

Received August 12, 2020, accepted August 25, 2020, date of publication August 31, 2020, date of current version September 11, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3020393

# A Fast and Accurate Algorithm for Nuclei Instance Segmentation in Microscopy Images

ZHIMING CHENG<sup>1</sup> AND AIPING QU<sup>1,2</sup>, (Member, IEEE)

<sup>1</sup>School of Computer, University of South China, Hengyang 421001, China

<sup>2</sup>Hunan Medical Big Data International Science and Technology Innovation Cooperation Base, Hengyang 421001, China

Corresponding author: Aiping Qu (nhqap@usc.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61701218, in part by the Natural Science Foundation of Hunan Province of China under Grant 2018JJ3449, in part by the Education Commission of Hunan Province of China under Grant 17B229, and in part by the Science and Technology Planning Project of Hengyang under Grant 2016KG90.

**ABSTRACT** Nuclei instance segmentation within microscopy images is a fundamental task in the pathology work-flow, based on that the meaningful nuclear features can be extracted and multiple biological related analysis can be performed. However, this task is still challenging because of the large variability among different types of nuclei. Although deep learning(DL) based methods have achieved state-of-the-art results in nuclei instance segmentation tasks, these methods are usually focus on improving the accuracy and require support of powerful computing resources. In this paper, we joint the detection and segmentation simultaneously, and propose a fast and accurate box-based nuclei instance segmentation method. Mainly, we employ a fusion module based on the feature pyramid network(FPN) to combine the complementary information of the shallow layers with deep layers for detection the nuclear location by bounding boxes. Subsequently, we crop the feature maps according to the bounding boxes and feed the cropped patches into an U-net architecture as a guide to separate clustered nuclei. The experiments show that the proposed approach outperforms prior state-of-the-art methods, not only on accuracy but also on speed. The source code will be released at: <https://github.com/QUAPNH/Nucleiseg>.

**INDEX TERMS** Nuclei instance segmentation, microscopy images, convolutional neural network.

## I. INTRODUCTION

Nuclei instance segmentation in microscopy images is a fundamental task in the pathology work-flow, based on that the nuclear morphometric and appearance features such as average size, density and pleomorphism can be extracted, and multiple biological related analysis such as classifying phenotypes and profiling treatments can be performed [1]. For instance, it is well-known that the Breast Cancer is the most common malignant tumor in women worldwide [2]. The gold standard for its diagnosis, grading and prognosis predicting remains examination of H&E stained tissue under a microscope [3]. Nuclear shapes and spatial arrangements, which are key factors of Nottingham grading system, are achieved by pathologists manual examine the H&E stained tissue sections [4]. With the development of the high throughput technologies, manual analysis of millions of tissue biopsies seems to be a bottleneck [5]. Therefore, automatic techniques

that accurately and fastly segment nuclei instance in histology images spanning diverse patients, organs and disease states, have significant contribution for developing computer aided systems of clinical and medical applications [6].

Nuclei instance segmentation, which aims to detect and segment every nucleus in a microscopy image simultaneously. Automatically accurate segmentation of nucleus at instance-level in microscopy images remains a challenging task due to several reasons [7]–[9]. Firstly, there are existing amount of nuclei occlusion, adhesion and clusters, which can easily cause over or under-segmentation and affect morphological feature measurements. Secondly, the blurred border and inconsistent staining make it is difficult to distinguish each nucleus in the images, and hence introduces mislabelling and incorrect annotations even accomplished by a specialized pathologist, which is challenging for obtaining objective and robust results. Thirdly, the images are acquired under a variety of pathological conditions. Nuclei enlarge, exhibit margination and prominent nucleoli that make nuclear appearance, magnitude, density vary among different cancer

The associate editor coordinating the review of this manuscript and approving it for publication was Kumaradevan Punithakumar<sup>1</sup>.

subtypes. Thus require the method possess good ability to generalize across these variations. The large differences of nuclear morphology and key challenges for segmentation of nuclei in the microscopy images are illustrated in Fig.1.

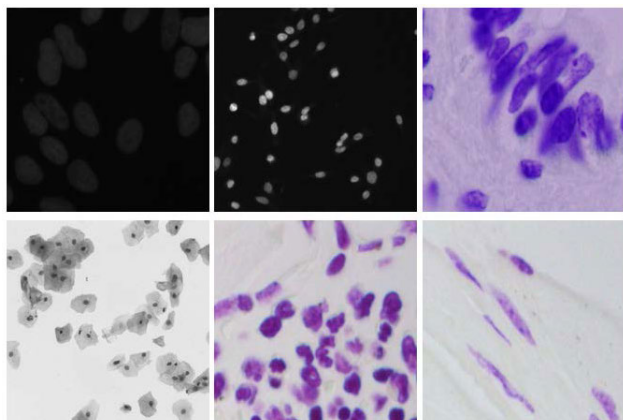


FIGURE 1. Examples of microscopy nuclei images.

In recent years, we have seen the rapid development of microscopic imaging techniques, lots of studies have developed various algorithms to study nuclei instance segmentation [1], [10]. Algorithms based on conventional image processing techniques such as thresholds, marker controlled watershed and morphological operations, are still employed for this complex and difficult task. They typically need the shape and spatial priors to guide the segmentation pipeline. A drawback of these classic algorithms is the need for prior knowledge to adjust the parameters, which cause the methods usually work well on one dataset but perform poorly when generalize to new datasets or experimental conditions change, thus the parameters need be returned. Recently, deep learning methods dominate the field of computer vision and medical image analysis [11], amount variants of deep Convolution Neural Network(DCNN) are applied in nuclei instance segmentation [12], [13]. Most algorithms exclusively focus on improving the accuracy of segmentation [14]. Thus, a lot of algorithms even with complicated combination of different techniques have been proposed, which hope to achieve state-of-the-art segmentation accuracy. In clinical situations, the computing resources are usually insufficient. Therefore, it is significant to develop efficient and effective algorithms that can perform fast and accurate instance segmentation on resource restricted situations.

Recently, in order to capture the nuclei instances accurately, Yi *et al.* [15] developed an attentive instance segmentation model called ANCIS which was built on a joint network that combined single shot multi-box detector (SSD) [16] and Unet [17]. The attention mechanisms are used in both the detection module and segmentation module of ANCIS to improve the detection and segmentation accuracy, but these attention mechanisms require more computations, which complicate the model and restrict its execution speed. In order to address this drawback, we propose a new architecture

for the encoding module and segmentation module. Inspired by the classical structure of U-Net, we design the encoding branch and utilize SSD to detect the bounding boxes, and leverage these boxes to crop the RoI patches for guiding the segmentation, particularly for the adhesion and clustered nuclei. The contributions of this work are:

- We joint the detection and segmentation simultaneously, and propose a simple box-based nuclei instance segmentation method, which utilizes detection branch to guide the segmentation.
- The fusion module based on FPN in the segmentation branch could aggregate the complementary information of the shallow layers with deep layers. Instance normalization is used in skip connection to remove the neighbor statistics for cropped patch and focus on the target.
- Experimental results demonstrate that the proposed approach achieves a remarkable performance, not only on accuracy but also on speed.

The rest of the paper is organized as follows: Section II reviews the relevant works on nuclei instance segmentation. Section III presents the proposed nuclei instance segmentation approach. In Section IV, we describe the datasets, experiments and results. Finally, we discuss the conclusion in Section V.

