Editor's Choice
Microscopy, 2022, **71(1)**, 50–59
DOI: https://doi.org/10.1093/jmicro/dfab033
Advance Access Publication Date: 21 August 2021
Article

MICROSCOPY

Morphological components detection for super-depth-of-field bio-micrograph based on deep learning

Xiaohui Du[1], Xiangzhou Wang[1], Fan Xu[2,*], Jing Zhang[1,*], Yibo Huo[1], Guangmin Ni[1], Ruqian Hao[1], Juanxiu Liu[1] and Lin Liu[1]

[1]MOEMIL Laboratory, School of Optoelectronic Science and Engineering, University of Electronic Science and Technology of China, No. 4, Section 2, North Jianshe Road, Chengdu 610054, China
[2]Department of Public Health, Chengdu Medical College, No. 783, Xindu Roda, Chengdu 610599, China
[*]To whom correspondence should be addressed. E-mail: xufan@cmc.edu.cn (F.X.); zhangjing@uestc.edu.cn (J.Z.)

Abstract

Accompanied with the clinical routine examination demand increase sharply, the efficiency and accuracy are the first priority. However, automatic classification and localization of cells in microscopic images in super depth of Field (SDoF) system remains great challenges. In this paper, we advance an object detection algorithm for cells in the SDoF micrograph based on Retinanet model. Compared with the current mainstream algorithm, the mean average precision (mAP) index is significantly improved. In the experiment of leucorrhea samples and fecal samples, mAP indexes are 83.1% and 88.1%, respectively, with an average increase of 10%. The object detection model proposed in this paper can be applied to feces and leucorrhea detection equipment, and significantly improve the detection efficiency and accuracy.

Key words: object detection, microscopy, Ritinanet, super-depth-of-field

Introduction

Fecal and leucorrhea microscopy are two routine examinations for pathological analysis in hospital laboratory. The world population is close to 7.9 billion, and the male-to-female ratio is about 1.02 [1]. According to the World Health Organization disease report, the incidence of digestive disease was 20–40% per year and 24.94% for gynecological diseases [2]. Apparently, there are abundant demands in routine clinical examinations on feces and leucorrhea. However, several challenges remain, including variance from manual operation, disgusting smell, aseptic operation, and inefficient and tedious operation [3]. With the extensive development and application of the visual detection technology of microscopy, a large number of detection images are generated during the detection process. It is inefficient to generate inspection reports by processing samples manually. Computer vision is the ability of a computer or machine to acquire human-like understanding from digital images or video [4]. Using machine vision is the latest development trend in medical human secretion detection. Although many models were developed to solve the problem of microcell object detection, several challenges, including complex feature extractors and preprocessed training processes, still remain. For example, the traditional machine vision method requires the design of complex feature extractors (such as morphological features and texture features), and a large number of images need to be preprocessed before training [5–8]. Previous studies mainly focused on the recognition and property retrieval of single-cell types [9,10], and few studies have focused on automatic recognition and localization of other common cells, such as epithelial cells (Epi cells), red blood cells (RBCs), white blood cells (WBCs) and molds.

Recently, deep learning achieved good application prospects in image classification, object detection and other computer vision tasks [11,12]. Compared with the traditional machine learning method, the deep learning method can automatically extract image features and simplify and avoid unnecessary image preprocessing; all of these merits can significantly improve the validity and accuracy of detection [13–15]. However, the application of convolutional neural networks (CNNs) for infrastructure inspection is still in its infancy; namely, it failed for multi-object cell detection.

In vitro diagnostic equipment greatly simplifies the process of sample detection. However, the bottleneck of this technique is mainly focused on the automatic identification of biological components [16]. The changing of the relative position between the cell center and the focus plane of microscope leads to variable cell morphological structure in the two-dimensional image, which is an important reason for the low accuracy of current deep learning object detection algorithms. For example, in Fig. 1, the morphological characteristics of cell 1 are obvious, while the morphology of cell 2 is in a fuzzy state in position 1. In contrary, the morphology of cell in position 2 is obvious while the cell in position

Received 20 July 2021; Revised 6 August 2021; Editorial Decision 19 August 2021; Accepted 20 August 2021

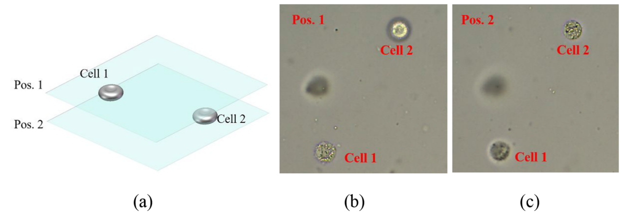


Fig. 1. Diagram of cell focusing position in leucorrhea samples. (a) The schematic diagram of focusing cells at different layers; (b) the image of leucorrhea sample at position 1 and (c) the image of the same sample with the same viewing field at position 2.

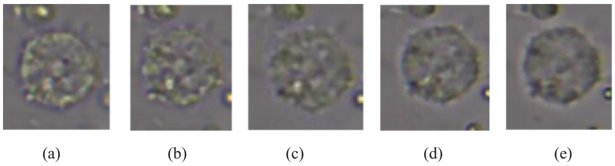


Fig. 2. 'Valid' and 'Invalid' examples for white blood cells in feces. (a), (b) and (c) are 'Valid' (d) and (e) are 'Invalid'.

1 is fuzzy. This is an application defect of the current object detection algorithm in the field of cell detection.

Cell component detection is mainly based on single image object detection in the current methods from which the target location and recognition can be achieved from a neural network. For example, Zhang et al. [17] combined Faster Rich feature hierarchies Convolutional Neural Network (Faster R-CNN) with the proposed circle scanning algorithm (CSA), which can effectively identify cancer cells. Hung et al. [18] used the Faster R-CNN to detect malaria parasites in bright field microscopy images of malaria-infected blood. Kang et al. [19] exploited two state-of-the-art CNN-based object detection methods, Faster R-CNN and SSD, as well as their variants for urine particle recognition. Although the detection in a single layer can achieve a high accuracy rate, it is still insufficient for clinical application. To confirm the cells in a field of view, it is necessary to repeatedly adjust the z-axis position of the microscope platform and check, given that multiple cells cannot be clearly displayed in the plane of focus.

