cross entropy are different, the three loss functions have to be balanced. We pre-trained the proposed model with $\alpha = 1, \beta = 1$ and $\gamma = 1$ to uncover the numerical size relation. The loss curves are shown in Fig. 4, from which we can learn that two cross entropy losses are about the same size while MSE loss is four times bigger than them. In order to balance all parts so that the total loss is not overwhelmed by MSE, we set $\alpha = 0.2, \beta = 1$ and $\gamma = 1$ in the subsequent experiments.

Then comes the selection of the optimizer and the color space. Stochastic gradient descent (SGD) algorithm has been widely used in CNN optimization while adaptive moment estimation (ADAM) algorithm has a better optimization performance. RGB color space is the basic storage form of image, but shows nonlinearity when color changes. We trained the following four models for 40 epochs with a batch size of 256, a weight decay of 0.0001 and an initial learning rate of 0.001 and decrease to a tenth every ten epochs: 1. RGB images with SGD optimizer; 2. RGB images with ADAM optimizer; 3. HSV images with ADAM optimizer; 4. Lab images with ADAM optimizer. The momentum of SGD is set to 0.9, and the parameters of ADAM are set as $\beta 1 = 0.9$, $\beta 2 = 0.99$ and $\varepsilon = 10 - 8$. Images are converted into specific color space then minus the statistical mean before input into the model. The final segmentation loss curves of these models are shown in Fig. 5. Apparently, the ADAM optimizer converges faster and performs better than SGD. The model trained by RGB space has a slight advantage over the others on training set. However, the pixel accuracy of independent test set is 0.966 for RGB, 0.967 for HSV, and 0.910 for Lab. Paired t-test manifests that the pixel-level segmentation accuracy of HSV is 0.001 higher than that of RGB with significance ($P = 0.02$). Actually, hues, saturation and brightness in HSV space are more in line with the eye's perception of color, which may contribute the slight improvement. As a result, we use ADAM optimizer and images of HSV color space to train the model in the subsequent experiments.

TABLE 4. Comparative results of independent testing and multiple cross-validation.

Pixel Acc	Independent Testing	Five-fold Cross Validation (7:1:2 for train: val: test)					
		Subset 1	Subset 2	Subset 3	Subset 4	Subset 5	Total
M±SD	0.967±0.021	0.950±0.031	0.968±0.024	0.967±0.028	0.959±0.036	0.964±0.025	0.962±0.030
P	-	0.004*	0.898	0.986	0.256	0.498	0.287

The mean values with the standard deviation are given in the second row, and the p values from t-test for independent testing and cross validation are given in the third row. *shows statistical significance

