International Journal of Computer Science and Information Technology
ISSN: 3005-9682 (Print), ISSN: 3005-7140 (Online) | Volume 4, Number 3, Year 2024
DOI: https://doi.org/10.62051/ijcsit.v4n3.04
Journal homepage: https://wepub.org/index.php/IJCSIT/index

YOLOv5-DCN: An Effective Improvement Based on YOLOv5 Detector

Xiaoxia Qi [1, 2, 3], Md Gapar Md Johar [4], Ali Khatibi [3], Jacquline Tham [3], Long Cheng [2, *]

[1] School of Big Data and Artificial Intelligence, Anhui Xinhua University, Hefei 230088, China

[2] School of Electronic and Information Engineering, Anhui Jianzhu University, Hefei 230601, China

[3] School of Graduate Studies, Management and Science University, Shah Alam, Malaysia

[4] Software Engineering and Digital Innovation Centre, Management and Science University, Shah Alam, Malaysia

*Corresponding Author: longcheng@stu.ahjzu.edu.cn

ABSTRACT

Urine sediment detection is of great significance for the diagnosis and monitoring of kidney diseases, urinary tract infections, stones, etc. This study aims to propose fast, high-precision and lightweight Urinary particles detection model based on YOLOv5, deformable convolution, and evaluate its performance in Urinary particles detection tasks.

KEYWORDS

Urine Sediment; YOLOv5; Deformable Convolution

1. INTRODUCTION

Urine sediment detection is a commonly used clinical test method. It reflects the pathological changes of the kidney and urinary system by observing the cells, tube type, crystals and other components in the urine, and is of great significance for the diagnosis and monitoring of kidney diseases, urinary tract infections, stones, etc.

In order to improve the efficiency and accuracy of urine sediment detection, some methods combining AI technology with urine sediment detection have emerged in recent years. This type of urine sediment detection method mainly utilizes convolutional neural networks (CNN) or other deep neural network (DNN) models to classify, segment, and detect urine sediment images, thereby achieving automated, fast, and accurate urine sediment analysis.

2. GENERIC OBJECT DETECTOR

Object detection, as one of the important tasks of computer vision, has been developed for over twenty years [1]. From the early traditional methods based on manual features to the later methods based on deep learning, object detection algorithms have continuously improved the accuracy and speed of detection.

The focus of development in the field of object detection varies over time, resulting in the emergence of many classic object detection algorithms.

(1) YOLO

YOLO [2] algorithm produced an impressive impression in terms of speed and accuracy in detecting and determining object coordinates. The YOLO algorithm can be used for unmanned aerial vehicles, military, autonomous driving, hospitals, and other computer vision (CV) tasks. Over the years, YOLO has developed many other variants.

(2) The Dataset for urine sediment

This paper is based on existing experimental platforms and focuses on studying the situation of Urinary particles data. At present, there is a lack of publicly available large-scale standard urine image datasets, which also affects the improvement of urine particle target detection performance. According to our research, in 2021, Goswami et al. [3] published a urine microscope image dataset UMID, which includes 3 categories of cells namely red blood cells, pus cells, and epithelial cells, with a total of 367 images. The USE dataset includes 8 categories, namely red blood cells, white blood cells, crystals, casts, molds, and epithelial cells.

(3) Evaluation Indicators

The indicators for evaluating the detection algorithm mainly include: precision, recall, Average Precision (AP), Mean Average Precision (mAP), F1 Score, and Frame Per Second (FPS), Floating-Point Operations (FLOPs).

(4) Summary of Previous Research

Existing methods for urinary particle recognition have achieved satisfactory results to a certain degree (Table 1 shows the summary of previous research on urinary particles detecting using the deep learning method, these may not be comprehensive as some papers may use different keywords.), but they still suffer from limitations. Most of the current methods can only handle a few types of urinary particles, while the variety of urinary particles is rich, with at least eight major categories and different subclasses within each category. Urinary particles from different subclasses exhibit distinct appearance characteristics and medical diagnostic significance, thus posing great challenges for the automatic recognition of urine sediment images.

