

Received March 26, 2022, accepted April 11, 2022, date of publication April 18, 2022, date of current version April 27, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3168721

# Semi-Supervised Nuclei Detection in Histopathology Images via Location-Aware Adversarial Image Reconstruction

CHENCHEN TIAN<sup>ID</sup>, LEI SU, ZHI WANG, AO LI<sup>ID</sup>, (Member, IEEE),  
AND MINGHUI WANG<sup>ID</sup>, (Member, IEEE)

School of Information Science and Technology, University of Science and Technology of China, Hefei 230027, China

Corresponding author: Minghui Wang (mhwang@ustc.edu.cn)

This work was supported by the National Natural Science Foundation of China under Grant 61871361, Grant 61971393, Grant 61471331, and Grant 61571414.

**ABSTRACT** Nuclei detection is a fundamental task for numerous downstream analysis of histopathology images. Usually, it requires a large number of labeled images for fully supervised nuclei detection to achieve optimal performance. However, the process of collecting sufficient and high-quality ground truth labels is extremely labor intensive. To alleviate this problem, in this paper, a novel semi-supervised learning framework is proposed for nuclei detection, which optimizes the detection network with the involvement of unlabeled image reconstruction. Specifically, we reconstruct unlabeled images from their detection maps representing detailed information about individual location of candidate nucleus, which will aid in regularizing the training process of the detection network by encouraging spatial consistency between original and reconstructed images. Moreover, to further facilitate image reconstruction, we adopt an adversarial learning scheme using image and instance level discriminators for the classification of original and reconstructed images. In this way, the capability of the detection network is successfully enhanced by taking advantage of both labeled and unlabeled images, thus leading to more accurate nuclei detection results. Extensive experiments show that we compare favorably with previous studies in various settings, which highlights the effectiveness of our proposed framework.

**INDEX TERMS** Nuclei detection, semi-supervised learning, histopathology image analysis.

## I. INTRODUCTION

Histopathology image analysis serves as the gold standard in the diagnosis of many diseases such as cancer [1]. Commonly, histopathology images are visualized with hematoxylin and eosin (H&E) stain, which can highlight the shape of nuclei [2], [3] and help pathologists to evaluate disease at the cellular level. It is well known that a single histopathological image may contain thousands of nuclei, and the histological characteristics of nuclei are quite critical in disease diagnosis, prognosis, and subsequent therapeutic approaches for patients [3]. Therefore, nuclei detection has become a core step in histopathological image analysis. Recent years have witnessed a growing interest in applying computational methods for systematic and objective analysis in histopathology images. These methods can relieve labor

intensity and enhance the efficiency since manual examinations require plenty of skills and experience of pathologists [4]. Furthermore, automatic detection and analysis of nuclei enable new perspectives to disease characterization, which cannot be gathered from manual assessments of tissue specimens.

Given the importance of cell-level information, researchers have been dedicated to exploring automatic methods to efficiently and accurately detect nuclei from histopathology images [5], [6]. Previous cell nuclei detection methods heavily depend on hand-crafted features, which have limited representation capabilities and tend to be sensitive to several changes such as cell morphology [7]. Recently, deep learning methods have attracted a great deal of interest in automatic nuclei detection. These methods typically use multi-layer convolutional neural network (CNN) that can automatically obtain discriminative feature representations for nuclei detection [8]–[10]. In comparison with previous

The associate editor coordinating the review of this manuscript and approving it for publication was Sudipta Roy<sup>ID</sup>.

methods, these deep learning based approaches have remarkable abilities to handle the variety of appearances of nuclei images and demonstrate significantly better detection results.

Despite the achievements of deep learning methods for nuclei detection, they usually require extensive amounts of labeled data to obtain satisfying performance [1], [3]. However, labeling nuclei is an extremely tedious and time-consuming job, as there are massive amounts of nuclei in each histopathology image. In consequence, it is very challenging to label every nucleus spreading over the images, which imposes high requirements on the involvement of expert pathologists [11]. Hence, there is an urgent need to develop semi-supervised learning methods for nuclei detection, which can exploit available unlabeled data to alleviate the need for effort-consuming labeling.

Recently, the value of reconstruction-based strategy has been recognized for semi-supervised learning approaches in various image analysis tasks [12], [13]. The key idea is to develop a framework attempting to effectively utilize unlabeled images during training by incorporating unsupervised reconstruction loss functions. For example, Chen *et al.* [12] adopt a semi-supervised method that performs unlabeled image reconstruction with combined reconstruction and segmentation loss for accurate medical image segmentation. In ultrasound image analysis, Zhang *et al.* [13] integrate the knowledge of unsupervised image reconstruction with the supervised lesion classification task, which shows high efficiency in improving breast diagnostic accuracy.

Inspired by the aforementioned studies, in this paper, a novel semi-supervised framework is proposed for nuclei detection in histopathology images. Aside from supervised training part with a portion of labeled images, we effectively incorporate unsupervised image reconstruction into the training of detection network as well. Importantly, as opposed to performing reconstruction directly from feature maps, we conduct image reconstruction from the pixel-wise detection results, i.e., detection maps. They are obtained to represent more detailed information on individual location of candidate nucleus. We hypothesize that this location-aware image reconstruction will aid in regularizing the training process of the detection network by encouraging spatial consistency between original and reconstructed images, and therefore successfully enhance the capability of the detection network. Besides, we also introduce an adversarial learning scheme to help minimize the difference of the original and reconstructed image on both image and instance (local nuclei regions) levels. In this way, our overall framework will predict more accurate results by simultaneously utilizing supervised and unsupervised knowledge for the nuclei detection task.

In general, we summarize the main contributions of this paper as follows: (i) we introduce an efficient semi-supervised learning framework based on location-aware image reconstruction, which enforces spatial consistency between original and reconstructed images to facilitate nuclei detection, (ii) we adopt an adversarial learning scheme using

image and instance level discriminators to further decrease the discrepancy between original and reconstructed images, (iii) experimental results on publicly available histopathology dataset demonstrate the proposed framework can achieve remarkable improvements for semi-supervised nuclei detection.

The organization of this paper is as follows. A review of previous works is given in Section II. The proposed semi-supervised nuclei detection framework is described in Section III. Experimental results and analysis are presented in Section IV. We discuss the potential future work and conclude the paper in Section V and VI.

## II. RELATED WORK

Automatic nuclei detection is the problem of determining the locations of nuclei without accurately delineating their boundaries [14]. Numerous research efforts [7], [15]–[20] have been dedicated to nuclei detection over the last years. Most of the early studies for nuclei detection involve customized feature extraction and morphological processing. The performances of these methods depend heavily on different manual features such as gradients, morphology and shapes [15]–[17]. With the advancements of feature learning, deep CNN has recently been employed for this problem to a large extent [7], [8], [10], [18]–[20]. Some of these methods take small image patches as input and conduct model inference in a sliding window manner. For example, Xu *et al.* [18] use a stacked sparse auto-encoder for high-level representation learning of patches from breast histopathology images in nuclei detection. Similarly, Xie *et al.* [19] introduce a structured regression CNN model (SR-CNN) to detect nuclei by predicting the probability of nucleus centroid. Sirinukunwattana *et al.* [8] further improve SR-CNN with a spatial-constrained layer to identify the position of nuclear center. Depending on patch-based inference, these methods may result in heavy computational costs for large histopathology images.

