

# Improved Immunohistochemistry Active Cell Counting Method for YOLOv5s

Xingyue Chen<sup>1,a</sup>, Ziyang Jia<sup>1\*</sup>, Qing Li<sup>2,b</sup>, Dachuan Zhang<sup>2,c</sup>, Lingjiao Pan<sup>1,d</sup>, Dawei Shen<sup>1,e</sup>

<sup>1</sup>School of Electrical and Information Engineering, Jiangsu Institute of Technology, Changzhou, Jiangsu, China

<sup>2</sup>Department of Pathology, Changzhou First People's Hospital, Changzhou, Jiangsu, China

**Abstract:** This article proposes an improved YOLOv5s counting method to address the problems of long-term manual counting of positive cells in immunohistochemical images and low consistency. First, by introducing the Triplet attention module, the model focuses on the positive cell area, reducing background interference and improving the network's ability to extract positive cell features; then, a small target detection layer is added to better utilize the semantic information of the network to improve positive cells recognition accuracy; then, the lightweight up-sampling operator CARAFE is used to improve the quality and accuracy of up-sampling; finally, the WIoU loss function is used to replace the original GIoU of YOLOv5 to enhance model detection performance. Experimental results show that the improved model has an average accuracy of 88.4%, which is 3.1% higher than the original YOLOv5 network model. It can count positive cells quickly and accurately, reducing the workload of doctors.

## 1. Introduction

Deep learning has achieved great success in the field of computer vision and has been widely used in the field of medical image analysis [1]. Target detection algorithms based on deep learning can automate the analysis of medical images to automatically identify and locate cells or lesions in the images, thus reducing the workload of doctors and improving the efficiency and accuracy of diagnosis. Cell counting, as a method of quantitatively detecting the number of cells, can be used to determine the proliferation rate of cells based on the number of cells, to explore the process of cell development, and to explain the development of diseases, etc. In 2019, Falk et al. [2] used the Unet network for the segmentation of cellular images, which in turn realizes the tasks of cell detection and counting. In 2020, Xu [3] proposed a YOLO-Dense multiscale fusion cell detection and counting model, which used the multiscale prediction idea of a feature pyramid, while combining the residual module and dense module to improve the accuracy and real-time performance of blood cell detection and counting. In 2021, Zingman et al. [4] applied a deep neural network-based single-shot detector to the task of detecting tip cells in retinal images and realized the automatic counting of tip cells. In 2022, Cui [5] combined a blood cell classification model combining the improved YOLOv5 network with the concave point detection method and a leukocyte classification model with an improved convolutional neural network to form a

blood cell counting model.

It can be seen that the development of artificial intelligence in the field of medical diagnosis is very rapid. Compared with the cells in the above studies, the positive cells in immunohistochemistry images have the characteristics of a small area and are difficult to identify, and there has not been any study for immunohistochemistry positive cell counting. This article improves the YOLOv5s model to realize the automatic counting of positive cells, solving the problem of time-consuming and poor consistency of manual counting, and providing new ideas and methods for research in the field of medical image analysis.

## 2 Improved positive cell counting model for YOLOv5s

### 2.1. YOLOv5 target detection algorithm

YOLO is a classic single-stage target detection model [6], which can transform the target detection problem into an end-to-end regression problem. YOLOv5 was developed by the Ultralytics team and is divided into four models by size, YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x. YOLOv5s has the smallest network depth and model width, given that only the category of positive cells is recognized in this study, YOLOv5s-6.0 version, which has less number of parameters and computation, is used as the base model, and its network structure is shown in Fig. 1, which mainly consists of four parts: the

<sup>a</sup>643395909@qq.com

\*Corresponding author: jiazhiyan@jst.edu.cn

<sup>b</sup>liqblk@163.com, <sup>c</sup>zhangdachuan@suda.edu.cn,

<sup>d</sup>jsjshedy@jst.edu.cn, <sup>e</sup>sdw@jst.edu.cn

input (Input), the backbone network (Backbone), the neck network (Neck), and the detection layer (Head). Among them, the Input is used to input the image into the network and perform preprocessing operations, including image resizing, Mosaic image enhancement, and adaptive anchor frame computation; the Backbone network adopts the CSPDarknet53 structure for image feature extraction; the Neck layer is used for further fusion and processing of feature maps extracted by the Backbone network; the Head layer contains three Detect detectors, which output respectively feature maps of sizes 20\*20, 40\*40 and 80\*80, corresponding to targets of 32\*32, 16\*16 and 8\*8 pixels. The detectors detect the feature maps at different scales and finally obtain the target location and category information.