## II. RELATED WORK

Within the prior literatures, methods based on the energy model such as the watershed algorithm, have been widely utilized for nuclear instance segmentation. Yang *et al.* [18] presented a marker-controlled watershed based on mathematical morphology to extract the nuclear instances with less over-segmentation. However, it often produces unreliable results because its morphological operator relied on consistent intensity difference between the foreground and background, which cannot well work in more complex situations. Lots of approaches have attempted to obtain an improved marker for marker-controlled watershed, such as active contours [19] and more complicated morphological operations to generate the energy landscape [20]. The results heavily dependent on the predefined geometry of the nuclei which determined the marker generation. Because obtaining sufficiently strong markers for guiding watershed algorithm is a challenge task, some methods have departed from the energy-based approach. Kwak *et al.* [21] tried to compute the concavity of nuclear clusters. Liao *et al.* [22] assumed the nuclear shape is eclipse and utilized ellipse-fitting to separate the clusters, which didn't encompass the natural diversity of the nuclei. In addition, the segmentation results of these methods were sensitive to the parameters, which usually determined by manual selection.

Recently, deep learning methods achieve outstanding performances than non-deep state-of-the-art methods and dominate the fields of computer vision tasks. Multiple researchers attempt to investigate the potential to employ deep learning in nuclei instance segmentation. One representative class of algorithms called box-free instance segmentation, which

segment nuclei instances directly without the aid of bounding boxes. For example, Chen *et al.* [23] presented a deep contour aware network(DCAN) by harnessing the nuclear appearance and contour information explicitly. They first utilized the complementary appearance and contour information to do semantic segmentation and extract contours under a unified multi-task learning framework. After that, they overlap the contours onto the semantic segmentation results in order to separate the touching and clustered nuclei. DCAN ranked the first in MICCAI nuclei segmentation challenge. However, it tends to produce over-segmentation when the boundaries between the touching nuclei unclear. Oda *et al.* [24] proposed a Boundary-Enhanced Segmentation Network (BESNet) that has an encoding path similar to Unet and two decoding paths to restore the resolution. One decoding path is designed to enhance the nuclear boundaries, which can be used to improve the quality of the entire nuclei segmentation achieved in the other decoding path. There are two limitations of BESNet, one is that it learns complementary information in nuclei branch but ignores the potential benefits from nuclei to contour, another is its computational complexity. Zhou *et al.* [25] proposed a Contour-aware Informative Aggregation Network (CIA-Net) with multilevel information aggregation module that leverage the merit of spatial and texture dependencies between nuclei and contour by bi-directionally aggregating task-specific features. Graham *et al.* [26] presented Hover-Net for nuclear segmentation and classification that leverages the instance-rich information encoded within both vertical and horizontal distances of nuclear pixels to their centers of mass. Spatial aware network (Spa-Net) [27] was proposed to predict the position information of each nuclei by combining a multi-scale dense unit. The final nuclei instances are separated by a clustering method through clustering the predicted location ordinates.

In the field of instance segmentation, to overcome the weakness of box-free methods which without global understanding of the objects, amounts of studies have sought to combine object detection and segmentation. These box-based methods locate the objects' bounding boxes from a global perspective and perform the instance segmentation within the region defined by the bounding boxes. Mask RCNN [28], which extended the Fast RCNN [29], perhaps the most common and successful two-stage method. The two-stage methods first detect the object bounding boxes by initial candidate regions, then crop patches according the bounding boxes and perform segmentation within the cropped patches for achieving more accurate results. However, these methods are computationally expensive and require support of powerful computing. Several approaches, as exemplified by SSD [16] and RetinaNet [30], which eschewed the second refinement stage in two-stages methods and focused on direct sliding-window prediction, and shown promising results. For nuclei instance segmentation, Wang *et al.* [31] designed a dilated residual network to solve the problem of information loss of small objects in deep neural network, and showed a well recognition and segmentation capability for nuclei

detection and segmentation in microscopic images. In order to improve the precision of neural cell instance segmentation, Yi *et al.* [15] constructed a joint network based on the SSD and Unet. They employed the attention mechanism in both detection and segmentation modules to refine the feature maps and make the model pay attention to the useful features. However, the attention modules complicate the model and need higher computing. In order to deal with adjacent neural cells, Yi *et al.* [32] proposed a context refined instance segmentation method to focus on the targets within cropped ROIs and suppress the background information. Yi *et al.* [33] proposed a keypoint-based object detector based on detecting the five pre-defined points of a nucleus and a separate segmentation branch to segment the nuclei instance, but it fails to localize the small objects. Chen *et al.* [34] presented a Boundary-assisted Region Proposal Networks (BRP-Net) for nuclei instance segmentation. A task specific feature encoding network is utilized to get the instance proposals in the first stage, and a instance segmentation network is used to segment proposals in the second stage.

### III. METHODS

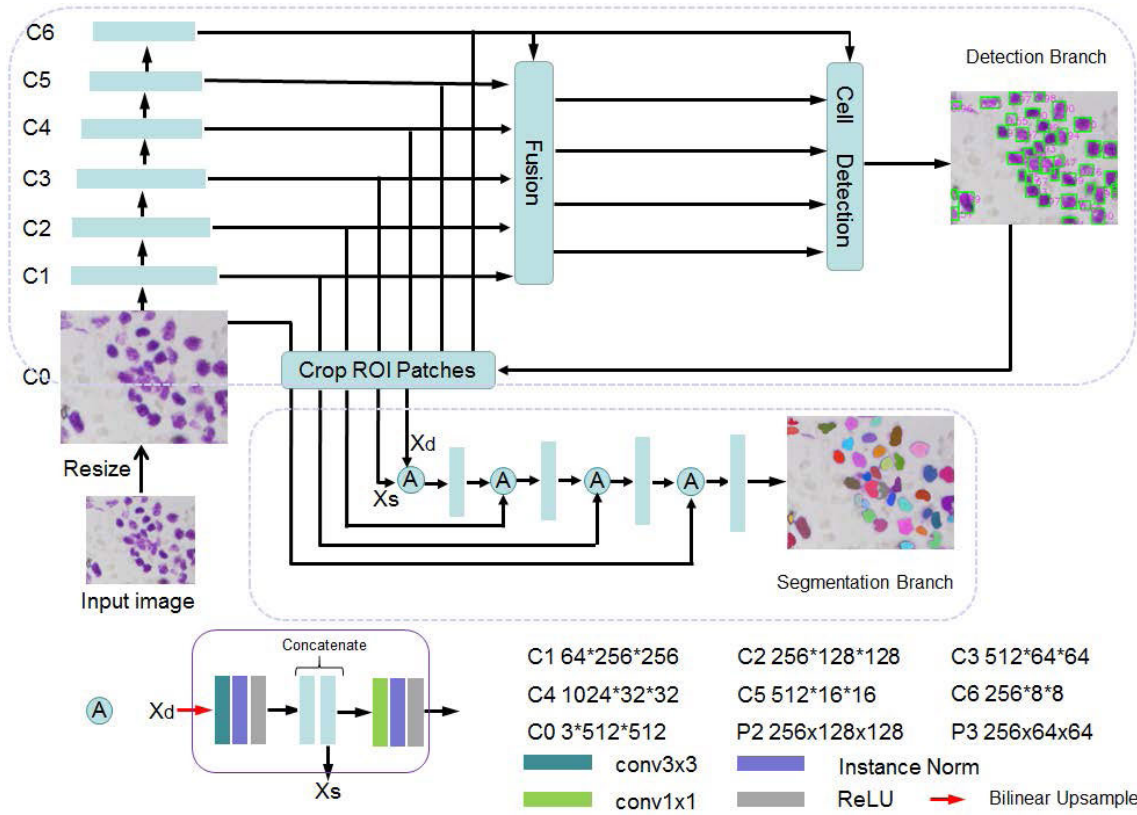
The flow chart of the proposed network is illustrated in Fig. 2. The proposed model is an unified and end-to-end trainable instance segmentation network which includes a detection branch to locate the bounding boxes of nuclei and a segmentation branch to segment nuclei. The segmentation is performed on the RoI patches which are cropped from the feature maps by employing the detected bounding boxes. Below we introduce our network in details.

