An Efficient Particle YOLO Detector for Urine Sediment Detection

Zejian Chen[1], Rong Hu[1], Fukun Chen[2], Haoyi Fan[2], Fum Yew Ching[3], Zuoyong Li[4], and Shimei Su[5(✉)]

[1] Fujian Provincial Key Laboratory of Big Data Mining and Applications, School of Computer Science and Mathematics, Fujian University of Technology, Fuzhou 350118, Fujian, China
[2] School of Computer and Artificial Intelligence, Zhengzhou University, Zhengzhou 450001, Henan, China
[3] School of Computer Sciences, Universiti Sains Malaysia, Penang 11800, Malaysia
[4] Fujian Provincial Key Laboratory of Information Processing and Intelligent Control, College of Computer and Control Engineering, Minjiang University, Fuzhou 350121, Fujian, China
[5] School of Electrical Engineering, Zhengzhou University, Zhengzhou 450001, Henan, China
smsu@zzu.edu.cn

Abstract. Urine sediment detection is an essential aid in assessing kidney health. Traditional machine learning approaches treat urine sediment particle detection as an image classification task, segmenting particles for detection based on information such as edges or thresholds. However, the segmentation of sediment particles is complex due to the low contrast and weak edge characteristics of urine sediment images. In this paper, we consider urine sediment particle detection as a object detection task and propose the YOLOv5s-CBL, a detector dedicated to particle detection. Specifically, to mitigate the impact of background noise on detection accuracy, we inherit CBAM on the YOLOv5s model to help the network filter useless noise information and find regions of interest to extract target features. Then, we expand the original three-scale feature layer to improve the sensitivity of the model to larger-scale target sediment particles. Finally, we use the BiFPN structure instead of the original PANet combined with the FPN structure, which can more effectively fuse multiple different scales of information to improve the detection performance of the model. We compared state-of-the-art methods on two real-world datasets, and the experimental results demonstrated the effectiveness of YOLOv5s-CBL.

Keywords: Urine sediment detection · Deep learning · YOLOv5

1   Introduction

Kidney disease is a major threat to human health and affects millions of people worldwide [20]. As an important method in analyzing kidney diseases, urine sediment detection can reflect the health of patients' kidneys through the changes in sediment particles, which is an important basis for subsequent clinical diagnosis. However, the diagnostic accuracy of urine sediment detection is dependent on the professionalism and clinical experience of the medical personnel. Meanwhile, it is easily affected by external factors such as visual bias and equipment malfunction. Therefore more and more automated urine analyzers based on machine learning methods are developed for the detection of urine deposits [9,13,15,19,30].

Traditional machine learning algorithms regard urine sediment detection as an image classification task. First, the data is preprocessed to segment the sediment particles from the urine image. Then the main features of the sediment are extracted by a CNN network. Finally, the extracted features are fed into a trained classifier for classification. However, most urine sediment images have low contrast and weak edge features. Therefore, segmenting the entire image's urine sediment particles is not easy. Recently, deep learning-based urine sediment detection has started to be used to solve the segmentation problem and achieved excellent results. So, similar to Liang et al. [10,11,26], we also consider urine sediment inspection as an object detection task.

In this paper, we propose the YOLOv5s-CBL model, an end-to-end object detection model that is more suitable for urine sediment detection. Specifically, we first considered the sparsity of urine images. To mitigate the impact of background noise on detection accuracy, we integrated CBAM [24] on top of the YOLOv5s model to help the network filter the noisy information from urine images to find the region of interest to extract target features. Then to improve the sensitivity of the model to larger-scale target sediment particles, we extended the original three-scale feature layer by using the added 64x downsampling as a fourth feature layer to detect larger-scale sediment particles in the urine images. Finally, to obtain better feature representation capability, we replace the original PANet [12] combined with the FPN [21] structure with the BiFPN [21] structure to fuse feature information of different scales in a targeted manner. The comparison with state-of-the-art methods on two urine sediment particle detection datasets demonstrates the effectiveness of the YOLOv5s-CBL.

The main contributions of this paper are as follows:

– We propose a urine sediment detection method based on an improved YOLOv5 model. By introducing a larger scale detection head and CBAM to help the network capture sediment particle targets at different scales in sparse urine images and using a BiFPN feature aggregation structure to fuse feature information at different scales to obtain better representation capability.
– We add a larger scale prediction head to detect larger sediment objects in the dataset, and filter the effect of background noise information with the integrated CBAM to help the network find the target region of interest from the whole image.

– We optimize the feature aggregation structure of the neck network by replacing the structure of PANet in YOLOv5 with BiFPN. While reducing the number of parameters, the network can learn the importance of distribution weights of each feature in a targeted manner to obtain better representation capability.
– We have conducted extensive experiments on two real-world datasets comparing DFPN, BCPNet, PVANet, YOLOv3-tiny, YOLOv3 and YOLOv5s models, and the experimental results demonstrate the effectiveness of the proposed method.

Our paper is structured as follows. We first discuss related work (Sect. 2) and present YOLOv5s-CBL in Sect. 3. Experimental results are presented in Sect. 4. Sections 5 and 6 are the conclusion and acknowledgements, respectively.