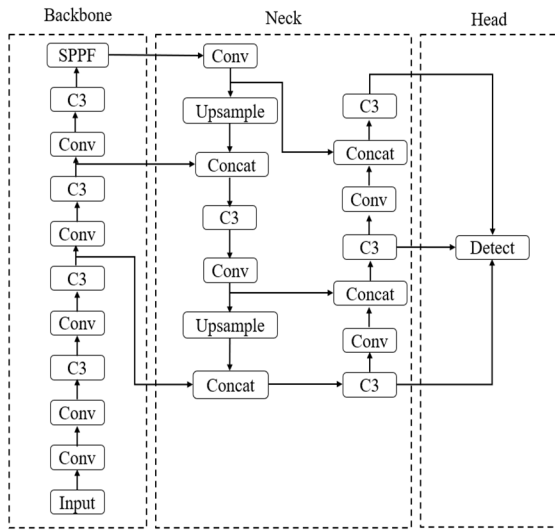


Fig. 1. YOLOv5s network architecture.

## 2.2. Addition of Triplet Attention Module

When pathology sections are stained immunohistochemically, the relatively small size of the positive cells and the background region occupying most of the image space, coupled with the lighter staining of certain cells similar to the background, make it difficult for positive cells to be accurately identified. In this paper, the Triplet attention module is introduced into the YOLOv5s neck network to help the model pay better attention to the region where the cells are located in the image and learn the features related to the positive cells, thus reducing the interfering factors and improving the detection accuracy of the positive cells. The structure of the Triplet module is shown in Fig. 2.

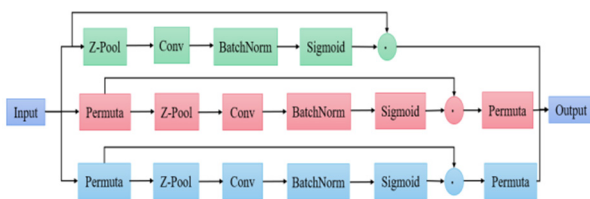


Fig. 2. Triplet module structure.

The Triplet attention module contains three parallel branches, two of which are used to capture information about interactions between channels and features in spatial dimensions H or W, while the last branch is used to construct spatial attention, and the outputs of these three branches are aggregated by taking the average. The Z-Pool layer in the figure serves for dimensionality reduction, feature extraction, and depth reduction by combining average pooling and maximum pooling on the tensor. Its formula is shown in (1-3):

$$Z - Pool = [AvgPool_{0d}(x), MaxPool(x)] \quad (1)$$

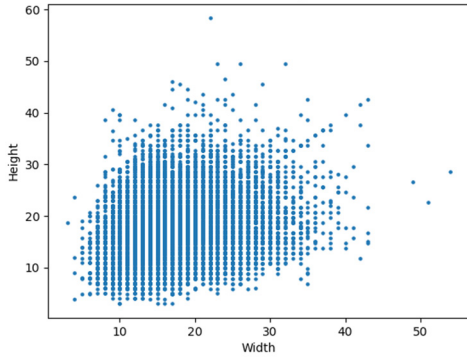
$$AvgPool(x) = \frac{1}{W \times H} \sum_{i=1}^H \sum_{j=1}^W x_{i,j} \quad (2)$$

$$MaxPool = \max_{H,W}(x) \quad (3)$$

Where, W represents the width of the feature map and H represents the height of the feature map. The Triplet Attention Module combines channel attention and spatial attention by introducing the concept of cross-dimensional interactions, thus compensating for the lack of independence between these two in the traditional attention mechanism. The richness enhancement of feature representations in different dimensions and the capture of cross-dimensional interaction information are realized through three branches that deal with the relationship between the channel dimension, the height dimension, and the width dimension, respectively. Ultimately, the outputs of these branches are combined by weighted averaging to improve the model's ability to learn and represent complex features.