#### A. NUCLEI DETECTION BRANCH

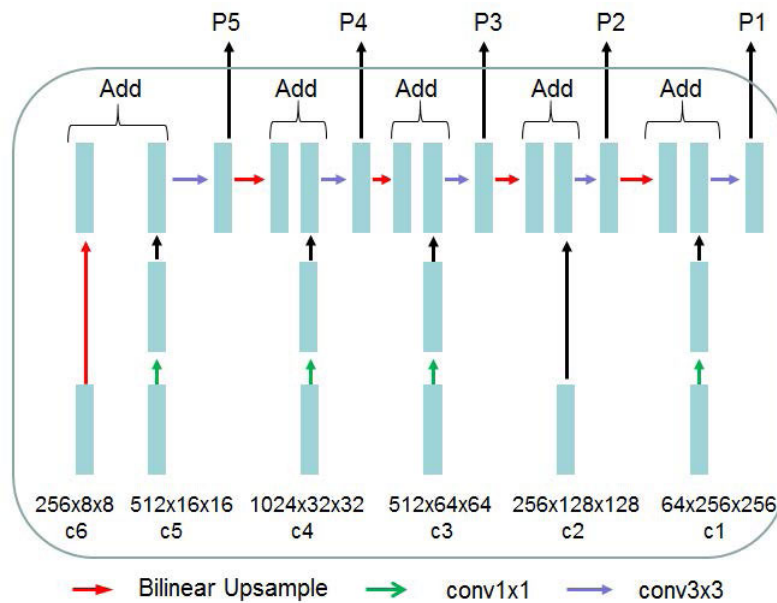
Inspired by ANCIS, we redesign the detection network as shown in the top half of Fig 2. We design a fusion module by fusing the shallow layers and deeper layers to build high-level semantic feature maps. We resize the input image to  $512 \times 512$  before being fed into fusion module. We use resnet50 Conv1-6 as the backbone network [35]. We set the parameters of Resnet50 same as Yi [15].

The feature pyramid network(FPN) [36] is a well-known structure in the field of object detection, which fuse the shallow feature map and the upsampled deep feature map, to detect different scale targets. In order to leverage the rich semantic information of small nuclei, we use FPN to fuse the shallow feature map and deep feature map of the backbone network. The details of fusion module are shown in Fig 3. Firstly, we convert the feature map  $c_1$ ,  $c_3$ ,  $c_4$  and  $c_5$  of backbone network to the channel numbers(256) by using a  $1 \times 1$  convolutional layer. Secondly, we resize feature map  $c_6$  using bilinear interpolation to  $16 \times 16$ , which is equal to  $c_5$ . We add the feature map from  $c_5$  and  $c_6$  to obtain output feature map  $p_5$  through a  $3 \times 3$  convolutional layer. Repeat similar strategy as the second step, we get output feature maps  $p_1$ - $p_4$  of fusion module.

In order to obtain the bounding boxes to guide the instance segmentation, we employ the single shot multibox



**FIGURE 2.** Flowchart of the proposed method. The above is the detection branch and below is the segmentation branch. ROI patches are cropped from the feature maps with the guidance of the bounding boxes obtained by the detection branch. The size of each layer is shown in *numberofchannels × height × width*. The input images are resized to 512 × 512.



**FIGURE 3.** Fusion module.

detection (SSD) model for nuclei detection. We pass the feature maps p1,p2, p3, p4, p5 and c6 into SSD for bounding box regression and confidence prediction. We utilize the same detection strategy as SSD [16]. Firstly, SSD spreads

the anchor boxes to multi-scale feature maps densely. The aspect ratios and scales in anchor boxes assignment process are set as (0.5,2), (0.5,2,1/3,3), (0.5,2,1/3,3), (0.5,2,1/3,3), (0.5,2,1/3,3), (0.5,2,1/3,3) and 0.01, 0.04, 0.1, 0.26, 0.42,

0.58 for p1, p2, p3, p4, p5 and c6, respectively. Secondly, we use the ground-truth boxes and labels to create the encoded anchors and their corresponding labels. Finally, we optimize the model weights by solving a total detection loss function.

Let  $L_{locs}$  and  $L_{conf}$  be the bounding box loss and confidence loss, respectively. The total detection loss function is a weighted combination of bounding box loss and confidence loss. It is defined as following:

$$L_{det} = \frac{1}{N_{pos}}(L_{locs} + \theta L_{conf}), \quad (1)$$

where  $N_{pos}$  is the number of positive anchors,  $\theta$  is a weight factor.

$L_{locs}$  indicates a smooth  $L_1$  loss. It is defined as following:

$$L_{locs} = \sum_{i \in pos} \sum_{m \in \{cx, cy, w, h\}} smooth_{L_1}(l_i^m - g_i^m), \quad (2)$$

$smooth_{L_1}(x)$  is a piecewise function:

$$smooth_{L_1}(x) = \begin{cases} 0.5x^2, & |x| \leq 1 \\ |x| - 0.5, & otherwise \end{cases} \quad (3)$$

where  $l_i^m$  and  $g_i^m$  are the predicted and encoded anchor boxes,  $m$  is parameter of coordinate.

$L_{conf}$  is defined as a binary cross entropy loss between the ground truth confidence and predicted box confidence. The form of  $L_{conf}$  is defined as following

$$L_{conf} = - \sum_i (x_i \log p_i + (1 - x_i) \log(1 - p_i)), \quad (4)$$

where  $x_i$  is the labeled confidence, and  $p_i$  is the predicted confidence,  $i$  refers to each position in a segmentation mask.

## B. NUCLEI SEGMENTATION BRANCH

After the detection operation, we obtain the bounding boxes for all nuclei in the input images, then we crop the instance nuclei regions from the encoder layer C0-C6 and perform the individual nuclei segmentation for each nuclei instance in these cropped regions. The process of nuclei segmentation module is shown in the bottom of Fig 2. Inspired by the classical structures of U\_Net, we design a skip connection to combine the feature maps from the shallow layers with the deep layers, which can utilize both high-level semantics and low-level image details.

The process of the skip connection is illustrated in the bottom\_left corner of Fig 2. A deep feature map  $x_d \in \mathbb{R}^{C_d \times W_d \times H_d}$  and a shallow feature map  $x_s \in \mathbb{R}^{C_s \times W_s \times H_s}$  are passed into skip connection, where  $C$ ,  $W$ ,  $H$  are the channel size, height and width of a feature map. Firstly, the feature map  $x_d$  is up-sampled to the same size as  $x_s$  through a bilinear interpolation. After a  $3 \times 3$  convolutional layer, instance normalization, and Rectified Linear Unit(ReLU) layer, the feature map is passed into concatenation operation. It should be noted that instance normalization is proved to be a perfect choice for our task. Instance normalization can be used to remove style statistics of instance for image generation [37].

