IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 13 May 2024, accepted 26 May 2024, date of publication 30 May 2024, date of current version 6 June 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3407066

RESEARCH ARTICLE

Enhancing Cervical Cell Detection Through Weakly Supervised Learning With Local Distillation Mechanism

JUANJUAN YIN[1], QIAN ZHANG[1], XINYI XI[1], MENGHAO LIU[1], WENJING LU[1], AND HUIJUAN TU[2]

[1]School of Information Science and Technology, Northwest University, Xi'an 710069, China
[2]Department of Radiology, Kunshan Hospital of Chinese Medicine, Suzhou 215399, China

Corresponding author: Huijuan Tu (20214132014@stu.suda.edu.cn)

This work was supported by the Kunshan City traditional Chinese medicine (TCM) science and technology development special fund (KZYY202302), and Key research and development projects of Kunshan Ministry of science and technology (KS1946).

ABSTRACT Cervical cancer is a malignancy that significantly impacts women's health. Liquid-based thin-layer cytology examination is presently the predominant method for cervical cancer cell detection. Traditional identification of pathological images of cervical cells mainly relies on professional physicians, which is time-consuming, labor-intensive, and has considerable limitations. The integration of deep learning with imaging showcases remarkable performance in medical-assisted diagnosis. Nevertheless, conventional fully supervised detection techniques face challenges in acquiring comprehensive annotated data samples. Moreover, the intricate cell categories within cervical cells present complexities, especially in small object detection. To address the aforementioned issues, we propose a weakly supervised model for cervical cell detection, named LD-WSCCD, based on a local distillation mechanism. First, our model extracts image features using single shot multibox detector (SSD). Then, leveraging the concept of knowledge distillation, a local distillation mechanism is designed to segregate foreground and complex background regions, directing the student network to concentrate on crucial pixels and channels. Finally, the detection of cervical cells is performed utilizing a multi-instance detector. Experimental results on a publicly accessible cervical cell dataset validate the effectiveness of our approach, boasting a mean average precision (mAP) value of 73.6%, surpassing other similar detection models. In future research, we aim to establish a comprehensive dataset of cervical pathological cells. Our focus is on enhancing the model's detection accuracy at the target boundary to effectively address the challenge of overlapping adhesive cells in cervical samples. Our goal is to achieve a well-balanced trade-off between the model's accuracy and speed.

INDEX TERMS Object detection, knowledge distillation, cervical cancer, weak supervision.

I. INTRODUCTION

Cervical cancer, also known as cervical cancer, represents a substantial threat to women's health globally, ranking as the second most commonly diagnosed cancer among females worldwide [1]. The diagnostic process for cervical cancer entails the meticulous analysis of pathological images containing myriad cells to accurately identify abnormalities and diseased cells. However, the inherent subjectivity and potential for errors in judgment stemming from factors such as physician experience and expertise underscore the pressing need for more precise and standardized diagnostic methodologies. The increasing production of pathological images, fueled by the implementation of national cancer screening initiatives, presents a formidable challenge that traditional manual assessment alone cannot effectively surmount. In recent years, the advent of artificial intelligence, harnessing advanced neural networks and visual algorithms, has emerged as a promising asset in the realm of medical image analysis. These cutting-edge technologies empower

The associate editor coordinating the review of this manuscript and approving it for publication was Mario Donato Marino.

the thorough examination and scrutiny of cervical pathological cells, playing a pivotal role in ensuring effective cancer screenings for women. By augmenting diagnostic accuracy and efficiency, these technological advancements significantly contribute to the prevention, early detection, and treatment of malignant tumors, especially in the context of cervical cancer. The deployment of artificial intelligence tools not only enhances the capabilities of healthcare professionals in analyzing complex pathological images but also holds the promise of revolutionizing the landscape of cancer diagnosis and treatment.

Traditional approaches to cervical cancer detection focus on segmenting and classifying cells within pathological images to facilitate diagnosis. Current studies, such as [2], [3], and [4], explore various feature selection techniques and utilize machine learning algorithms like support vector machines (SVM) and Adaboost [5] to enhance the identification and classification of cell targets in cervical pathology images. However, acknowledging the constraints of traditional methodologies and the intricacies involved in detecting cervical pathological cells, recent research has shifted towards leveraging deep learning techniques for increased detection accuracy. Cutting-edge methods, such as the optimized YOLOv3 model [6], have been tailored to detect abnormal cell targets, showcasing notable enhancements in detection precision. Moreover, advancements in multi-instance learning networks, exemplified in the research by Pal et al. [7], have further elevated the detection accuracy of cervical pathological cell images. These developments underscore the ongoing evolution in diagnostic technologies, paving the way for more efficient and effective detection and classification of cervical cancer markers through the integration of sophisticated deep learning methodologies.

Current challenges in cervical cancer detection arise from the labor-intensive and error-prone manual annotation processes essential for model training. In response to these challenges, researchers are actively exploring weakly supervised learning methods as a more efficient and resource-saving alternative for training models in medical image analysis. By decreasing the reliance on precisely annotated data, weak supervision strategies exhibit promise in alleviating the adverse effects of noisy labels on detection performance. This study aims to advance the field of cervical cancer detection by investigating the application of weakly supervised learning methods to enhance the accuracy and efficiency of identifying cervical pathological cells. Through the utilization of cutting-edge artificial intelligence techniques, this research endeavors to amplify current diagnostic capabilities and elevate the overall effectiveness of cancer screenings in combating cervical cancer. The primary contributions of our method are as follows:

- Leveraging a local distillation mechanism, we propose a novel weakly supervised cervical cell detection model named LD-WSCCD, designed to detect cervical cell categories and their respective positions.
- Our local distillation mechanism effectively segregates foreground and intricate background regions, directing student networks to concentrate on crucial pixels and channels, thus maximizing the utilization of detailed information from local features.
- Experimental results on the publicly accessible cervical cell dataset validate the efficacy of our approach, showcasing a model mAP value of 73.6%, surpassing the performance of comparable detection networks.

TABLE 1. Related work.