## 2.3. Add a small target detection layer

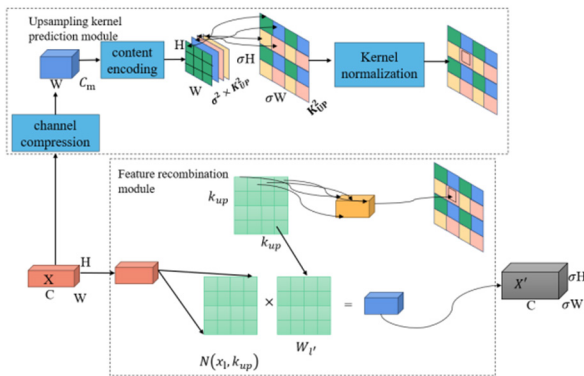
The size of the YOLOv5 output feature layer is usually 1/32, 1/16 and 1/8 of the size of the input feature map, i.e. if the size of the input feature map is 640× 640, then the corresponding output feature layers are 20, 20, 40, and 20 respectively. × 20, 40× 40 and 80× 80. In this paper, the size distribution of the a priori frames of the dataset obtained by the K-Means clustering method is shown in Fig. 3, and it can be seen that the heights of the a priori frames are mainly concentrated in the smaller size regions, indicating that the positive cells are tiny targets. In this case, the original YOLOv5 detection scale may not perform well because relying on deep features for small target detection may loss of some detailed information. Therefore, based on the a priori information of the length and width of the positive cells, this paper adds a small target feature layer with an input size of 1/4, corresponding to a 160× 160 detection feature map for detecting targets of size 4× 4 and above targets. This can better utilize the semantic information of the network to improve the recognition accuracy of positive cells, and further optimize the detection effect by fusing with the feature maps of other layers.



**Fig. 3.** Distribution of target sizes in the dataset.

### 2.4. Using the lightweight up-sampling operator CARAFE

CARAFE is an up-sampling operator [7]. Compared with the original interpolation method of YOLOv5, the CARAFE operator can effectively transmit and diffuse the information in the feature map and better capture positive cells by expanding the range of the receptive field. Detailed information to improve the positioning and classification capabilities of small targets. The CARAFE operator consists of an up-sampling core prediction module and a feature reorganization module, as shown in Fig. 4. The up-sampling core prediction module is used to up-sample the input feature map and predict and calculate features. So that the subsequent feature reorganization module can better utilize this information; the feature reorganization module accepts the feature map from the up-sampling core prediction module and reorganizes and integrates the features to ensure that the feature map can better retain semantic information.

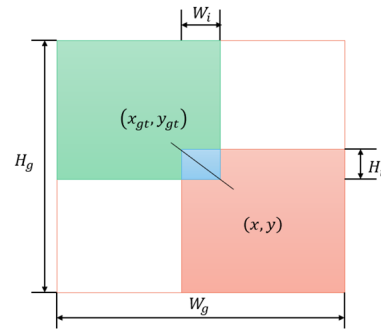


**Fig. 4.** CARAFE module structure.

### 2.5. Introduction of the WIoU loss function

YOLOv5 adopts GIoU as the loss function, although compared with the traditional IoU loss function, it introduces the minimum outer bounding box, which solves the problem that the loss is equal to zero when there is no overlap at all between the prediction box and the real box, but when there is a containment relationship between the two boxes, the GIoU loss function still fails to change the loss value. In the case of a dense distribution of positive cells and more overlapping

bounding boxes, the GIoU loss function can not effectively differentiate, resulting in the model's difficulty in accurately locating and identifying positive cells during training. In this regard, this paper adopts the WIoU loss function [8] to replace the original loss function, and by introducing the weights, the loss between different targets can be balanced to solve the limitations of the GIoU loss function in dealing with the inclusion relationship and dense targets. The WIoU loss function for bounding box regression is shown in Figure 5:

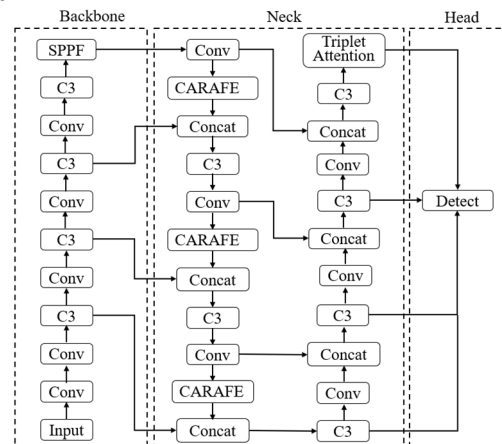


**Fig. 5.** Plot of WIoU loss function.

The formula is shown in (4):

$$\mathcal{R}_{WIoU} = \exp \left( \frac{(x - x_{gt})^2 + (y - y_{gt})^2}{(W_g^2 + H_g^2)^*} \right) \quad (4)$$

Where  $W_g$  and  $H_g$  denote the width and height of the smallest outer rectangle of the bounding box of the real target,  $W_i$  and  $H_i$  are the width and height of the intersection part of the predicted bounding box and the bounding box of the real target, respectively, and  $x_{gt}$  and  $y_{gt}$  denote the values of the horizontal and vertical coordinates of the centre point of the bounding box of the real target. The WIoU integrally takes into account the intersection of the real frame with the predicted frame and at the same time can flexibly adjust the size and position of the predicted frame according to the needs. and position, thus reducing the error between the predicted frame and the real frame. The network structure of the improved YOLOv5s model is shown in Fig. 6.



**Fig. 6.** Improved network structure.

### 3. Experimental data preparation

#### 3.1. Data set production

A total of 1012 CD8 immunohistochemistry staining images were collected, and after converting the image saving format to JPG, the images were randomly divided into training set, validation set and test set according to 7:2:1, with the resolution size of 1916×995. Rectangular boxes were drawn manually using the annotation tool Labeling to label the positive cells in the captured images, to get the positive cells' number and location information in the immunohistochemistry images. A sample image of labeling is shown in Fig. 7:



Fig. 7. Labeling diagram of positive cells.

#### 3.2. Data pre-processing

In this paper, the data preprocessing methods of color normalization and data enhancement are used to improve the image quality and expand the dataset, respectively, and the specific flowchart is shown in Fig. 8:

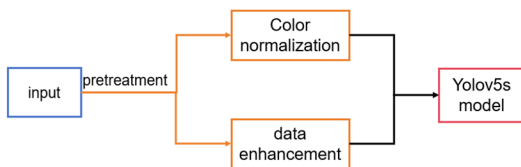


Fig. 8. Flowchart of pathology section image preprocessing.

All the pathology section images collected in this study were stained based on immunohistochemistry, different hospitals may use different production processes and operation methods, resulting in some degree of difference in the staining results. Therefore, in this paper, color normalization is adopted for pathology slice images to make different pathology slides have similar color characteristics, to reduce the occurrence of misdetection and missed detection. The method used in this paper is based on an unsupervised decomposition-based color normalization method proposed by A. Vahadane et al. The method decomposes the image into sparse and non-negative stain density maps and combines the normalized stain density maps with the stain color base of the target image presumed by the pathologist to achieve the effect of changing the image color while keeping its structure unchanged. As shown in Fig. 9:

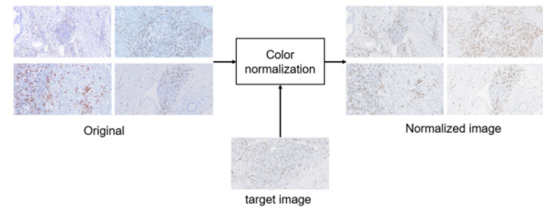


Fig. 9. Comparison before and after color normalization.