For our segmentation task, each cropped patch is supposed to include mainly one nucleus. Thus, we can formalize our problem as removing the neighbor statistics for each cropped patch, so it is feasible to utilize instance normalization. Let  $x \in \mathbb{R}^{H \times W}$  be a cropped ROI patch, the instance normalization can be defined as following:

$$x'_{h,w} = \gamma \left( \frac{x_{h,w} - \mu}{\sigma} \right) + \beta, \quad (5)$$

where  $\mu$  and  $\sigma$  are respectively the mean and variance of the cropped patch,  $\gamma$  and  $\beta$  are scaling factors to control the extent of cropped patch normalization. Secondly, we concatenate map from the previous operation and a shallow feature map  $x_s$ . Finally, we transfer the concatenation feature map through a  $1 \times 1$  convolutional layer, instance normalization, and a Rectified Linear Unit(ReLU) layer, to obtain the output feature map.

We use binary-cross entropy loss as the objective loss of the segmentation. It is given by

$$L_{seg} = - \frac{1}{N} \sum_j \frac{1}{M_j} \sum_i (t_{ij} \log p_{ij} + (1 - t_{ij}) \log(1 - p_{ij})), \quad (6)$$

where  $p$  and  $t$  represent the predicted and ground-truth segmentation probability map, respectively.  $i$  is the positions of pixels,  $j$  is positive predicted bounding box,  $M$  is number of pixels in the map, and  $N$  is the number of positive predicted bounding boxes.

## IV. EXPERIMENTS AND RESULTS

### A. DATA SETS

We employ two public data sets: Data Science Bowl 2018(DSB2018) and MoNuSeg to verify the effectiveness of the proposed method. The two datasets are described in detail as follows:

#### 1) DATA SCIENCE BOWL 2018

DSB2018 is released by the Kaggle competition, which aims to develop nuclei segmentation methods in a variety of microscopy images without manual interaction or adjustment [38]. The dataset contains a large number of segmented nuclei images. The images were extracted under a variety of microscopic imaging conditions, microscopy instruments, operators and staining protocols, thus there are vary in cell types, magnification and imaging modality. The nuclei particle size distribution is in the range of 21 to 1037 pixels, which brings great difficulties for nuclear detection and segmentation in some images. The dataset is usually utilized to measure an algorithm's generalization ability across variations.

#### 2) MoNuSeg DATA SET

MoNuSeg [6] is released to encourage the development of nucleus segmentation algorithms on a diverse set of Hematoxylin-Eosin (H&E) stained tissue images. The dataset contains 30 images with 21623 manually annotated nuclear boundaries in total. Each image is download from the Cancer

Genomic Atlas(TCGA) of an individual patient. Since the size of whole slide image is too large, the image is cropped to  $1000 * 1000$  to reduce the computational requirement. The dataset collected from 18 different hospitals. Since the staining process and image scanning equipments across labs aren't exactly the same, there exist large appearance variations among them. Furthermore, in order to ensure the abundance to nuclear types, 7 different organs viz., breast, liver, kidney, prostate, bladder, colon and stomach, including both benign and diseased tissue samples, are represented.

## B. EVALUATION METRICS

We utilize mask-level average precision(AP) at IoU (intersection over union) threshold of  $\alpha$ (term as  $AP@_\alpha$ ) to evaluate the nuclear instance segmentation accuracy [15]. AP is formulated as following:

$$AP = \frac{1}{11} \sum_{l \in \{0, 0.1, \dots, 1\}} \max(p(r \geq l)) \quad (7)$$

where  $p$  is the mean precision at 11 recall( $r$ ) levels ( $\{0, 0.1, \dots, 1\}$ ). We also use  $AIoU@_\alpha$  to measure the average IoU at thresholds  $\alpha$  to measure the quality of the segmentation following the work of [15]. It is defined as following:

$$AIoU@a = \frac{1}{N_\alpha} \left( \sum_{i=1}^{N_\alpha} (IoU)_i \right), \quad (8)$$

where  $N_\alpha$  is the total number of predictions that satisfy  $IoU \geq \alpha$  ( $\alpha = 0.5$  and  $0.7$ ).  $IoU$  is calculated between the predicted segmentation mask and all the ground-truth masks for each prediction:

$$IoU = \max_i \left( \frac{x \cap y_i}{x \cup y_i} \right), \quad x \in \mathbb{R}^{W \times H}, \quad y_i \in \mathbb{R}^{W \times H} \quad (9)$$

where  $x$  is the predicted mask,  $y_i$  is the  $i$ -th ground-truth mask,  $W$  and  $H$  are the width and height of the image, respectively.

In addition to accuracy, another important evaluation index of target detection and segmentation algorithm is speed. Frame Per Second(FPS), which is the number of images that can be processed per second, is a common index to evaluate the speed. Therefore, we use FPS, which is calculated both the inference time and post-processing time, to measure the speed of our method.

## C. IMPLEMENTATION DETAILS

We implement the model in the open source software library Pytorch, and perform it on a high performance computer with 35.4816 Tflops CPU and 18.8 Tflops GPU. In order to avoid over-fitting and improve the generalization ability, we utilize several data augmentation methods such as random expanding, cropping, flipping, contrast distortion and brightness distortion. We initialize the parameters of the backbone network from Resnet50 which is pretrained on Imagenet, and set the other parts with a standard Gaussian distribution. Since the detection branch is used to crop the RoI patches for guiding the segmentation branch, we first train the detection network,

then froze its parameters when training the segmentation network. We initialize the learned scaling factors  $\beta = 0$  and  $\gamma = 1$  in Formula (1). We set the mini-batch size to 100 and the learning rate to 0.001. We use the stochastic gradient descent(SGD) to solve the optimize problem in our method.

## D. ABLATION STUDY

To verify the fusion module is effective for improving the SSD's detection accuracy of bounding boxes, we compare the results of the detection module with SSD on kaggle 2018 Data Science Bowl and MoNuSeg. We randomly select some images from two datasets, and display the detailed results in Fig.4 and Fig.5, respectively. From the results, we could see that the detection module used in our method could improve the detection accuracy of bounding boxes.

To verify the mechanisms of fusion and instance normalization are effective for improving the model's performance, we compare the quantitative results of our method with different modules on kaggle 2018 Data Science Bowl and MoNuSeg dataset. The results are showed in Table 1 and Table 2. Baseline means the method without Fusion module and instance normalization. Its segmentation branch is a standard Unet structure. Fusion means the method with fusion mechanism. InstanceNorm means skip connection with instance normalization operation. Fusion+InstanceNorm means the proposed method with fusion mechanism and instance normalization. The results showed in Table 1 and Table 2 appear to suggest that the fusion and instance normalization mechanism can improve the performance of segmentation, and the combination method with both mechanisms achieves the best performance.