Classification	Model
Two-stage detection model	R-CNN SS [8] SVM [9] SPP-Net [10] Fast R-CNN [11] VGG-16 [12] Faster R-CNN [13]
Single stage detection model	Yolo series: Yolov1 [14]; Yolov2 [15]; Yolov3 [16]; Yolov4 [17] SSD series: SSD [18]; RSSD [19]; DSSD [20]; FSSD [21]; DSOD [22] RetinaNet [23] CornerNet [24] CenterNet [25] EfficientDet [26]

Our research primarily focuses on analyzing cervical cell images by leveraging deep neural networks to develop object detection models that classify and localize different types of cells in these images. This work is structured into five chapters to address this task comprehensively. The first section serves as the introduction, elaborating on the research background and significance. section II discusses related work, introducing the current research status of cervical cell detection. section III details the methodology, providing a comprehensive description of our proposed LD-WSCCD model. section IV conducts experiments, showcasing a series of comparative and ablation experiments. Lastly, in section V, the conclusion provides a summary of our work and prospects for future research endeavors.

II. RELATED WORK

In the domain of computer vision, object detection continues to be a topic of paramount interest. The aim of object detection is to accurately determine the positional coordinates of each target within a complex image, distinguishing them from the background, and categorizing the target accordingly. In recent years, the advancement of deep convolutional neural networks (CNN) has significantly improved the effectiveness of object detection. Utilizing deep learning object detection algorithms allows for superior feature extraction and the transformation of original image data into abstract semantic information through network models. This semantic information performs exceptionally well in complex real-world scenarios, notably enhancing the accuracy of object detection. Currently, deep learning-based object detection algorithms are primarily classified into two categories: two-stage object

detection models and single-stage object detection models. The related research is presented in Table 1.

The two-stage detection algorithm treated object detection as a classification problem. The R-CNN model proposed by Girshick et al. [27] made a significant breakthrough in object detection, and subsequent work in object detection drew substantial inspiration from this approach. The fundamental process of R-CNN involved initially identifying 2000 candidate boxes through selective search (SS) [8]; these candidate boxes were then standardized in size and subjected to feature extraction using AlexNet. Subsequently, SVM [9] was employed for classification, and non-maximum suppression was applied during filtration to obtain candidate boxes. Following fine-tuning, the ultimate target box was derived. While R-CNN exhibited notable performance improvements over previous algorithms, it incurred substantial computational overhead in obtaining candidate regions via SS, leading to redundant feature computations and slower model training speeds. The SPP-Net target pricing model introduced by He et al. [10] eliminated the need for selecting candidate regions, opting instead to feed images directly into the convolutional network to mitigate computational complexity. Through the utilization of spatial pyramid pooling layers, the model achieved image size normalization, addressing discrepancies in input sizes and reducing computational overhead, consequently greatly enhancing detection speed. In response to the spatial complexity of SPP-Net, Girshick [11] proposed the Fast R-CNN object detection algorithm, amalgamating the strengths of SPP-Net with enhancements to R-CNN. Notably, VGG-16 [12] replaced AlexNet as the backbone network, the SVM classifier was substituted with a softmax classifier, and a multitasking mode was integrated, collectively enhancing network performance and detection speed. Following the introduction of Fast R-CNN, Ren et al. [13] formulated the Faster R-CNN object detection framework to address the challenges associated with generating candidate regions. This algorithm introduced the region proposal network (RPN) to supplant the SS for candidate box generation. By applying a sliding window operation at each point, CNN was directly employed for network feature extraction, resulting in substantial performance enhancements. Nonetheless, Faster R-CNN exhibited limitations in detecting small targets.

In comparison to two-stage detection algorithms, single-stage detection algorithms processed multiple tasks on a single network and provided detection results directly, achieving an end-to-end solution mode. The detection speed of the models was significantly enhanced. Single-stage detection algorithms, also referred to as regression-based detection algorithms, could sacrifice some accuracy to improve speed. The YOLO series comprised a collection of object detection networks rooted in single-stage detection, including YOLOv1 [14], YOLOv2 [15], YOLOv3 [16], and YOLOv4 [17]. These networks divided the image into multiple grids, each predicting target boxes and category probability scores. The primary advantage of the YOLO series was its rapid speed, making it suitable for real-time applications. YOLOv1 introduced by Joseph Redmon et al. in 2015 merged classification and regression tasks into a single CNN, enhancing the algorithm's speed by eliminating the step of generating candidate boxes. However, the grid-based approach led to insufficient accuracy in detecting small targets. YOLOv2 introduced a normalization layer, DarkNet19 as the backbone network, multi-scale training, fine-grained features, and an anchor box mechanism to enhance performance. YOLOv3 employed DarkNet53 as the backbone network, multi-scale prediction, and 9 anchor boxes to improve accuracy while maintaining real-time performance. YOLOv4 integrated various research techniques, enhanced the backbone network, and introduced CutMix, Mosaic, DropBlock regularization, Mish activation function, and other strategies for a balanced trade-off between accuracy and speed. Moreover, single shot multibox detector (SSD) [18] also excelled in single-stage object detection tasks, addressing speed and accuracy issues. Various modifications such as RSSD [19], DSSD [20], FSSD [21], DSOD [22], RetinaNet [23], CornerNet [24], CenterNet [25], EfficientDet [26], and others further improved single-stage detection performance in the field.

Distillation was a concept commonly used in the field of chemistry. Since Hinton et al. [28] introduced this concept to deep learning in 2015, knowledge distillation had been successfully applied in image classification tasks. Faced with increasingly large network structures, the aim of knowledge distillation was to transfer knowledge from large-scale teacher network models to shallow student models to enhance the performance of the shallow networks. Serving as an effective solution for model compression, knowledge distillation could harness rich information from large teacher networks to direct new small-scale student models, thereby conserving resources. Knowledge distillation based on transfer learning had facilitated mutual learning between cross-domain data and the disentanglement of models and knowledge. During the training process, a large teacher network could be perceived as a "black box" and could also safeguard sensitive data. As a potent technology for compressing and expediting deep neural networks, knowledge distillation had found widespread application across various artificial intelligence domains, encompassing visual recognition [29], speech recognition [30], natural language processing (NLP) [31], and recommendation systems. Additionally, knowledge distillation had utility in other contexts such as data privacy [32] and defense against adversarial attacks [33]. In weakly supervised object detection tasks, Zeng et al. [34] had advanced a novel framework, WSOD2, incorporating object distillation. This framework integrated adaptive linear combinations to jointly evaluate the objectivity and CNN confidence from bottom-up (BU) and top-down (TD) sources, derived from low-level measurements, to ascertain multiple regression targets. Zeni and Jung [35] had proposed an additional refinement step known as refinement knowledge distillation, designed to enhance the accuracy of the detector.