Due to the large number of positive cells in the immunohistochemistry images, the annotation workload is large, and the number of datasets is limited at the early stage of the experiment, so this paper adopts the image enhancement operation to expand the dataset. As shown in Fig. 10, it mainly includes two operations, rotation and mirroring. Rotation is to rotate the original image and its labeled data randomly, with a rotation angle between 0 and 30°. Mirroring is flipping the original image and its labeled data in left and right mirrors.

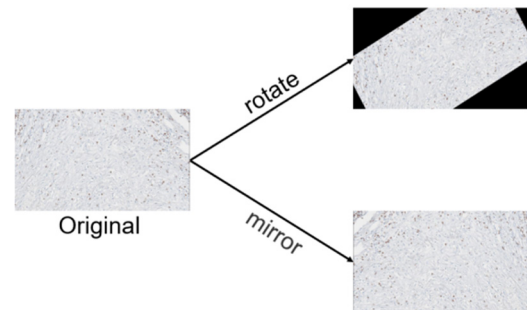


Fig. 10. Image enhancement effect diagram.

### 4. Experimental analysis

#### 4.1. Experimental parameters and evaluation indexes

The version of the experimental server used in this paper is Windows Server 2016 Standard, the processor is Intel(R) Xeon(R) Silver 4310 CPU @2.10GHz 3.30GHz (2 processors), the language is Python 3.8.5, and the deep learning network framework is Pytorch. The model The training epochs are set to 300 rounds, the batch\_size is 8, and the initial learning rate is 0.05. In this paper, we use average precision (AP) as the evaluation index of the model performance, and the AP value is the area of the curve plotted with the axes of the accuracy rate P(precision) and the recall rate R(recall). P, R is calculated by the following formula:

$$Precision = \frac{TP}{FP + TP} \times 100\% \quad (5)$$

$$Recall = \frac{TP}{TP + FN} \times 100\% \quad (6)$$

TP is the number of correctly identified positive cells, FP is the number of background impurities recognized as positive cells, and FN is the number of positive cells

recognized as background.

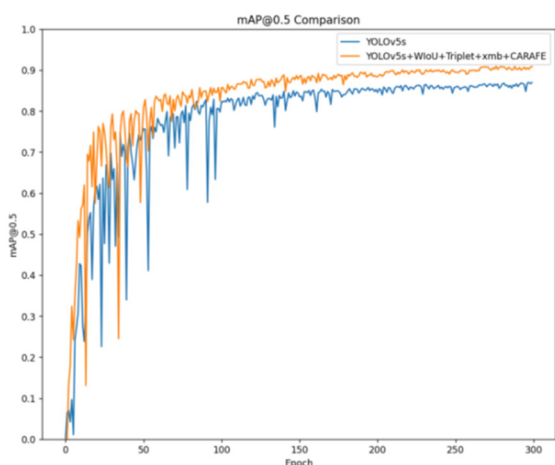
### 4.2. Ablation experiments

To verify the effectiveness of the improved model, ablation experiments were conducted based on the YOLOv5s model, and higher AP values indicated better performance of the target detection algorithm. From Table 1, it can be seen that after adding the Triplet attention mechanism, the model mAP value is improved by 1.4%, and the model can recognize positive cells more accurately. Adding the small target detection layer has the greatest impact on the accuracy improvement, which reaches 3.2%. After using the WIoU loss function, the

recall and accuracy of the model detection improved by 3.9% and 0.3%, respectively, and the mAP value improved by 1.8%. Up-sampling by the CARAFE operator enhanced the extraction of cell detail information, and the model AP/% was improved by 0.9%. The average accuracy of the improved model reached 89.1%, the detection accuracy was improved by 3.8% compared with that of the original YOLOv5s, and the detection speed reached 29.48 frame-s<sup>-1</sup>. This indicates that the improved model can effectively assist doctors in carrying out immunohistochemical counting. The map@0.5 comparison curve of the two model training before and after improvement is shown in Fig. 11.