As an alternative approach, fully convolutional neural network (FCN) [21] and its variants have been successfully adopted for nuclei detection. They do not require fixed size of input images and can avoid repeated inference for overlapping patches to improve their efficiency. For example, Xie *et al.* [10] propose a FCN-based detection model to directly output probability maps with higher values near cell centers. Zhou *et al.* [7] develop a sibling FCN architecture for simultaneous nuclei detection and fine-grained classification. Li *et al.* [20] present a position of interest (POI) detection network by considering the information of nuclei positions, which has shown compelling performance for fully supervised nuclei detection.

Although great success has been obtained by the aforementioned deep learning methods, they heavily rely on large quantities of labeled data and in consequence require tremendous time and labor efforts for pathologists. Compared to fully supervised approaches, semi-supervised methods show better capabilities to reduce manual labeling efforts

by exploiting unlabeled data for improved detection performance. For example, for the task of signet ring cell detection, Li *et al.* [22] propose an efficient self-training (ST) approach by generating pseudo labels for unlabeled images and then re-training the detection models with these pseudo labels. Although promising nuclei detection results are obtained by this strategy, it may underestimate the side effects of pseudo labels. Given that the pseudo labels may be incorrectly predicted, the potential of ST will be limited for improvements from unlabeled images [12]. Instead of retraining the detection model with pseudo labels, our method leverages a location-aware reconstruction network that utilizes unlabeled images to facilitate the training process of our semi-supervised detection framework.

Recently, reconstruction-based strategy has received considerable attention for various image analysis tasks. In natural images, Lu *et al.* [23] perform image reconstruction as an auxiliary task during training for depth completion. For medical image analysis, Chen *et al.* [12] use a multi-task learning approach including a segmentation network and a reconstruction network for semi-supervised medical image segmentation. Zhang *et al.* [13] adopt a reconstruction network using unlabeled images for accurate diagnosis of breast cancer in a semi-supervised way. Hou *et al.* [24] use an unsupervised sparse autoencoder for nuclei detection and feature extraction with unlabeled images reconstruction and achieve favorable performance as well.

Compared with the aforementioned reconstruction-based methods, our method is technically different. First, instead of performing reconstruction from feature maps, we reconstruct the unlabeled images from full sized detection maps. These maps contain detailed spatial information of each candidate nucleus, which can guide our semi-supervised framework to efficiently learn spatial consistency, i.e., similarity of the spatial distribution of nuclei locations, between original and reconstructed images. Second, we further explore multi-level adversarial learning to facilitate image reconstruction. To be concrete, in addition to using a typical image-level discriminator to differentiate original images from the reconstructed ones as a whole [25], we also design an instance-level discriminator to focus on local nuclei regions with the guidance of candidate nuclei locations. Under this setting, the outputs of the reconstruction network are encouraged to be as close as possible to the original images both globally and locally, thus leading to better regularization for the learning process of the whole semi-supervised framework.

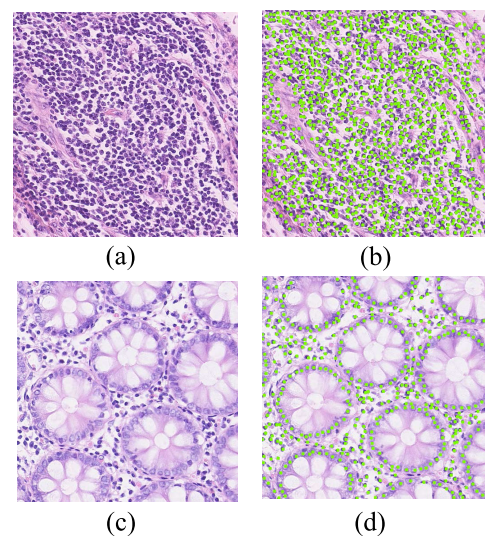
### III. MATERIALS AND METHODS

In this section, we first introduce the dataset used for nuclei detection and then we present the architecture of our proposed semi-supervised framework as well as the learning schemes in details.

#### A. DATASET

In this study, extensive experiments are conducted on a public cell nuclei dataset [8] to evaluate the performance of

the proposed framework. This dataset consists of 100 H&E stained histopathology images of colorectal adenocarcinomas, which are obtained from the Department of Computer Science, University of Warwick. Firstly, by using an Omnyx VL120 scanner, 10 whole-slide images from 9 patients of colorectal adenocarcinomas were collected at the resolution of  $0.55 \mu\text{m}/\text{pixel}$ . Then all images of this dataset were extracted from the non-overlapping areas of these whole-slide images with a common size of  $500 \times 500$  pixels. The selected areas contain artifacts and over-staining to represent real-world challenges. Finally, experienced pathologists were invited to conduct the manual annotation of nuclei in the 100 extracted images, and a total number of 29 756 nuclei were marked at the center for detection task. Visualization of histopathology images and their corresponding labels can be seen in Figure 1. For this dataset, we randomly split it into training, validation and testing sets at a ratio of 7:1:2 by following previous study [7]. That is to say, there are 70, 10 and 20 images in our training, validation and testing sets, respectively. Similar to [26], we keep testing set unchanged and randomly remove the labels of a portion of training data to validate the effectiveness of our semi-supervised framework.



**FIGURE 1. Visualization of histopathology images and their corresponding labels. Green dots in (b) and (d) represent the labels for nuclei detection.**

#### B. MODEL DESCRIPTION

The overall design of our semi-supervised framework is illustrated in Figure 2, which consists of a detection network, a reconstruction network and multi-level discriminators. Given a small number of labeled images with ground truth labels, we aim to further improve the performance of the detection network by taking advantage of unlabeled images without extra labor efforts. Specifically, our method allows both the labeled and unlabeled images being fed into a detection network, which can be divided into supervised and unsupervised learning parts respectively. For the supervised part, given a labeled image  $X^l$ , the FCN-based detection

