IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

SPECIAL SECTION ON DEEP LEARNING FOR COMPUTER-AIDED MEDICAL DIAGNOSIS

Received November 21, 2018, accepted January 22, 2019, date of publication February 1, 2019, date of current version February 27, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2896920

RIC-Unet: An Improved Neural Network Based on Unet for Nuclei Segmentation in Histology Images

ZITAO ZENG[1,2], WEIHAO XIE[3], YUNZHE ZHANG[3], AND YAO LU[1,2,4]

[1]School of Data and Computer Science, Sun Yat-sen University, Guangzhou 510006, China
[2]Guangdong Province Key Laboratory of Computational Science, Guangzhou 510275, China
[3]PVmed Inc., Guangzhou 510275, China
[4]Department of Ultrasound, The fifth Affiliated Hospital, Sun Yat-sen University, Guangzhou 510275, China

Corresponding author: Yao Lu (luyao23@mail.sysu.edu.cn)

This work was supported in part by the Ministry of Science and Technology of China under Grant 2018YFC1704206 and Grant 2016YFB0200602, in part by National Science Foundation of China under Grant 81801809, Grant 81830052, and Grant 11401601, in part by the Guangdong Province Frontier and Key Technology Innovative Grant 2015B010110003 and 2016B030307003, in part by the Guangdong Province Applied Science and Technology Research Grant 2015B020233008, in part by the Guangzhou Cooperative and Creative Key Grant 201604020003, and in part by the Guangzhou Science and Technology Creative Key Grant 2017B020210001.

ABSTRACT As a prerequisite for cell detection, cell classification, and cancer grading, nuclei segmentation in histology images has attracted wide attention in recent years. It is quite a challenging task due to the diversity in staining procedure, cell morphology, and cell arrangement between different histopathology images, especially with different color contrasts. In this paper, an Unet-based neural network, RIC-Unet (residual-inception-channel attention-Unet), for nuclei segmentation is proposed. The techniques of residual blocks, multi-scale and channel attention mechanism are applied on RIC-Unet to segment nuclei more accurately. RIC-Unet is compared with two traditional segmentation methods: CP and Fiji, two original CNN methods: CNN2, CNN3, and original U-net on The Cancer Genomic Atlas (TCGA) dataset. Besides, in this paper, we use Dice, F1-score, and aggregated Jaccard index to evaluate these methods. The average of RIC-Unet and U-net on these three indicators are 0.8008 versus 0.7844, 0.8278 versus 0.8155, and 0.5635 versus 0.5462. In addition, our method won the third place in the computational precision medicine nuclei segmentation challenge together with MICCAI 2018.

INDEX TERMS Computational pathology, nuclei segmentation, residual block, deep learning.

I. INTRODUCTION

Pathology images serve as a gold standard for doctors in disease diagnosis and are widely recognized by authorities. For pathological images, the difference in the morphology of the nuclei is the main basis for the current tumor diagnosis. In-depth understanding of the morphological changes of the nucleus is very meaningful for the diagnosis and identification of tumors. Therefore, accurate segmentation of the nuclei in pathological images is an emergent and fundamental part for further analysis. When accurately segment the outline of the nuclei, we will then get some basic morphological information of the nuclei, such as the size and the shape of the nucleus, etc. In addition, we can also count the number of nuclei and record the arrangement of nuclei in the whole image. The ratio of nucleus and cytoplasm is also an important indicator which can help pathologists make more accurate diagnosis.

Semantic segmentation is a classic problem in the field of computer vision, and is also very essential in medical imaging research. In the past few decades, many classic segmentation algorithms have been applied in the area of medical imaging. Otsu [1] is a classic method based on dynamic thresholding. However, the selection of the threshold depends on the histogram distribution. In digital pathology images, histogram distribution is complicated, the difference in foreground and background is not obvious, and the performance of Otsu is therefore reduced. Watershed algorithm [2], which performs region base growing is based on the initial seed point and often encounters problems of over- segmentation. Therefore, to obtain ideal results, watershed algorithm usually needs some complex post-processing methods. Besides, active contour method such as gradient vector flow(GVF) [3] has also been widely used on CT and MR images for organ

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

segmentation, but this method needs a priori contour to iteratively fit ground truth. As the number of nuclei in pathological images is too large, it is difficult for GVF to realize automatic detection. In recent years, the applications of Deep Convolutional Neural Networks(DCNNs) have proved the powerful performance on image classification, object detection and semantic segmentation [4]–[6]. Deep convolution network has the advantage of automatic feature extraction and can be trained end-to-end compared with traditional image processing methods which use hand-crafted features. Now it can be recognized that deep learning method has become an emerging need for well-annotated labels, especially for semantic segmentation tasks.

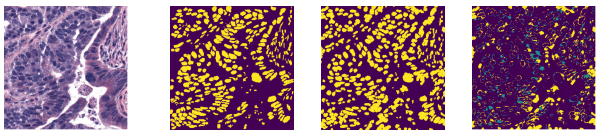

FIGURE 1. (a): original image. (b): ground truth label. (c): predicted result by U-net. (d): the difference between ground truth and predicted result. The yellow pixels represent the result of the miss-segmentation, and the green pixels represent the result of the missed segmentation. As can be seen, U-net fails on some border areas between closing and touching cells, result in some inaccuracy for instance segmentation. Best viewed on screen with zoom in.