**TABLE 1. Effects of fusion and instance normalization mechanisms on kaggle 2018 Data Science Bowl.**

Method	AP@0.5	AP@0.7	AIoU@0.5	AIoU@0.7
Baseline	62.49	60.31	82.17	84.31
InstanceNorm	62.55	60.50	<b>82.70</b>	<b>84.90</b>
Fusion	71.31	60.76	82.06	84.51
Fusion+InstanceNorm	<b>71.38</b>	<b>60.80</b>	82.62	<b>84.90</b>

**TABLE 2. Effects of fusion and instance normalization mechanisms on MoNuSeg.**

Method	AP@0.5	AP@0.7	AIoU@0.5	AIoU@0.7
Baseline	68.32	46.23	76.64	80.47
InstanceNorm	68.69	46.30	<b>77.53</b>	<b>80.99</b>
Fusion	76.62	53.77	76.72	80.59
Fusion+InstanceNorm	<b>77.65</b>	<b>55.05</b>	77.02	80.83

## E. RESULTS

Recently, a lot of literatures are committed to develop nuclei instance segmentation methods and got state-of-the-art performance in microscopy images, such as Mask RCNN [28], DCAN [23], BESNet [24], CIA-Net [25], ANCIS [15], CRN-CIS [32], SPA-Net [27] and BRP-Net [34], Since the source

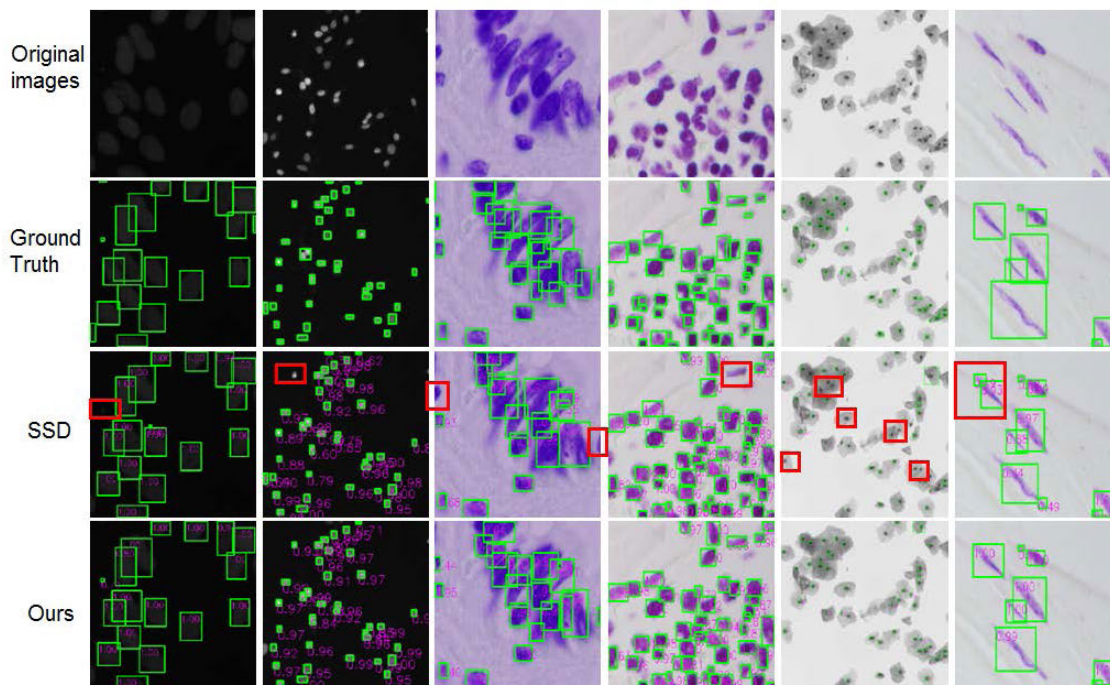


FIGURE 4. Effects of the detection module on kaggle 2018 Data Science Bowl. The red rectangles show the undetected regions.

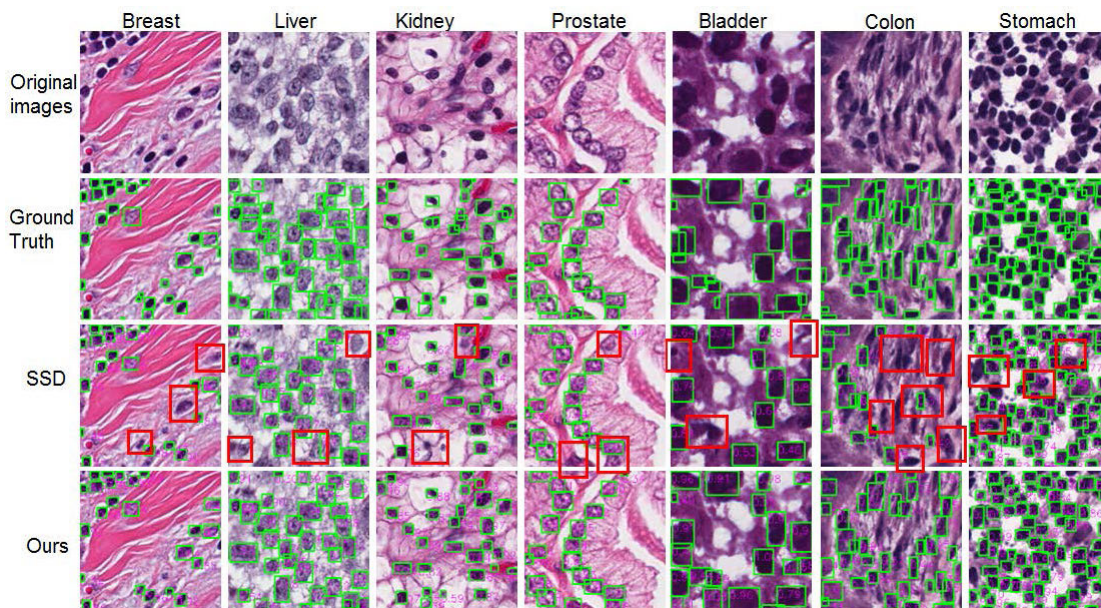
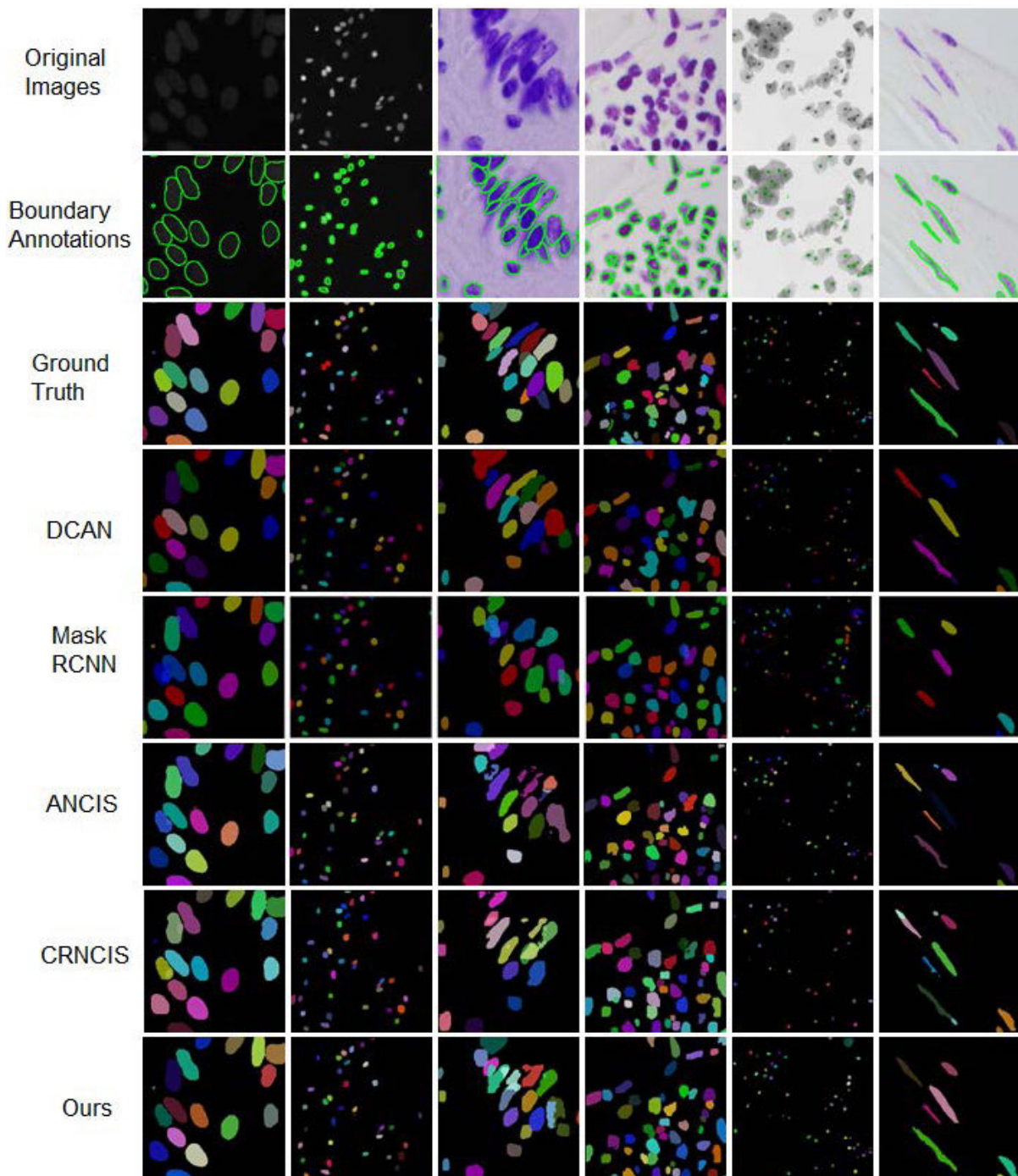