In the task of cervical pathological cell detection, as each pathological image may contain thousands of cells, the labeling of cancerous and diseased cells is often incomplete. The dataset comprises numerous noisy labels, and the labeled bounding boxes may enclose single or multiple cells that require identification. Leveraging this dataset and integrating insights from prior studies, we introduce a weakly supervised cervical cell detection model named LD-WSCCD, based on a local distillation mechanism. This mechanism separates foreground from complex background regions and directs student networks to concentrate on critical pixels and channels, effectively leveraging detailed information from local features. Treating the cervical cell detection task as a weakly supervised object detection challenge, inspired by the OICR approach [36], we devise a multi-instance detector, namely the MIL detector, to accomplish the detection of cervical cell categories and positions.

III. METHOD

A. DATASETS

We use the cervical cell dataset [37], which contains a total of 7086 images, including 6667 images in the training set and 419 images in the testing set. These object examples are labeled by experienced pathologists and divided into 11 cell categories, namely: ASC-US (ascus), ASC-H (asch), low-grade squamous intraepithelial lesions (lsil), high-grade squamous intraepithelial lesions (hsil), squamous cell carcinoma (scc), atypical glandular cells (agc), trichomonas (tric), candida (cand), bacterial flora (flora), herpes (herps), and actinomycetes (actin).

B. IMPLEMENTATION DETAILS

We have completed the development on the Ubuntu 16.04 system, using Python 3.6 language to build the environment and Nvidia RTX 3080 GPU to accelerate model training. We use Pytorch to build a deep network model in our experiment, and the input cervical cell image size was $512 \times 512 \times 3$, the epoch was set to 120, the batch size was set to 10, the optimizer weight attenuation coefficient was set to 0.0005, and the initial learning rate was set to 0.0001.

C. EVALUATING INDICATOR

For the detection task of cervical pathological cells, the precursor task is the classification task, which aims to determine whether a pathological image contains target cells. For such binary classification tasks, four basic scenarios are defined in machine learning: true positive (TP), false positive (FP), true negative (TN), and false negative (FN). In theory, the sum of these four scenarios should be equal to the total number of samples in the test case.

1) PRECISION

In the predicted positive cervical pathological images, the proportion of actual labels also being positive can be expressed mathematically as follows:

$$P = \frac{TP}{TP + FP} \tag{1}$$

2) RECALL

Allow the model to have a small number of false positives, which are actually negative but mistakenly judged as positive by the model. However, for positive samples, which are samples from positive patients, the model should strive to ensure that these samples are correctly classified and avoid incorrect predictions that could lead patients to miss the optimal treatment window. The mathematical expressions for this scenario are as follows:

$$R = \frac{TP}{TP + FN} \tag{2}$$

3) PR CURVE

The precision-recall (PR) curve illustrates the trade-off between precision and recall. The precision-recall relationship is contradictory, as higher precision typically corresponds to lower recall, and vice versa. In the task of cervical cell detection, a higher precision rate is prioritized over the recall rate due to the direct impact of detection results on the treatment timeline for positive patients. Early detection and treatment are recommended to ensure timely intervention.

4) AP AND MAP

Furthermore, we utilize the average precision (AP) and mean average precision (mAP) metrics to assess the detection capabilities of the model concerning cervical cells. AP serves as a holistic metric that considers the accuracy across varying recall rates, enabling the evaluation of the model's detection performance across different target categories. The computation of AP involves calculating the area under the PR curve. The mean average precision, denoted as mAP, is determined by averaging the AP values across all categories, expressed as $\text{mAP} = \frac{1}{K}\Sigma AP(K)$. Typically, mAP serves as the primary metric for evaluating the detection performance.

5) IOU

intersection over union (IoU) represents the similarity between the predicted box and the true box of the detection model, which is the ratio of the overlapping area of these two regions to the total area of the two. The mathematical expression is as follows:

$$IoU = \frac{Area(PreBox) \cap Area(GTBox)}{Area(PreBox) \cup Area(GTBox)} \tag{3}$$

If the predicted box completely overlaps with the true box, then IoU=1; On the contrary, IoU=0. Therefore, the value of IoU is within the range of [0,1]. In object detection tasks, if the IoU between the predicted box and the true box of the model is greater than a certain threshold, it can indicate

that the task model has successfully detected. Generally, this threshold is 0.5.

D. LD-WSCCD MODEL

We introduce a novel weakly supervised object detection model aimed at tackling the task of cervical cell detection. The LD-WSCCD model's overall architectural design is depicted in Figure 1. The data is inputted into the base network SSD, where multi-scale features are fused through SSD convolutions, and rich feature information within cervical cell images is extracted through local distillation feature loss and additional processes. Subsequently, feature vectors for detection and classification flows are generated via two fully connected layers, and then processed using MIL detectors to derive the classification and localization outcomes of cervical cells.

1) SSD FEATURE EXTRACTION NETWORK

SSD detects targets in images by utilizing multiple feature maps of different scales as inputs, merging prior boxes and confidence predictions. SSD exhibits the following advantages:

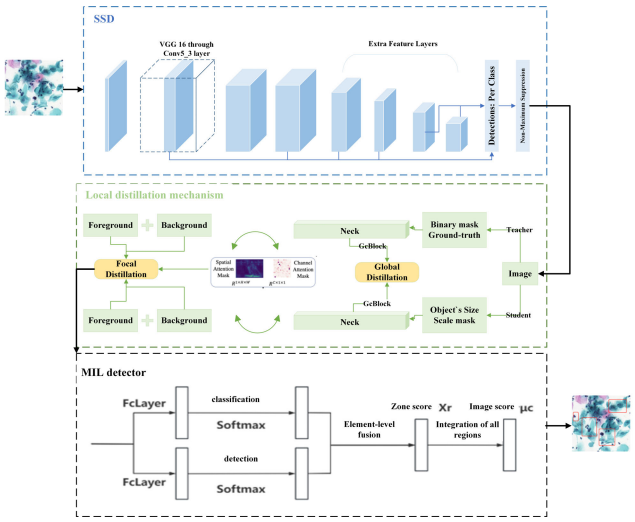

FIGURE 1. LD-WSCCD model structure.