The detection of multi-layer images is dominantly in service with tomography image detection, computed tomography (CT) and magnetic resonance imaging (MRI) [20,21]. Unfortunately, it is not applicable because of the differences in micro-image detection applications and clinical significance. The object distance of the target is different when a field of vision is imaged in a microscopic imaging system. In the same field of vision, the targets in different positions of the image cannot be combined into a three-dimensional (3D) complete target, which is different from CT. Therefore, 3D object detection methods such as 3D Faster R-CNN cannot be used to realize the object detection in microscopic scene. Here, we demonstrated an end-to-end deep learning method for micro multi-object detection in super-depth-of-field (SDoF) micrographs, based on Retinanet [22]. To compare with other mainstream object detection algorithms, we collected fecal and leucorrhea sample libraries. The samples were collected using the viewing field, and each field had multiple clear images. The training and testing sets were split by the field; see the experimental part for details.

The major novel aspects of the current research work are as follows:

(i) An object detection algorithm in the state of SDoF is proposed, which has better performance than the single image object detection algorithm.
(ii) We applied our proposed method to the leucorrhea and fecal datasets we collected and demonstrated a significant improvement over previous methods.
(iii) The effectiveness of the composition of the network we proposed is verified by the ablation study.

The remainder of the paper is organized as follows. Materials and methods Section describes the proposed models and materials used in this work in detail. Section Result is dedicated to the experiments and results. The discussion is provided in the next section, and concluding remarks are presented in the final section.

Materials and methods

In this section, we first introduce our materials, and the whole proposed architecture is briefly illustrated in Fig. 3. Details of the network are described in the following subparagraph.

Materials

All images were collected by a microscopic imaging system. Briefly, the samples were flow diluted, stirred, placed and imaged with a microscopic imaging system. The process of autofocusing was operated in each field of view of samples, and a group of microscopic image was captured when the z-axis of microscope platform moved continuously. The clearest image was selected from each image group through the image definition algorithm. The clearest 3 images and 5 images for feces for leucorrhea were chosen to input the object detection model. The image definition algorithm named Tenengrad [23] can be described as follows:

$$D = \frac{1}{C}\sum Sobel[I(x, y)], \quad 19 < Sobel[I(x, y)] < 120 \qquad (1)$$

where D is the definition value, I is the gray microscopic image, which is filtered by Sobel operator and C is the number of pixels that meet the conditions.

For algorithm training and testing, experienced laboratory experts annotated the cells of all the images in the development dataset as the ground-truth with rectangular boxes of different colors. We divided the dataset into a training and test set in the ratio 8:2 randomly based on the viewing field. Detailed leucorrhea sample information and the dataset split are summarized in Table 1. The fecal sample information is shown in Table 2.

The hardware and software platform of the experiment was a Windows 7 system with Intel Core i7-7820X CPU @ 3.6 GHz × 16, an NVIDIA GeForce RTX 2080 Ti Graphic Processing Unit (GPU), CUDA 10.0 and cuDNN v7. Motic B1Digital microscope is used for fecal sample imaging with a 40× objective lens (numerical aperture: 0.65, material distance: 0.6 mm). The resolution of the microscope is 1600 × 1200. As for the leucorrhea sample, we used an OLYMPUS CX31 biological microscope with a 40× objective lens samed as the lens used in Motic for leucorrhea imaging. An EXCCD01400KMA Charge-coupled Device (CCD)

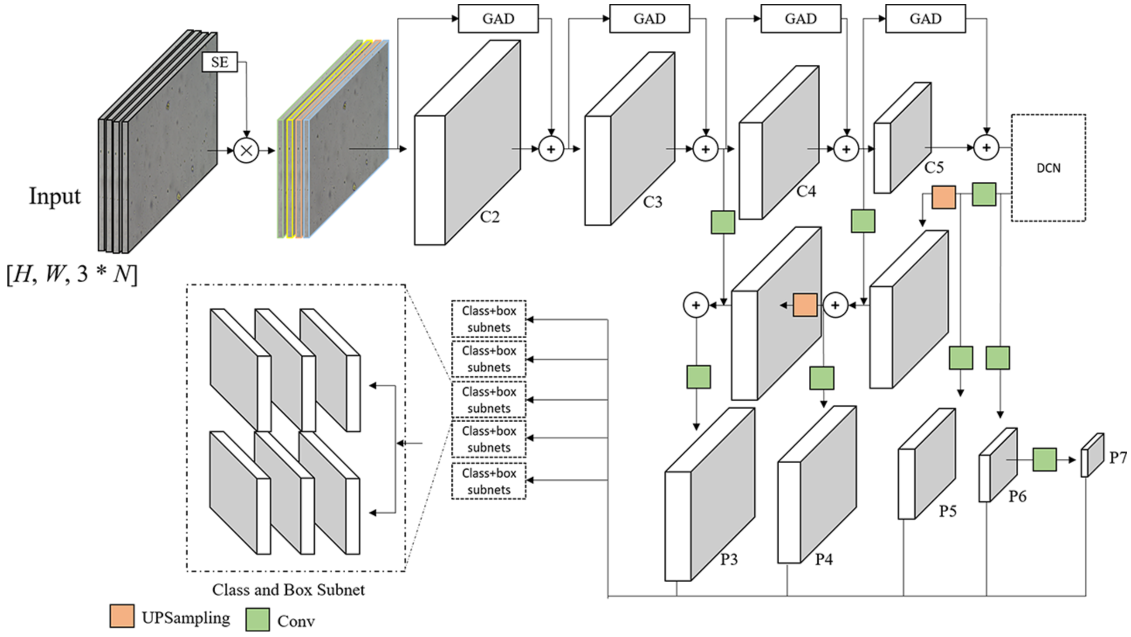


Fig. 3. The architecture of the proposed model. *N* images of the same field of view are taken as one input, which is different from Retinanet. The input dimension is $H \times W \times 3 \times N$.

Table 1. Leucorrhea sample quantity statistics

Contents	Information	Dataset A: training	Dataset B: testing
Data acquisition	2018.9-2019.12	NA	NA
# views	1552	1241	311
# samples	162	NA	NA
RBCs	638	519	119
WBCs	3341	2659	682
Molds	781	544	237
Epi cells	2693	2165	528
Pyos	461	364	97
Tris	17	9	8

Resolution of images are 1920 × 1200; #, numbers; NA, not applicable; RBCs, red blood cells; WBCs, white blood cells; Epi cells, epithelial cells; Pyos, pyocytes; Tris, trichomonads.

Table 2. Fecal sample quantity statistics

Contents	Information	Dataset A: training	Dataset B: testing
Data acquisition	2018.9-2019.12	NA	NA
# views	10 670	8536	2134
# samples	1885	NA	NA
RBCs	7448	6208	1240
WBCs	1691	1362	329
Molds	6437	5131	1306
Pyos	148	115	33

Resolution of images are 1600 × 1200; #, numbers; NA, not applicable; RBCs, red blood cells; WBCs, white blood cells; Pyos: pyocytes.

camera with a pixel size of $6.45\,\mu m \times 6.45\,\mu m$ is used for exposure (resolution 1920×1080).