Fully convolution Network (FCN) [5] achieved promising results on several benchmarks with relatively simple architecture. Following this common architecture, several improved networks further improved the segmentation result, such as U-net [7] which utilized features from different scales and had a better performance on biomedical images. Apart from the inspiring accuracy achieved by U-net, for some challenging histology images, U-net could hardly distinguish touching or overlapping cells, as well as some confusing background areas, which could be seen in Fig 1. Besides, the original U-net architecture has strong representation ability, it also requires a large number of parameters. When the dataset size is relatively small, it is prone to a certain degree of over-fitting to influence the generalization ability of the model. The popularity of FCN-based model arises from its fully convolution architect which is quite computation efficiency. It could be widely applied regardless of input image size, an annoying problem in some networks containing fully connected layers. FCN has already proved the effectiveness of utilizing features from different scales, which could produce more accurate segmentation results. Based on the development of FCN, several other popular architectures achieved state-of-art results by utilizing multi-scale features in different ways, such as Seg-Net [8], Deep Lab [9], Refine-Net [10]. However, the difficulties of nuclei segmentation task for histology images arises from two aspects, the heterogeneity in nuclear appearance and the condition of touching nucleus in overlapping cells. Although these methods have achieved great success in natural image scenes, but for multi-organ pathological image nuclei segmentation, these methods are rarely used. To tackle the touching objects in biomedical images, topological information is important, and contour prediction is one of the common method [11]. Chen *et al*. [12], developed deep contour network based on FCN to predict both gland and contour combining multi-level features. Kumar *et al*. [13] modified the prediction output into three class classifications including the contour, which is proved to have better accuracy. Although these methods introduce contour information to accurately segmentation result, but in the main part of the segmentation network, these methods make insufficient use of the semantic information of the image, because most of these segmentation networks only extract the semantic information of the former layer, they do not pay much attention to the shallower layers' semantic information. For the reason that this paper proposed a revised network architecture to incorporate more discriminative features as well as multi-scale features. Following the basic network architecture of U-net, we revised the down-sampling module to contain both residual module [14] as well as inception module [15] to extract more powerful features. Residual connection has been proved to better represent image features, which has been first experimented on image classification task. Here we incorporated residual block in down sampling part to extract more representative features for segmentation. Inception module is well known for its computational efficiency while incorporating multi-scale features with different kernel sizes, which has also been proved in image classification. So we included inception module together with residual block in down-sample part. We named the new revised block as RI-block(Residual-Inception-block), which contained both larger reception field and better feature representations. Recently, channel attention mechanism has been proved to be quite effective in extracting more representative features [16] for image classification, as well as semantic segmentation [17]. Channel attention mechanism can focus the parameter training on the region of interest, taking into account the correlation between the channels, and alleviate the overfitting problem. Therefore, we incorporated this module [17] in our up-sampling part to handle the heterogeneity of nuclei appearances called DC-block(Deconvolution-Channel-block). To better segment touching nuclei in instance level, we used a separate up-sampling module to predict contours, which was quite effective for improvement in instance level evaluation metrics. Experiments on the public TCGA dataset [18] and computational precision medicine (CPM) nuclei segmentation challenge [19] validate the effectiveness of our proposed algorithm.

II. MATERIALS AND METHOD

A. THE DATASET

In this paper, we use TCGA(The Cancer Genomic Atlas) dataset which contains 30 whole slide images(WSIs) [18], and use only one WSI per patient to maximize nuclear appearance variation. Since the computational requirements for processing WSIs are high, the WSI images are cropped into sub-images of size 1000 * 1000 from regions dense in nuclei, keeping only one cropped image per WSI and

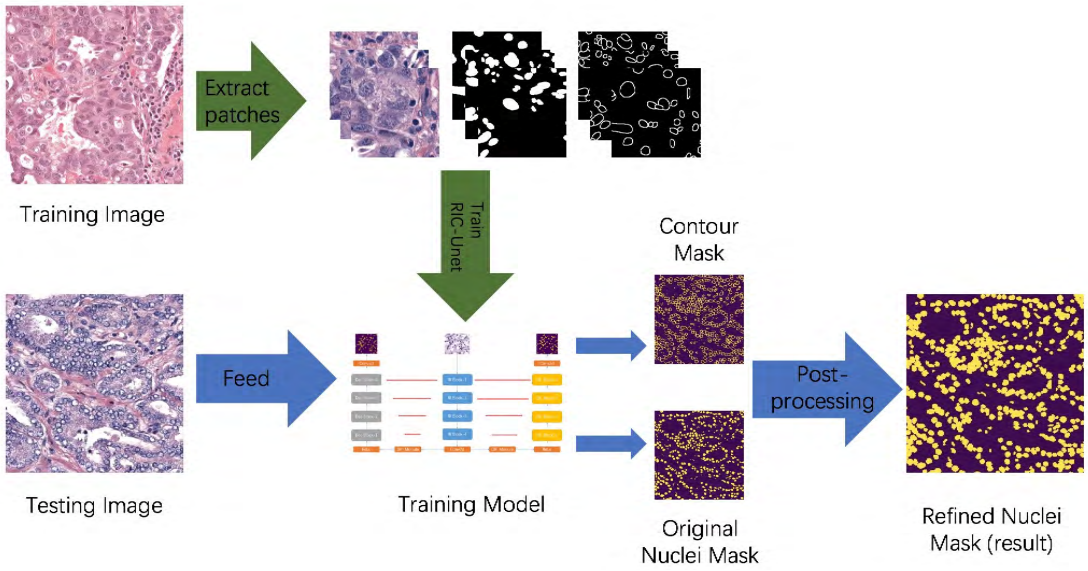

FIGURE 2. Algorithm flow chart.

per patient. Each sub-images are annotated the boundary, nuclei and background separately. To increase the diversity of nuclear appearances, these images have covered seven different organs viz. breast, liver, kidney, prostate, bladder, colon and stomach, including both benign and diseased tissue samples. In addition, the dataset split 16 sub-images as training and validation set which contains 13372 nuclei, and 14 sub-images as testing set which contains 8251 nuclei, the number of nuclei boundaries that are well-annotated is going to be totally 21623.