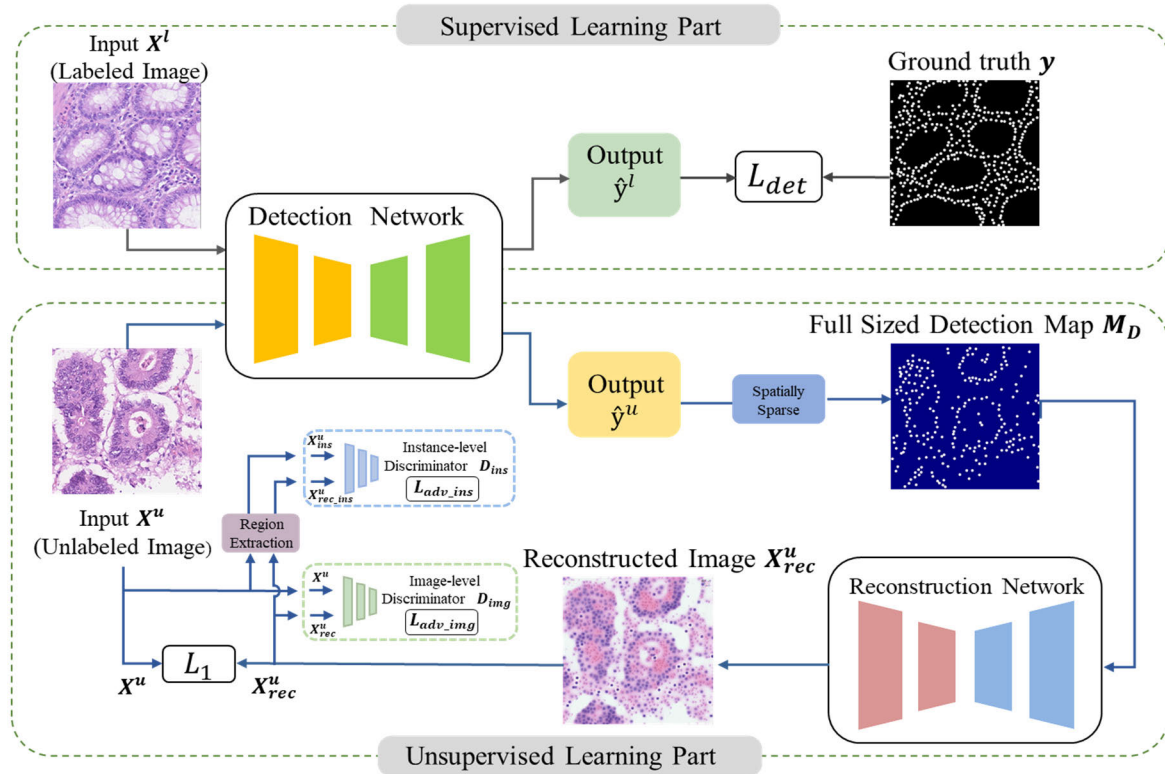


FIGURE 2. Overview of the proposed semi-supervised framework for nuclei detection.

network will output a probability map  $\hat{y}^l$  representing the nuclei locations. By using a detection loss  $L_{det}$ , we aim to minimize the difference of its output  $\hat{y}^l$  and the ground truth label  $y$  in a supervised way. For the unsupervised part, our method incorporates unlabeled images into the training process of the detection network as well. We first feed an unlabeled image  $X^u$  into the detection network to generate a probability map  $\hat{y}^u$ . Then  $\hat{y}^u$  is spatially sparsified to obtain a full sized detection map  $M_D$ , which indicates individual location of candidate nucleus in  $X^u$ . After that, the reconstruction network performs a location-aware reconstruction of  $X^u$  by generating  $X_{rec}^u$  from  $M_D$  by using  $L_1$  loss. In this way, our framework can be encouraged to learn spatial consistency between original and reconstructed images, which will aid in regularizing the training process of the detection network for improved detection performance. In addition, we further enhance image reconstruction with multi-level discriminators  $D_{img}$  and  $D_{ins}$  by using an image-level adversarial loss  $L_{adv\_img}$  and an instance-level adversarial loss  $L_{adv\_ins}$ .  $D_{img}$  is designed to differentiate  $X^u$  and  $X_{rec}^u$  as a whole, and  $D_{ins}$  focuses on extracted local nuclei regions in  $X_{rec}^u$  and  $X^u$ , denoted as  $X_{rec\_ins}^u$  and  $X_{ins}^u$  respectively. The details of our proposed method are given below.

### 1) NETWORK ARCHITECTURE

#### a: DETECTION NETWORK

Our detection network architecture is built on a widely applied neural network U-Net [27]. As a variant of FCN,

it consists of one down-sampling path that encodes the input to high-level features and one up-sampling path decoding these features to the output of the same size as the input. During the down-sampling path, we replace the conventional convolution connections in U-Net with residual learning [28], which helps to learn more powerful representations for nuclei detection. The down-sampling process contains three residual blocks and a  $2 \times 2$  max pooling layer is used for reducing half between each block. In each block, we utilize a stack of two  $3 \times 3$  convolutional layers followed by batch normalization (BN) and rectified linear unit (ReLU). In the up-sampling path, we also employ the residual connections instead of conventional convolution operations to benefit the optimization process of the detection network. The difference between the down-sampling and up-sampling path is that the max pooling layer is replaced with a deconvolution operation. Then we apply a  $1 \times 1$  convolutional layer followed by a softmax operation to obtain the probability map for nuclei detection.

#### b: RECONSTRUCTION NETWORK

As illustrated in Figure 3, we propose an encoder-decoder architecture with several inception blocks for image reconstruction. Given the scale variation of nuclei, the inception blocks are introduced for enough fusion of multi-scale features from different receptive fields to better learn representations of the input image. Similar to GoogleNet [29], each inception block involves four branches

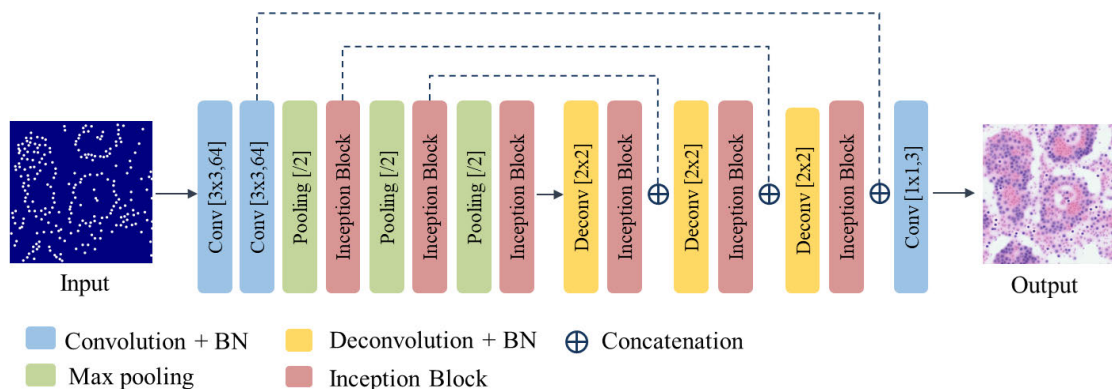


FIGURE 3. Overall flowchart of the proposed reconstruction network.

and each branch is composed of one or more convolutional layers of different kernel sizes, such as  $1 \times 1$  or  $3 \times 3$ . For the encoder part, the input image initially passes two  $3 \times 3$  convolutional layers with 64 filters followed by a BN and ReLU layer respectively, and then a  $2 \times 2$  max-pooling layer is used for down-sampling. After that, three inception blocks with pooling operations are applied to obtain high-level features. For the decoder part, we also utilize three inception blocks but with deconvolution operations to up-sample feature maps to a desired size. In addition, for the purpose of aggregating information from the hierarchy of feature maps, high-resolution information from the encoder part is copied and concatenated to corresponding up-sampled output for successive learning. Finally, after passing through a  $1 \times 1$  convolution layer and a sigmoid function, the network outputs a reconstructed image of the same size as the input generated from an unlabeled original image.

c: DISCRIMINATOR NETWORK

In order to further enhance the training process for image reconstruction, we adopt multi-level adversarial learning by an image-level discriminator  $D_{img}$  and an instance-level discriminator  $D_{ins}$ . Similar to [25], both of them contain five  $4 \times 4$  convolution layers with different number of filters including 64,128,256,512 and 1, respectively. And each convolutional layer is followed by a leaky ReLU [30] with a slope of 0.2 except the last one.