① Fast performance: In comparison to algorithms like Faster R-CNN, SSD demonstrates quicker detection speeds. SSD can conduct image detection in a single pass, whereas other algorithms necessitate multiple scans of the image, significantly reducing detection time. Consequently, SSDs are well-suited for applications demanding high real-time performance.

② High accuracy: SSD achieves enhanced accuracy relative to other models. This is attributed to SSD's utilization of multi-scale feature maps and prior boxes for target detection. These features facilitate a more precise capture of detailed target information in the image, enhancing detection accuracy.

③ Strong flexibility: SSD can accommodate targets of diverse sizes by employing a multi-scale feature map strategy. This approach enables the network to better adjust to various target sizes and shapes.

④ No requirement for candidate region extraction: SSD eliminates the need for extracting candidate regions from images, leading to a notable reduction in computational costs and enhanced detection speed. Moreover, SSDs can minimize missed detections as they encompass the entire image for detection.

The SSD utilizes VGG16 as the base feature extraction layer, with a notable feature of VGG being the use of several consecutive small convolution kernels in place of larger ones. For instance, by substituting a 5×5 convolution kernel with two 3×3 small convolution kernels, an activation function is applied after each convolution operation. This approach not only reduces the model's parameters but also facilitates more effective extraction and enlargement of the model's receptive field, enabling the extraction of more robust features. VGG16 comprises 5 convolutional layers and 3 fully connected (FC) layers, with the FC layers primarily utilized for classification. Given that the base network is responsible for feature map extraction, the SSD replaces the FC6 and FC7 layers within VGG16 with convolutional layers. Consequently, all dropout and FC8 layers are omitted, while new convolutional layers 6, 7, 8, and 9 are introduced. The parameters for this modification and the convolutional layer parameters of VCG16 are acquired through transfer learning.

a: EXTRA FEATURE LAYERS

In comparison to the prior Faster R-CNN and YOLO series networks, we propose a multi-scale feature fusion structure for feature extraction. Feature extraction occurs at various scales, leveraging feature maps of varying sizes for detection purposes. Simultaneously, softmax classification and position regression are executed across multiple feature maps to capture global information and enhance detection accuracy. Semantic information characteristic of each layer is embedded in the feature map obtained post each convolution operation. Semantic richness increases with higher feature layers. Diverse feature maps denote information utilization at different hierarchical levels, and the integration of multi-scale features inherently boosts detection outcomes. Moreover, as convolution transitions from shallow to deep layers, the receptive field expands from small to large, advocating for the advantages of multi-scale features for multi-scale object detection.

b: ANCHOR

In contrast to two-stage networks like Faster R-CNN, which predetermine candidate boxes and conduct classification tasks prior to detection tasks, SSD operates as a one-stage network and does not predefine candidate boxes. Hence, we introduced the concept of Anchors. Before training, a set of prior boxes/default boxes is established for each grid, serving as reference points for iteratively adjusting to the actual target location through variations in displacement and aspect ratio. Subsequently, the true target location is

determined through softmax classification and bounding box regression. The rules for defining the prior boxes include:

Given the objective of predicting m feature maps, the formula for determining the scale of the prior box for each feature map is as follows:

$$S_k = S_{min} + \frac{S_{max} - S_{min}}{m-1}(k-1), k \in (1, m) \tag{4}$$

where S_{min} is 0.2 and S_{max} is 0.9, indicating that the scale ranges from 0.2 at the lower level to 0.9 at the higher level, with intermediate scales spaced according to specified rules. Computing the width ($w_k^a = s_k\sqrt{a_r}$) and height ($h_k^a = s_k\sqrt{a_r}$) of a particular layer's prior box involves setting different ratios. For a ratio of 1, a specific prior box is added with a scale of $s_k^{'} = \sqrt{s_k s_{k+1}}$. The center position for each prior box is set to $(\frac{i+0.5}{|f_k|}, \frac{j+0.5}{|f_k|})$, where f_k denotes the size of the k-th feature map.

c: MULTIBOXLOSS

To locate multiple object categories, $x_{ij}^p = 1$ is used to indicate that the i-th prior box matches the j-th ground truth box of category p; otherwise, $x_{ij}^p = 0$. Based on this matching strategy, we obtain $\sum_i x_{ij}^p \geq 1$, indicating multiple prior boxes match the j-th ground truth box. The overall objective loss function is the weighted sum of localization loss (loc) and confidence loss ($conf$).

$$L(x, c, l, g) = \frac{1}{N}\left(L_{conf}(x, c) + \alpha L_{loc}(x, l, g)\right) \tag{5}$$

where N represents the number of matched prior boxes. When the matched prior box value is 0, the loss value is set to 0. The localization loss (loc) is the smooth $L1$ loss between the predicted box (l) and the ground truth box (g). The regression offset between the center (cx, cy) of a default bounding box d and its width (w) and height (h) is calculated. The mathematical expressions are as follows:

$$\begin{aligned} L_{loc}(x, l, g) &= \sum_{i \in Pos}^{N} \sum_{m \in (cx, c\ y, w, h\}} x_{ij}^k \,\mathrm{smooth}_{\mathrm{L1}}\left(l_i^m - g_j^m\right) \\ g_j^{cx} &= \frac{g_j^{cx} - d_i^{cx}}{d_i^w g_j^{cy}} = \frac{g_j^{cy} - d_i^{cy}}{d_i^h} \\ g_j^w &= \log\left(\frac{g_j^w}{d_i^w}\right) g_j^h = \log\left(\frac{g_j^h}{d_i^h}\right) \end{aligned} \tag{6}$$

Confidence loss (($conf$) is a softmax loss based on multi-class confidence. The mathematical expressions are as follows:

$$L_{conf}(x, c) = -\sum_{i \in Pos}^{N} x_{ij}^p - \sum_{i \in Neg} \log\left(\hat{c}_i^0\right), \hat{c}_i^p = \frac{\exp\left(c_i^p\right)}{\sum_p \exp\left(c_i^p\right)} \tag{7}$$