As for the ground truth of data sets, images in the same field are annotated at the same time, as the position of the same cell in different images is fixed, and the difference is only the clarity. The same cell in different images is annotated with attribute 'Valid'. If a cell is recognizable by experts in the current image, it is annotated as 'Valid'; otherwise, it is 'Invalid'. A cell will have different 'Valid' attributes in different images. If and only if the cell attribute is 'Valid' in one image, it is a positive sample for the whole field, which is regarded as the ground truth. The 'Valid' and 'Invalid' examples for WBCs are shown in Fig. 2.

Architecture overview

The network structure we designed is based on Retinanet [22], as shown in Fig. 3.

In our detection samples, images are divided by the field, which is obtained from different object distances. Doctors judge the category by observing the cells at different distances between the target and the focusing position, but it is difficult to distinguish the category only by observing the single form at a single position, especially when it is in the out–of-focus state. Fig. 2a–e show multiple defocused WBCs. Among them, the cell morphology of (a)–(b) is clear and identifiable, which is marked as 'Valid', while the cell morphology of (c)–(d) is ambiguous and marked as 'Invalid'. It is observed that the morphological differences of the five images are small. Thus, there is a certain subjective error in labeling the tangible components with fuzzy morphology when labeling the 'Valid'/'Invalid' attribute of dataset. This error will further affect the accuracy of convolutional neural network (CNN) model. Inspired by the doctor's manual detection method, we use multiple images to form the depth of field for detection. The input of the detection network is motivated by multi-inputs. The input of the traditional object detection algorithm is a single image with a size of $H \times W \times 3$. The input of the model we designed is $H \times W \times 3 \times N$, where N represents the number of images obtained by field, and 3 is the red green blue (RGB) channel of each image. N is 3 for feces, while 5 for leucorrhea. Our goal is to fuse the feature information of different input images as much as possible for leucorrhea

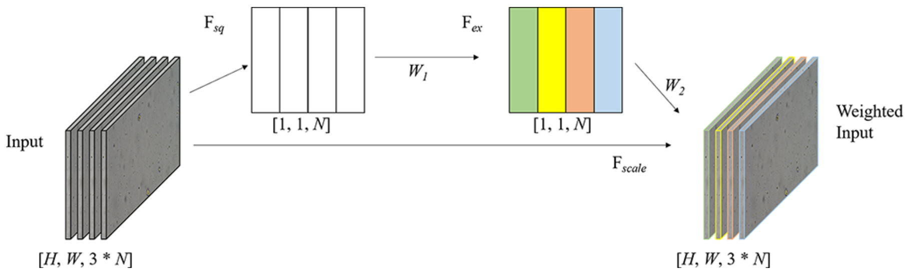


Fig. 4. Improved Squeeze-and-Excitation for input.

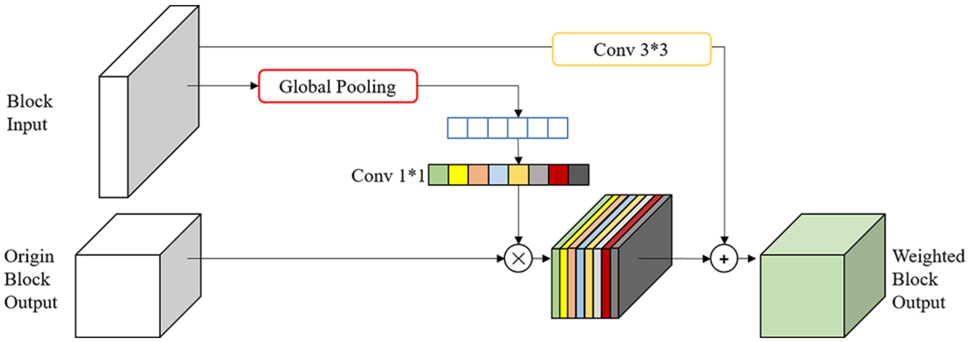


Fig. 5. Global attention down-sample (GAD) block.

and feces object detection, which can be achieved by the subsequence channel fusion. Consequently, we concatenate N images along the channel direction while inputting.

The closer the relative distance between the objects and the focus plane, the clearer the shape mapped on the collected microscopic image and the easier it is to identify. The image region where it is located has a positive contribution to target recognition. On the contrary, the shape is fuzzy, the image in defocus state should be regarded as negative sample when training CNN model, otherwise it will make a negative contribution to target recognition. Therefore, the weights of each formed element in different images selected by the definition algorithm are different. We propose to use the Squeeze-and-Excitation (SE) [24] to assign weights to images in different z-positions, which is shown in Section SE block for input weighting.

The model we proposed is established with a shared Resnet [25] extractor. The target to be detected is usually small in microscopic images. After different stages of pooling, the feature information of the target will be lost. Thus, we need to fuse the large-scale information with the small-scale feature information. In each block, we propose an improved version of the global attention down-sample (GAD) to fuse high-order features and low-order features, as shown in Section Global attention down-sample block, which is motivated by a global attention up-sample [26] model.

Besides, as the deeper layer of Resnet contains more feature information of the whole image, the representation ability of small targets is not good enough. We added Deformable ConvNets (DCN) [27] to the part of the C5 layer connected by the feature pyramid, which is shown in Section Deformable ConvNets for low-level features.

The remainder of the network we proposed is the same as the Retinanet. We constructed five pyramid feature maps with different stride sizes (P3, P4, P5, P6 and P7), and the pyramid features are used to connect the classification and regression sub-networks as the object output of the model.

SE block for input weighting

The input of our model is composed of N images with different focus positions, and the corresponding channel number is $3 \times N$. Different from the SE model [24], SE performs squeeze operation on all channels, while our variant se module divides $3 \times N$ channels into N groups for squeeze operation. The variant model is shown in Fig. 4.

In the squeeze stage, an average group pooling is adopted as the layer-level feature. Specifically, each layer is an image with RGB channels. When performing average pooling, the average of all RGB channels of the image is taken to generate a global feature map. Then, the global feature is operated by excitation to learn the relationship between each channel and get the weights of the different channels. Finally, the weighted input is obtained by multiplying the original input. The squeeze operation can be described as:

$$F_{sq}(f_c) = \frac{1}{H \times W \times 3} \sum_{i=1}^{H} \sum_{j=1}^{W} \sum_{k=1}^{3} f_c(i, j, k), \quad f_c \in R \tag{2}$$

where f_c is the pixel of the input. The excitation operation is as follows:

$$F_{ex}(F_{sq}, W_1, W_2) = \sigma(W_2 \cdot \mathrm{RELU}(W_1 \cdot F_{sq})) \tag{3}$$

where σ is the sigmoid function, and W_1, W_2 are the parameters to be trained. RELU is the RELU activate function.