B. ALGORITHM FRAMEWORK

Our algorithm contains three parts: image pre-processing, model architecture and image post-processing. The easy flow chart can be seen in Fig 2. In section C, the network architecture is showed, and in section D, the methods of pre-processing and post-processing are briefly introduced. More specifically, because the training set of pathological images we acquired is limited, in order to have sufficient data to train the segmentation network, we extracted each original training set of 1000 * 1000 to sub-patches (which size is 224 * 224), each original image is cut into 100 smaller ones with overlap, and then feed them into the network for training.

C. MODEL ARCHITECTURE

Differ from classification task, segmentation needs to incorporate more local information based on high resolution image as well as global features in low resolution to differentiate foreground and background. Common segmentation models utilize features of different levels with different resolutions, which can be seen in encoder-decoder framework. As mentioned above, for pathological cells, the diversity of cell shapes and dense cell overlap have posed new challenges for the segmentation of pathological cells. Inspired by inception-net, U-net and some pioneered research works in nature image semantic segmentation and pathological image segmentation tasks, we proposed our network to tackle the problems mentioned above. In order to enable our model to identify different types of cell shapes and cells of different scales, it was subjected to better extraction of global context and local context to solve the problem of diverse cell shapes. Meanwhile, in order to deal with the problem of dense cell overlap, we followed the idea proposed by Chen *et al.* [12] in the segmentation of gland instances, adopting a multi-task learning framework, allowing the network to learn to segment the nucleus and cell contour at the same time. This network will use the cell contour as auxiliary information assist in the differentiation of dense cells and reduce errors at the object level. In particular, our model contains two stages. In encoding phase, we used a series of RI blocks to extract the multi-scale local details. Besides, we used convolution operators with stride 2, kernel size 3 * 3 to instead of max-pooling operators to retain more details. We thought that this combined down-sample network architecture could better extract features of different resolutions. In decoding phase, we use a structure similar to U-net to concatenate high level features with low level features in up-sampling procedure. In order to better select the high resolution features, inspired by Yu *et al*. [17], DC blocks which contained the CAB(Channel-Attention-Block) module to give attention weight to high resolution feature channels based on low resolution features are used in up-sampling part. In detail, the high level features is transformed to attention vector through average pooling, then weights contained in this vector is multiplied by low level features to give different attention weights to channels. RIC-Unet(Residual-Inception-Channel attention-Unet) architecture is shown in Fig 3, where it takes an input patch as well as outputs of the contour segmentation result, and an input patch as well as outputs of the nuclei segmentation result.

1) RI BLOCK

Traditional U-net only contains basic convolution layers. Our method replaces convolution layers with RI-Blocks, which includes residual module and inception module.

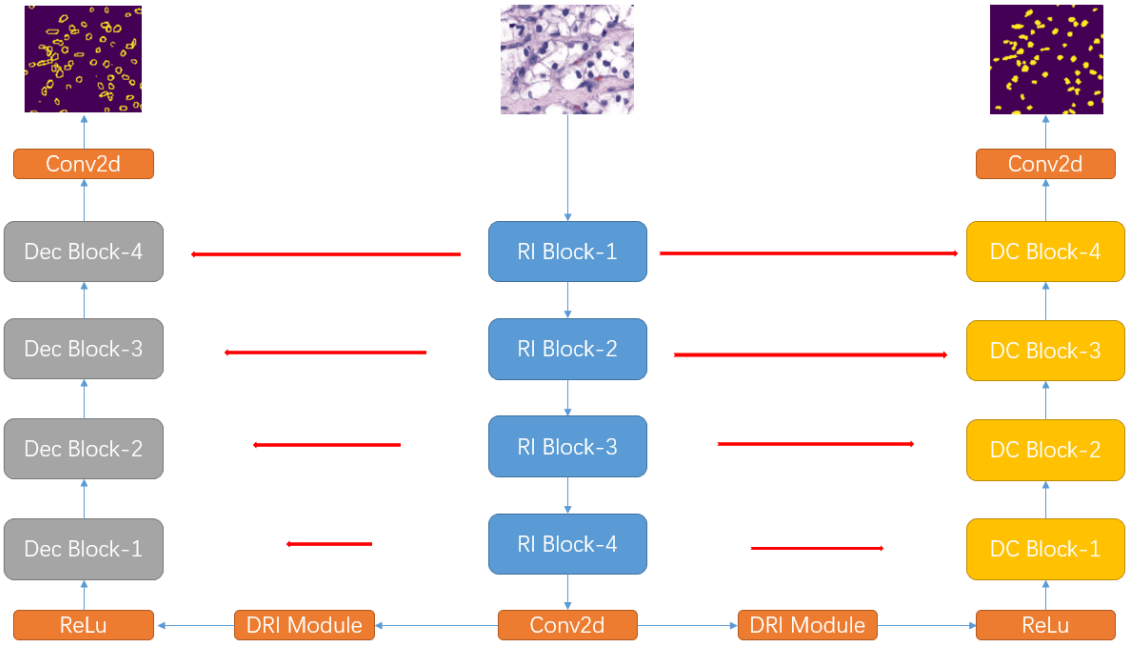

FIGURE 3. Model Architecture: this model contains three parts and two outputs, the left part is the same architecture as U-net up-sampling part, and the output is nuclei's contour mask, the middle part contains RI block, which details can see Fig. 4(a), the right part contains DC block, which detail can see Fig. 4(b), the output is original nuclei mask.