2) LOCAL DISTILLATION MECHANISM

Knowledge distillation is a model compression method that preserves the network model structure. By amplifying the dark knowledge, the student network learns to replicate the dark knowledge of the teacher network, facilitating transfer learning. Common detectors predominantly leverage FPN [38] for integrating multi-scale semantic information. FPN combines features extracted from various levels of the backbone network, enhancing the student network's utilization of multi-scale information and boosting its performance significantly. In a standard SSD, multi-scale feature fusion is employed to derive feature maps from multiple network layers. Although these feature maps are used for softmax classification and position regression to capture global information simultaneously, the shallow layers may lack adequate feature extraction due to a limited number of convolutional layers. Transfer of feature knowledge from teacher networks can substantially improve student performance, with the feature distillation formula being expressed as follows:

$$L_{fea} = \frac{1}{CHW}\sum_{k=1}^{C}\sum_{i=1}^{H}\sum_{J=1}^{W}\left(F_{k,i,j}^T - f\left(F_{k,i,j}^S\right)\right)^2 \tag{8}$$

where F^T represents the teacher's features, F^S represents the student's features, and f reshapes the teacher feature F^T into an adaptation layer of the student feature F^S in the same dimension. Here, H and W denote the height and width of the feature map, while C denotes the number of channels. Conventional feature distillation tends to overlook the inter-pixel correlations. To address this issue, we introduce a local distillation mechanism that emphasizes the distinction between foreground and background images. This mechanism guides the student networks to concentrate on crucial pixels and channels by leveraging insights from the teacher networks.

The research conducted on SENet [39] and CMAM [40] demonstrated that emphasizing key pixel regions and employing special channel attention mechanisms during the network learning process benefited the network in learning features effectively. This approach also contributed to the enhanced performance of CNN models. In their work, Zagoruyko and Komodakis [41] improved the efficacy of knowledge distillation by incorporating a spatial attention mask mechanism. Building upon the aforementioned research findings, we apply similar methods to give special attention to local pixels and channels, extracting corresponding attention masks.

In cervical pathological cell images, the difference in appearance between the detected cells and normal cells is not significant, and there are no prominent distinctions between the foreground and background areas of target detection. Additionally, the collected images lack clarity, resulting in a more complex image background. To tackle this issue, a binary mask M is used to separate the foreground and background regions. If the horizontal and vertical coordinates (i, j) of the feature map lie within the ground truth box,

$M_{i,j}$=1; otherwise, it is set to 0.

$$M_{i,j} = \begin{cases} 1, if(i,j) \in r \\ 0, O\ therwise \end{cases} \tag{9}$$

Then, we calculate the absolute average of pixels and channels:

$$G^S(F) = \frac{1}{C} \cdot \sum_{c=1}^{C} |F_c|$$
$$G^C(F) = \frac{1}{HW} \cdot \sum_{i=1}^{H} \sum_{j=1}^{W} |F_{i,j}| \tag{10}$$

where H, W, and C represent the height, width, and number of channels of the feature map, respectively. G^S and G^C represent spatial and channel attention maps, respectively. By introducing the temperature hyperparameter T to regulate the distribution, the spatial attention mask A^S and channel attention mask A^C are calculated:

$$A^S(F) = H \cdot W \cdot \text{softmax}(G^S(F)/T)$$
$$A^C(F) = C \cdot softmax(G^C(F)/T) \tag{11}$$

During the training process, there exists a disparity between the student's mask and the teacher's mask, necessitating the utilization of the teacher's mask for guiding the student's mask. Utilizing the spatial attention mask A^S and channel attention mask A^C of the teacher detector, the feature loss calculation formula is as follows:

$$L_{fea} = \alpha \sum_{k}^{C} \sum_{i}^{H} \sum_{j}^{W} M_{i,j} A^S_{i,j} A^C_k \left(F^T_{k,i,j} - f\left(F^S_{k,i,j}\right)\right)^2 + \beta \sum_{k}^{C} \sum_{i}^{H} \sum_{j}^{W} \left(1 - M_{i,j}\right) A^S_{i,j} A^C_k \left(F^T_{k,i,j} - f\left(F^S_{k,i,j}\right)\right)^2 \tag{12}$$

where F^T and F^S represent the feature maps of the teacher detector and the student detector, respectively, we introduce hyperparameters α and β to balance the loss between foreground and background. To facilitate the successful learning of the spatial and channel attention masks in the teacher detector by the student detector, an attention loss function L_{att} is incorporated:

$$L_{att} = \gamma \cdot \left(l\left(A^S_t, A^S_S\right) + l\left(A^C_t, A^C_S\right)\right) \tag{13}$$

where l denotes the average absolute error function, and γ is employed to balance the two losses. Utilizing the aforementioned calculation results, the final loss function L_{dis} is derived as follows:

$$L_{dis} = L_{fea} + L_{att} \tag{14}$$

Train the student detector using the loss function L_{dis}, guide the student detector in learning from the teacher's network, and implement a local distillation mechanism.

3) MIL DETECTOR

In the task of cervical pathological cell detection, as there may be thousands of cells in each pathological image, the labeling of cancerous and diseased cells is often incomplete. The dataset contains numerous noisy labels, and the labeled bounding boxes may encompass single or multiple cells that require detection. Treating the cervical cell detection task as a weakly supervised object detection problem, a multi-instance detector, *i.e.*, MIL detector, is designed based on the concept of OICR. This MIL detector is utilized to perform the detection of cervical cell categories and their respective positions.

The features extracted from the convolutional layer are separated into a detection flow and a classification flow through two fully connected layers. These flows are processed by softmax layers to produce two corresponding matrix scores, denoted as $\sigma(x^{det})$ and $\sigma(x^{cls})$. By multiplying these two matrices, the predicted score $x_r = \sigma(x^{det}) \cdot \sigma(x^{cls})$ for the r-th region in the proposal box is generated. The predicted score $\mu_c = \sum_{r=1}^{r} x_{cr}$ for the entire image is determined by aggregating the predicted scores across all regions. During the training phase, the fundamental multi-instance detector is supervised by the predicted scoreμ_c and the ground truth y_c of the class annotations. This supervision guides the calculation of the cross-entropy loss L_{cls}.

$$L_{cls} = -\sum_{C=1}^{C} \{y_c \log \mu_c + (1 - y_c)(1 - \log \mu_c)\} \tag{15}$$