Table 1. Summary of improvements on YOLO

YOLO family	Improvement	Results
YOLOv1(Redmon et al., 2016)	Single shot detector combines and solves the problem of drawing boundary boxes and class identification.	Higher accuracy and speed compared to two-stage object detector such as Faster R-CNN.
YOLOv2(Redmon & Farhadi, 2017)	Iterative improvements on Batch Normalization, higher resolution detection and use of anchor boxes.	Reduction in architecture, faster detection and higher accuracy and better detection of high resolution images.
YOLOv3 (Redmon & Farhadi, 2018)	Addition of objectness score to bounding box prediction, added connections to the backbone network layers and predictions at three separate granularities.	Improves detection of smaller objects.
YOLOv4(Alexey et al., 2020)	Improved feature aggregations, bag of freebies with mosaic augmentations and use of mish activation.	Achieved improved accuracy and ease of training, high quality performance and accessibility.
YOLOv5 (Nepal & Eslamiat, 2022)	Reduced network parameters, use of Cross Stage Partial Network (CSPNet) for the head, PANet for the neck of the architecture, residual structure and auto-anchor. It also utilizes mosaic augmentations.	Extremely easy to train, inference on individual, batch images, video feed and webcam ports. Ease of transfer and use of weights. Faster and more lightweight than previous YOLO.
YOLOv6 (C. Li et al., 2022)	Redesigned network backbone and neck to EfficientRep Backbone and Rep-PAN Neck. The Network head is decoupled separating different features from the final head.	Improvement in detecting small objects, anchor free training of model. Less stable and flexible as compared to YOLOv5.
YOLOv7(C.-Y. Wang et al., 2020)	Layer aggregation using E-ELAN, trainable bag of freebies, 35% fewer network parameters. Model scaling for concatenation-based model.	Increase in speed and accuracy, ease of training and inference.
YOLOv8(G. Jocher et al., 2023)	The Head section use the current mainstream decoupling hea d structure, separating classification and detection heads, and also replacing it from Ancho Based to Anchor Free	Supports high -resolution image detection, Improving detection accuracy at the cost of FLOPs

3. THE APPLICATION VALUE OF DEEP LEARNING METHODS IN URINE PARTICLE DETECTION

This paper presents a deep learning method for detecting urinary particles. We designed a preliminary urinary particle detection algorithm based on the general object detection framework and analyzed its strengths and weaknesses. To address the challenges of detecting urinary particles, such as small size, low contrast, and similar morphology, we proposed some improvements to the algorithm and evaluated their effectiveness. We compared our improved algorithm with other object detection methods on different datasets and demonstrated its superiority and generality.

The object of this study is urinary particles, including red blood cells, white blood cells, epithelial cells, tubular cells, crystals, etc. The method of this study is based on YOLOv5, combined with deformable convolution Urinary particles detection model. The specific steps are as follows:

(1) Using YOLOv5 as the basic model, utilizing its efficient and accurate target detection capabilities;

(2) Replace some standard convolutions in YOLOv5 with deformable convolution to enhance the model's transformation modeling and feature fusion capabilities;

There are deformation and attitude change phenomena commonly existing in the detection target in the Urinary particles dataset. This study introduces deformable convolution into the backbone network of YOLOv5. The deformable convolution is used to enhance the adaptability of the network to Geometric transformation, improve the expression ability of the feature map, and better capture the fine-grained features of the target object, thus improving the detection accuracy.

Given the processing mode of most detection algorithms at present, the backbone network will use different size blocks to extract and process features according to each down sampling stage. A backbone network is a neural network composed of multiple convolutional modules used to extract features from input images. Among them, there are three convolutional modules called C3 modules, which are located in the second, third, and fourth positions of the backbone network. Each C3 module contains three convolutional layers that perform convolution operations on input features to obtain output features. The core of convolution operation is a convolution kernel, which is a small matrix used for weighted summation of a local region of input features. Different convolutional kernels can have different sizes, namely the number of rows and columns in the matrix. In each C3 module, the convolutional kernel size of the first and second convolutional layers is 1x1, which means that only weighted summation is performed on each position of the input features, without considering the relationship between positions. The convolution kernel size of the third convolutional layer is 3x3, which means that each position of the input feature and its surrounding eight positions are weighted and summed to consider the relationship between positions.

In order to further improve the detection effect of YOLOv5,based on the YOLOv5, we replaced the ordinary convolution in the second, third and fourth c3 modules of the backbone network structure to a deformable convolution operator that can better meet the diversity requirements of visual task objectives, that is, the sampling position of the convolution kernel can be shifted adaptively according to the shape and content of the input features, so as to improve the flexibility and expression ability of convolution, as shown in figure 1.