2   Related Work

2.1   Urine Sediment Detection

Traditional urine sediment detection uses manual microscopy to count the sediment of centrifuged urine samples. However, the accuracy of detection depends on the skill level of the cytologist and increases the amount of labor. Therefore it cannot be applied on a large scale. Machine learning methods have been widely used for urine sediment image analysis to improve standardization and detection accuracy. Ranzato et al. [15] developed a simple and generalized bioparticle identification system. Using a hybrid Gaussian classifier to identify 12 urine sediment particles achieved 93% accuracy. Liu et al. [13] used an SVM classifier to classify a variety of urine sediment particles, including RBC, WBC, Cast, and Crystal, and achieved 91% accuracy. Liang et al. [9] used SVM and decision tree to construct a classification filter, which improved the detection accuracy to 93.72%. Shen et al. [19] constructed a multiclass classifier based on the AdaBoost learning algorithm and SVM, which effectively used Harr wavelet features to improve the recognition accuracy. Zhou et al. [30] used an artificial neural network model to classify 12 classes of sediment images after segmentation, and the recognition accuracy of red blood cells reached 96.19%. Although traditional machine learning methods achieve excellent performance, they can only detect limited types of urine sediment particles. Moreover, the low contrast and weak edge features of urine sediment images make segmentation to extract features more difficult.

Recently deep learning-based urine sediment detection has been proposed to solve the problem mentioned above. Zhang et al. [28] used a pre-trained Faster R-CNN [18] model to detect blood cells in urine sediment images and achieved an F1 score of 91.4%. However, only two classes of red blood cells and white blood cells could be detected. Pan et al. [14] used convolutional neural networks to identify three classes of urine sediment particles and achieved 98.07% accuracy. Ji et al. [7] developed a particle recognition system based on AlexNet and the area feature algorithm (AFA). The network was trained by 300,000 images and finally achieved 97% accuracy. Liang et al. [10] identified urine sediment particles

in an end-to-end manner based on trimmed-SSD, Faster R-CNN [18], PVANet [8], vand Multi-Scale Faster R-CNN [4] models. The mAP reached 84.1% in 7 categories of urine sediment particles. Liang et al. [11] proposed DFPN based on FPN [21] by combining DenseNet, which can effectively eliminate the category confusion problem in urine sediment images and achieved 86.9% mAP. Yan et al. [26] used the bidirectional context propagation network BCPNet [2] for urine sediment particle detection, improving the localization and classification ability of the model. However, although the deep learning-based approach effectively overcomes the problem of complex segmentation and feature extraction, it requires a large number of images with manual annotation. In addition, these methods have many problems in practical application, such as operation speed and endpoint deployment.

2.2   Object Detection

Object detection is a core problem in the field of computer vision, where the goal is to isolate a specific object from the input image and obtain both the corresponding class and location information. Compared with classification tasks, object detection requires more stringent recognition ability of deep learning models, which need to fully understand the foreground and background information of the image to find the target of interest in the background.

To solve the above problem, Ross Girshick et al. [1] proposed a representative two-stage object detection model, R-CNN. R-CNN [1] divides the detection process into two independent phases, first extracting several candidate regions from the target image and then scaling each candidate region to a fixed size and feeding it into the CNN network for classification and recognition. Although R-CNN [1] significantly improves the performance of target detection tasks, the complex detection process and slow inference speed prevent it from being applied to terminals for real-time detection.

To achieve end-to-end object detection and improve detection speed, Joseph Redmon et al. [16] proposed the one-stage target detection model YOLO. It treats the object detection task as a regression problem, predicting the position and class of the object directly from the entire input image through a unified framework, significantly improving the detection speed. Compared to other object detection models, the YOLO series is faster and more scalable for a wider range of applications. Wang et al. [23] proposed a YOLOv5-CHE model dedicated to leukocyte image recognition based on YOLOv5, which achieved 99.3% mAP. Wang et al. [22] used the improved YOLOv5-P2 model to detect rebar ends, significantly improving detection accuracy on dense small targets. Yang et al. [27] developed a face recognition system using the YOLOv5 model to screen whether a mask is worn. The results substantially outperformed other classical object detection models to achieve 97.90% accuracy. Although the YOLO series is widely used in industrial and medical fields, there is still a gap in the study of YOLOv5-based urine sediment detection. Therefore, we improved the YOLOv5 model on two urine sediment particle datasets to make it suitable for urine sediment detection.
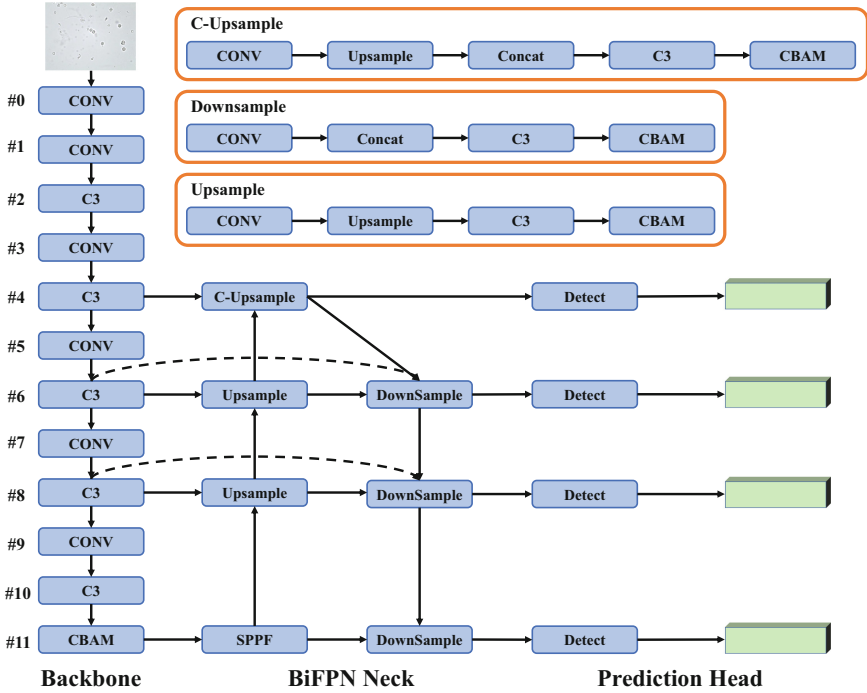
Fig. 1. The architecture of the proposed YOLOv5s-CBL. (1) CSPDarknet53 backbone with a CBAM module. (2) Neck using the structure of the BiFPN. (3) Feature map of four prediction heads using the CBAM module in neck.