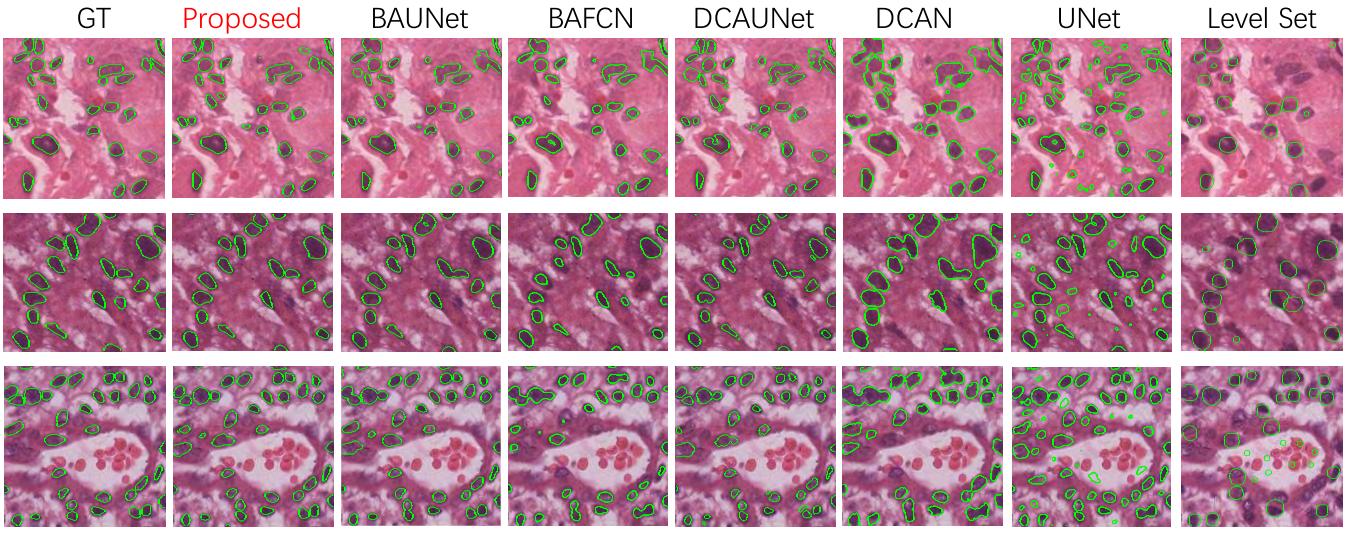

FIGURE 6. Comparison of the segmentation results of the deep-learning based methods. The green segmentation results from left to right are produced by the following: 1) Ground Truth, 2) Proposed (ours), 3) BAU-Net (from BAFCN), 4) DCAU-Net (from DCAN), 5) BAFCN [29], 6) DCAN [28], 7) U-Net for ternary segmentation, 8) Level Set algorithm [38].

C. DEPENDENCY ON DATA

Deep learning model is data driven and the performance decrease when the training samples are insufficient or non-homologous with the test sample. In order to verify that the division of the dataset is random, to prove that the model is robust when training data is reduced, and to make the results of subsequent comparative experiments convincing, we conduct the following experiments. First we randomly extracted about 10% images from glioma nuclei dataset as independent testing set and divided the patches augmented from the rest images into 9:1 for training and validation. Then we re-divided the dataset into 7:1:2 for training, validation and testing, and conducted five-fold cross validation. Six models were trained by the optimal settings in Part B, while the number and the content of the training samples are not the same, and the test results are listed in Table 4. The different division ratio of dataset has little impact on the final performance. The overall results of cross validation have no significant change comparing with independent testing. Only one subset of cases shows a significant decrease in pixel accuracy. The above results show that 70% of the samples are sufficient for model training, the performance of our model is convincing and robust. In the subsequent experiments, we use 70% for training, 10% for validation and 20% for testing and the results are acquired by cross-validation.

D. IMPROVEMENT EFFECTIVENESS

The proposed method was compared with other deep-learning based methods such as the simple U-Net [13], DCAN [28], boundary-aware FCN [29]. We conducted this experiment to prove the effectiveness of the structural improvements and the combination of the regression model. The training and testing samples are from glioma nuclei dataset and set according to Part B&C. For the simple U-Net, we trained a ternary segmentation model by annotation *A* and *B*.Because the contour/boundary-aware idea was first proposed in [28], [29], we replicated them on our dataset for comparison and abbreviate them as DCAN and BAFCN, respectively. In addition, we derived two other models by replacing the FCN structure with U-Net in former two models and abbreviate them as DCAU-Net and BAU-Net, respectively. Moreover, traditional nuclei segmentation methods such as level set algorithm [38] is also involved in the comparative experiment. All approaches for comparison purposes have been trained for enough epochs using groups of parameter settings to ensure that they converge. Besides, no post-processing method is involved in these approaches.

In terms of the visual results, as illustrated in Fig.6, all deep learning based methods can detect the approximate region of the nuclei. But traditional method like level set fails to locate nuclei with irregular shapes (Fig.6, column 8). Due to