Global attention down-sample block

Many researchers have proved that combining CNNs with a well-designed network module can obtain excellent performance and feature information [26]. We consider that the low-level features with abundant details can be used to weight the high-level features to improve the resolution.

The GAD module we designed is shown in Fig. 5. GAD takes the global context information as the guidance for high-level features in the global pooling operation. Specifically, the global context information generated from low-level features goes through 1 × 1 convolution, batch normalization and nonlinear transformation and then multiplies with high-level features, which is the same as the SE block. Finally, the low-level features are convoluted (kernel: 3 × 3; stride: according to the block scale) and added to the weighted high-level features.

Deformable ConvNets for low-level features

As the deeper feature map of Resnet contains more global information, it is not good enough to represent small targets. Therefore, we added DCN [27] to the part of the C5 layer connected by the feature pyramid, as shown in Fig. 6.

Compared with the traditional CNN, DCN can be deformation, which improves the effective range of the receptive field. Even though the introduction of DCN may require more detection time, the precision was improved significantly.

Other tricks

Data augmentation

The data augmentation method was used in the training process. The optical structure and acquisition system are relatively fixed in the microscopic images compared with other image acquisition systems. Therefore, the data augmentation methods adopted random rotation, random flipping and 0.8–1.2 scaling without considering color adjustment and other strategies.

Transfer learning and fine-tuning

Transfer learning and fine-tuning can significantly accelerate the speed of model training. In the experiment, we set the initialization parameters of the model to the parameters trained on COCO [28], which can be download from the website and then fine-tune it [29].

Training

In the training process, the network we designed is an end-to-end object detection model. The training batch was set to 2, which is limited by the memory size of the Graphic Processing Unit (GPU). The image size of 1200 × 1920 (leucorrhea) and 1200 × 1600 (feces) were compressed as 800 × 1280 (leucorrhea) and 800 × 1067 (feces) for inputs by the method of bilinear interpolation. We used the focal loss function of Retinanet to reduce the influence of the imbalance of positive and negative samples in the process of solving. Moreover, the Adam method was used to update the parameters of this model.

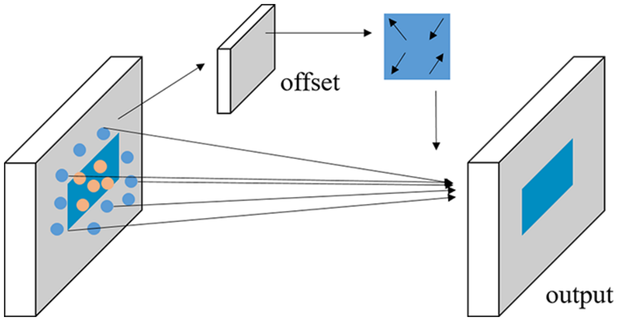


Fig. 6. Deformable ConvNets.

Inference

Similarly, during inference, we compressed the size of the input group of images to 800 × 1280 (leucorrhea) and 800 × 1067 (feces). We decoded the box prediction from a maximum of 1000 top-score predictions per level after setting the detector confidence threshold to 0.05 to improve the inference speed. Non-maximum suppression (NMS) with a threshold of 0.5 was adopted to yield the predictions.

Table 3. The result for the cell detection algorithm in the datasets

Datasets	Items	AP	F_1	mAP	mF_1	Fps
Leucorrhea	RBCs	0.826	60.38			
	WBCs	0.776	324.01	0.838	141.29	8.16
	Molds	0.773	106.72			
	Epi cells	0.973	314.35			
	Pyos	0.693	37.94			
	Tris	0.986	4.34			
Fecal	RBCs	0.946	622.29	0.881	314.70	10.04
	WBCs	0.878	113.14			
	Molds	0.857	503.65			
	Pyos	0.843	19.71			

AP, average precision; mAP, mean average precision; Pyo, pyocytes; Epi cells, epithelial cells; RBCs, red blood cells; WBCs, white blood cells; Tris, trichomonads. Fps represents frames per second. F1: F1 score = 2*Precision*Recall/(Precision+Recall) mF1: mean f1 score.

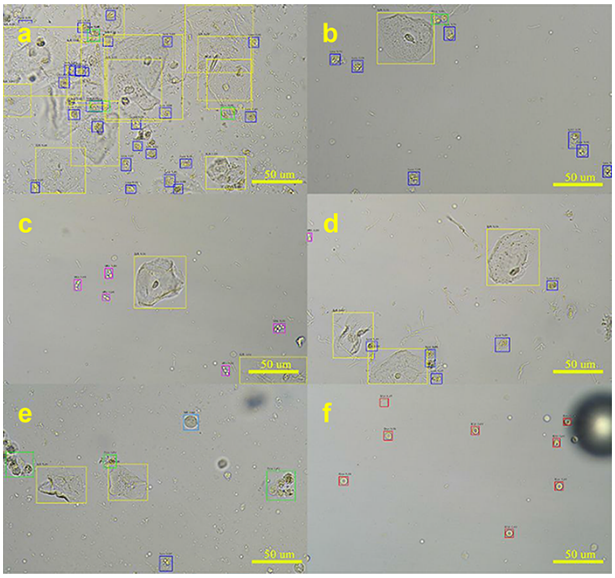


Fig. 7. Curated examples of this model on our leucorrhea dataset. A score threshold of 0.4 was used for displaying. Red rectangles represent red blood cells (RBCs), blue rectangles represent white blood cells (WBCs), yellow rectangles represent epithelial cells (Epi cells), green rectangles represent pyocytes (Pyos), purple rectangles represent molds and cyan rectangles represent suspected trichomonads (Tris). (a), (b) and (d) The detection images of Epi cells and WBCs; (c) the detection image of molds; (e) the detection image of Tris, Epi cells and Pyos and (f) the detection image of RBCs.