3   Method

3.1   Overview of YOLOv5s-CBL

This work aims to develop an end-to-end urine sediment detection system. Therefore, we optimized the YOLOv5s model based on two urine sediment particle examination datasets and proposed a YOLOv5s-CBL model that is more suitable for urine sediment detection. The structure of YOLOv5s-CBL is shown in Fig. 1.

Prediction Head for Larger Object. The original YOLOv5 uses a three-scale feature layer to detect large, medium, and small targets in the dataset. However, we found that Dataset1 contains many larger-scale urine sediment particles. Therefore, the detection performance of YOLOv5 on these targets is not satisfactory. To improve the detection accuracy, we extended the original three-scale feature layer. We added 64x downsampling after 32x downsampling as a fourth prediction head to detect urine sediment particles at a larger scale. Figure 1 shows that the prediction head has a larger perceptual field and is more

sensitive to larger-scale targets. With the four-scale feature layer, our model makes full use of both shallow feature information and high-level semantic information to improve the detection performance significantly.
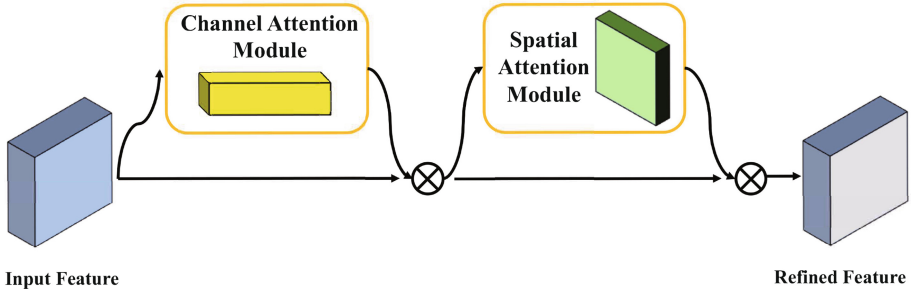

Fig. 2. The overview of CBAM.

CBAM. We found that urine sediment particles occupy only a tiny area of the image, so a large amount of background information would negatively affect the detection accuracy of the model. We integrated the CBAM [24] into YOLOv5 to help the model extract useful target information from the whole image and filter the interference of noisy information.

CBAM [24] is a simple and effective lightweight attention module that can be integrated into any CNN architecture for end-to-end training. As shown in the Fig. 2, CBAM [24] contains two submodules, CAM (Channel Attention Module) and SAM (Spatial Attention Module), which perform attention operations in the channel and spatial dimensions, respectively. The CAM module compresses the input image in the spatial dimension to make the model more focused on the meaningful information part of the image. The channel attention is calculated as follows:

$$
\begin{aligned}
M_c(F) &= \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F))) \\
&= \sigma(W_1(W_0(F_{avg}^c)) + \sigma(W_1(W_0(F_{\max}^c))
\end{aligned}
\tag{1}
$$

where σ denotes the sigmoid activation function, $W_0 \in R^{C/r \times C}$ and $W_1 \in R^{C \times C/r}$ denote the two weights of the MLP, respectively. Unlike the CAM module, the SAM module compresses the image in the channel dimension to obtain the target's location information and help the model find the region of interest from the whole image. The spatial attention is calculated as follows.

$$
\begin{aligned}
M_s(F) &= \sigma(f^{7 \times 7}([AvgPool(F); MaxPool(F)])) \\
&= \sigma(f^{7 \times 7}([F_{avg}^s; F_{\max}^s]))
\end{aligned}
\tag{2}
$$

where σ denotes the sigmoid activation function and $f^{7 \times 7}$ represents the convolution kernel size of 7×7.