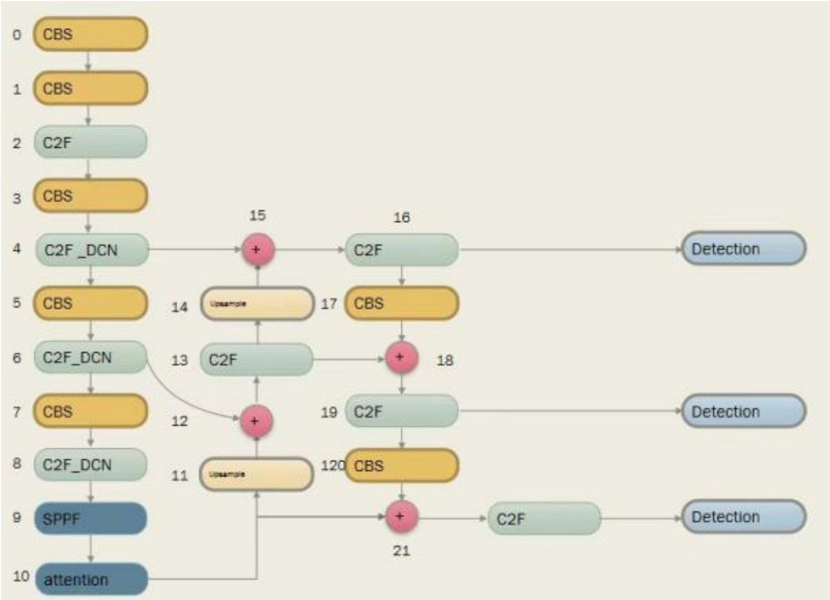

Figure 1. The Network Structure of Deformable YOLOv5

(4) The data source of this study is the urine image dataset, which includes normal and abnormal samples, with each sample containing urine images and corresponding annotation files;

(5) Train, validate and test the proposed YOLOv5-DCN model using the constructed dataset;

(6) Evaluate the proposed Model performance.

4. EXPERIMENTAL ANALYSIS

Our experiment was conducted on a Windows 10 desktop computer equipped with a GeForce RTX 4070 Ti GPU (24G graphics memory). The software environment is Windows 10, Intel (R) Core (TM) i7-14700KF, Python 3.11, and PyTorch 1.12.1. Using the SGD optimizer for training, the bounding box regression loss is selected as CIoU Loss, and the batch size for each graphics card is set to 12. The default configuration of Yolov5 is basically used: the initial learning rate is 0.01, the learning rate attenuation is set to 0.0005, the training cycle is 150 rounds, and the input image size is 640 x 640.

In our experimental environment, the mAP50-95 of the original YOLOv5s reached 51.8, and the mAP50-95 value of YOLOv5-DCN increased by 0.3, indicating that the YOLOv5-DCN model performs more robustly and has better tolerance for different IoU thresholds. In addition, our model's values increased by 1.6 and 0.8 at IoU thresholds of 0.5 and 0.75, respectively. This means that YOLOv5-DCN can provide higher accuracy than YOLOv5s when the predicted bounding box overlap rate with the actual target is 50% or even higher. The introduction of deformable convolution and double-layer routing attention mechanism has significantly improved the overall performance of our model. Although the addition of both increases the computational and parameter complexity of the model, they are both within a reasonable range.

However, YOLOv5-DCN has achieved significant improvement in FPS, with an increase of 11.88 compared to YOLOv5s, which can meet the requirements of high-performance real-time detection and has more advantages in real-time processing applications compared to YOLOv5s.

5. CONCLUSION

Overall, the application prospects of computer vision are broad, and object detection algorithms still face enormous challenges in solving practical problems.

In order to verify the universality of the YOLOv5-DCN model, we conducted effective validation on the Urine particle dataset and found a significant improvement in detection accuracy and efficiency, particularly in detection speed, which has met the detection speed requirements of high-performance devices.

In the future, we will further research the application of our method on more detection datasets and explore optimization methods for object detection algorithms. Of course, we will continue to expand or subdivide the urine residue dataset, providing more comprehensive analytical basis for clinical diagnosis in the medical field in the future.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the financial support from Discipline Professional Leader Cultivation Project (DTR2024056), Anhui Province Quality Engineering "101 Plan" Project (2023ylyjh059).

REFERENCES

[1] Zou Z, Shi Z, Guo Y, et al. Object Detection in 20 Years: A Survey [J]. 2019.

[2] Redmon J, Divvala S, Girshick R, et al. You Only Look Once: Unified, Real-Time Object Detection[C]//Computer Vision & Pattern Recognition. IEEE, 2016.

[3] Goswami D, Aggrawal H O, Gupta R, et al. Urine Microscopic Image Dataset [J]. 2021.

[4] Dai J, Qi H, Xiong Y, et al. Deformable Convolutional Networks [J]. IEEE, 2017.

[5] Ji Q, Jiang Y, Qu Z W L. An Image Recognition Method for Urine Sediment Based on Semi-supervised Learning [J]. Innovation and research in biomedical engineering: IRBM, 2023, 44(2):100739.1-100739.9.

[6] Terven J, Cordova-Esparza D M, Romero-Gonzalez J A. A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS [J]. Machine Learning and Knowledge Extraction, 2023, 5(4):1680-1716.