FIGURE 5. Effects of the detection module on MoNuSeg. The red rectangles show the undetected regions.

codes of Mask RCNN, DCAN, CRNCIS and ANCIS are public available, we just compare our method with these methods for fairness. It should be noted that we use exactly the same network architecture of the proposed method for both the two different datasets.

Firstly, we test the performance of the proposed method on DSB dataset. We split the original training datasets with

annotations into three different sub-datasets: training set with 402 images, validation set with 134 images and testing set with 134 images. Table 3 shows the quantitative results of nuclei instance segmentation on the testing set. From the results presented in Table 3, we can see that the proposed method achieves a best performance with 71.38 AP@0.5 and 60.80 AP@0.7, which indicate that the proposed method



**FIGURE 6.** Results of nuclei instance segmentation of our method on kaggle 2018 Data Science Bowl.

can improve the instance segmentation performance. To demonstrate the segmentation, we also detailedly display the results of six images, which are randomly selected from the testing set in Fig. 6. We could see that Mask RCNN can segment the tiny structures of small nuclei, but it fails to capture the correct structure when the nuclei instance is large and cell adhesion. When the boundaries is clear, DCAN can obtain a satisfactory results, but it is less effective

at separating the overlapping and clustered nuclei because the unclear nuclear boundaries. ANCIS can segment the tiny and slender structures, but it tends to make several over-segmentation. Compared to these methods, our model exhibits remarkable performance in segmentation of both tiny and adhesion structures.

Secondly, we test the proposed method on the MoNuSeg dataset. We use the same strategy described in [6] to split



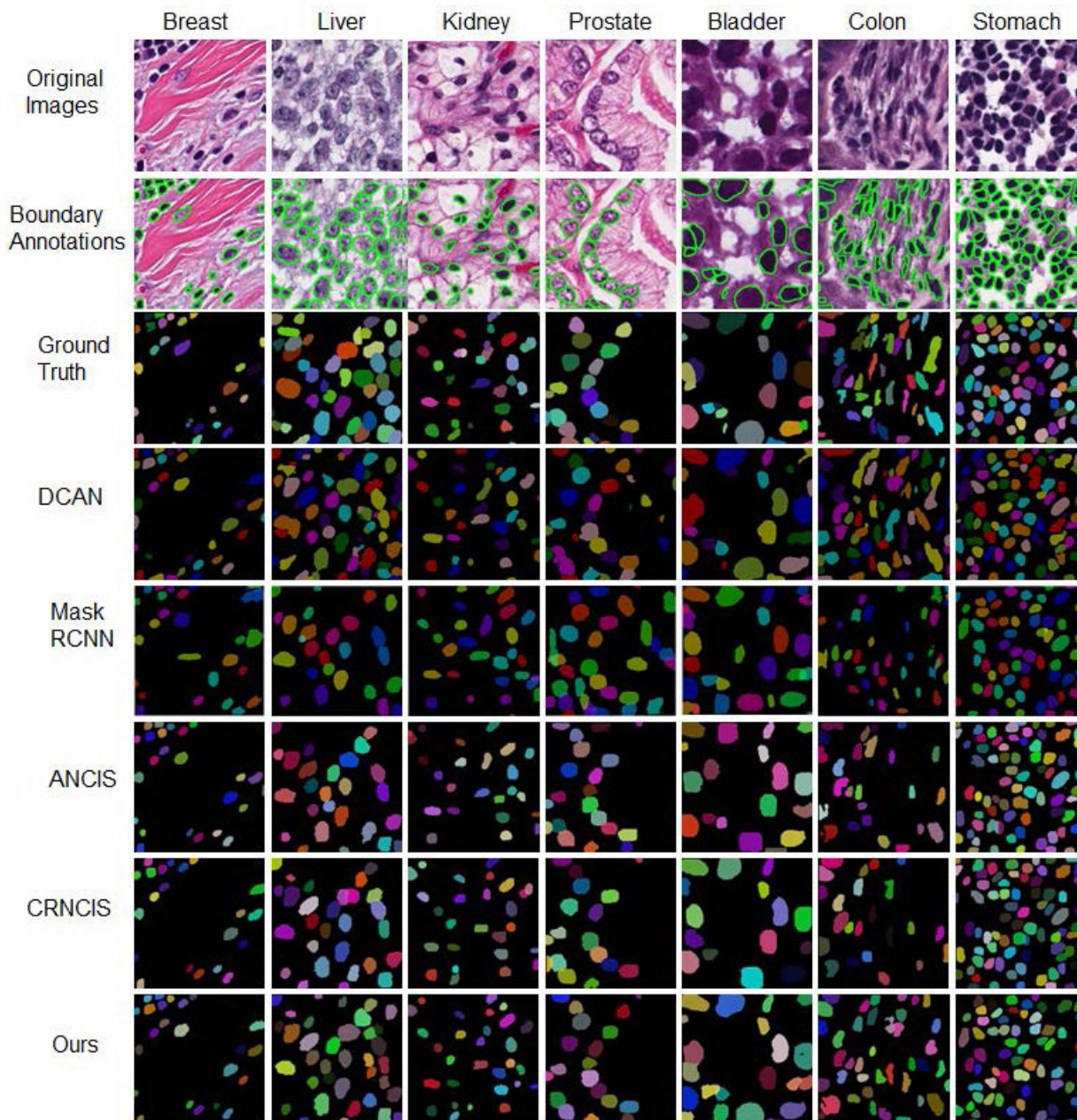


FIGURE 7. Results of nuclei instance segmentation of our method on MoNuSeg.