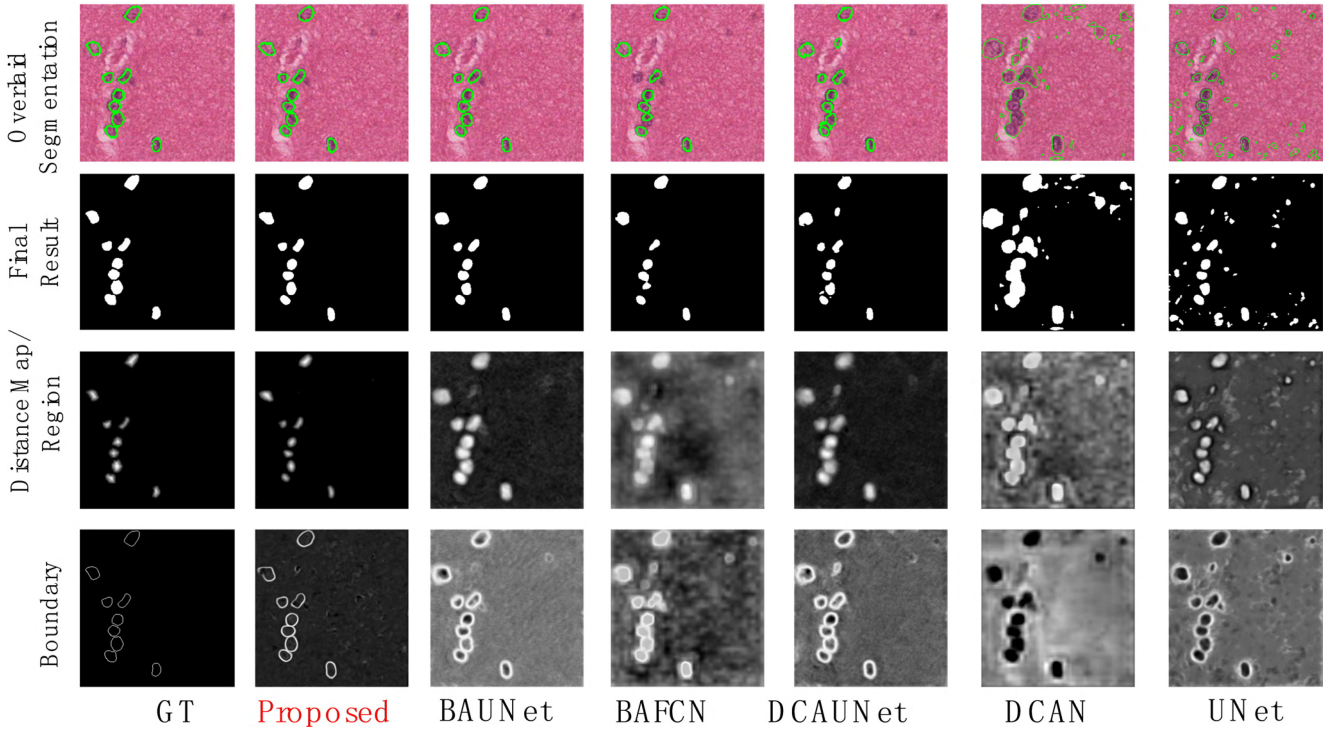

FIGURE 7. Detailed comparison of the individual image segmentations. From top to bottom are 1) the original images overlaid by segmentations, 2) the binary segmentation results, 3) the region information presented as the distance maps or the probability maps, and 4) the boundary probability maps.

the competition relationship between three labels, ternary U-Net mistakenly detects some objects (Fig.6, column 7). Because of the inconsistent between boundary segmentation and nuclei segmentation, artifacts are introduced in the results of DCAU-Net and DCAN (Fig.6, column 5&6). Three models embedded with fusion layers produce smoother contours (Fig.6, column 2,3&4), indicating the effectiveness of the fusion layers. Moreover, the contours produced by U-Net based models are more consistent with the ground truth than FCN based ones (Fig.6, column 3&4 versus 5&6), which demonstrates the advantage of using U-Net as the basic structure.

In high grade gliomas, nuclei connect with each other due to high proliferation and atypia, which results in the undersegmentation of several models. However, for both sparsely distributed and touching nuclei, the results of the proposed model are highly matched with the ground truth (Fig.6, column 2), thus indicating that the combination of regression model and classification model is effective. Some intermediate results are shown in Fig.7. For region detection task (Fig.7 row 4), the probability maps of U-Net based models have deep contrast between the object and the background (column 3&5) while the ones of FCN based models are vague (column 4&6). As for the distance map regression, the above characteristic is more obvious and the model concentrates on the core region of object (Fig.7, column 2). For the boundary segmentation task (Fig.7, row 3), the performances are highly related with the region detection task. Models with clear objects produce distinct boundary probability maps. The two intermediate results together determine the final segmentation performance. The proposed method obtains the thinnest boundary lines and the clearest region and hereby performs best.

In terms of the quantitative evaluation, the overall comparison of the evaluation metrics is provided in Table 5 and Fig.8. According to the results, models with the U-Net structure have a little improvement in their pixel-level indexes compared to those with the FCN structure. The application of the distance map makes the object-level segmentation more accurate. Thus the proposed dual U-Net quantitatively outperforms others. For the averages of the three measurements, our method achieved 0.81, 0.66 and 3.95 for F1, AJI and HD, respectively. In the pixel-level evaluation, the proposed network and the BAU-Net are almost identical with respect to their F1-scores. In the object-level evaluation, the proposed network has the lowest HD. By combining these two aspects, the AJI shows that our method has the best segmentation performance.