The block architecture is illustrated in Fig 4(a). Input feature is processed by a convolution layer, followed by DRI module. Then another convolution layer with stride of 2 to squeeze the feature map size is used, following the paradigm of encoder framework. At last, we use the residual block to extract more representative features. DRI module is a more complicated version of inception module which can extract features within different receptive fields with various kernel size and this is proposed in [20]. By building this block, our intuition is to incorporate more representative features within different reception fields, which could improve the overall segmentation accuracy. Relu layers are used after each feature extraction module as a common paradigm.

2) DC BLOCK

As for the up sampling part, DC block is used to be built to better utilize features from different levels to get the final prediction result. By selecting more powerful features with attention weights, the attention mechanism is used in several deep learning tasks [21]–[23]. Inspired by the attention mechanism, CAB module presented in [16], is first designed to be used in up-sample part in semantic segmentation task. This module is included in our up-sample block to better utilize features extracted from down-sample part. The basic idea of CAB module is to use high level features produce attention vector by average pooling, then this vector serves as guidance to let model pay attention to different channels of low level features, which could also be viewed as a kind of feature refinement by assigning weights to different channels. The architecture of DC block is shown in Fig 4(b). The CAB module are inserted before the deconvolution layer to combine features from different resolutions by attention mechanism.

3) LOSS FUNCTION

Focal loss is proposed by Lin *et al.* [24], to complete better mining dense object detection task in difficult samples and prevent the vast number of easy negatives from overwhelming the detector during training. This loss plays an important role in our model training. During the training process, focal loss can adjust the sample weight according to the training loss, so that the difficult samples can be assigned higher weights during the model training. This loss can be defined as:

$$FL(p_t) = -\alpha_t(1-p_t)^{\gamma}\log(p_t) \tag{1}$$

For notational convenience, we define p_t:

$$p_t = \begin{cases} p & if\ y=1 \\ 1-p & otherwise, \end{cases} \tag{2}$$

where p is the probability that the classification is positive, and α_t, γ are both hyper-parameter. We followed the focal loss and modified the weights according to the image information to better learn the contour and the background information.

D. IMPLEMENTATION

1) PRE-PROCESSING METHOD

Since the staining procedure varies significantly among the training and testing images, color normalization has become an essential step in pre-processing. We took the method proposed by Macenko *et al.* [25], to normalize the H and E channel by extracting feature vectors. The result of color normalization is demonstrated in Fig 5. As for the source image in color normalization, we chose it based on simple observation that nuclei and contour were more clear in the

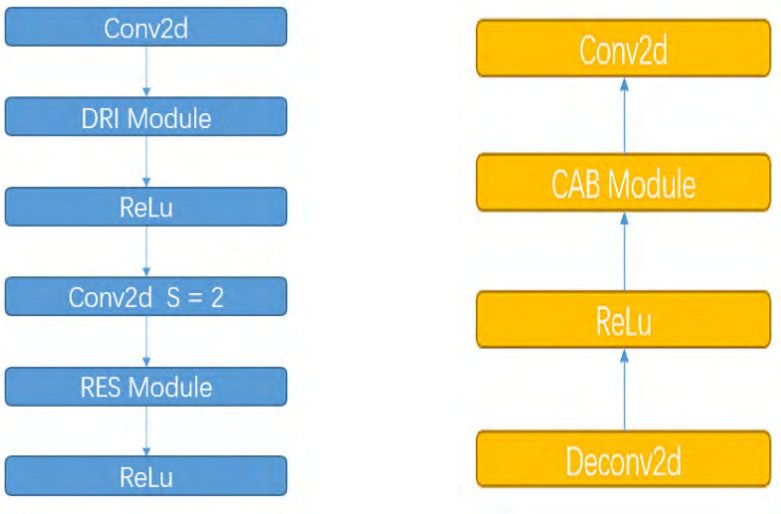

FIGURE 4. RI block and DC block, (a): RI Block, (b): DC block.

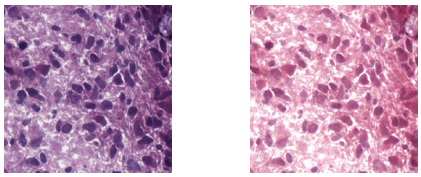

FIGURE 5. Color normalization visualization, (a): original image, (b): normalized image.

source image. Then we applied data normalization by the mean and standard deviation of the input image. Besides, in order to avoid over-fitting as much as possible, we conducted some data augmentation operations for training set, such as image flip, image crop, image translation, etc.