BiFPN. Although the feature fusion approach of PANet [12] combined with FPN [21] has been widely used in deep learning models, it simply sums features at different scales after fixing them to a specific size. This approach ignores the difference in the degree of contribution of different resolutions to feature fusion. Inspired by EfficientDet, we replace the feature aggregation structure of YOLOv5 with the BiFPN [21] structure. Similar to PANet [12], BiFPN [21] is also based on a bidirectional feature aggregation path structure. By fusing the features of both bottom-up and top-down paths to enhance the characterization ability of the backbone network. Moreover, to differentially learn different input features, BiFPN [21] introduces an additional weight for each input to the network, removes nodes with only one input edge, and adds a jump connection between the input and output nodes at the same scale. Based on such improvements, the model can learn the importance of different input features in a targeted manner and obtain better feature representation capability. The results on two urine sediment particle detection datasets show that BiFPN [21] has better performance in urine sediment detection. The BiFPN [21] structure is shown in Fig. 3.
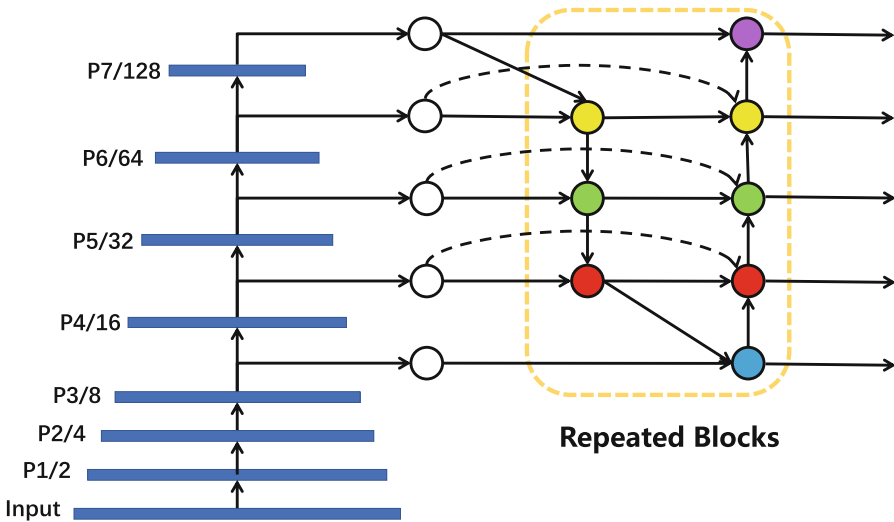

Fig. 3. The structure of BiFPN.

4 Experimental Results and Analysis

In this section, we first present the dataset, evaluation metrics, baseline methodology, and implementation details. Then, we compare with the latest baseline on Dataset1 and Dataset2 to validate the effectiveness of the proposed method.
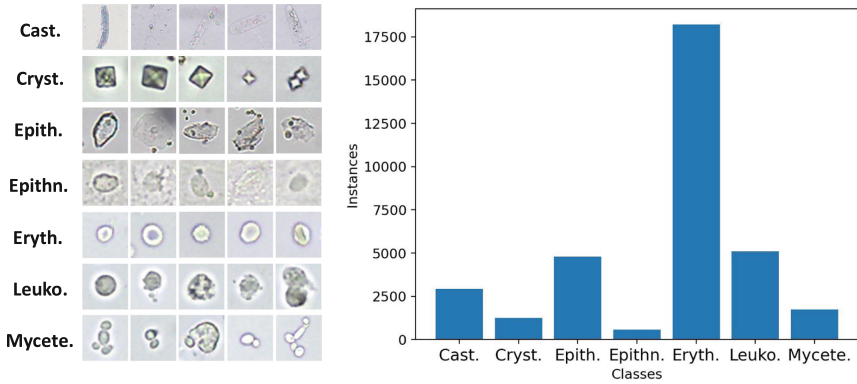
Fig. 4. Samples images and the number of labels of each category in Dataset1.

4.1   Dataset and Evaluation Metric

We evaluate the performance of the proposed method by using two real-world urinary sediment datasets, termed Dataset1 and Dataset2. Dataset1 was from the USE public dataset. [10,11,25] and consisted of 5646 images with a resolution of 800×600, containing seven cell categories: cast, cryst, epith, epithn, eryth, leuko, and mycete. The images and the corresponding number of labels of each category in Dataset1 are shown in Fig. 4; Dataset2 was from a self-built corporate dataset that contains 3200 urinary sediment images with the image size of 1024×1024 and 8 predefined example categories: RBC, WBC, SQEP, CAOX, OCRY, BACI, FUNGI, and MUCS. Figure 5 shows the eight categories of Dataset2 and the number of labels of each category.

In this paper, we use the mAP (mean Average Precision) as the evaluation metric. The mAP is the mean value of the average precision of all categories, which is often used to evaluate the detection accuracy of a model in object
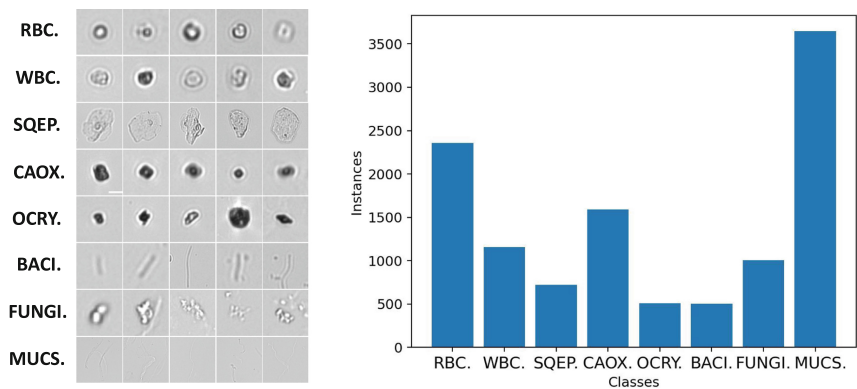