E. OPEN DATASET COMPARISON

Finally, we used MoNuSeg dataset to train the dual U-Net and validated its advanced performance. The data published in [31], [32], [35] were used for benchmarking against their methods. The results are listed in Table 6. DIST method refers to distance map regression, whose basic structure is U-Net. RIC is abbreviation of residual-inception-channel

TABLE 5. Comparative analysis of individual images form the independent dataset.

Method	F1 score (Pixel based)		HD (Object based)		AJI (Object based)	
	M ± SD	Paired t	M ± SD	Paired t	M ± SD	Paired t
Proposed	0.8196±0.0521	-	3.948±1.647	-	0.6591±0.0811	-
BAU-Net	0.8158±0.0501	0.5267	4.281±1.775	0.1054	0.6332±0.8980	0.01099*
DCAU-Net	0.7338±0.0576	4.858e-10*	4.536±1.918	0.006543*	0.5527±0.0833	3.322e-09*
BAFCN	0.7544±0.0891	1.645e-05*	5.293±2.213	4.612e-06*	0.5444±0.1187	2.089e-07*
DCAN	0.7118±0.0703	3.200e-11*	10.171±5.518	4.741e-08*	0.4082±0.1270	3.925e-14*
U-Net	0.6951±0.1158	3.577e-08*	4.096±1.626	0.4444	0.5008±0.1210	1.329e-10*
Level Set	0.5328±0.0584	7.191e-14*	8.6866±1.852	4.695e-7*	0.3153±0.0635	3.174e-11*

The mean values with the standard deviation are given in the left columns, and the p values from the paired t-test for the proposed method and the others are given in right columns. *shows statistical significance

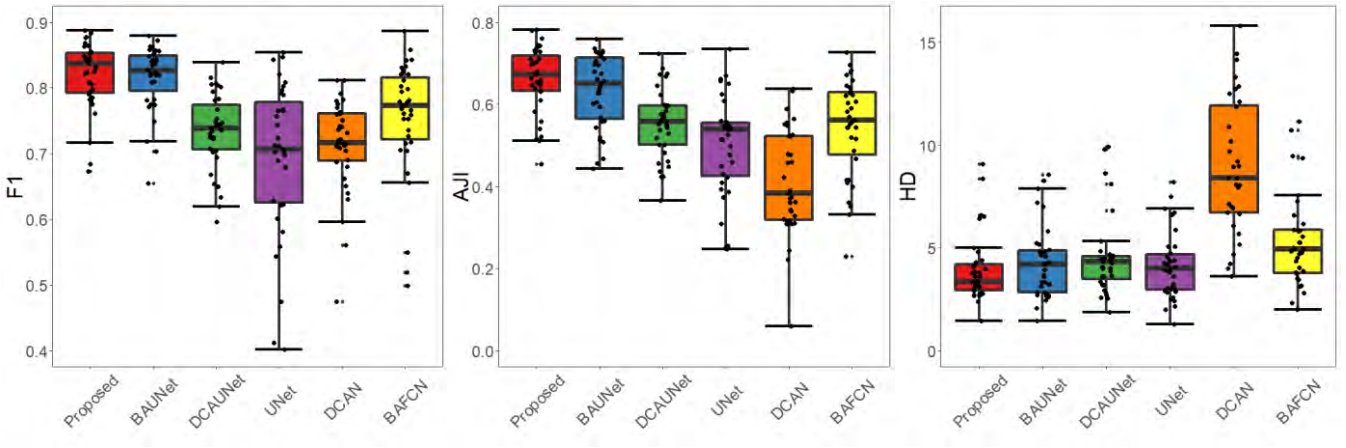

FIGURE 8. Boxplot of the overall comparison results. From left to right are the F1 score, the AJI and the HD. The proposed method is shown in red.

TABLE 6. Comparative analysis for individual images from MoNuSeg dataset.