2) POST-PROCESSING METHOD

The predicted original nuclei results will have many overlapping cells, so we need to post- process the predicted results and refine the segmentation results. Firstly, we obtained the contour predict map and nuclei predict map, and for each pixel, we took the pixel as a nuclear pixel when the probability of the foreground exceeded 0.5 and the probability of the contour was below 0.5. After that, the overlapping of nuclei was alleviated, but the nuclei mask was smaller than ground truth at the same time. Therefore, we then dilated each cell by using the disk template whose size was 3. This dilation would stop until the nuclei pixels became closed enough to the contour to get the final segmentation results.

3) PARAMETER SETTINGS

The initial learning rate is 0.0001, and then the learning rate is reduced by ten percent per 1000 iterations. The batch size is 2. The optimization method uses Adam, and the weight attenuation coefficient is set as 0.0005. We set the weight of the front and background as 1 and 2. The weights of the outline and the background are set as 5.0 and 1.0 respectively. Epoch is performed 100 times, and the model is selected through data verification which is evaluated for each epoch. Since our training loss is slightly different from our evaluation index, we also evaluate Dice [27], AJI [13], and F1-score [26]. Finally, we select the optimal model based on the scores of the three indicators, and then use the model to evaluate the test data. A simple weighting method was applied in this process with the weight of Dice, AJI and F1-score being set as 0.25, 0.5 and 0.25 respectively.

E. EVALUATION CRITERION

The evaluating nuclei segmentation methods should penalize both object-level (nucleus detection) and pixel-level (nucleus shape and size) errors. A commonly used object detection metric is the F1-score [26]. For ground truth objects G_i indexed by i and segmented objects S_j indexed by j, the F1-score is based on true positives TP (count of all ground truth objects G_i with an associated segmented (detected) object S_j), false positives FP (count of all segmented objects S_j without a corresponding ground truth object G_i), and false negatives FN (count of all ground truth objects G_i without a corresponding detected object S_j) [13]. F1-score is defined as follows:

$$F1 = \frac{2TP}{2TP + FP + FN} \tag{3}$$

To make the results more convincing, Dice's coefficient [27] and aggregated Jaccard index(AJI) [13] are also used to evaluate our results. Dice coefficient is a statistic method

used for comparing the similarity of two samples. In image segmentation, this formula can be refined as:

$$DSC = \frac{2|G \cap S|}{|G| + |S|} \tag{4}$$

where $|G|$ and $|S|$ represent the ground truth image's pixels and associated segmented image's pixels. $|G \cap S|$ represents the intersect pixels of two images. Aggregated Jaccard index is also a vital evaluation indicator which was first proposed by Kumar *et al*. [13]. Original Jaccard index can be defined as:

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} \tag{5}$$

Different from original Jaccard index, for each ground truth nucleus G_i in an image, when a segmented nucleus S_j, is associated to this image, Kumar *et al*. [13] add the contributions to the aggregated Jaccard index by adding the pixel of $G_i \cap S_j$ to AJIs numerator, and that of $G_i \cup S_j$ to the denominator. Besides, pixels of false negatives and false positives are also added to the denominator, which is a necessary performance, as otherwise, it is likely that the mean Jaccard index will be relatively low due to the bias of the segmentation system towards slight under-segmentation.

TABLE 1. Evaluation result by different method.

Methods	Average AJI	Average Dice	Average F1-score
CP	0.1232	0.5974	0.4046
Fiji	0.2733	0.6493	0.6649
CNN2	0.3842	0.6928	0.7358
CNN3	0.5083	0.7623	0.8267
U-net	0.5462	0.7807	0.8155
U-net + DRI	0.5629	0.7844	0.8023
U-net + DRI + CAB	0.5604	0.7915	0.8013
Proposed Method	**0.5635**	**0.8008**	**0.8278**

III. RESULTS

A. EXPERIMENTAL RESULTS

Based on evaluation indicators, we set the previous methods (CP [28], Fiji [29], CNN2 [30], CNN3 [13]) as baseline. In addition, we use all test images for original U-net, U-net with DRI module, U-net with DRI and CAB module and ours RIC-Unet. The evaluation results are shown below. (See Table 1). From Table 1, it is obviously that original U-net obtained higher average AJI and average Dice than CNN3, but average F1-score is lower than CNN3. After adding modules based on U-net, we found whether introducing DRI module or introducing DRI and CAB module, the average AJI and average Dice both higher than original U-net, but average F1-score both lower than U-net. Besides, DRI module introduced separately on average AJI and average F1-score performs better than simultaneous introduction of DRI and CAB module, but by adding DRI and CAB module can get higher average Dice. However, among all of these methods, RIC-Unet can get the highest score for all indicators.

Besides, we focus on comparing the performance of original U-net and RIC-Unet on test images. In order to better show the results of these two methods, a three images which nuclei are sparse, overlapping are not serious to be selected, another case also contain three images which nuclei are dense, overlapping are serious. These six images are named *simple_image* 1,2,3 and *complex_image* 1,2,3 respectively. The sub-image segmentation results of these six large images are shown in Fig 6, Fig 7, the simple case's and complex case's evaluation result are shown in Table 2, Table 3. From Fig 6, It can be observed that our method and U-net both can have better segmentation results for pathological images with sparse nuclei. Even on some of the evaluation indicators of the whole image, U-net's results are slightly higher than our method. In spite of our method can mitigate miss-segmentation, but the situation of missed segmentation is serious than U-net. As for complex case, it's clear that original U-net cause a serious over-segmentation, but RIC-Unet can properly suppress the occurrence of this situation. And original results of U-net have shown a lot of overlap. The results obtained through our method also have some overlaps, but they are more similar to the ground truth.