Fig. 5. Samples images and the number of labels of each category in Dataset2.

detection tasks, which is defined as follows:

$$mAP = \frac{\sum\limits_{i=1}^{C} AP_i}{C} \tag{3}$$

where AP denotes the average precision, defined as the area enclosed by the PR curve and the coordinate axis, C denotes the total number of categories detected.

4.2   Baseline Methods

We compare the proposed YOLOv5-CBL with the following state-of-the-art methods:

- **DFPN** [11] is an FPN [21] network with DenseNet as the backbone network, which is used to detect sediment particles in urine images.
- **BCPNet** [2] is a bi-directional context propagation network for real-time semantic segmentation, which enhances the localization and differentiation ability of the model by building a hybrid feature pyramid architecture that complements the spatial information of the higher-level features and the semantic information of the bottom-level features.
- **PVANet** [8] is an improved end-to-end model based on Faster R-CNN [18], which further improves the speed of detection while maintaining the accuracy of Faster R-CNN [18].

We also conducted experimental comparisons with YOLOv3 [17], YOLOv3-SPP [17], and YOLOv5s mainstream object detection models to validate the effectiveness of the YOLOv5s-CBL further.

4.3   Implement Details

In this paper, the experimental environment is Windows 11 64-bit operating system with the 12th Gen Intel(R) Core(TM) i9-12900H@2.50 GHz,1TB SSD, and NVIDIA GEFORCE RTX 3070Ti with 8 GB memory. All our models are based on Python 3.7 runtime environment, Pytorch 1.11.0, CUDA 11.6, and CUDNN 8.3.

In the training phase, we used stochastic gradient descent and cosine learning rate decay strategies to train the network, with initial learning rate 0.01, weight decay factor 0.0005, momentum factor 0.937, and batch size 16. Depending on the dataset we used a different number of epochs and input sizes, where the number of epochs and input size on Dataset1 is 200 and 800×800, while the number of iterations and input size on Dataset2 is 70 and 1024×1024.

In addition, we used Mosaic and image perturbation data enhancement strategies such as HSV-Hue augmentation, HSV-Saturation augmentation, HSV-Value augmentation, translate, scale, and flip during the training of all models.

Table 1. The comparison of the performance in Dataset1.

Method	mAP	Eryth	Leuko	Epith	Cryst	Cast	Mycete	Epithn
PVANet [8]	84.10	88.40	84.30	87.10	87.70	76.50	89.00	76.00
DFPN [11]	86.90	93.80	92.70	87.10	83.90	75.90	90.40	84.40
BCPNet [2]	88.20	**94.85**	**94.38**	87.99	84.66	74.57	90.71	90.21
YOLOv3-tiny [17]	88.50	76.10	90.70	88.60	86.20	91.10	95.70	90.90
YOLOv3 [17]	90.20	80.20	93.20	88.60	89.80	91.40	96.50	91.70
YOLOv5s	90.60	81.20	91.40	90.50	**90.80**	91.70	96.60	91.70
YOLOv5s-CBL	**91.70 0.17**	83.10 0.50	**91.970.45**	**92.67 0.32**	89.93 1.03	**91.87 0.25**	**97.40 0.10**	**95.00 0.96**

Table 2. The comparison of the performance in Dataset2.

Method	mAP	RBC	WBC	SQEP	CAOX	OCRY	BACI	FUNGI	MUCS
YOLOv3-tiny [17]	60.60	78.60	66.20	91.90	80.40	36.60	44.20	48.20	38.70
YOLOv3 [17]	66.50	76.80	**75.00**	92.7	**85.70**	39.40	56.60	59.10	46.40
YOLOv5s	67.4	**82.40**	67.50	90.50	83.60	40.90	**63.70**	**59.30**	51.10
YOLOv5s-CBL	**68.40 0.62**	80.53 1.01	73.60 1.47	**93.37 0.21**	81.53 0.38	**45.37 1.63**	63.57 3.43	56.9 3.36	**52.50 1.14**

4.4   Experimental Results and Analysis

In this section, we show the results of YOLOv5s-CBL on two real-world datasets and compare them with the state-of-the-art methods to demonstrate the effectiveness of the proposed method. The results of the comparison experiments are shown in Table 1 and Table 2. We conducted several experiments on Dataset1 and Dataset2 and took the average value as our final result.

Comparisons with the State-of-the-Art. Table 1 shows the results of state-of-the-art methods on Dataset1. It can be seen that the detection accuracy of the YOLOv5s-CBL is better than the existing state-of-the-art methods on Dataset1, reaching 91.7% mAP. Compared with PVANet [8], DFPN [11], and BCPNet [2], the mAP of YOLOv5s-CBL is improved by 7.6%, 4.8%, and 3.5%, respectively, while the mAP is improved by 1.1% compared with YOLOv5s. Note that the results in PVANet [8], DFPN [11], and BCPNet [2] were copied from the original papers.