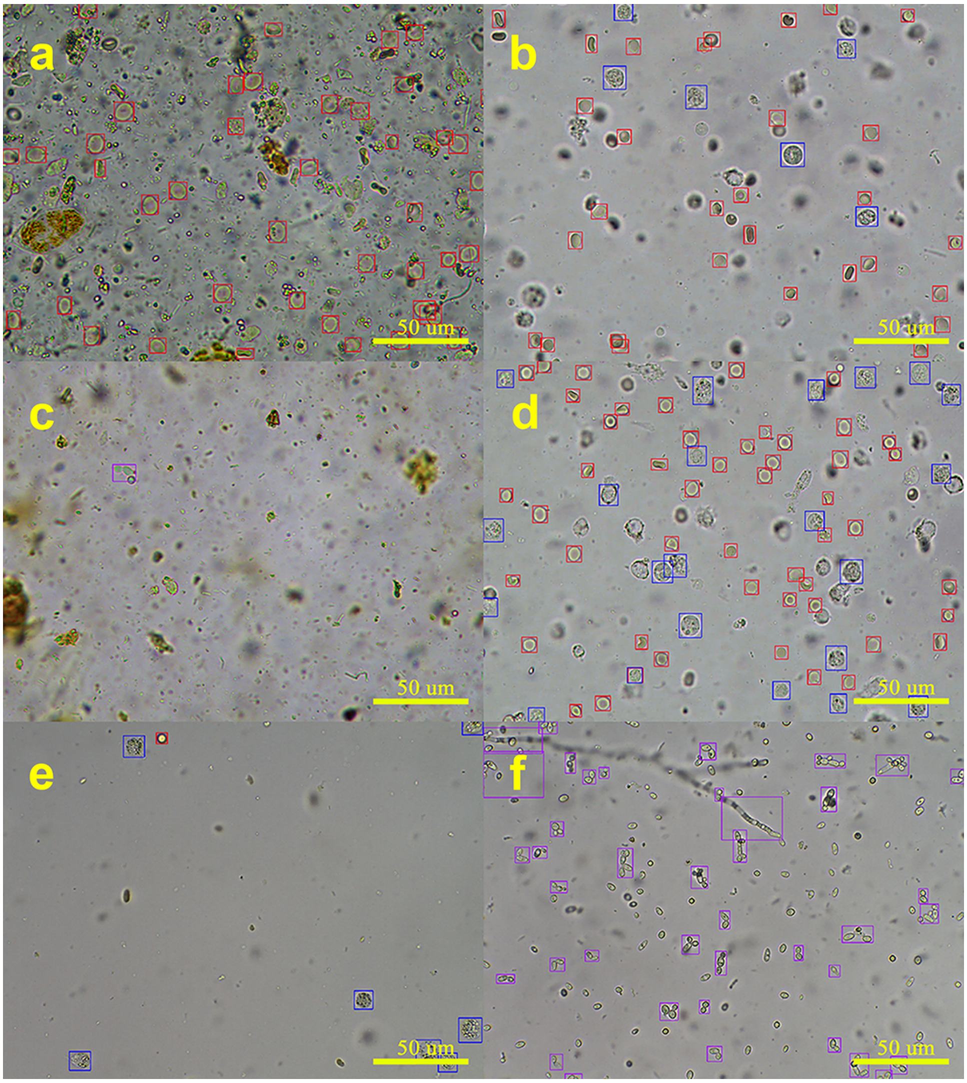


Fig. 8. Curated examples of this model on our fecal dataset. A score threshold of 0.4 was used for displaying. Red rectangles represent red blood cells (RBCs), blue rectangles represent white blood cells (WBCs), purple rectangles represent molds and green rectangles represent pyocytes (Pyos). (a) Detection result of RBCs. (b), (d) and (e) the detection images of RBCs and WBCs, (c) the mold image with light color and (f) the detection image of molds.

Table 4. Performance of ablations in leucorrhea samples

SE	GAD	DCN	AP_{RBC}	AP_{WBC}	AP_{Mold}	AP_{Epi}	AP_{Pyo}	AP_{Tri}	mAP	Fps
√			0.807	0.742	0.737	0.975	0.717	0.935	0.819	9.06
	√		0.815	0.748	0.714	0.972	0.732	0.892	0.813	9.01
		√	0.778	0.751	0.723	0.975	0.716	0.986	0.822	9.08
√	√	√	0.826	0.776	0.773	0.973	0.693	0.986	0.838	8.16

SE, Squeeze-and-Excitation; GAD, global attention down-sample; DCN, Deformable ConvNets; Pyo, pyocytes; Epi, epithelial cell; RBC, red blood cell; WBC, white blood cell; AP, average precision; mAP, mean average precision; Tri, trichomonad. Fps represents frames per second.

Results

Metrics

N images in each field of vision correspond to a group of ground truth tags used in our detection task, which is different from object detection with single image. The mean average precision (mAP) was applied to assess the performance of prediction methods for each group of images. In detail, precision is the ratio of correctly detected objects to all positive objects detected, and recall is the ratio of correctly detected objects to all objects with basic authenticity. Whether the detection is correct depends on the value of the intersection over union (IOU). The targets detected by the prediction model were ranked by confidence. Then, different precision and recall were obtained by sorting IOU from high to low, which is called the precision–recall (PR) curve. The area under the recall curve was the average precision (AP). mAP is calculated by averaging different levels of AP.

Table 5. Performance of ablations in fecal samples

SE	GAD	DCN	AP_{RBC}	AP_{WBC}	AP_{Mold}	AP_{Pyo}	mAP	Fps
√			0.940	0.874	0.821	0.852	0.872	10.53
	√		0.942	0.844	0.833	0.831	0.863	10.33
		√	0.946	0.855	0.825	0.847	0.868	10.14
√	√	√	0.946	0.878	0.857	0.843	0.881	10.04

SE, Squeeze-and-Excitation; GAD, global attention down-sample; DCN, Deformable ConvNets; RBC, red blood cell; WBC, white blood cell; AP, average precision; Pyo, pyocytes; mAP, mean average precision. Fps represents frames per second.

Table 6. Comparison of detection results of leucorrhea samples

Method	Backbone	AP_{RBC}	AP_{WBC}	AP_{Mold}	AP_{Epi}	AP_{Pyo}	AP_{Tri}	mAP	Fps
FR-CNN	ResNet-50	0.382	0.681	0.708	0.860	0.538	0	0.528	5.4
SSD300	VGG-16	0.158	0.191	0.017	0.814	0.123	0.564	0.311	13.7
SSD512	VGG-16	0.567	0.529	0.277	0.811	0.362	0.646	0.532	9.4
YOLOV3	Darknet-53	0.535	0.486	0.335	0.533	0.216	0.495	0.433	7.6
Casc. R-CNN	ResNet-50	0.224	0.715	0.711	0.868	0.639	0	0.526	4.0
Retinanet	ResNet-50	0.750	0.662	0.639	0.914	0.567	0.661	0.699	5.8
Ours	ResNet-50	0.826	0.776	0.773	0.973	0.693	0.986	0.838	8.2

AP, average precision; mAP, mean average precision; RBC, red blood cell; WBC, white blood cell; Epi, epithelial cell; Pyo, pyocytes; Tri, trichomonad. Fps represents frames per second.