Apart from TGCA, we use our method to participate in the computational precision medicine (CPM) nuclei segmentation challenge. By using image tiles from whole slide tissue images, we aim to reduce the requirements of computation and memory. The image tiles are rectangular regions extracted from a set of Glioblastoma and Lower Grade Glioma whole slide tissue images. Nuclei in each image tile in the training set has been subjected to manual segmentation.

The scoring for this sub-challenge is completed by using two variants of the DICE coefficient: the first is the traditional Dice coefficient ($DICE_1$) to measure the overall overlapping between the reference/human segmentation and the participant segmentation, and the second is the Ensemble Dice ($DICE_2$) to capture mismatch in the way the segmentation regions are split, while similarities exist in a large degree in the overall region. The two DICE coefficients of each image tile will be calculated in the test dataset. The score for the image tile will be the mean value of the two dice coefficients and the score for the entire test dataset will be the mean value of the scores for the image tiles.

Although it is to be noticed that there is no public test set in this challenge, the results we submitted have shown that we can obtain 0.8968 in $DICE_1$ and 0.8280 in $DICE_2$, the average DICE is 0.8624, which win the third place in this challenge. The challenge's rank can be found in *"http://miccai.cloudapp.net/competitions/83#results"*.

B. DISCUSSION

From Table 1, We found that adding DRI module to the original U-net increases the receptive field, which leads to a certain improvement in the segmentation results. Especially for average AJI, there was a significant improvement. With the introduction of the CAB module, we hope to focus more on the effective features to improve the segmentation results, however it is not difficult to find a slight decrease in average AJI and average F1-score, but Dice has been further improved. However, compared with the original U-net, adding DRI module or adding DRI and CAB module, average AJI and average Dice can both improved, while average

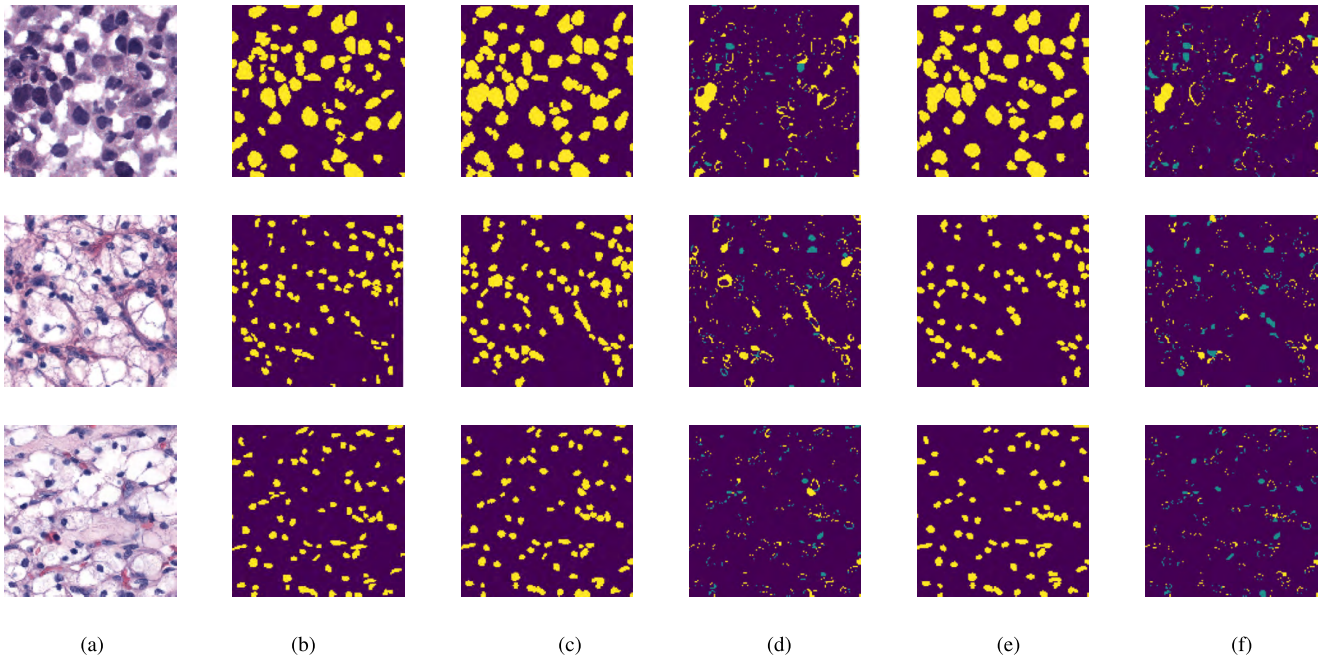

FIGURE 6. Three sub-images result on simple case, the original image named *simple_image* 1, *simple_image* 2 and *simple_image* 3. In (d) and (f), the yellow pixels represent the result of the miss-segmentation, and the green pixels represent the result of the missed segmentation. (a) Sub Image. (b) Sub Image GT. (c) U-net. (d) Difference between GT and U-net. (e) RIC-Unet. (f) Difference between GT and RIC-Unet.