Due to the varying sizes of cells in cervical images, basic multi-instance detectors might exhibit a bias towards larger cells, potentially leading to the localization of entire image content by multiple detectors. Building upon the concept of OICR, we have developed a staged optimization multi-instance classifier to enhance the performance of detectors. Each optimization stage comprises a fully connected layer and a softmax layer. The recommendation scores from the i-th stage optimizer serve as the supervision signal for the subsequent $i + 1$ layer, encompassing $C + 1$ categories including a background class. The aggregated output from k optimization stages is utilized as the supervised information for the feature distillation process. The supervision information y^k_{cr} guides the optimization output at the $k - 1$ stage for the r-th region and category c, yielding the recommendation score x^k_{cr}. Furthermore, to mitigate noise stemming from previous predictions, a weighted term $w^k_j = \chi^{k-1}_{Cjc}$ is introduced to refine the loss function throughout the iterative optimization stages. The network is trained to minimize the weighted cumulative optimization loss L_{ref} over the stepwise optimization iterations.

$$L_{ref} = -\frac{1}{|R|} \sum_{j=1}^{|R|} \sum_{c=1}^{C+1} w^k_r y^k_{cr} \log x^k_{cr} \tag{16}$$

The complete model undergoes training while being guided by the distillation loss L_{dis}, the basic multi-instance detector loss L_{cls}, and the optimization loss L_{ref}.

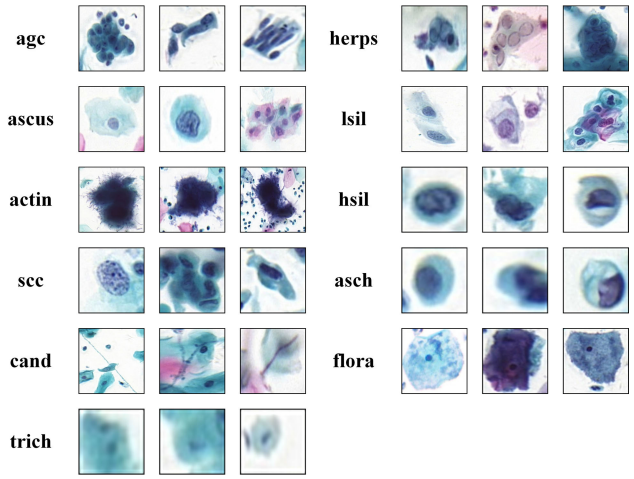

FIGURE 2. Detection visualization.

TABLE 2. Experimental results of LD-WSCCD on different cells.

LD-WSCCD	AP value	Precision	Recall	IoU	mAP value
lsil	78.4%	77.8%	78.6%	78.2%	
hsil	74.6%	74.8%	74.9%	74.7%	
scc	57.9%	58.1%	57.2%	56.9%	
agc	54.4%	53.9%	54.1%	53.8%	
asch	70.9%	71.2%	70.3%	71.1%	
ascus	73.2%	72.9%	72.7%	73.5%	73.6%
tric	76.5%	77.1%	76.9%	76.2%	
cand	80.1%	79.8%	80.2%	79.8%	
flora	83.3%	83.1%	82.6%	83.2%	
herps	78.7%	79.3%	78.5%	78.8%	
lsil	81.6%	81.5%	82.1%	81.2%	

IV. EXPERIMENTS

A. COMPARATIVE EXPERIMENT

As depicted in Table 2, cells exhibiting low-grade squamous intraepithelial lesions (lsil) showcase larger nuclei compared to those without lesions, achieving a commendable AP value of 78.4%. High-grade squamous intraepithelial lesions (hsil) present more pronounced abnormalities, characterized by a higher nuclear to cytoplasmic ratio than typical cell structures, yielding an AP value of 74.6%. Squamous cell carcinoma (scc) displays multiple cell adhesions, resulting in a relatively lower detection performance with an AP of 57.9%. Atypical glandular cells (agc), primarily comprising small targets, exhibit inferior model performance compared to larger cells, with an AP of 54.4%. ASC-US (ascus) and ASC-H (asch) cells boast clear cell image boundaries and are sizeable targets that are easily distinguishable. The model demonstrates good detection accuracy for these cells, with AP values of 73.2% and 70.9%, respectively. For other cell types not necessitating special attention in cervical cancer detection, notable distinctions exist among the cells. The model excels in detecting these cell types, with AP values of 76.5% for trichomonas (tric), 80.1% for candida (cand), 83.3% for bacterial flora (flora), 78.7% for herpes (herps), and 81.6% for actinomycetes (actin). The overall model's mAP for detection performance stands at 73.6%, with the detection efficacy illustrated in Figure 2.

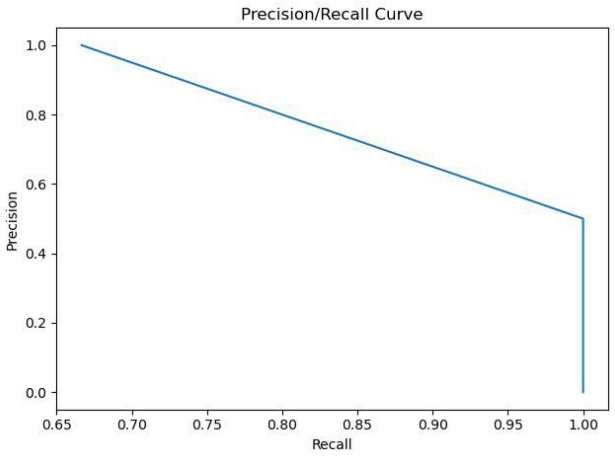

FIGURE 3. Precision and recall relation curve.

TABLE 3. Comparison of the results of different models for five types of key cells.

Model	Cell name	AP value	Precision	Recall	IoU	mAP value
Faster R-CNN	lsil	69.6%	68.4%	69.2%	69.7%	56.9%
	hsil	67.8%	65.9%	66.5%	67.1%	
	scc	44.9%	46.2%	45.9%	45.7%	
	agc	37.7%	38.6%	37.5%	36.9%	
	hsil	64.4%	65.1%	63.5%	67.1%	
SSD	lsil	70.3%	70.1%	70.2%	71.3%	59.8%
	hsil	69.7%	68.5%	67.4%	68.6%	
	scc	48.8%	48.6%	47.5%	49.2%	
	agc	49.6%	49.7%	48.6%	48.9%	
	hsil	60.8%	60.1%	61.2%	60.3%	
YOLOv4	lsil	72.1%	72.6%	71.8%	73.1%	60.8%
	hsil	68.6%	67.5%	69.3%	68.4%	
	scc	52.7%	52.1%	53.4%	53.2%	
	agc	45.4%	44.9%	45.3%	46.1%	
	hsil	65.4%	66.2%	65.1%	65.8%	
LD-WSCCD	lsil	78.4%	77.8%	78.6%	78.2%	67.2%
	hsil	74.6%	74.8%	74.9%	74.7%	
	scc	57.9%	58.1%	57.2%	56.9%	
	agc	54.4%	53.9%	54.1%	53.8%	
	asch	70.9%	71.2%	70.3%	71.1%	

Our LD-WSCCD model is trained and utilized for predictions on the cervical cell dataset. To visually represent the outcomes of the model training, Precision-Recall (PR) curves for each cell category are plotted, as depicted in Figure 3. The x-axis of the PR curves represents recall, while the y-axis represents precision.