2) LOSS FUNCTION

a: DETECTION NETWORK

the goal of our detection network is to learn a mapping relationship between a histopathology image and its nuclei locations. We consider the scenario of using a labeled  $x^l$  when training the detection network in a supervised way. Formally, for every pixel  $i$  in  $x^l$ , the detection model outputs  $\hat{y}_i^l$  indicating its probability of being a nucleus, and the detection error

is calculated by binary cross-entropy as follows:

$$L_{det} = -\frac{1}{N} \sum_{i=0}^N (1 - y_i) \log(1 - \hat{y}_i^l) + y_i \log \hat{y}_i^l, \quad (1)$$

where  $y_i$  is the ground truth label of pixel  $i$ , with  $y_i = 0$  being the background and  $y_i = 1$  being the nucleus.  $N$  denotes the number of all pixels in the input image  $X^l$ . Note that the dataset only provides the coordinates of nuclei centroids, and by following [20] we employ a  $5 \times 5$  rectangular mask centered at each centroid as the ground truth label. We do not apply  $L_{det}$  for an unlabeled image  $X^u$  since there is no ground truth label.

b: RECONSTRUCTION NETWORK

In addition to the above supervised learning part, we propose to effectively feed unlabeled images into the training of the detection network via location-aware adversarial image reconstruction. It is noteworthy that if we reconstruct unlabeled images from the high-level feature maps extracted from the encoder of a detection network, it may be difficult to fully explore intrinsic spatial association between reconstructed and original images. This is more problematic for accurate detection of nuclei in histopathology images, as the reconstruction process may fail to focus on candidate nuclei locations bothered by noisy background, given the fact that nuclei are relatively small and often surrounded by massive background pixels.

To encourage the reconstruction network to produce output  $X_{rec}^u$  that is spatially consistent with the original image  $X^u$ , in this study we first feed  $X^u$  into the detection network and the output  $\hat{y}^u$  is then spatially sparsified to obtain a full sized detection map  $M_D$  indicating individual location of candidate nucleus in  $X^u$ , which is completed by using a  $9 \times 9$  max pooling operation with stride 1 and the ‘SAME’ padding scheme [31] to match the size of the  $M_D$  and  $\hat{y}^u$ . After that, the reconstruction network performs a location-aware reconstruction of  $X^u$  by generating  $X_{rec}^u$  from  $M_D$ . To match  $X^u$  and  $X_{rec}^u$ , the reconstruction network is trained using  $L_1$

loss defined as:

$$L_1 = \frac{1}{N} \|X_{rec}^u - X^u\|_1, \quad (2)$$

where  $N$  denotes the total number of pixels in  $X^u$  or  $X_{rec}^u$  as mentioned before.

Intuitively, by leveraging the full sized detection map  $M_D$ , our proposed reconstruction strategy is able to capture more detailed information for nuclei locations, and therefore provides spatial consistency constraint to the training of the detection network for improving the detection performance.

### c: DISCRIMINATOR NETWORK

To further help the reconstruction network to generate  $X_{rec}^u$  that is as close to  $X^u$  as possible, we employ multi-level adversarial losses with discriminators  $D_{img}$  and  $D_{ins}$ . We employ  $D_{img}$  to differentiate original images from the reconstructed ones as a whole, and adopt  $D_{ins}$  to focus on local nuclei regions in  $X_{rec}^u$  and  $X^u$ , denoted as  $X_{rec\_ins}^u$  and  $X_{ins}^u$  respectively, with the guidance of candidate nuclei locations. Specifically, we consider location of each candidate nucleus when the output of  $X^u$  or  $X_{rec}^u$  exceeds a predefined threshold  $T_p$ . Then  $X_{rec\_ins}^u$  and  $X_{ins}^u$  are obtained with max pooling operations on these locations with the size of  $p \times p$ , followed by element-wise multiplication of  $X_{rec}^u$  and  $X^u$  respectively. Finally,  $X_{rec\_ins}^u$  and  $X_{ins}^u$  are fed into  $D_{ins}$  for instance-level differentiation. Accordingly, the multi-level discriminator networks are trained by an image-level adversarial loss  $L_{adv\_img}$  and an instance-level adversarial loss  $L_{adv\_ins}$ , which are defined as:

$$\begin{aligned} L_{adv} &= L_{adv\_img} + L_{adv\_ins} \\ &= -(\log(1 - D_{img}(X^u)) + \log(D_{img}(X_{rec}^u))) \\ &\quad -\lambda(\log(1 - D_{ins}(X_{ins}^u)) + \log(D_{ins}(X_{rec\_ins}^u))), \end{aligned} \quad (3)$$

where  $\lambda$  leverages the importance of adversarial losses. We train both discriminator networks to minimize  $L_{adv}$ , whereas the reconstruction network is trained to fool the discriminators by maximizing  $L_{adv}$ .

### 3) OVERALL OBJECTIVE FOR OUR PROPOSED FRAMEWORK

Taken together, the overall loss of our framework can be written as:

$$L_{total} = L_{det} + \alpha L_1 + \beta L_{adv}, \quad (4)$$

where  $\alpha$  and  $\beta$  denote the weights of corresponding losses, respectively. We optimize our semi-supervised framework with the loss  $L_{total}$  for better detection accuracy. In this manner, our framework will enforce spatial consistency of candidate nuclei between original and reconstructed images, which can positively influence the parameter optimization for the detection network and vice versa. Collectively, our method introduces location-aware image reconstruction while keeping the detection network updated simultaneously to boost

nuclei detection accuracy. Moreover, the training process for image reconstruction is further enhanced via a multi-level adversarial learning scheme both globally and locally. Under such semi-supervised setting, the detection network can be trained not only on the well labeled images, but also on unlabeled images with the involvement of location-aware adversarial image reconstruction, leading to improved detection performance.