To further verify the effectiveness of YOLOv5s- CBL, we compare YOLOv3-tiny [17], YOLOv3 [17], and YOLOv5s on Dataset 2. Unlike Dataset 1, Dataset 2 contains more small objects and the feature information is more difficult to capture, which poses a greater challenge to the model's performance. As can be seen from the results in Table 2, our method still achieves the best performance even on the tiny target urine sediment particle dataset. Compared with YOLOv5s, the mAP improved by 1.0%, proving that YOLOv5s-CBL performs more accurately and robustly than other models. We show the confusion matrix on Dataset1 and Dataset2 respectively, as shown in Fig. 6.

Ablation Studies. We evaluated the importance of each component by ablation experiments on Dataset1, as shown in Table 3.
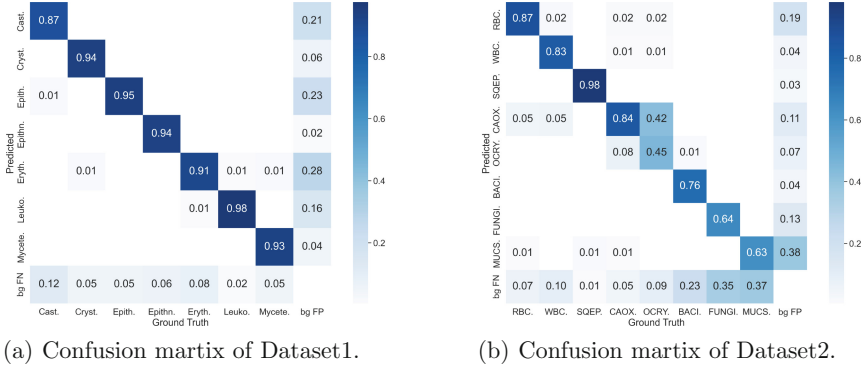
(a) Confusion martix of Dataset1.   (b) Confusion martix of Dataset2.

Fig. 6. Confusion matrix. (a)Confusion martix of Dataset1. (b)Confusion martix of Dataset2.

Table 3. Ablation on dataset1 with YOLOv5s.

larger-head	BiFPN	CBAM	mAP	cast	cryst	epith	epithn	eryth	leuko	mycete
			90.60	81.20	91.40	90.50	90.80	91.70	96.60	91.70
✓			91.40	83.20	**91.90**	92.30	88.80	91.90	97.00	94.40
	✓		90.80	80.50	92.00	91.30	89.10	91.60	97.20	93.90
		✓	91.30	80.80	90.70	92.00	**92.00**	**92.20**	96.40	95.30
✓	✓		91.50	82.30	91.50	**93.00**	89.70	92.10	97.50	94.60
✓		✓	91.40	82.70	91.20	91.70	90.30	91.70	97.40	94.60
	✓	✓	91.50	82.60	90.80	92.40	89.90	92.00	97.50	95.00
✓	✓	✓	**91.90**	**83.60**	91.50	92.90	90.20	92.10	**97.40**	**95.70**

Effect of Extra Prediction Head. We added 64x downsampling to YOLOv5s to extract deeper semantic information and used it as a larger-scale prediction head to capture larger-size urine sediment particles in the input image. As can be seen from Table 3, the larger-scale prediction head improves the mAP of the model by 0.7% and outperforms the YOLOv5s in almost all categories.

Effect of BiFPN. We replaced the feature aggregation structure of YOLOv5s with a lightweight BiFPN [21]. BiFPN [21] optimizes the feature fusion and enables the network to learn more critical feature information in a targeted manner by increasing the corresponding weights. The results in Table 3 show that the mAP of YOLOv5s with BiFPN [21] improves by 0.2% compared to YOLOv5s without BiFPN [21].

Effect of Attention Module. To reduce the impact of redundant background information on the detection accuracy of the model, we added the CBAM [24] to the YOLOv5s network for extracting the region of interest on the input image. The results proved that the CBAM [24] could effectively reduce the interference of background information and improve the mAP of the model, as shown in Table 3.

To further explore the impact of attention modules on the network, we compared the current mainstream attention modules SE [6] and Coordinate Attention (CA) [5]. The results in Table 4 show that CBAM [24] can locate and identify objects of interest in urine images more effectively than SE [6] and CA [5].

Table 4. The comparison of the performance in different attention module.

Method	mAP	Cast	Cryst	Epith	Epithn	Eryth	Leuko	Mycete
YOLOv5s-SE	91.30	**82.80**	**93.40**	92.10	89.10	92.10	97.30	92.20
YOLOv5s-CBAM	**91.50**	82.60	90.80	**92.40**	**89.90**	92.00	**97.05**	**95.00**
YOLOv5s-CA	90.90	81.10	90.50	92.00	88.80	91.80	97.30	94.90

Table 5. The comparison of the performance in different IoU loss function.