Table 7. Comparison of detection results of fecal samples

Method	Backbone	AP_{RBC}	AP_{WBC}	AP_{Mold}	AP_{Pyo}	mAP	Fps
FR-CNN	ResNet-50	0.553	0.588	0.731	0.805	0.669	5.2
SSD300	VGG-16	0.475	0.615	0.621	0.821	0.633	12.8
SSD512	VGG-16	0.630	0.752	0.654	0.812	0.743	8.1
YOLO-V3	Darknet-53	0.548	0.628	0.655	0.486	0.579	6.7
Cascade R-CNN	ResNet-50	0.551	0.629	0.775	0.815	0.693	3.8
Retinanet	ResNet-50	0.750	0.792	0.793	0.882	0.804	5.8
Ours	ResNet-50	0.946	0.878	0.857	0.843	0.881	10.0

AP, average precision; RBC, red blood cell; WBC, white blood cell; mAP, mean average precision; Pyo, pyocytes. Fps represents frames per second.

Detection results of the method proposed

According to the different sample types of the two datasets, we trained and tested them separately. All models were trained on the corresponding training sets and tested on the testing sets. Within the training process, the mean and other evaluation information was output in the validation set in each epoch. The model was tested after the training, and the performance details are shown in Table 3. Figures 7 and 8 display the renderings of leucorrhea samples and fecal samples, respectively.

Ablation study

We extended our experiments to validate the proposed network architecture components. From the base Retinanet framework, four additional ablations were studied: with improved SE module, GAD, DCN and proposed model. The results (Tables 4 and 5) showed that the mAP of improved SE + GAD + DCN was the best among other ablations.

Comparison with other models

We compared our model with the state-of-the-art detectors in Tables 6 and 7. To compare the state-of-the-art object detection methods with our proposed methods, we made some variants on these methods because the detection objects are multi-images in a single field, as shown in Fig. 9.

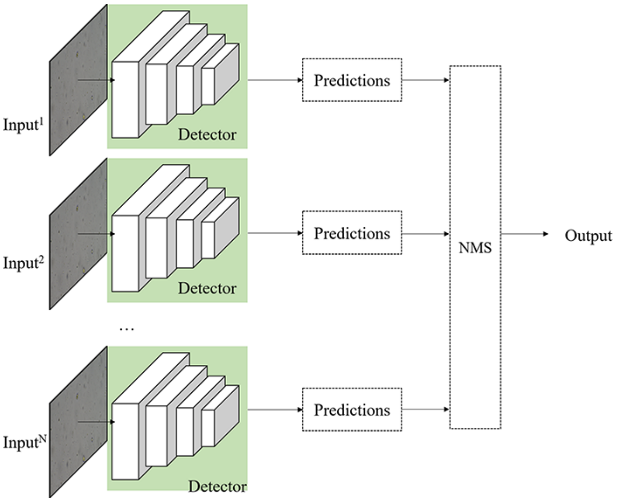
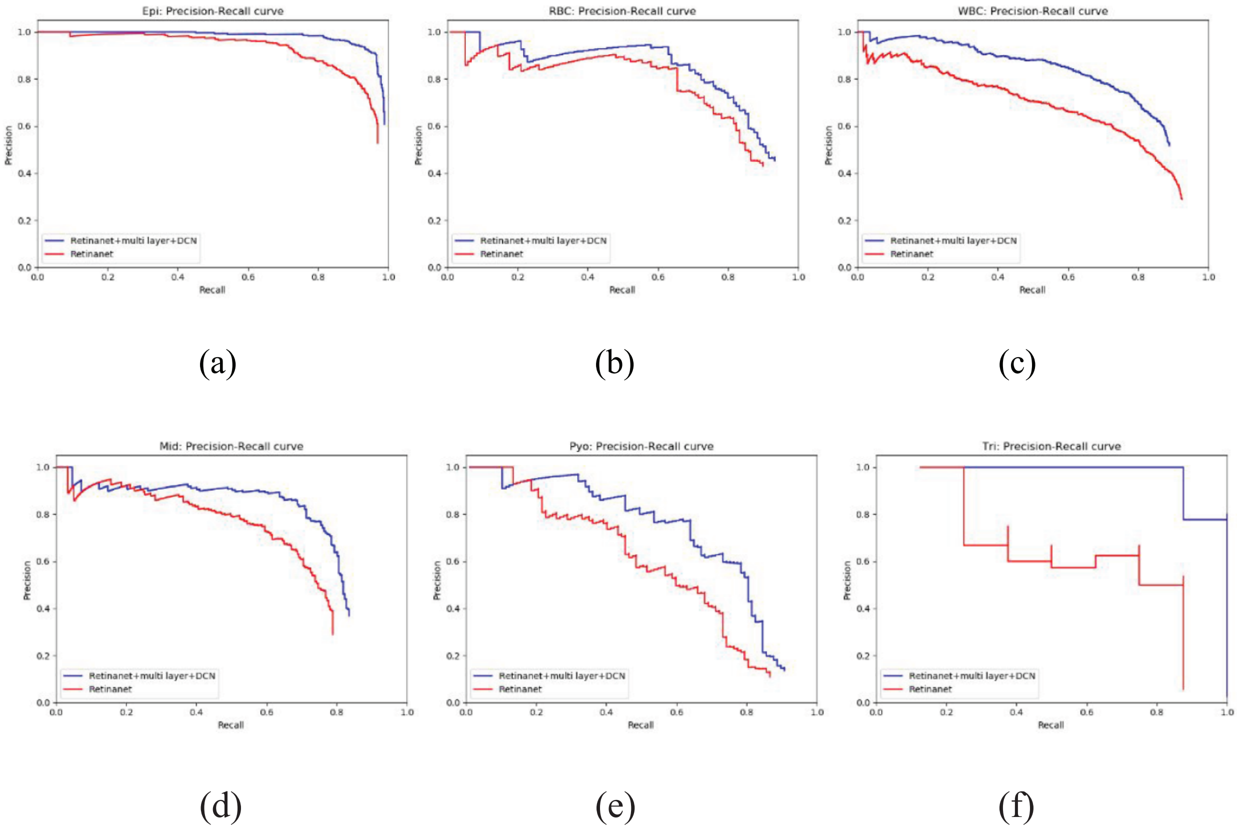


Fig. 9. The architecture of the comparison model.