### C. TRAINING DETAILS

During training, we adopt data augmentation such as rotation, horizontal and vertical flipping to reduce overfitting [20]. The framework is implemented using PyTorch under Linux with CPU Intel Xeon 4110 @ 2.10GHz and GPU NVIDIA GeForce RTX 2080 Ti. For detection network, we update its parameters using Adam strategy [32] with initial learning rate 0.01 and momentum {0.5, 0.999}, where the learning rate is decayed to 0.001 at 30 epochs and 0.0001 for the remaining 40 epochs. For reconstruction network and two discriminators, we trained them by the Adam optimizer with the learning rate of 0.1 and  $10^{-5}$ , respectively. We set the hyperparameter values as:  $\lambda = 0.1$  in (3) and  $\alpha = 0.1, \beta = 0.005$  in (4). For nuclei local regions in  $D_{ins}$ , the pooling window value  $p$  is 11, and the threshold  $T_p$  is 0.85. Similar to [33], to prevent the reconstruction network suffering from initial noisy detection maps, we pretrain the detection network with labeled images before training the overall semi-supervised framework, and then update both the supervised and unsupervised part jointly to boost the detection performance.

## IV. EXPERIMENTAL RESULTS

In this section, detailed experiments are carried out to assess the proposed method and investigate its effectiveness in different settings.

### A. EVALUATION CRITERION

To assess the detection performance of our method, common evaluation criteria are adopted, such as precision (P), recall (R) and F1 score (F1) [8]. They are defined as:

$$P = \frac{TP}{TP + FP}, \quad (5)$$

$$R = \frac{TP}{TP + FN}, \quad (6)$$

$$F1 = \frac{2PR}{P + R}, \quad (7)$$

where TP, FP, FN denote the number of true positives, false positives and false negatives, respectively. Similar to [7], we conduct non-maximum suppression (NMS) on the output of our detection network to obtain final detection result. By following [20], if a detected location lies inside a region of 6 pixels around a nuclear center of the label, it is considered to be TP. All detected locations outside these regions are viewed as FP and the ones are not matched by any of these regions are FN.

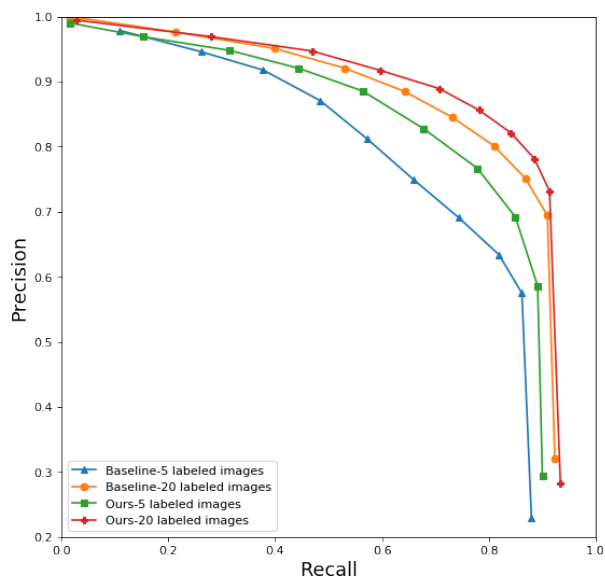
**B. EFFECTIVENESS OF THE PROPOSED SEMI-SUPERVISED METHOD**

To evaluate the effectiveness of our method, we train the detection network in a purely supervised way only with labeled images to serve as the baseline of our framework. To explore how the baseline behaves when the number of labeled images changes, we compare F1 scores of the cases where 5, 10, 20 and 40 labeled images are used respectively. As reported in Table 1, using only a small number of labeled images indicates poor detection results. As the number of labeled images increases from 5 to 40, the overall detection performances become better, which is reasonable since the detection error decreases with more labels.

**TABLE 1. F1 score of the baseline and our method with different numbers of labeled images.**

Method	Labeled images			
	5	10	20	40
Baseline	0.716	0.763	0.805	0.832
Ours	0.773	0.796	0.828	0.845

In contrast to the purely supervised baseline, our semi-supervised framework introduces obvious improvements on F1 scores in each case with different number of labeled images. In particular, for cases with less labeled images, our method outperforms the baseline by a larger margin. For example, with only 5 labeled images, the performance is improved from 0.716 to 0.773, which confirms the effectiveness of our semi-supervised method in lack of labeled images. As shown in Figure 4, the precision-recall curve is generated by using different threshold values for the detection



**FIGURE 4. Precision-recall curve for nuclei detection.**

output. It can be observed that the curve of our method is always closer to the upper-right corner compared to the baseline while using 5 and 20 labeled images, respectively. That is to say, under the same number of labeled images, the precision of our method is always higher than the baseline given the same recall value. Meanwhile, given the same precision value, the recall of our method is still higher than the baseline. Besides, while using 5 labeled images, the area under the precision-recall curve of our method is larger than that of the baseline, which also exists in the case of 20 labeled images. Taken together, it can be observed that our method consistently outperforms the baseline with different number of labeled images.

**C. COMPARISON TO PREVIOUS WORK**

We compare the proposed method against previous studies for nuclei detection. For purely supervised learning, we introduce the widely applied U-Net [27] and the start-of-the-art POI [20], which are trained only on labeled images. Besides, the unlabeled images are also utilized for improved performance by semi-supervised self-training (ST) method [22]. For a fair comparison, the ST algorithm is implemented with the same architecture as our detection network described in Section III. In detail, the pseudo labels of nuclei locations we use to retrain the detection network are of high confidence, that is to say, their predicted probabilities need to exceed a predefined value  $\theta$  to alleviate noises of large background regions to a certain extent [34]. We assess the performance between our method and ST with two different values of  $\theta$ , i.e., 0.8 and 0.9, for cases with 5,10,20,40 labeled images.

As listed in Table 2, compared with U-Net, POI yields better detection results in a purely supervised way. While for semi-supervised learning, our method can improve the detection precision and recall substantially, leading to a higher F1 score and detection accuracy in contrast to POI in all cases. Meanwhile, to achieve comparable performance with the supervised methods, our method requires a smaller number of labeled images in the training set. For example, with only 5 labeled images our method can acquire higher detection accuracy than that of POI using twice the number of labeled images. With 10 labeled images, our method (0.796) can achieve 95.6% of the performance of POI (0.833) on F1 score using 40 labeled images. These results demonstrate that our method can take advantage of unlabeled images and requires less labeled images for training to achieve a promising result.