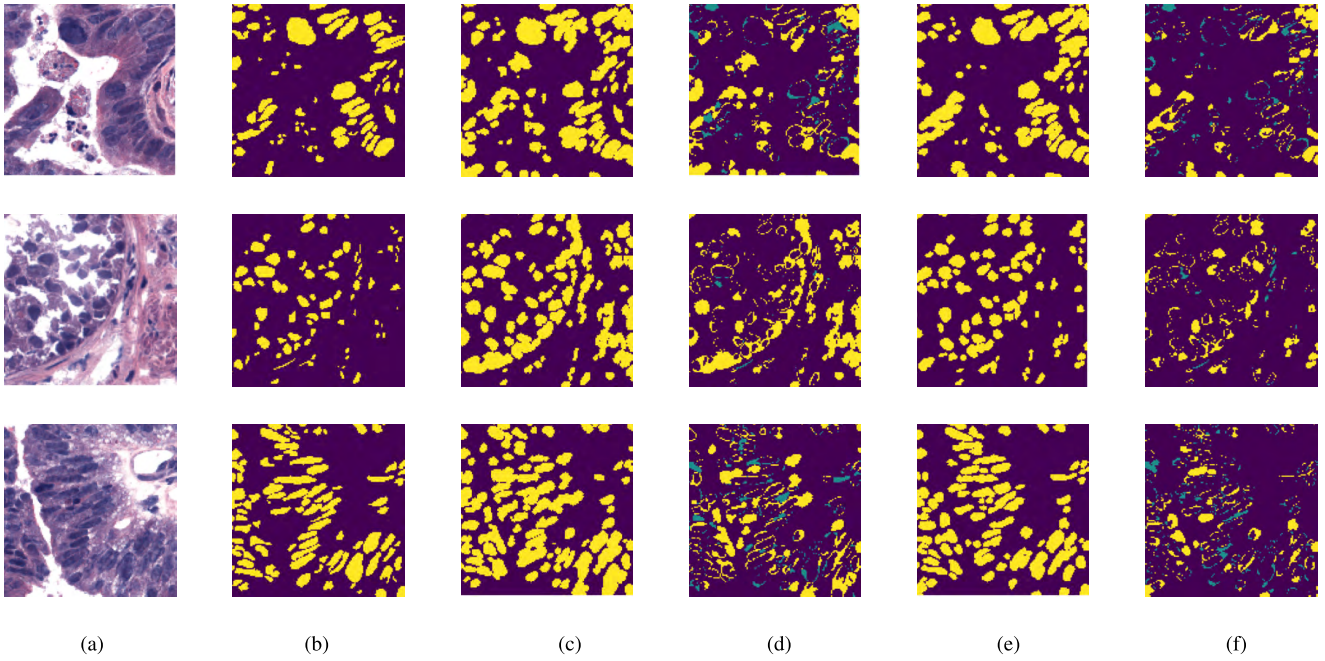

FIGURE 7. Three sub-images result on complex case, the original image named *complex_image* 1, *complex_image* 2 and *complex_image* 3. In (d) and (f), the yellow pixels represent the result of the miss-segmentation, and the green pixels represent the result of the missed segmentation. (a) Sub Image. (b) Sub Image GT. (c) U-net. (d) Difference between GT and U-net. (e) RIC-Unet. (f) Difference between GT and RIC-Unet.

F1-score decreased. The reason for these results is probably that the network is too deep and the number of images is insufficient to cause a certain degree of overfitting. Therefore, in order to alleviate the overfitting and improve the average F1-score to increase the detection rate of the nuclei, in our RIC-Unet, we also use residual blocks, so that the network can not only deepen to learn more detailed features, but also enrich the semantics through skip connection, thus reducing the error rate of pixel-level classification.

Our network incorporates information from different resolutions and separating touching nucleus with extra contour prediction. And our method is higher than other methods in related indicators. Besides, the test time of one image with size 1000 * 1000 is less than 1 second on one

TABLE 2. Simple case evaluation results by original U-net and RIC-Unet.

Image	Method	AJI	Dice	F1-score
simple_image 1	U-net	0.6398	0.8732	**0.8567**
	RIC-Unet	**0.6486**	**0.8865**	0.8465
simple_image 2	U-net	0.5294	**0.7425**	**0.8394**
	RIC-Unet	**0.5352**	0.7408	0.8284
simple_image 3	U-net	**0.6274**	**0.8181**	**0.8866**
	RIC-Unet	0.6191	0.8069	0.8756

TABLE 3. Complex case evaluation results by original U-net and RIC-Unet.

Image	Method	AJI	Dice	F1-score
complex_image 1	U-net	0.4927	0.7546	0.7503
	RIC-Unet	**0.5002**	**0.7801**	**0.7862**
complex_image 2	U-net	0.2691	0.5468	0.7081
	RIC-Unet	**0.3911**	**0.6956**	**0.7746**
complex_image 3	U-net	0.3898	0.6854	0.7313
	RIC-Unet	**0.4004**	**0.7305**	**0.7663**

GPU GTX 1080TI, including both pre-processing and post-processing time, which is relatively computational efficient.

Multi-organ nuclei segmentation is useful in the research field of digital pathology, and can serve a series following tasks including detection, cell counting and cancer classification. Although our research has improved the nuclei segmentation results, large improvement margin still needs to be explored, especially for some complicated cases with nuclei and contour which are not quite clear. Although our method has a stronger discrimination effect on some deeper backgrounds which color are not much different from the color of the nuclei than U-net, there are still some miss-segmented results.

IV. CONCLUSIONS

In this paper, we aim to enable the network better learn effective features and get more accuracy nuclei segmentation result, we propose a revised network architecture which use RI blocks and DC blocks. This network can be better than other methods not only in evaluating indicators, but also cost-effective enough to better assist doctors in better diagnosis of the details of these histology images. Future research topic in nuclei segmentation would be reinforcement learning to automatically select features from different resolutions, or adversarial training including test images would be considered to utilize features from test images without annotations. Besides, design more effective post-processing methods to deal with the problem of cell overlap is also meaningful.