Organ	Image	AJI(Object based)				F1 score(Pixel based)			
		Proposed	DIST[32]	RIC[35]	CNN[31]	Proposed	DIST[32]	RIC[35]	CNN[31]
Bladder	1	0.6814	0.6475	-	0.5465	0.8625	0.8623	-	0.9312
	2	0.5765	0.5467		0.4968	0.7790	0.7768		0.6304
Colorectal	1	0.5007	0.4240	-	0.4891	0.7176	0.7212	-	0.7679
	2	0.5352	0.4484		0.5692	0.7343	0.7360		0.7118
Stomach	1	0.6107	0.6408	-	0.4538	0.8446	0.8547	-	0.8913
	2	0.6342	0.6550		0.4378	0.8414	0.8520		0.8982
Breast	1	0.6033	0.5334	-	0.4974	0.8044	0.7761	-	0.6885
	2	0.6753	0.5884		0.5796	0.8117	0.8380		0.7476
Kidney	1	0.6101	0.5648	-	0.4792	0.7904	0.7805	-	0.6606
	2	0.5778	0.5420		0.6672	0.7650	0.7606		0.7837
Liver	1	0.5236	0.5466	-	0.5175	0.7516	0.7877	-	0.6726
	2	0.4790	0.4432		0.5148	0.7268	0.6684		0.7036
Prostate	1	0.7140	0.6273	-	0.4914	0.8366	0.8030	-	0.8306
	2	0.5272	0.6294		0.3761	0.8122	0.7903		0.7537
Overall		0.5899	0.5598	0.5635	0.5083	0.7913	0.7809	0.8278	0.7623

In [31], AJI, HD, F1 and Dice of individual images were provided. In [32], AJI and F1 of individual images were provided. In [35], only AJI, F1 and Dice of overall image were provided. We listed these results in the form of the table above.

attention-U-Net specially designed for nuclei segmentation. CNN from [31] is a ternary classification CNN model for nuclei segmentation. The proposed method obtains 0.5899 for AJI and 0.7913 for F1-score. Compared with the latest methods, our method ranks second in F1-score which reflects the pixel level accuracy. The advantage of our method is separating overlapping nuclei, and the AJI which focus on the object level segmentation ranks first. It's worth mention that

the proposed method surpasses DIST because our model utilizes additional boundary information. The boundary information on one hand constrains the contour of the final results and on the other hand reinforces the regression model. The experiment with open dataset indicates that the proposed method is advanced.

V. DISCUSSION

In clinical practice and research, computer-assisted pathology has gradually shown its advantages over the years. The segmentation of nuclei is the first and most important step in automatic histopathological analysis. However, many segmentation methods suffer from challenges such as irregular shapes, inter- or intra-nuclei color variations, and overlapping or touching nuclei.

In this study, we utilized both the region and boundary information and proposed a novel network to overcome these difficulties. From the comparative experiments on glioma nuclei dataset and multi-organ nuclei dataset, we show that the combination of U-Net, the boundary-aware structure and the distance map regression is effective and advanced in overlapping nuclei segmentation problems.

The basic idea of this paper is integrating boundary information and region information to separate overlapping nuclei. Plenty of researches have focused on this thought and our work is based on some existing foundations such as boundary-aware structure, which is not designed for nuclei segmentation, yet suitable for separating touching nuclei. In these models, the final segmentation result is determined by the performance of the region segmentation and boundary segmentation. However, the final performance is restricted by the poor intermediate results. Therefore, it's necessary to modify the internal structure.

One structural modification in this paper is replacing FCN structure with U-Net structure. Thanks to the multiple convolutional and concatenating layers in the upsampling process, the results produced by U-Net structure are more delicate and accurate than FCN. This modification was verified by experiments in Section IV Part D. But the improvement of the model is limited due to the characteristic of classification model. The classification model is insensitive to the pattern of the boundary patches and several wrongly classified boundary pixels contribute little to the global loss. Both the score map of the boundary and the region is rough and vague. So new ideas are needed to optimize the model.

We embedded the distance map regression in the previous model. The most important innovation in this paper is to use the regression model combining with U-Net to enhance the regional detail prediction. Looking at the advantage of the distance map regression alone, the distance map is more in line with the description of the fuzzy boundary, and since the regression model is applied to fit the distance map, the trend of the numerical distribution is retained, avoiding the aliasing of touching regions. The basic principle is that the regression of the distance map implicitly imposes a priori shape on the network and enables the network to learn some object-level knowledge. In addition, the network is forced to generate a spatial smooth output, so the shape of the results will be more regular. Furthermore, the combination of regression model and classification model enhances their respective advantages. As we put the regression model and the boundary classification model into one network and trained them together, two branches interact with each other. The high correlation between distance map and boundary annotation makes the synchronous training and the reinforcing relationship possible. The boundary classification model forces the network to be boundary sensitive, which regularize the contour of the distance map. The regression model forces the network to learn the difference between boundary and core, which enhance the ability of features to express boundary. From the results we can see that the proposed model has the most accurate perception of the boundary. With the increasing of two coarse segmentation accuracy, the fine result is improved.