Method	mAP	Cast	Cryst	Epith	Epithn	Eryth	Leuko	Mycete
YOLOv5s-CBL+CIoU	**91.90**	83.60	91.50	**92.90**	90.20	**92.10**	97.40	**95.70**
YOLOv5s-CBL+EIoU	91.40	82.70	90.90	92.50	90.30	91.50	97.10	94.80
YOLOv5s-CBL+α-IoU	91.00	81.80	90.50	90.70	91.90	91.50	97.00	93.60
YOLOv5s-CBL+α-CIoU	91.40	82.20	**92.20**	90.10	**92.50**	91.60	97.20	93.90
YOLOv5s-CBL+α-EIoU	90.60	80.30	90.80	91.30	**92.50**	91.80	96.80	90.80

Effect of IoU. We explored the impact of IoU Loss on the network detection accuracy, and compared CIoU Loss, EIoU Loss [29], α IoU-Loss, α CIoU Loss, and α-EIoU Loss on the YOLOv5s-CBL, respectively, where α-CIoU Loss and α-EIoU Loss are replacing the IoU Loss in α-IoU Loss with CIoU Loss and EIoU Loss [29]. CIoU Loss is the default bounding box regression loss in the YOLOv5s, which further considers the overlapping area, centroid distance, and aspect ratio of the bounding box based on DIoU Loss, effectively improving the regression accuracy of the bound box; EIoU Loss [29] splits the aspect ratio influence factor based on CIoU Loss and calculates the width-height difference value separately, optimizing the problem of the difference between the width-height of the bounding box and the confidence; α-IoU is a Power IoU loss function proposed by Jiabo He et al. [3] at NeurIPS 2021, which contains a Power IoU term, a Power canonical term, and a Power parameter α. By adjusting α, the detector can adaptively increase the loss of high IoU objects and the weighting of the gradient to improve bounding box regression accuracy. Experiments on multi-target detection benchmarks and models show that α-IoU Loss [3] can significantly outperform existing IoU-based losses. Note that the Power parameter α of α-IoU Loss [3] is taken as 3 for all experiments in this paper. The results of the IoU Loss comparison experiments are shown in Table 5.

We found that the detection accuracy of YOLOv5-CBL based on CIoU Loss outperformed the models based on EIoU Loss [29] and α IoU Loss [3] on Dataset1. Therefore, we finally use CIoU Loss as the bounding box regression loss function.

Detection Result on Dataset1 and Dataset2. We have selected some images from the test set as the display of the detection results, as shown in Fig. 7.
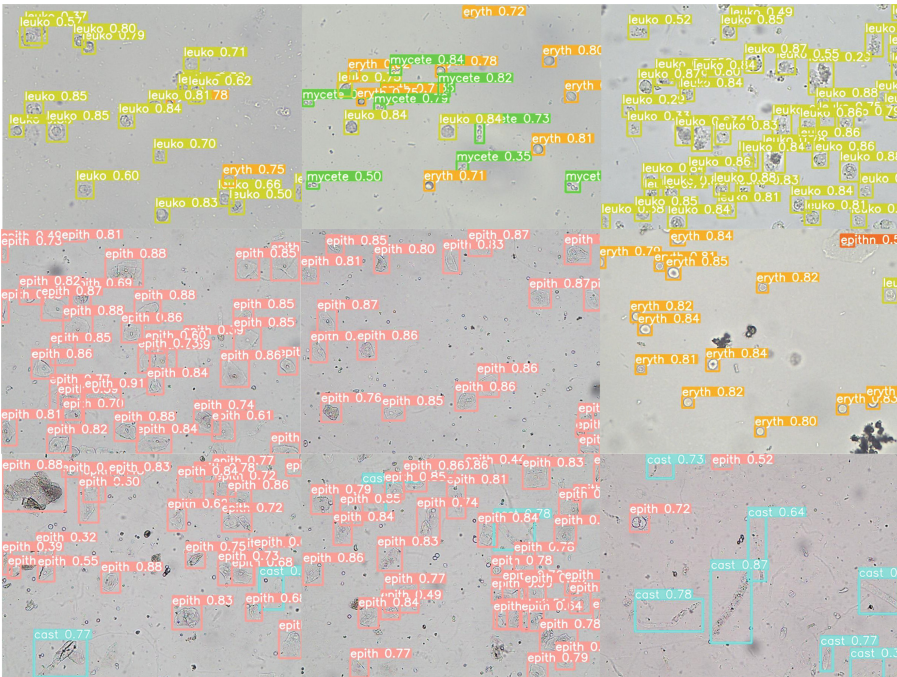

Fig. 7. Some visualization results from our YOLOv5s-CBL.

5   Conclusion

In this paper, we present the YOLOv5s-CBL model dedicated to urine sediment detection. The model integrates a larger-scale prediction head and CBAM to capture sediment particles in urine images. The features at different scales are then fused by BiFPN to improve the accuracy of urine sediment detection. We compared state-of-the-art methods on two urine sediment particle datasets, and the experimental results demonstrated the effectiveness of YOLOv5s-CBL.