To compare the performance of the LD-WSCCD model, we conduct experiments utilizing mainstream object detection network models on the same dataset partition and within the same experimental environment. The AP values of five key cell types are compared across each model, along with their corresponding mAP values, as illustrated in Table 3.

Based on the analysis of the experimental results mentioned above, the following conclusions can be drawn:

(1) In comparison to the Faster R-CNN model, our model exhibits a 10.3% increase in mAP value. When contrasted with the two-stage detection model, our single-stage detection model consolidates multiple tasks into a unified process, adopting an end-to-end solution approach that enhances both the speed and accuracy of detection.

(2) Our model demonstrates a 7.4% increase in mAP value over the SSD model. Similarly, when compared to

the YOLOv4 model, our model achieves a 6.4% uptick in mAP value. This highlights how our model not only addresses the speed limitations of SSD models but also enhances the comparatively lower accuracy of YOLO series networks. By leveraging VGG16 as the backbone network and implementing the multi-scale feature fusion strategy, the model matches large targets with large-scale feature maps and small targets with small-scale feature maps. This approach enhances detection accuracy while maintaining model speed.

(3) Our model effectively distinguishes between foreground and background images through the Local distribution mechanism, and successfully accomplishes the final detection task of categorizing and locating cervical cells using the MIL detector, leading to superior performance.

TABLE 4. Ablation results. LDM express Local distillation mechanism.

Model	SSD	LDM	MIL	mAP value
SSD	✓	-	-	66.2%
SSD+LDM	✓	✓	-	69.3%
SSD+MIL detector	✓	-	✓	71.5%
LD-WSCCD	✓	✓	✓	73.6%

B. ABLATION EXPERIMENT

To verify the effectiveness of our proposed three components in improving cervical cell detection performance, ablation experiments were designed. The mAP results of 11 cells are shown in Table 4.

The mAP value of a single SSD model is 66.2%. When only the Local distribution mechanism is added, the mAP increases by 3.1%, and when only the MIL detector is added, the mAP increases by 5.3%. When both the Local distribution mechanism and MIL detector are added, the mAP increases by 7.4%. The experimental results demonstrate that SSD uses multi-scale feature maps and prior boxes to detect targets, which can better capture the detailed information of targets in the image. The added Local Distillation mechanism pays special attention to local pixels and channels, extracts corresponding attention masks, and further improves detection accuracy. Finally, the detection task of cervical cell category and location is completed through MIL detector.

V. CONCLUSION

A. SUMMARY

Cervical cell pathology images encompass millions of cell tissues, making manual annotation a time-consuming and labor-intensive task that is prone to erroneous or inaccurate labeling. This leads to a significant presence of noisy labels, which adversely impacts the model's detection performance. Leveraging this dataset and building on prior research, we introduce a weakly supervised cervical cell detection model named LD-WSCCD, based on a local distillation mechanism. The data is fed into the base network SSD, where convolution and multi-scale feature fusion extract rich feature information from the cervical cell images through processes like local distillation feature loss. Subsequently, two fully connected layers generate feature vectors for detection and classification pathways. Following processing with the MIL detector, we obtain classification and localization results for cervical cells. Experimental findings on our dataset demonstrate that LD-WSCCD can accomplish cervical cell detection tasks under weakly supervised conditions, exhibiting superior accuracy and performance compared to existing algorithms. This enables the successful auxiliary diagnosis of cervical cancer.

B. PROSPECT

In auxiliary medical image diagnosis, there are pressing challenges in cervical cell detection tasks. Future advancements can focus on the following aspects:

(1) Establishment of a comprehensive, specialized, and adequately large dataset for cervical pathological cell analysis is essential. Despite ongoing efforts to promote cancer screening, the privacy concerns and dispersed nature of generated cervical pathology images pose challenges. Annotating cervical pathology images demands significant resources, hindering the creation of a professional dataset for research purposes.

(2) Enhancing the accuracy of detecting overlapping adherent cells within cervical samples is crucial. The ambiguous boundaries resulting from the adhesion between cell nucleus and cytoplasm, along with the stacking of multiple targets, significantly impact detection precision. Improving the detection accuracy of these complex targets is a key research focus for the future.

(3) Striking a balance between accuracy and speed in model performance remains a critical issue. The growing complexity of neural network models has resulted in excessively large networks, hindering their deployment and practical implementation. While enhancing network performance by incorporating various modules can be beneficial, it often comes at the cost of reduced operational speed. Developing lightweight network models that are easily deployable and offer tangible real-world value is paramount for advancing research in this field.

ACKNOWLEDGMENT

(Juanjuan Yin and Qian Zhang are co-first authors.)

REFERENCES

[1] C. R. Clark, N. Baril, M. Kunicki, N. Johnson, J. Soukup, K. Ferguson, S. Lipsitz, and J. Bigby, "Addressing social determinants of health to improve access to early breast cancer detection: Results of the Boston REACH 2010 breast and cervical cancer coalition women's health demonstration project," *J. Women's Health*, vol. 18, no. 5, pp. 677–690, 2010.

[2] B. Ashok and P. Aruna, "Comparison of feature selection methods for diagnosis of cervical cancer using SVM classifier," *Int. J. Eng. Res. Appl.*, vol. 6, no. 1, pp. 94–99, 2016.

[3] Y. Marinakis, G. Dounias, and J. Jantzen, "Pap smear diagnosis using a hybrid intelligent scheme focusing on genetic algorithm based feature selection and nearest neighbor classification," *Comput. Biol. Med.*, vol. 39, no. 1, pp. 69–78, 2009.