ACKNOWLEDGEMENT

Zitao Zeng and Weihao Xie contributed equally to this work.

REFERENCES

[1] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-9, no. 1, pp. 62–66, Jan. 1979.

[2] L. Vincent and P. Soille, "Watersheds in digital spaces: An efficient algorithm based on immersion simulations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 6, pp. 583–598, Jun. 1991.

[3] C. Xu and J. L. Prince, "Snakes, shapes, and gradient vector flow," *IEEE Trans. Image Process.*, vol. 7, no. 3, pp. 359–369, Mar. 1998.

[4] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. NIPS*, 2012, pp. 1097–1105.

[5] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Dec. 2015, pp. 3431–3440.

[6] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam. 2017). "Rethinking atrous convolution for semantic image segmentation." [Online]. Available: https://arxiv.org/abs/1706.05587

[7] 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.

[8] V. Badrinarayanan, A. Handa, and R. Cipolla. (2015). "SegNet: A deep convolutional encoder-decoder architecture for robust semantic pixel-wise labelling." [Online]. Available: https://arxiv.org/abs/1505.07293

[9] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille. (2017). "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs." [Online]. Available: https://arxiv.org/abs/1606.00915

[10] G. Lin, A. Milan, C. Shen, and I. Reid, "RefineNet: Multi-path refinement networks for high-resolution semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 5168–5177.

[11] A. BenTaieb and G. Hamarneh, "Topology aware fully convolutional networks for histology gland segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2016, pp. 460–468.

[12] H. Chen, X. Qi, L. Yu, and P.-A. Heng, "DCAN: Deep contour-aware networks for accurate gland segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2016, pp. 2487–2496.

[13] 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. 1550–1561, Jul. 2017.

[14] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[15] C. Szegedy *et al.*, "Inception-v4, inception-resnet and the impact of residual connections on learning," in *Proc. AAAI*, 2017, pp. 1–7.

[16] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2018, pp. 7132–7141.

[17] C. Yu, J. Wang, C. Peng, C. Gao, G. Yu, and N. Sang. (2018). "Learning a discriminative feature network for semantic segmentation." [Online]. Available: https://arxiv.org/abs/1804.09337

[18] *The Cancer Genome Atlas (TCGA)*. Accessed: May 14, 2016. [Online]. Available: http://cancergenome.nih.gov/

[19] *MICCAI 2018 Computational Precision Medicine Competition: Digital Pathology: Segmentation of Nuclei in Images*. Accessed: Apr. 14, 2018. [Online]. Available: http://miccai.cloudapp.net/competitions/83

[20] Y. Li and L. Shen, "A deep residual inception network for HEP-2 cell classification," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Springer, 2017, pp. 12–20.

[21] L. Chen *et al.*, "SCA-CNN: Spatial and channel-wise attention in convolutional networks for image captioning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6298–6303.

[22] F. Wang *et al.*, "Residual attention network for image classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2017, pp. 3156–3164.

[23] V. Mnih, N. Heess, and A. Graves, "Recurrent models of visual attention," in *Proc. Neural Inf. Process. Syst.*, 2014, pp. 2204–2212.

[24] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár. (2018). "Focal loss for dense object detection." [Online]. Available: https://arxiv.org/abs/1708.02002

[25] M. Macenko *et al.*, "A method for normalizing histology slides for quantitative analysis," in *Proc. IEEE Int. Symp. Biomed. Imag., From Nano Macro*, Jun./Jul. 2009, pp. 1107–1110.

[26] C. J. van Rijsbergen, *Information Retrieval*. London, U.K.: Butterworths, 1979.

[27] L. R. Dice, "Measures of the amount of ecologic association between species," *Ecology*, vol. 26, no. 3, pp. 297–302, 1945.

[28] A. E. Carpenter *et al.*, "CellProfiler: Image analysis software for identifying and quantifying cell phenotypes," *Genome Biol.*, vol. 7, no. 10, p. R100, 2006.

[29] F. Dong *et al.*, "Computational pathology to discriminate benign from malignant intraductal proliferations of the breast," *PLoS ONE*, vol. 9, no. 12, p. e114885, 2014.

[30] 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.

ZITAO ZENG received the B.E. degree in computer science from Center China Normal University, Wuhan, China, in 2016. He is currently pursuing the M.E. degree in computer science with Sun Yat-sen University, Guangzhou, China. His research interests include medical image processing and computer-aided diagnosis.

WEIHAO XIE received the M.E. degree in computer science from the Guangdong University of Technology, China, in 2016. He is currently an Algorithmic Engineer with PVmed Inc., Guangzhou, China. His research interests include computer vision and computer-aided diagnosis.

YUNZHE ZHANG received the master's degree in electrical computer engineer from Carnegie Mellon University, Pittsburgh, PA, USA, in 2016. He is currently a Research Engineer with PVmed Inc., Guangzhou, China. His research interests include medical image analysis and medical text processing.

YAO LU was a Postdoctoral Research Fellow and Research Investigator with the Medical School, University of Michigan. He is currently a Professor with the School of Data and Computer Science, Sun Yat-sen University, Guangzhou, China. His research interests include inverse problem, medical image processing, and computer-aided diagnosis.