We show the detection results using different values of  $\theta$  for ST in Table 2, and it can be observed that  $\theta = 0.9$  yields slightly better results than  $\theta = 0.8$ , accordingly we will refer to ST with  $\theta = 0.9$  simply as ST in the next. It can be seen that our method achieves consistently higher efficiency across different numbers of labeled images. For example, the precision, recall and F1 score of ST are 0.756, 0.737 and 0.746 using 5 labeled images. In comparison, our method successfully brings 1.1, 4.2 and 2.7 points of improvements respectively. When there are 10 and 20 labeled images used, our method can further improve the F1 score by 1.6 and 1.1 points,

**TABLE 2.** Comparison with previous work using different numbers of labeled images.

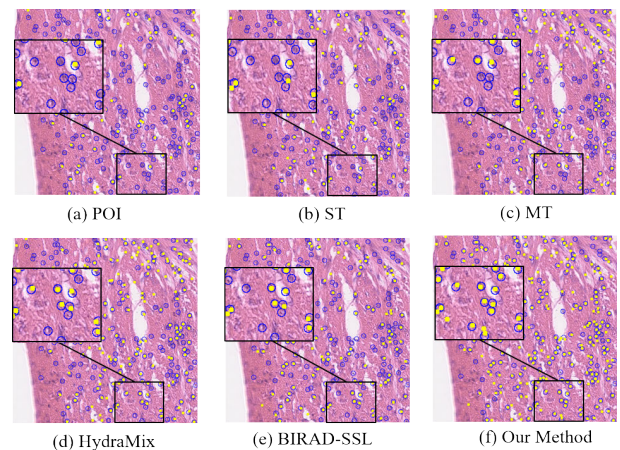
Method	Labeled images											
	5			10			20			40		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1
U-Net	0.676	0.611	0.659	0.676	0.729	0.702	0.729	0.776	0.752	0.803	0.786	0.796
POI	0.718	0.708	0.713	0.769	0.760	0.764	0.805	0.810	0.808	0.828	0.838	0.833
ST ( $\theta = 0.8$ )	0.755	0.733	0.744	0.767	0.784	0.776	0.819	0.810	0.814	0.832	0.846	0.839
ST ( $\theta = 0.9$ )	0.756	0.737	0.746	0.773	0.787	0.780	0.819	0.816	0.817	0.831	0.850	0.840
MT	0.754	0.748	0.751	0.791	0.765	0.778	0.823	0.815	0.819	0.829	0.847	0.838
HydraMix	0.742	0.771	0.756	0.808	0.760	0.783	0.827	0.805	0.816	0.843	0.835	0.839
BIRAD-SSL	0.730	0.775	0.752	0.776	0.782	0.779	0.811	0.827	0.819	0.826	0.849	0.837
Ours	0.767	0.779	0.773	0.795	0.798	0.796	0.825	0.830	0.828	0.830	0.860	0.845

respectively. When the number of labeled images further increases to 40, the unlabeled images may provide limited benefits in improving performance, as the ratio of labeled images in training set is relatively high. However, our method still manages to yield slightly better results than ST, which along with the other results, suggests the effectiveness of our method.

In addition, we further compare the proposed method against other state-of-the-art semi-supervised learning methods including Mean Teacher (MT) [35] and HydraMix [36]. Besides, we also compare with the other recent reconstruction-based semi-supervised image classification method BIRAD-SSL [13] to further validate the effectiveness of our method. To make the comparisons fair, we reproduce this method by using the same architecture as our detection network described in Section III. Table 2 displays that our method compares favorably against these methods, demonstrating the strength of our method for semi-supervised nuclei detection. Taken together, we can conclude that compared to the aforementioned methods, our method can make better use of unlabeled images especially when labeled images are scarce. The underlying reason is that, our framework encourages spatial consistency between original and reconstructed images via location-aware adversarial reconstruction, which in turn positively influences the parameter optimization for the detection network with enhanced performance. We further provide some typical results of detection for these approaches and our method using 5 labeled images in Figure 5, and the magnified regions show that our method successfully detects more nuclei that cannot be well tackled by other approaches.

#### D. ABLATION STUDY

In this section, we present an ablation study to highlight the contribution of the proposed location-aware reconstruction



**FIGURE 5.** Typical results of nuclei detection with sample test images, where yellow dots are detected centers of the nuclei and blue circles represent ground-truth areas. The magnified regions enclosed with a black box clearly show that our method successfully detects more nuclei that cannot be well tackled by other approaches.

and multi-level adversarial learning. In order to provide additional insight into their individual performance and their combined effectiveness, we conduct the ablation experiments with 5 labeled images and discuss the results in detail next.

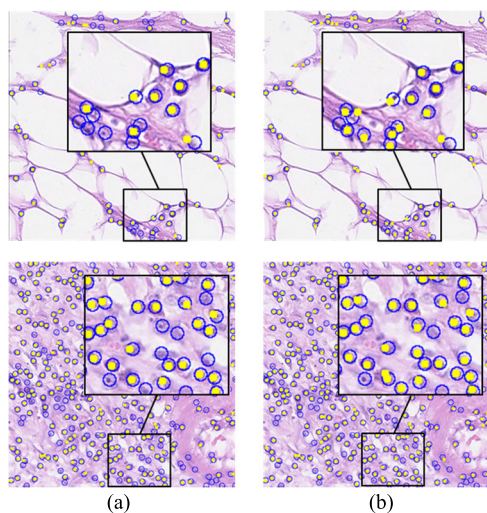
We confirm the effectiveness of the components in the proposed method and the results are summarized in Table 3, where “Baseline” refers to a detection network trained only using the labeled images and is abbreviated as “B”. Based on it, “B + Rec” denotes the incorporation of the unlabeled image reconstruction but without multi-level adversarial learning. Similarly, this pipeline adding the image-level discriminator  $D_{img}$  is denoted as “B + Rec +  $D_{img}$ ”. It is observed that “B + Rec” delivers a significant and consistent improvement over the baseline on both precision and recall, and obtains a remarkably higher F1 score of 0.763, which



demonstrates the effectiveness of our location-aware image reconstruction. Moreover, from the results in Table 3 we can see that adversarial learning brings additional benefit to detection performance. Especially, by incorporating both  $D_{img}$  and  $D_{ins}$ , we further achieve 2.1 and 1.0 points of improvements in recall and F1 score over the “B + Rec” respectively, which confirms that combining image-level and instance-level discriminators is beneficial for detection task. In Figure 6, we also provide some visualizations of detection results with and without the adversarial learning of our method, and the magnified regions further highlight the effect of using adversarial learning for the proposed method.

**TABLE 3. Ablation study on the components of our framework.**

Method	P	R	F1
Baseline(B)	0.691	0.743	0.716
B + Rec	0.767	0.758	0.763
B + Rec+ $D_{img}$	0.753	0.784	0.768
Ours	0.767	0.779	0.773



**FIGURE 6. Example results for nuclei detection. (a) represents the detection result of our method without adversarial learning; (b) reflects an image detected using our method with adversarial learning.**

We also evaluate the effect of using feature maps as the input of the reconstruction network based on the same architecture described in Section III. Following previous study [12], the high-level feature map is extracted from the encoder of the detection network that has a smaller size of  $62 \times 62$  than the input image. In comparison, we achieve a performance boost of 2.0 points by using a full sized detection map as illustrated in Table 4, which verifies the benefits of our location-aware reconstruction for nuclei detection. The

**TABLE 4. Ablation study on the input of reconstruction network.**

Input	P	R	F1
Feature Map	0.748	0.758	0.753
Detection Map	0.767	0.779	0.773

**TABLE 5. Performance when choosing different size of local nuclei regions.**

$p$	7	11	15	19	23
F1	0.764	<b>0.773</b>	0.769	0.765	0.760

**TABLE 6. Performance when choosing different  $T_p$ .**

$T_p$	0.75	0.80	0.85	0.90	0.95
F1	0.763	0.767	<b>0.773</b>	0.766	0.762

reason is that the detection map we used contains more detailed information for nuclei locations and thus our detection framework is developed to encourage spatial consistency of candidate nuclei between original and reconstructed images, giving rise to higher detection accuracy.