References

1. Girshick, R., Donahue, J., Darrell, T., Malik, J.: Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2014
2. Hao, S., Zhou, Y., Guo, Y., Hong, R.: Bi-direction context propagation network for real-time semantic segmentation. arXiv preprint arXiv:2005.11034 (2020)
3. He, J., Erfani, S., Ma, X., Bailey, J., Chi, Y., Hua, X.S.: α-iou: a family of power intersection over union losses for bounding box regression. In: Advances in Neural Information Processing Systems 34 (2021)
4. Hoang Ngan Le, T., Zheng, Y., Zhu, C., Luu, K., Savvides, M.: Multiple scale faster-rcnn approach to driver's cell-phone usage and hands on steering wheel detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 46–53 (2016)
5. Hou, Q., Zhou, D., Feng, J.: Coordinate attention for efficient mobile network design. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 13713–13722 (2021)
6. Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7132–7141 (2018)
7. Ji, Q., Li, X., Qu, Z., Dai, C.: Research on urine sediment images recognition based on deep learning. IEEE Access **7**, 166711–166720 (2019)
8. Kim, K.H., Hong, S., Roh, B., Cheon, Y., Park, M.: Pvanet: deep but lightweight neural networks for real-time object detection. arXiv preprint arXiv:1608.08021 (2016)
9. Liang, Y., Fang, B., Qian, J., Chen, L., Li, C., Liu, Y.: False positive reduction in urinary particle recognition. Expert Syst. Appl. **36**(9), 11429–11438 (2009)
10. Liang, Y., Kang, R., Lian, C., Mao, Y.: An end-to-end system for automatic urinary particle recognition with convolutional neural network. J. Med. Syst. **42**(9), 165 (2018)
11. Liang, Y., Tang, Z., Yan, M., Liu, J.: Object detection based on deep learning for urine sediment examination. Biocybernetics Biomed. Eng. **38**(3), 661–670 (2018)
12. Liu, S., Qi, L., Qin, H., Shi, J., Jia, J.: Path aggregation network for instance segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 8759–8768 (2018)
13. Liu, X., Sun, Z.: A kind of computer microscopic urinary sediments analyzer by svm. In: 2008 International Workshop on Education Technology and Training & 2008 International Workshop on Geoscience and Remote Sensing, vol. 1, pp. 483–486. IEEE (2008)
14. Pan, J., Jiang, C., Zhu, T.: Classification of urine sediment based on convolution neural network. In: AIP Conference Proceedings, vol. 1955, p. 040176. AIP Publishing LLC (2018)
15. Ranzato, M., Taylor, P., House, J.M., Flagan, R., LeCun, Y., Perona, P.: Automatic recognition of biological particles in microscopic images. Pattern Recogn. Lett. **28**(1), 31–39 (2007)
16. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: unified, real-time object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 779–788 (2016)
17. Redmon, J., Farhadi, A.: Yolov3: an incremental improvement. arXiv preprint arXiv:1804.02767 (2018)

18. Ren, S., He, K., Girshick, R., Sun, J.: Faster r-cnn: Towards real-time object detection with region proposal networks. In: Advances in Neural Information Processing Systems 28 (2015)
19. Shen, M.l., Zhang, R.: Urine sediment recognition method based on svm and adaboost. In: 2009 International Conference on Computational Intelligence and Software Engineering, pp. 1–4. IEEE (2009)
20. Suhail, K., Brindha, D.: A review on various methods for recognition of urine particles using digital microscopic images of urine sediments. Biomed. Signal Process. Control **68**, 102806 (2021)
21. Tan, M., Pang, R., Le, Q.V.: Efficientdet: scalable and efficient object detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 10781–10790 (2020)
22. Wang, C., ZHANG, Y., Sun, S., Zhang, H.: Steel-bar end face detection based on improved yolov5 algorithm. Computer Systems and Applications, pp. 68–80 (2022)
23. Wang, J., Sun, Z., Guo, P., Zhang, L.: Improved leukocyte detection algorithm of yolov5. Computer Engineering and Applications, pp. 134–142 (2022)
24. Woo, S., Park, J., Lee, J.-Y., Kweon, I.S.: CBAM: Convolutional Block Attention Module. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds.) ECCV 2018. LNCS, vol. 11211, pp. 3–19. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-01234-2_1
25. Yan, M., Liu, Q., Yin, Z., Wang, D., Liang, Y.: A bidirectional context propagation network for urine sediment particle detection in microscopic images. In: ICASSP 2020–2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 981–985 (2020)
26. Yan, M., Liu, Q., Yin, Z., Wang, D., Liang, Y.: A bidirectional context propagation network for urine sediment particle detection in microscopic images. In: ICASSP 2020–2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 981–985. IEEE (2020)
27. Yang, G., Feng, W., Jin, J., Lei, Q., Li, X., Gui, G., Wang, W.: Face mask recognition system with yolov5 based on image recognition. In: 2020 IEEE 6th International Conference on Computer and Communications (ICCC), pp. 1398–1404. IEEE (2020)
28. Zhang, X., Chen, G., Saruta, K., Terata, Y.: Detection and classification of rbcs and wbcs in urine analysis with deep network (2018)
29. Zhang, Y.F., Ren, W., Zhang, Z., Jia, Z., Wang, L., Tan, T.: Focal and efficient iou loss for accurate bounding box regression. arXiv preprint arXiv:2101.08158 (2021)
30. Zhou, X., Xiao, X., Ma, C.: A study of automatic recognition and counting system of urine-sediment visual components. In: 2010 3rd International Conference on Biomedical Engineering and Informatics, vol. 1, pp. 78–81. IEEE (2010)