TABLE 3. Quantitative results of nuclei instance segmentation of our method on kaggle 2018 Data Science Bowl.

Method	AP@0.5	AP@0.7	AIoU@0.5	AIoU@0.7	FPS
MaskRCNN	68.66	51.30	82.65	83.52	1.8970
DCAN	55.13	44.23	80.52	82.87	0.1545
ANCIS	70.83	60.44	82.53	84.86	5.0762
CRNCIS	62.57	59.41	79.67	82.30	4.3577
Ours	<b>71.38</b>	<b>60.80</b>	<b>82.62</b>	<b>84.90</b>	<b>7.4786</b>

the dataset into 3 sub-datasets: training set, validation set and testing set. Compared with the DSB dataset, images in MoNuSeg dataset exist larger number of nuclei, more adhesion and heavy noisy information from the background.

Thus, it is much difficult to segment nuclear instance in these histopathological images. The quantitative results on the testing set are reported in Table 4. The results demonstrate that the proposed method also obtain the best performance with 77.65 AP@0.5 and 55.05 AP@0.7, and comparable result of AIoU among these methods. We also display the detailed results for 7 random selected images of different organs in Fig. 7. The results illustrate that MASK RCNN fails to segment the large instance and DCAN fails to segment the adhesion instance. The ANCIS can capture the tiny and slender structures, but miss several nuclei that similar to the background. Our method has greatly improve the metric of AP@0.5 and AP@0.7.

Regarding the processing time, the average FPS of the proposed method is 7.4786 and 8.8527 for two datasets, respectively, that means the proposed method only takes average of 0.1337 and 0.1130 seconds to complete the segmentation of an image from two datasets, respectively. From Table 3 and Table 4, we also know that the average processing times for ANCIS are 0.1970 and 0.1857 seconds and Mask-RCNN are 0.5271 and 0.5142 seconds, respectively. Therefore, the speed of the proposed method is faster than the other methods.

**TABLE 4. Quantitative results of nuclei instance segmentation of our method on MoNuSeg.**

Method	AP@0.5	AP@0.7	AIoU@0.5	AIoU@0.7	FPS
MaskRCNN	69.50	35.89	74.60	80.95	1.9446
DCAN	51.54	32.61	<b>77.23</b>	<b>81.14</b>	0.2282
ANCIS	69.22	45.44	75.70	80.01	5.3841
CRNCIS	68.17	45.55	76.45	80.15	4.8604
Ours	<b>77.65</b>	<b>55.05</b>	77.02	80.83	<b>8.8527</b>

## V. DISCUSSION AND CONCLUSION

Analysis of nuclei in microscopic images is the first step towards developing automated computer aided methods for diagnosis and prognosis of cancer. A lot of researches attempted to propose various algorithms focusing on accurate nuclei instance segmentation which enable subsequent exploration of nuclear features to predict clinical outcome. For example, DCAN [23] first utilized the complementary appearance and contour information to improve the discriminative capability of intermediate features for separating the adhesion and clustered nuclei. However, its performance heavily depends on the quality of the nuclear boundaries. When the boundaries are unclear such as the adhesion and clustered nuclei, the segmentation results is unsatisfactory. Besides, it is easy to segment the adhesion and clustered nuclei as a large individual one, this phenomenon causes that it has a good performance of AIoU. Mask RCNN [28] fails to capture the nuclei instance who is large and cell adhesion, because it adopts ROI align to restrict the object feature patch to  $7 \times 7$  that lead to ignore the high-resolution details in the shallow layers. ANCIS [15] utilized attention mechanism both in segmentation branch and detection branch, it performs well on tiny and slender neural cells, but it tends to make mistakes to segment touching instances who have the same textures and classes. In addition, the attentive mechanism make it computing complex.

In this paper, we have designed a fusion module to utilize the multi-scale features of backbone network for nuclei instance bounding boxes detection, and utilized a skip connection mechanism to combine the features maps from the shallow layers with the deep layers of standard Unet. We have shown that the proposed method obtains the state-of-the-art nuclei segmentation results compared to a lot of recently published methods across two well-known competition datasets. Regarding the processing time, Mask-RCNN utilizes a very large arrays in memory to store the nuclear instances because it inherently stores a single instance per

channel and there are many nuclei in a image, which causes that it requires much longer processing time. ANCIS utilizes attentive mechanism both in the detection and segmentation branches, which makes the computing procedure more sophisticated. In our method, we abandon the attentive module and use some simple mechanisms to optimize the method. In addition, the fusion module could improve the detection accuracy of bounding boxes, and the instance segmentation is performed within the patches that are cropped according the bounding boxes. These are the main reasons that the proposed method could improve the nuclei instance segmentation accuracy.

A major bottleneck for the development of satisfactory nuclei instance segmentation methods is the limitation of training data, particularly with accurate annotations. In this paper, we use the datasets from two well-known competitions, which could be seen as ‘optimized’ version of real data. The real data in clinical environment could contain artificial damages, noise, non-uniform staining and artifacts, etc. Accurate annotation of real data is time consuming and requires professional prior knowledge. Sometimes, annotations are quite different even for professional annotators. In addition, the datasets used in this manuscript include thousands of nuclei, but they are obtained from limited patients. Therefore, these competition datasets are insufficient reflect the characters of real clinical data. Thus, we will test our algorithm in real clinical data in our future work.

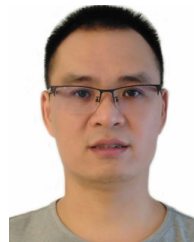
## REFERENCES

- [1] 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.
- [2] R. L. Siegel, K. D. Miller, and A. Jemal, “Cancer statistics, 2019,” *CA, Cancer J. Clinicians*, vol. 69, no. 1, pp. 7–34, 2019.
- [3] E. A. Rakha, J. S. Reis-Filho, F. Baehner, D. J. Dabbs, T. Decker, V. Eusebi, S. B. Fox, S. Ichihara, J. Jacquemier, S. R. Lakhani, J. Palacios, A. L. Richardson, S. J. Schnitt, F. C. Schmitt, P.-H. Tan, G. M. Tse, S. Badve, and I. O. Ellis, “Breast cancer prognostic classification in the molecular era: The role of histological grade,” *Breast Cancer Res.*, vol. 12, no. 4, p. 207, Aug. 2010.
- [4] A. Qu, J. Chen, L. Wang, J. Yuan, F. Yang, Q. Xiang, N. Maskey, G. Yang, J. Liu, and Y. Li, “Two-step segmentation of hematoxylin-eosin stained histopathological images for prognosis of breast cancer,” in *Proc. IEEE Int. Conf. Bioinf. Biomed. (BIBM)*, Nov. 2014, pp. 218–223.
- [5] A. Khademhosseini and R. Langer, “A decade of progress in tissue engineering,” *Nature Protocols*, vol. 11, no. 10, p. 1775, 2016.
- [6] 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.
- [7] H. Irshad, L. Montaser-Kouhsari, G. Waltz, O. Bucur, J. Nowak, F. Dong, N. W. Knoblauch, and A. H. Beck, “Crowdsourcing image annotation for nucleus detection and segmentation in computational pathology: evaluating experts, automated methods, and the crowd,” in *Proc. Pacific Symp. Biocomput. Co-Chairs*. Singapore: World Scientific, 2014, pp. 294–305.
- [8] N. Kumar, R. Verma, D. Anand, and Y. Zhou, “A multi-organ nucleus segmentation challenge,” *IEEE Trans. Med. Imag.*, vol. 39, no. 5, pp. 1380–1391, May 2020.
- [9] 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.
- [10] Z. Hu, J. Tang, Z. Wang, K. Zhang, L. Zhang, and Q. Sun, “Deep learning for image-based cancer detection and diagnosis—A survey,” *Pattern Recognit.*, vol. 83, pp. 134–149, Nov. 2018.