Further, we investigate the sensitivity of the hyperparameters of the pooling window value  $p$  and the threshold  $T_p$  used in  $D_{ins}$ . Table 5 shows the detection results with different settings of  $p$ , and it can be observed that when  $p = 11$ , our method achieves the best performance on F1 score for nuclei detection. Smaller size of local nuclei region leads to poorer performance, which may be caused by insufficient information for instance-level classification. Besides, as the value is larger than 11, it shows no substantial benefit and a too large region results in decreased performance probably due to the background noises. We also discuss the effect of the threshold  $T_p$  in Table 6, and our method reaches the best F1 score when  $T_p = 0.85$ .

## V. DISCUSSION

Automatic nuclei detection is critical for several histopathology image analysis algorithms. However, obtaining the ground truth for nuclei detection is extremely labor intensive. To overcome this challenge, we develop a novel semi-supervised framework by taking unlabeled image reconstruction into account. As nuclei are often sparse in a histopathology image and surrounded by massive background pixels, we propose to leverage the detection map for image reconstruction and update the detection network simultaneously. The advantage is that the reconstruction

process only needs to particularly focus on candidate nuclei locations without being bothered by noisy background. Accordingly, our framework successfully enforces spatial consistency of candidate nuclei between unlabeled original and reconstructed images, which can positively influence the parameter optimization for the detection network. In this manner, our method is capable of simultaneously utilizing both labeled and unlabeled images for improved nuclei detection performance.

Although our method achieves promising nuclei detection results with limited labeled images, there is still room for further improvement. In terms of network architecture, we employ the detection network with residual learning for combination of features from different layers, and it can be further extended by more sophisticated network components such as self-attention mechanism [37], which can capture rich contextual relationships for better feature representations of nuclei detection. In future work, one possible direction is to explore the usage of our proposed framework in other related image analysis tasks such as nuclei segmentation. In addition, although our FCN-based framework has already reduced the computational cost to a certain extent compared with patch-based nuclei detection methods [18], [19]. We realize that further improvements in time efficiency are needed to enhance the practicality of our method. The latest multi-focus fusion schemes such as [38], [39] have achieved promising results in reducing time complexity, which can be incorporated into the proposed framework in future work to enhance the robustness and efficiency for nuclei detection.

## VI. CONCLUSION

In this paper, a novel location-aware adversarial image reconstruction method is proposed for semi-supervised nuclei detection in histology images to deal with insufficient labeled data. Aside from supervised training part with a portion of labeled images, we effectively incorporate unlabeled image reconstruction into the training of detection network as well. Furthermore, we facilitate the training process for image reconstruction via multi-level adversarial learning. In this way, our framework is developed to encourage spatial consistency between original and reconstructed images for enhancing the capability of the detection network. Experimental results suggest that the proposed framework helps to enhance the performance significantly, which demonstrates the effectiveness of location-aware adversarial image reconstruction in semi-supervised nuclei detection.

## REFERENCES

- [1] B. Hu, Y. Tang, E. I.-C. Chang, Y. Fan, M. Lai, and Y. Xu, "Unsupervised learning for cell-level visual representation in histopathology images with generative adversarial networks," *IEEE J. Biomed. Health Informat.*, vol. 23, no. 3, pp. 1316–1328, May 2019.
- [2] J. Adur, M. Bianchi, V. B. Pelegati, S. Viale, M. F. Izaguirre, H. F. Carvalho, C. L. Cesar, and V. H. Casco, "Colon adenocarcinoma diagnosis in human samples by multicontrast nonlinear optical microscopy of hematoxylin and eosin stained histological sections," *J. Cancer Therapy*, vol. 5, no. 13, pp. 1259–1269, 2014.
- [3] H. Qu, P. Wu, Q. Huang, J. Yi, Z. Yan, K. Li, G. M. Riedlinger, S. De, S. Zhang, and D. N. Metaxas, "Weakly supervised deep nuclei segmentation using partial points annotation in histopathology images," *IEEE Trans. Med. Imag.*, vol. 39, no. 11, pp. 3655–3666, Nov. 2020.
- [4] H. Yang, J.-Y. Kim, H. Kim, and S. P. Adhikari, "Guided soft attention network for classification of breast cancer histopathology images," *IEEE Trans. Med. Imag.*, vol. 39, no. 5, pp. 1306–1315, May 2020.
- [5] H. Chen, Q. Dou, X. Wang, J. Qin, and P. A. Heng, "Mitosis detection in breast cancer histology images via deep cascaded networks," in *Proc. AAAI Conf. Artif. Intell.*, Phoenix, AZ, USA, 2016, pp. 1160–1166.
- [6] J. Xu, L. Xiang, Q. Liu, H. Gilmore, J. Wu, J. Tang, and A. Madabhushi, "Stacked sparse autoencoder (SSAE) for nuclei detection on breast cancer histopathology images," *IEEE Trans. Med. Imag.*, vol. 35, no. 1, pp. 119–130, Jan. 2016.
- [7] Y. Zhou, Q. Dou, H. Chen, J. Qin, and P. A. Heng, "SFCN-OPI: Detection and fine-grained classification of nuclei using sibling FCN with objectness prior interaction," in *Proc. AAAI Conf. Artif. Intell.*, New Orleans, LA, USA, 2018, pp. 2652–2659.
- [8] K. Sirinukunwattana, S. E. A. Raza, Y.-W. Tsang, D. R. J. Snead, I. A. Cree, and N. M. Rajpoot, "Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1196–1206, May 2016.
- [9] A. A. Cruz-Roa, J. E. A. Ovalle, A. Madabhushi, and F. A. G. Osorio, "A deep learning architecture for image representation, visual interpretability and automated basal-cell carcinoma cancer detection," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, Nagoya, Japan, 2013, pp. 403–410.
- [10] Y. Xie, F. Xing, X. Shi, X. Kong, H. Su, and L. Yang, "Efficient and robust cell detection: A structured regression approach," *Med. Image Anal.*, vol. 44, pp. 245–254, Feb. 2018.
- [11] K. Pasupa, S. Tungjitnob, and S. Vatathanavaro, "Semi-supervised learning with deep convolutional generative adversarial networks for canine red blood cells morphology classification," *Multimedia Tools Appl.*, vol. 79, nos. 45–46, pp. 34209–34226, Dec. 2020.
- [12] S. Chen, G. Bortsova, A. G.-U. Juárez, G. van Tulder, and M. de Bruijne, "Multi-task attention-based semi-supervised learning for medical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, Shenzhen, China, 2019, pp. 457–465.
- [13] E. Zhang, S. Seiler, M. Chen, W. Lu, and X. Gu, "BIRADS features-oriented semi-supervised deep learning for breast ultrasound computer-aided diagnosis," *Phys. Med. Biol.*, vol. 65, no. 12, Jun. 2020, Art. no. 125005.
- [14] 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.
- [15] S. Ali and A. Madabhushi, "An integrated region-, boundary-, shape-based active contour for multiple object overlap resolution in histological imagery," *IEEE Trans. Med. Imag.*, vol. 31, no. 7, pp. 60–1448, Jul. 2012.
- [16] 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.
- [17] M. Veta, J. P. W. Pluim, P. J. van Diest, and M. A. Viergever, "Breast cancer histopathology image analysis: A review," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 5, pp. 11–1400, May 2014.
- [18] J. Xu, L. Xiang, R. Hang, and J. Wu, "Stacked sparse autoencoder (SSAE) based framework for nuclei patch classification on breast cancer histopathology," in *Proc. IEEE 11th Int. Symp. Biomed. Imag. (ISBI)*, Beijing, China, Apr. 2014, pp. 999–1002.
- [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.*, Munich, Germany, 2015, pp. 358–365.
- [20] X. Li, W. Li, and R. Tao, "Staged detection-identification framework for cell nuclei in histopathology images," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 1, pp. 183–193, Jan. 2020.
- [21] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Boston, MA, USA, Jun. 2015, pp. 3431–3440.
- [22] J. Li, S. Yang, X. Huang, Q. Da, X. Yang, Z. Hu, Q. Duan, C. Wang, and H. Li, "Signet ring cell detection with a semi-supervised learning framework," in *Proc. Int. Conf. Inf. Process. Med. Imag.*, Hong Kong, 2019, pp. 842–854.