In spite of these advantages, the proposed method has some shortcomings. For instance, when we attempt to solve the under-segmentation problem, the over-segmentation problem arises. As we can see in the 2nd row of Fig. 6, a nucleus is divided into two parts by our method. although it may be due to the non-uniform color distribution of the nucleus, it does show that the proposed method is not perfect. Besides, there is still a big gap between the proposed method and the latest method in F1-score.

In conclusion, our method takes advantage of several deep-learning strategies such as the U-Net structure, the boundary-aware strategy and deep regression of the distance map. By the acceleration of GPU TITAN Xp, The proposed approach can segment an image with a size of 1000×1000 in 0.5 seconds. Hence, our method is capable of solving overlapping nuclei and outperforms the state-of-art methods.

VI. CONCLUSION

Nuclear segmentation is a challenging and important task in computer-assisted histopathological analysis. In this paper, we presented a novel network called the dual U-Net in which the boundary information and distance map are combined to help nuclear segmentation. The network has a special dual-upsampling structure for multiple task training and a combination layer for refined segmentation. The regression branch for predicting the distance map enables the network to detect the nuclear core region while the classification branch for boundary segmentation enables the network to separate touching objects. By using the convolution operation of the fusion layer, accurate nuclear segmentation results are extracted from the coarse segmentation. The experimental results on the glioma nuclei dataset demonstrate that the combination of these structure is effective and proposed method is competent for glioma cytopathology research. Furthermore, the comparative experiments on open dataset verify the advanced nature of the proposed method. Not limited to gliomas, the proposed network is easy to apply to segmenting various kinds of nuclei of other tumors.

Although the dual U-Net has outstanding performance in glioma cases, the accuracy of the touching nuclei with serious deformations still has space for improvements and the over-segmentation problem remains to be solved.

ACKNOWLEDGMENT

The authors would like to thank for the glioma dataset prepared and provided at Huashan Hospital of Fudan University.

REFERENCES

[1] H. Irshad, A. Veillard, L. Roux, and D. Racoceanu, "Methods for nuclei detection, segmentation, and classification in digital histopathology: A review—Current status and future potential," *IEEE Rev. Biomed. Eng.*, vol. 7, pp. 97–114, 2014.

[2] K. S. Beevi and G. R. Bindu, "Analysis of nuclei detection with stain normalization in histopathology images," *Indian J. Sci. Technol.*, vol. 8, no. 23, pp. 547–551, 2015.

[3] F. Xing and L. Yang, "Robust nucleus/cell detection and segmentation in digital pathology and microscopy images: A comprehensive review," *IEEE Rev. Biomed. Eng.*, vol. 9, pp. 234–263, 2016.

[4] L. Vincent, "Morphological grayscale reconstruction in image analysis: Applications and efficient algorithms," *IEEE Trans. Image Process.*, vol. 2, no. 2, pp. 176–201, Apr. 1993.

[5] L. A. D. Cooper, J. Kong, D. A. Gutman, F. Wang, J. Gao, C. Appin, S. Cholleti, T. Pan, A. Sharma, L. Scarpace, T. Mikkelsen, T. Kurc, C. S. Moreno, D. J. Brat, and J. H. Saltz, "Integrated morphologic analysis for the identification and characterization of disease subtypes," *J. Amer. Med. Inform. Assoc.*, vol. 19, no. 2, pp. 317–323, 2012.

[6] C. Wählby, I.-M. Sintorn, F. Erlandsson, G. Borgefors, and E. Bengtsson, "Combining intensity, edge and shape information for 2D and 3D segmentation of cell nuclei in tissue sections," *J. Microsc.*, vol. 215, pp. 67–76, Jul. 2004.

[7] S. K. Nath, K. Palaniappan, and F. Bunyak, "Cell segmentation using coupled level sets and graph-vertex coloring," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI)*, vol. 9, 2016, pp. 101–108.

[8] H. Chang, J. Han, A. Borowsky, L. Loss, J. W. Gray, P. T. Spellman, and B. Parvin, "Invariant delineation of nuclear architecture in glioblastoma multiforme for clinical and molecular association," *IEEE Trans. Med. Imag.*, vol. 32, no. 4, pp. 670–682, Apr. 2013.

[9] Y. Al-Kofahi, W. Lassoued, W. Lee, and B. Roysam, "Improved automatic detection and segmentation of cell nuclei in histopathology images," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 4, pp. 841–852, Apr. 2010.

[10] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[11] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: https://arxiv.org/abs/1409.1556

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

[13] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 640–651, Apr. 2017.

[14] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI)*, 2015, pp. 234–241.