In the training process, the input of these detectors is a single microscopic image, and the ground truth of this image is composed of the elements with the attribute 'Valid'. We trained all the images in the training set on the comparison models, which is the same as the origin training process. In the inference process, all images in the test set were detected by



(a) (b) (c)

(d) (e) (f)

Fig. 10. Precision–recall (PR) curves of different kinds of cells in leucorrhea datasets; the red line represents the original Retinanet, and blue represents our model. (a) Epithelial cell (Epi) PR curves, (b) red blood cell (RBC) PR curves, (c) white blood cell (WBC) PR curves, (d) Mildew cureves (Mid) PR curves, (e) pyocyte (Pyo) PR curves and (f) trichomonad (Tri) PR curves.

comparison models at first. After that, we combined the detection results of microscopic images that belong to the same field of view and then used the NMS method to merge the duplicate detection boxes in each field of view. In this way, a unified detection result for each field of view was obtained. The training parameters of the original Retinanet and the model we proposed are consistent, except for the network structure. Other model parameters are the same as in the original paper. For the model evaluation, we evaluate the performance of the model according to the field of view. Tables 6 and 7 show the comparison between the AP of different types of cells in leucorrhoea samples and fecal samples, respectively.

To demonstrate that our model has a better detection performance, we selected the leucorrhea datasets for comparison and drew the PR curves of the origin Retinanet and improved Retinanet (Fig. 10). Compared with the PR curves of Epi cells, RBCs, WBCs, molds and *Trichomonas*, our proposed model (blue curve) presented closer to the upper right corner of the coordinate axis. Consequently, the detection effectiveness of our proposed method was better. Fig. 11 shows the comparison examples of manual annotation and detection results of origin Retinanet and our proposed model.

Discussion

The experimental results disclosed that the proposed model could guarantee the detection frame rate, and the index of mAP was significantly improved at the same time. The mAP of four types of targets increased by 9% in fecal datasets and 14% in five types of targets in leucorrhea datasets, which is far better than other mainstream object detection models. Meanwhile, with the addition of the DCN + GAD module, mAP can be improved slightly when detecting small targets (e.g. WBC and RBC), while the detection index of large targets remains unchanged.

Currently, the mainstream object detection algorithms are mainly applied to the target location and extraction in a single image. To adapt to the situation of multiple images in SDoF, we input the images of each field of view into the corresponding object detection model one by one and combine the results of them. The performance of mainstream object detection algorithms on leucorrhea and fecal datasets is shown in Tables 6 and 7, respectively.

In the leucorrhea samples, the SSD300 model [13] had the highest frame rate, but its detection precision was the worst. The detection precision of the Faster R-CNN [15] model and Cascade R-CNN [30] model in *Trichomonas* was poor, while the detection accuracy of the Retinanet model was higher than that of other models, which was used as the baseline for our proposed model. The improved model we proposed achieved good detection results in different types of cells. At the same time, the detected frame rate was higher than other models, which is only next to the SSD300 model.

In the detection results of fecal samples, the frame rate of SSD300 was better, and the accuracy of detection results was higher than that of leucorrhoea samples. The detection precision of the YOLOV3 [14] model was poor in both

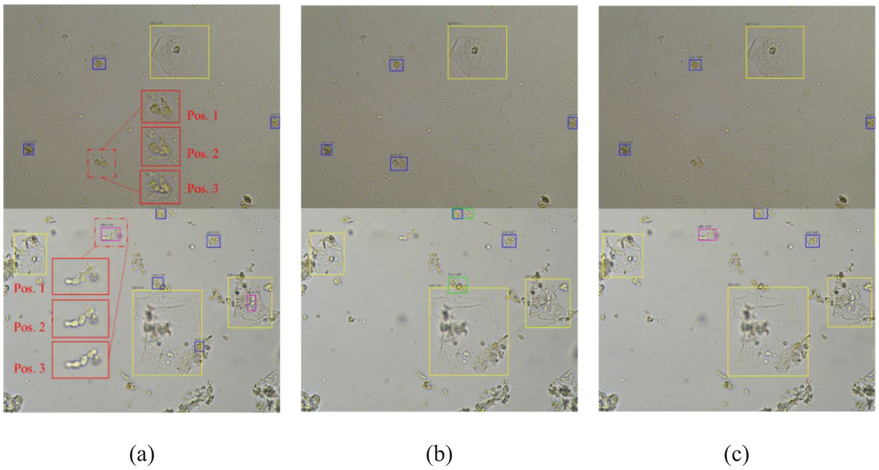


Fig. 11. Detection comparison on the leucorrhea dataset. (a) Ground truth; (b) detection results of origin Retinanet and (c) results of our proposed model. The target in the dotted box is the image at different focusing positions in the field.

microscopic image sets, although the YOLOV3 model performed well in object detection in natural scenes. Similarly, Retinanet [29] also achieved good detection accuracy. Compared with the model we proposed, the detection accuracy of our model is also greatly improved, and our model also performs well in terms of the frame rate.

For the manual annotation, the criterion of 'Valid'/'Invalid' of cells, especially for the slightly out-of-focus cells, is subjective according to the experienced laboratory experts, as shown in Fig. 2. This error will affect the accuracy of the CNN model for single image detection, which leads to the poor effect of these methods. The proposed method uses multiple images to form the depth of field for training and detection, which can effectively avoid this error. At the same time, considering that the number of leucorrhea samples is less than fecal samples, the detection accuracy of the leucorrhea sample model is lower than fecal samples in the previous object detection models.

The contrast precision recall curves (PR curves) with Ritinanet were drawn (Fig. 9). The PR curve of our model (blue) achieved a significantly improved effect in different types of cells; the *R* value was higher at the same *P* value. By comparing the PR curves of Epi cells and RBCs, the improvement of the model was small, and the interval between red and blue curves was small, while the improvement of WBCs, pyocytes and *Trichomonas* was larger. All these phenomena correspond to the results obtained in Table 4. The detection improvement of molds was significantly improved when a higher score threshold was set, indicating that the original Retinanet has relatively low confidence, while the model we proposed has higher confidence.

Taken together, the evaluation results further confirm the advantages of our model. The mAP of the leucorrhea and fecal datasets was 83.8% and 88.1%, respectively, which was better than other detection models. It shows that our model has a better feature expression ability for SDoF micrograph and is adaptable to different microscopic image application scenarios.

Conclusions

In this study, we proposed an object detection algorithm for SDoF micrographs. The detection index is much higher than the mainstream object detection algorithm based on testing the leucorrhea and the fecal datasets separately. The mAP increased by nearly 10%. The detection efficiency was equivalent to the Retinanet, which can fully meet the needs of the automated online detection. The proposed algorithm has reference significance for the application of object detection in SDoF microscopy system. The code is available at Github: https://github.com/xiaohuilang/Morphological-components-detection-for-super-depth-of-field-bio-micrograph-based-on-deep-learning.

Funding

This research was supported partly by the National Science Foundation of China (No. 61905036), the Fundamental Research Funds for the Central Universities (University of Electronic Science and Technology of China) (No. ZYGX2019J053), China Postdoctoral Science Foundation (2019M663465) and Chengdu Science and Technology Bureau (2019-YF09-00097-SN).