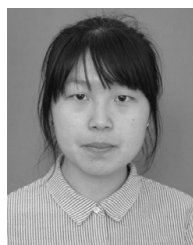
- [23] K. Lu, N. Barnes, S. Anwar, and L. Zheng, "From depth what can you see? Depth completion via auxiliary image reconstruction," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Seattle, WA, USA, Jun. 2020, pp. 11303–11312.
- [24] L. Hou, V. Nguyen, A. B. Kanevsky, D. Samaras, T. M. Kurc, T. Zhao, R. R. Gupta, Y. Gao, W. Chen, D. Foran, and J. H. Saltz, "Sparse autoencoder for unsupervised nucleus detection and representation in histopathology images," *Pattern Recognit.*, vol. 86, pp. 188–200, Feb. 2019.
- [25] T.-H. Vu, H. Jain, M. Bucher, M. Cord, and P. Perez, "ADVENT: Adversarial entropy minimization for domain adaptation in semantic segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Long Beach, CA, USA, Jun. 2019, pp. 2517–2526.
- [26] P. Haeusser, A. Mordvintsev, and D. Cremers, "Learning by association—A versatile semi-supervised training method for neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 626–635.
- [27] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Med. Image Comput. Comput.-Assisted Intervent.*, Munich, Germany, 2015, pp. 234–241.
- [28] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 770–778.
- [29] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Boston, MA, USA, May 2015, pp. 1–9.
- [30] B. Xu, N. Wang, T. Chen, and M. Li, "Empirical evaluation of rectified activations in convolutional network," 2015, *arXiv:1505.00853*.
- [31] M. Tofighi, T. Guo, J. K. P. Vanamala, and V. Monga, "Prior information guided regularized deep learning for cell nucleus detection," *IEEE Trans. Med. Imag.*, vol. 38, no. 9, pp. 2047–2058, Sep. 2019.
- [32] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Represent.*, San Diego, CA, USA, 2015, pp. 1–15.
- [33] W.-C. Hung, Y.-H. Tsai, Y.-T. Liou, Y.-Y. Lin, and M.-H. Yang, "Adversarial learning for semi-supervised semantic segmentation," in *Proc. Brit. Mach. Vis. Conf. (BMVC)*, Newcastle, U.K., 2018, pp. 1–17.
- [34] Q. Xie, M.-T. Luong, E. Hovy, and Q. V. Le, "Self-training with noisy Student improves ImageNet classification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Seattle, WA, USA, Jun. 2020, pp. 10684–10695.
- [35] A. Tarvainen and H. Valpola, "Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results," in *Proc. Adv. Neural Inf. Process. Syst.*, Long Beach, CA, USA, 2017, pp. 1195–1204.
- [36] R. M. S. Bashir, T. Qaiser, S. E. A. Raza, and N. M. Rajpoot, "HydraMix-Net: A deep multi-task semi-supervised learning approach for cell detection and classification," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, Lima, Peru, 2020, pp. 164–171.
- [37] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, Long Beach, CA, USA, 2017, pp. 5998–6008.
- [38] R. Lai, Y. Li, and J. Guan, "Multi-scale visual attention deep convolutional neural network for multi-focus image fusion," *IEEE Access*, vol. 7, pp. 114385–114399, 2019.
- [39] S. Liu, J. Ma, L. Yin, H. Li, S. Cong, X. Ma, and S. Hu, "Multi-focus color image fusion algorithm based on super-resolution reconstruction and focused area detection," *IEEE Access*, vol. 8, pp. 90760–90778, 2020.



**CHENCHEN TIAN** received the B.S. degree in electronic science and technology from the School of Information Science and Technology, University of Science and Technology of China (USTC), in 2019, where she is currently pursuing the M.S. degree. Her research interests include deep learning and medical imaging.



**LEI SU** received the B.S. degree in intelligent science and technology from Xidian University, China, in 2017. She is currently pursuing the Ph.D. degree with the University of Science and Technology of China (USTC). She is a member of the Center for Biomedical Engineering. Her research interests include deep learning, bioinformatics, and medical image.



**ZHI WANG** received the B.S. degree in electronic science and technology from the School of Information Science and Technology, University of Science and Technology of China (USTC), in 2018, where she is currently pursuing the Ph.D. degree. She is a member of the Center for Biomedical Engineering. Her research interests include deep learning and medical image.



**AO LI** (Member, IEEE) received the B.S. degree in biophysics from the School of Life Science, University of Science and Technology of China (USTC), in 2000, and the Ph.D. degree in biomedical engineering from the School of Information Science and Technology, USTC, in 2005. He is currently an Associate Professor with the School of Information Science and Technology and the Center for Biomedical Engineering, USTC. His research interests include biomedical information processing and artificial intelligence.



**MINGHUI WANG** (Member, IEEE) received the B.S. degree from the School of Gifted Youth, University of Science and Technology of China (USTC), and the Ph.D. degree in biomedical engineering from the School of Information Science and Technology, USTC, in 2006. She is currently an Associate Professor with the School of Information Science and Technology and the Center for Biomedical Engineering, USTC. Her research interests include bioinformatics, biomedical informatics, and artificial intelligence in medicine.

...