[15] D. Ciresan, A. Giusti, L. M. Gambardella, and J. Schmidhuber, "Deep neural networks segment neuronal membranes in electron microscopy images," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 2843–2851.

[16] F. Xing, Y. Xie, and L. Yang, "An automatic learning-based framework for robust nucleus segmentation," *IEEE Trans. Med. Imag.*, vol. 35, no. 2, pp. 550–566, Feb. 2016.

[17] Y. Xie, X. Kong, F. Xing, F. Liu, H. Su, and L. Yang, "Deep voting: A robust approach toward nucleus localization in microscopy images," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI)*, 2015, pp. 374–382.

[18] A. Giusti, D. C. Cireşan, J. Masci, L. M. Gambardella, and J. Schmidhuber, "Fast image scanning with deep max-pooling convolutional neural networks," in *Proc. IEEE 20th Int. Conf. Image Process.*, Sep. 2013, pp. 4034–4038.

[19] Y. Xie, F. Xing, X. Kong, H. Su, and L. Yang, "Beyond classification: Structured regression for robust cell detection using convolutional neural network," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI)*, 2015, pp. 358–365.

[20] A. Janowczyk, S. Doyle, H. Gilmore, and A. Madabhushi, "A resolution adaptive deep hierarchical (RADHicaL) learning scheme applied to nuclear segmentation of digital pathology images," *Comput. Methods Biomech. Biomed. Eng., Imag. Vis.*, vol. 6, no. 3, pp. 270–276, 2018.

[21] X. Zhou, F. Li, J. Yan, and S. T. Wong, "A novel cell segmentation method and cell phase identification using Markov model," *IEEE Trans. Inf. Technol. Biomed.*, vol. 13, no. 2, pp. 152–157, Mar. 2009.

[22] M. Maška, O. Daněk, S. Garasa, A. Rouzaut, A. Muñoz-Barrutia, and C. Ortiz-de-Solorzano, "Segmentation and shape tracking of whole fluorescent cells based on the Chan–Vese Model," *IEEE Trans. Med. Imag.*, vol. 32, no. 6, pp. 995–1006, Jun. 2013.

[23] F. Liu and L. Yang, "A novel cell detection method using deep convolutional neural network and maximum-weight independent set," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI)*, 2017, pp. 63–72.

[24] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2274–2282, Nov. 2012.

[25] H. Yang and N. Ahuja, "Automatic segmentation of granular objects in images: Combining local density clustering and gradient-barrier watershed," *Pattern Recognit.*, vol. 47, pp. 2266–2279, Jun. 2014.

[26] Y. Song, L. Zhang, S. Chen, D. Ni, B. Lei, and T. Wang, "Accurate segmentation of cervical cytoplasm and nuclei based on multiscale convolutional network and graph partitioning," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 10, pp. 2421–2433, Oct. 2015.

[27] L. Zhang, M. Sonka, L. Lu, R. M. Summers, and J. Yao, "Combining fully convolutional networks and graph-based approach for automated segmentation of cervical cell nuclei," in *Proc. IEEE 14th Int. Symp. Biomed. Imag.*, Apr. 2017, pp. 406–409.

[28] H. Chen, X. Qi, L. Yu, Q. Dou, J. Qin, and P.-A. Heng, "DCAN: Deep contour-aware networks for object instance segmentation from histology images," *Med. Image Anal.*, vol. 36, pp. 135–146, Feb. 2017.

[29] H. Shen, R. Wang, J. Zhang, and S. J. McKenna, "Boundary-aware fully convolutional network for brain tumor segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI)*, 2017, pp. 433–441.

[30] M. Khoshdeli and B. Parvin, "Deep learning models delineates multiple nuclear phenotypes in H&E stained histology sections," 2018, *arXiv:1802.04427*. [Online]. Available: https://arxiv.org/abs/1802.04427

[31] N. Kumar, R. Verma, S. Sharma, S. Bhargava, A. Vahadane, and A. Sethi, "A dataset and a technique for generalized nuclear segmentation for computational pathology," *IEEE Trans. Med. Imag.*, vol. 36, no. 7, pp. 1550–1560, Jul. 2017.

[32] P. Naylor, M. Laé, F. Reyal, and T. Walter, "Segmentation of nuclei in histopathology images by deep regression of the distance map," *IEEE Trans. Med. Imag.*, vol. 38, no. 2, pp. 448–459, Feb. 2019.