Acknowledgements

We would like to express our thanks to Professor Yu-Tang Ye and the staff at the MOEMIL laboratory, who counted the samples used in this study.

Conflict of Interest

The authors declare that they have no conflict of interest.

References

1. World population. https://countrymeters.info/en/World. 19 June 2020.
2. Dossett M L, Cohen E M, and Cohen J (2017) Integrative medicine for gastrointestinal disease. *Prim. Care* 44: 265–280.
3. Abraham B P (2018) Fecal lactoferrin testing. *Gastroenterol. Hepatol. (NY)* 14: 713–716.
4. Ballard D H, and Brown C M (1982) *Computer Vision*, (Prentice Hall Professional Technical Reference, New York, USA).
5. Ghosh P, Bhattacharjee D, and Nasipuri M (2016) Blood smear analyzer for white blood cell counting: a hybrid microscopic image analyzing technique. *Appl. Soft. Comput.* 46: 629–638.

6. Manik S, Saini L M, and Vadera N (2016) Counting and classification of white blood cell using Artificial Neural Network (ANN). In: *2016 IEEE 1st International Conference on Power* Electronics, *Intelligent* Control *and Energy Systems (ICPEICES)*, pp 1–5 (IEEE, Delhi, India).
7. Piuri V and Scotti F (2004) Morphological classification of blood leucocytes by microscope images. In: *2004 IEEE International Conference onComputational Intelligence for Measurement Systems and Applications*, pp 103–108 (CIMSA).
8. Wang M, Zhou X B, Li F H, Huckins J, King R W, and Wong S T C (2008) Novel cell segmentation and online SVM for cell cycle phase identification in automated microscopy. *Bioinformatics* 24: 94–101.
9. Sunarko B, Djuniadi Bottema M, Iksan N, Hudaya K A N, and Hanif M S (2020) Red blood cell classification on thin blood smear images for malaria diagnosis. *J. Phys. Conf. Ser* 1444: 012036. https://iopscience.iop.org/article/10.1088/1742-6596/1444/1/012036.
10. Zhang J, Zhong Y, Wang X Z, Ni G M, Du X H, Liu J X, Liu L, and Liu Y (2017) Computerized detection of leukocytes in microscopic leukorrhea images. *Med. Phys.* 44: 4620–4629.
11. Deng J, Pan Y, Yao T, Zhou W, Li H, and Mei T (2019) Relation distillation networks for video object detection. In: *2019 IEEE/CVF International Conference on Computer Vision,* ed. ICCV, pp 7022–7031 (IEEE, Seoul, Korea).
12. He K, Gkioxari G, Dollar P, and Girshick R (2020) Mask R-CNN. *IEEE Trans. Pattern Anal. Mach. Intell.* 42: 386–397.
13. Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu C-Y, and Berg A C (2016) SSD: single shot multibox detector. In: *14th European Conference on Computer Vision*, ed. ECCV, pp 21–37 (Springer Verlag, Amsterdam, the Netherlands).
14. Redmon J and Farhadi A (2018) YOLOv3: an incremental improvement. In: *Internaltional Conference on Computer Vision and Pattern Recognition*, ed. CVPR, (IEEE, Salt Lake City, Utah).
15. Ren S, He K, Girshick R, and Sun J (2017) Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* 39: 1137–1149.
16. Leng L, Yang Z, Kim C, and Zhang Y (2020) A light-weight practical framework for feces detection and trait recognition. *Sensors* 20: 2664.
17. Zhang J K, Hu H G, Chen S Y, Huang Y J, and Guan Q (2016) Cancer cells detection in phase-contrast microscopy images based on faster R-CNN. Int. *Sym. Comput. Intel.* 1: 363–367.
18. Hung J, Goodman A, Lopes S, Rangel G, Ravel D, Costa F, Duraisingh M, Marti M, and Carpenter A (2017) Applying Faster R-CNN for object detection on malaria images. In: *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, ed. CVPR, pp 2160–2174 (IEEE, Honolulu, Hawaii, USA).
19. Kang R, Liang Y, Lian C, and Mao Y (2018) An end-to-end system for automatic urinary particle recognition with convolutional neural network. *J. Med. Sys.* 42: 165.
20. Lapa P, Castelli M, Goncalves I, Sala E, and Rundo L (2020) A hybrid end-to-end approach integrating conditional random fields into CNNs for prostate cancer detection on MRI. *Appl. Sci.* 10: 338.
21. Shakeel P M, Burhanuddin M A, and Desa M I (2020) Automatic lung cancer detection from CT image using improved deep neural network and ensemble classifier. *Neural Comput. Appl.* https://link.springer.com/article/10.1007%2Fs00521-020-04842-6.
22. Lin T Y, Goyal P, Girshick R, He K M, and Dollar P (2020) IÈEE Transactions on Pattern Analysis And Machine Intelligence. *IEEE Int. Conf. Comp. Vis.* 42: 318–327.
23. Yeo T T E, Ong S H, Jayasooriah, and Sinniah R (1993) Autofocusing for tissue microscopy. *Image Vis. Comput.* 11: 629–639.
24. Hu J, Shen L, Albanie S, Su G, and Wu E (2020) Squeeze-and-Excitation networks. *IEEE Trans. Pattern Anal. Mach. Intell.* 42: 2011–2023.
25. He K, Zhang X, Ren S, and Sun J (2016) Deep residual learning for image recognition. In: *29th IEEE Conference on Computer Vision and Pattern Recognition*, ed. CVPR, pp 770–778 (IEEE, Las Vegas, NV, USA).
26. Li H, Xiong P, An J, and Wang L (2019) Pyramid attention network for semantic segmentation. In: 29th *British Machine Vision Conference,* ed. BMVC, (Amazon; et al.; Microsoft; *NVIDIA;* SCANs; SCAPE. BMVA Press, London, UK).
27. Zhu X Z, Hu H, Lin S, and Dai J F (2019) Deformable ConvNets v2: more deformable, better results. In: *2019 IEEE/Cvf Conference on Computer Vision And Pattern Recognition*, ed. CVPR by CVPR, pp 9300–9308, (IEEE, Long Beach, CA, USA).
28. COCO: Common Object in Context. https://cocodataset.org/. 12 November 2020.
29. Retinanet download. https://github.com/fizyr/keras-retinanet/releases. 12 November 2020.
30. Cai Z W, and Vasconcelos N (2018) Cascade R-CNN: delving into high quality object detection. In: *2018 IEEE/CVFConference on Computer VisionandPattern Recognition,* ed. CVPR, pp 6154–6162 (IEEE, Salt Lake City, UT, USA).