[4] V. Chandran, "The genetics of psoriasis and psoriatic arthritis," *Clin. Rev. Allergy Immunol.*, vol. 44, pp. 149–156, Jan. 2013.

[5] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *J. Comput. Syst. Sci.*, vol. 55, no. 1, pp. 119–139, 1997.
[6] R. H. Kaufman, K. Schreiber, and T. Carter, "Analysis of atypical squamous (glandular) cells of undetermined significance smears by neural network–directed review," *Obstetrics Gynecol.*, vol. 91, no. 4, pp. 556–560, 1998.
[7] A. Pal, Z. Xue, K. Desai, A. A. F. Banjo, C. A. Adepiti, L. R. Long, M. Schiffman, and S. Antani, "Deep multiple-instance learning for abnormal cell detection in cervical histopathology images," *Comput. Biol. Med.*, vol. 138, Nov. 2021, Art. no. 104890.
[8] Y. Li and R. L. Stevenson, "Multimodal image registration with line segments by selective search," *IEEE Trans. Cybern.*, vol. 47, no. 5, pp. 1285–1298, May 2017.
[9] J. L. Balcazar, Y. Dai, and O. Watanabe, "Provably fast training algorithms for support vector machines," in *Proc. IEEE Int. Conf. Data Mining*, Dec. 2001, pp. 43–50.
[10] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1904–1916, Sep. 2015.
[11] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1440–1448.
[12] H. Qassim, A. Verma, and D. Feinzimer, "Compressed residual-VGG16 CNN model for big data places image recognition," in *Proc. IEEE 8th Annu. Comput. Commun. Workshop Conf. (CCWC)*, Jan. 2018, pp. 169–175.
[13] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.
[14] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 779–788.
[15] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6517–6525.
[16] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," 2018, *arXiv:1804.02767*.
[17] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal speed and accuracy of object detection," 2020, *arXiv:2004.10934*.
[18] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single shot multibox detector," in *Proc. 14th Eur. Conf. ECCV*, Amsterdam, The Netherlands. Switzerland: Springer, 2016, pp. 21–37.
[19] J. Jeong, H. Park, and N. Kwak, "Enhancement of SSD by concatenating feature maps for object detection," 2017, *arXiv:1705.09587*.
[20] C.-Y. Fu, W. Liu, A. Ranga, A. Tyagi, and A. C. Berg, "DSSD: Deconvolutional single shot detector," 2017, *arXiv:1701.06659*.
[21] Z. Li, L. Yang, and F. Zhou, "FSSD: Feature fusion single shot multibox detector," 2017, *arXiv:1712.00960*.
[22] Z. Shen, Z. Liu, J. Li, Y.-G. Jiang, Y. Chen, and X. Xue, "Object detection from scratch with deep supervision," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 2, pp. 398–412, Feb. 2020.
[23] 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. 2999–3007.
[24] H. Law and J. Deng, "CornerNet: Detecting objects as paired keypoints," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 734–750.
[25] X. Zhou, D. Wang, and P. Krähenbühl, "Objects as points," 2019, *arXiv:1904.07850*.
[26] M. Tan, R. Pang, and Q. V. Le, "EfficientDet: Scalable and efficient object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 10778–10787.
[27] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 580–587.
[28] G. Hinton, O. Vinyals, and J. Dean, "Distilling the knowledge in a neural network," 2015, *arXiv:1503.02531*.
[29] M. Zhang, G. Song, H. Zhou, and Y. Liu, "Discriminability distillation in group representation learning," in *Proc. 16th Eur. Conf. Comput. Vis.*, Glasgow, U.K. Switzerland: Springer, 2020, pp. 1–19.
[30] P. Shen, X. Lu, S. Li, and H. Kawai, "Interactive learning of teacher–student model for short utterance spoken language identification," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2019, pp. 5981–5985.
[31] V. Sanh, L. Debut, J. Chaumond, and T. Wolf, "DistilBERT, a distilled version of BERT: Smaller, faster, cheaper and lighter," 2019, *arXiv:1910.01108*.
[32] J. Wang, W. Bao, L. Sun, X. Zhu, B. Cao, and S. Y. Philip, "Private model compression via knowledge distillation," in *Proc. AAAI Conf. Artif. Intell.*, 2019, pp. 1190–1197.
[33] X. Chen, Y. Zhang, H. Xu, Z. Qin, and H. Zha, "Adversarial distillation for efficient recommendation with external knowledge," *ACM Trans. Inf. Syst.*, vol. 37, no. 1, pp. 1–28, 2018.
[34] Z. Zeng, B. Liu, J. Fu, H. Chao, and L. Zhang, "WSOD2: Learning bottom-up and top-down objectness distillation for weakly-supervised object detection," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 8292–8300.
[35] L. F. Zeni and C. R. Jung, "Distilling knowledge from refinement in multiple instance detection networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2020, pp. 768–769.
[36] P. Tang, X. Wang, X. Bai, and W. Liu, "Multiple instance detection network with online instance classifier refinement," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2843–2851.
[37] Y. Liang, Z. Tang, M. Yan, J. Chen, Q. Liu, and Y. Xiang, "Comparison-based convolutional neural networks for cervical cell/clumps detection in the limited data scenario," 2018, *arXiv:1810.05952*.
[38] 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.
[39] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7132–7141.
[40] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 3–19.
[41] S. Zagoruyko and N. Komodakis, "Paying more attention to attention: Improving the performance of convolutional neural networks via attention transfer," 2016, *arXiv:1612.03928*.

JUANJUAN YIN is currently pursuing the master's degree with the School of Information Science and Technology, Northwest University. Her research interests include computer vision and medical image processing.

QIAN ZHANG is currently pursuing the master's degree with the School of Information Science and Technology, Northwest University. Her research interests include computer vision and medical image processing.

XINYI XI is currently pursuing the master's degree with the School of Information Science and Technology, Northwest University. Her research interests include computer vision and medical image processing.

MENGHAO LIU is currently pursuing the master's degree with the School of Information Science and Technology, Northwest University. His research interests include computer vision and medical image processing.

WENJING LU received the master's degree from the School of Information Science and Technology, Northwest University. His research interests include computer vision and medical image processes.

HUIJUAN TU is currently a Radiologist with Kunshan Hospital of Chinese Medicine. Her research interest includes medical images.