[33] S. Loffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," 2015, *arXiv:1502.03167*. [Online]. Available: https://arxiv.org/abs/1502.03167

[34] B. Xu, N. Wang, T. Chen, and M. Li, "Empirical evaluation of rectified activations in convolutional network," 2015, *arXiv:1505.00853*. [Online]. Available: https://arxiv.org/abs/1505.00853

[35] Z. Zeng, W. Xie, Y. Zhang, and Y. Lu, "RIC-Unet: An improved neural network based on unet for nuclei segmentation in histology images," *IEEE Access*, vol. 7, pp. 21420–21428, 2019.

[36] D. P. Huttenlocher, G. A. Klanderman, and W. J. Rucklidge, "Comparing images using the Hausdorff distance," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 15, no. 9, pp. 850–863, Sep. 1993.

[37] A. Vedaldi and K. Lenc, "MatConvNet: Convolutional neural networks for MATLAB," in *Proc. 23rd ACM Int. Conf. Multimedia*, 2015, pp. 689–692.

[38] X. Qi, F. Xing, D. J. Foran, and L. Yang, "Robust segmentation of overlapping cells in histopathology specimens using parallel seed detection and repulsive level set," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 3, pp. 754–765, Mar. 2012.

[39] Y. Liu, P. Zhang, Q. Song, A. Li, P. Zhang, and Z. Gui, "Automatic segmentation of cervical nuclei based on deep learning and a conditional random field," *IEEE Access*, vol. 6, pp. 53709–53721, 2018.

[40] T. Totu, R. M. Buga, A. Dumitru, M. Costache, N. Sladoje, and S. G. Stanciu, "An objective scoring framework for histology slide image mosaics applicable for the reliable benchmarking of image quality assessment algorithms," *IEEE Access*, vol. 6, pp. 53080–53091, 2018.
[41] Y. Li and L. Shen, "cC-GAN: A robust transfer-learning framework for HEp-2 specimen image segmentation," *IEEE Access*, vol. 6, pp. 14048–14058, 2018.
[42] X. Du, W. Zhang, H. Zhang, J. Chen, Y. Zhang, J. C. Warrington, G. Brahm, and S. Li, "Deep regression segmentation for cardiac bi-ventricle MR images," *IEEE Access*, vol. 6, pp. 3828–3838, 2018.
[43] J. Ker, L. Wang, J. Rao, and T. Lim, "Deep learning applications in medical image analysis," *IEEE Access*, vol. 6, pp. 9375–9389, 2018.

XIELI LI was born in Zhejiang, China, in 1994. He received the B.Sc. degree in electronic engineering from Fudan University, Shanghai, China, in 2016, where he is currently pursuing the master's degree with the Department of Electronic Engineering. His current research interests include medical image processing, digital pathology, and deep-learning application.

YUANYUAN WANG received the B.Sc., M.Sc., and Ph.D. degrees in electronic engineering from Fudan University, Shanghai, China, in 1990, 1992, and 1994, respectively.

From 1994 to 1996, he was a Postdoctoral Research Fellow with the School of Electronic Engineering and Computer Science, University of Wales, Bangor, U.K. In 1996, he went back to the Department of Electronic Engineering, Fudan University, as an Associate Professor, and then promoted to Full Professor, in 1998. He is currently the Director of the Biomedical Engineering Center, Fudan University. He has authored or coauthored six books and 500 research papers. His research interests include medical ultrasound techniques and medical image processing.

QISHENG TANG received the B.M. degree in general practice from the Shanghai Medical College, Fudan University, China, in 2011, where he is currently pursuing the master's degree with the Department of Neurosurgery, Huashan Hospital. His current research interests include neuropathology, neuroscience, and deep learning for neuro-oncology.

ZHEN FAN received the master's degree in surgery from Anhui Medical University, Hefei, China, in 2010. He is currently pursuing the Ph.D. degree in neurosurgery with the Huashan Hospital, Fudan University, Shanghai, China, in 2016. From 2010 to 2015, he completed the resident and chief resident training of neurosurgery at the Anhui Provincial Hospital and became an attending Neurosurgeon, in 2016. He focuses on mapping eloquent area and epileptic foci for the neurosurgical plan using multimodal images. His research interest includes deep-learning application in brain function and brain diseases.

JINHUA YU received the Ph.D. degree in electronic engineering from Fudan University, Shanghai, China, in 2008.

From 2008 to 2010, she was a Postdoctoral Fellow with the Department of Bioengineering, University of Missouri, Columbia, MO, USA. She is currently a Full Professor with the Electronic Engineering Department, Fudan University. Her current research interests include medical signal analysis, radiomics, and ultrasound imaging.