- [11] M. Zahangir Alom, T. M. Taha, C. Yakopcic, S. Westberg, P. Sidike, M. Shamima Nasrin, B. C. Van Esesn, A. A. S. Awwal, and V. K. Asari, "The history began from AlexNet: A comprehensive survey on deep learning approaches," 2018, *arXiv:1803.01164*. [Online]. Available: <http://arxiv.org/abs/1803.01164>
- [12] E. Moen, D. Bannon, T. Kudo, W. Graf, M. Covert, and D. Van Valen, "Deep learning for cellular image analysis," *Nature Methods*, vol. 16, pp. 1233–1246, May 2019.
- [13] T. Wollmann, M. Gunkel, I. Chung, H. Erfle, K. Rippe, and K. Rohr, "GRUU-net: Integrated convolutional and gated recurrent neural network for cell segmentation," *Med. Image Anal.*, vol. 56, pp. 68–79, Aug. 2019.
- [14] P. Naylor, M. Lae, F. Reyat, 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.
- [15] J. Yi, P. Wu, M. Jiang, Q. Huang, D. J. Hoepfner, and D. N. Metaxas, "Attentive neural cell instance segmentation," *Med. Image Anal.*, vol. 55, pp. 228–240, Jul. 2019.
- [16] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single shot multibox detector," in *Proc. Eur. Conf. Comput. Vis. Cham, Switzerland: Springer*, 2016, pp. 21–37.
- [17] 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.*, 2015, pp. 234–241.
- [18] X. Yang, H. Li, and X. Zhou, "Nuclei segmentation using marker-controlled watershed, tracking using mean-shift, and Kalman filter in time-lapse microscopy," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 53, no. 11, pp. 2405–2414, Nov. 2006.
- [19] J. Cheng and J. C. Rajapakse, "Segmentation of clustered nuclei with shape markers and marking function," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 3, pp. 741–748, Mar. 2009.
- [20] V. Mitko, D. P. J. Van, K. Robert, A. Huisman, M. A. Viergever, J. P. W. Pluim, and M. Konradin, "Automatic nuclei segmentation in H&E stained breast cancer histopathology images," *PLoS ONE*, vol. 8, no. 7, 2013, Art. no. e70221.
- [21] T. K. Jin, S. M. Hewitt, X. Sheng, P. A. Pinto, and B. J. Wood, "Nucleus detection using gradient orientation information and linear least squares regression," *Proc. SPIE*, vol. 9420, Mar. 2015, Art. no. 94200N.
- [22] M. Liao, Y.-Q. Zhao, X.-H. Li, P.-S. Dai, X.-W. Xu, J.-K. Zhang, and B.-J. Zou, "Automatic segmentation for cell images based on bottleneck detection and ellipse fitting," *Neurocomputing*, vol. 173, pp. 615–622, Jan. 2016.
- [23] 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.
- [24] H. Oda, H. R. Roth, K. Chiba, J. Sokolić, T. Kitasaka, M. Oda, A. Hinoki, H. Uchida, J. A. Schnabel, and K. Mori, "BESNet: Boundary-enhanced segmentation of cells in histopathological images," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Berlin, Germany: Springer, 2018, pp. 228–236.
- [25] Y. Zhou, O. F. Onder, Q. Dou, E. Tsougenis, H. Chen, and P.-A. Heng, "Cia-net: Robust nuclei instance segmentation with contour-aware information aggregation," in *Proc. Int. Conf. Inf. Process. Med. Imag.* Hong Kong: Springer, 2019, pp. 682–693.
- [26] S. Graham, Q. D. Vu, S. E. A. Raza, A. Azam, Y. W. Tsang, J. T. Kwak, and N. Rajpoot, "Hover-net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images," *Med. Image Anal.*, vol. 58, Dec. 2019, Art. no. 101563.
- [27] N. A. Koohbanani, M. Jahanifar, A. Gooya, and N. M. Rajpoot, "Nuclear instance segmentation using a proposal-free spatially aware deep learning framework," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2019, pp. 622–630.
- [28] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2017, pp. 2961–2969.
- [29] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 91–99.
- [30] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2980–2988.
- [31] E. K. Wang, X. Zhang, L. Pan, C. Cheng, A. Dimitrakopoulou-Strauss, Y. Li, and N. Zhe, "Multi-path dilated residual network for nuclei segmentation and detection," *Cells*, vol. 8, no. 5, p. 499, May 2019.
- [32] J. Yi, P. Wu, Q. Huang, H. Qu, D. J. Hoepfner, and D. N. Metaxas, "Context-refined neural cell instance segmentation," in *Proc. IEEE 16th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2019, pp. 1028–1032.
- [33] J. Yi, P. Wu, Q. Huang, H. Qu, B. Liu, D. J. Hoepfner, and D. N. Metaxas, "Multi-scale cell instance segmentation with keypoint graph based bounding boxes," in *Proc. 22nd Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI)*, 2019, pp. 369–377.
- [34] S. Chen, C. Ding, and D. Tao, "Boundary-assisted region proposal networks for nucleus segmentation," 2020, *arXiv:2006.02695*. [Online]. Available: <http://arxiv.org/abs/2006.02695>
- [35] 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.
- [36] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2117–2125.
- [37] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Instance normalization: The missing ingredient for fast stylization," 2016, *arXiv:1607.08022*. [Online]. Available: <http://arxiv.org/abs/1607.08022>
- [38] J. C. Caicedo, A. Goodman, K. W. Karhohs, B. A. Cimini, J. Ackerman, M. Haghghi, C. Heng, T. Becker, M. Doan, C. McQuinn, M. Rohban, S. Singh, and A. E. Carpenter, "Publisher correction: Nucleus segmentation across imaging experiments: The 2018 data science bowl," *Nature Methods*, vol. 17, p. 241, Jan. 2020.



**ZHIMING CHENG** was born in Wuhan, Hubei, China, in 1994. He received the B.S. degree in electronic and information engineering from the Wuhan University of Engineering Science. He is currently pursuing the M.S. degree in software engineering with the University of South China. His research interests include the detection and segmentation of pathological organs in medical images.



**AIPING QU** (Member, IEEE) was born in Hunan, China, in 1982. He received the B.S. and M.S. degrees in mathematics science from Hunan University, in 2005 and 2010, respectively, and the Ph.D. degree in computer science from Wuhan University, in 2015. He is currently an Associate Professor of computer science with the University of South China. He has published over 20 papers in journals and conferences. His research interests include optimization, machine learning, and medical image analysis.

...