**Table 1.** Ablation test results.

basic model	WIoU	Triplet	Small target detection layer	CARAFE	P/%	R/%	AP/%
YOLOv5s					85.9	77.6	85.3
√	√				86.2	81.5	87.1
√		√			86.7	79.4	86.7
√			√		87.4	83.1	88.5
√				√	86.8	79.3	86.2
√	√	√	√	√	87.8	84.4	89.1



**Fig. 11.** Comparison of mAP@0.5.

### 4.3. Comparison experiments of different detection models

The model in this paper is compared with the current mainstream target detection models on a homemade immunohistochemistry image dataset, and the results are shown in Table 2. From Table 2, it can be seen that SDD, Fast RCNN and YOLOv4 models generate weight files with large size and low accuracy, which are not suitable for cell counting; The AP values detected by the yolov7 and yolov8 models are higher than those of the previous models, but the weights are larger from Table 2, it can be seen that the original YOLOv5s model has an average accuracy of 85.3%, which is higher than the other models. From the test results, it can be seen that the model detection results in this paper have been significantly improved, with an average accuracy rate of 89.1%.

**Table 2.** Performance comparison of multiple algorithms.

mould	backbone	AP/%	size/MB
SDD	VGG16	62.14	110.67
Fast-Rcnn	ResNet-50	60.06	108.01
YOLOv4	CSPDarknet53	55.73	244.58
YOLOv7	CSPDarknet53	76.3	71.32
YOLOv8	CSPDarknet53	68.2	21.47
Article model	CSPDarknet53	89.1	17.16

### 4.4. Cell count consistency and correlation test

To verify the validity of the model interpretation results, 60 CD8 immunohistochemical staining images were randomly selected for experiments: the median was selected as the cut-off value for grouping, and 30 cases were in each of the high- and low-expression groups, and the consistency and correlation between the improved YOLOv5s model counting interpretation and the manual interpretation, and ImageJ area interpretation of the CD8 staining results were observed. The results are shown in Table 3:

**Table 3.** Statistics of reading results.

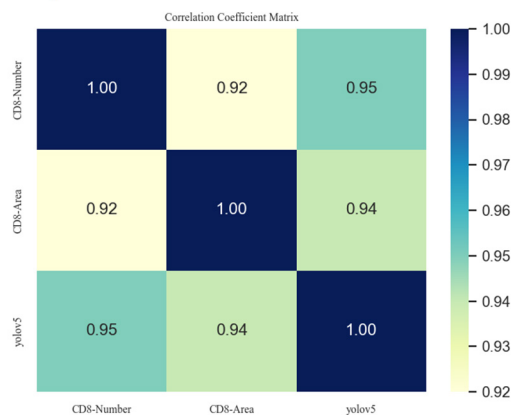
	model of this paper	Manual counting of CD8 numbers	CD8 area
minimum value	48.00	42.00	2,410.00
maximum values	1,345.00	1,457.00	120,874.00
upper quartile	419.00	217.40	17,514.00

SPSS25.0 was applied to analyze the data, and the consistency was judged by the Kappa coefficient, which was shown in Table 4 that the consistency of the scores between the model of this paper and the CD8-Number interpretation was better ( $\kappa=0.867$ ,  $P<0.01$ , Table 4) and higher than that of the CD8-Area ( $\kappa=0.756$ ,  $P<0.01$ , Table 4).

**Table 4.** Improved YOLOv5s model counts versus manual interpretation and ImageJ interpretation of the test of oneness.

		Manual counting of CD8 numbers			
		high	low	Kappa	p-value
model of this paper	high	28	2	0.867	<0.01
	low	2	28		
CD8 area	high	26	4	0.732	<0.01
	low	4	26		

The figures shown in the box of Fig. 12 are the intra-group correlation coefficients (r) between the model interpretation results of this paper and the manual interpretation and Imagej area interpretation, and from the figure, it can be seen that the improved YOLOv5s model has a significant positive correlation with the manual interpretation ( $r=0.95$ ,  $P<0.01$ , Fig. 12) and the correlation coefficient is higher than that of the ImageJ area interpretation.



**Fig. 12.** Correlation analysis of improved YOLOv5s model count interpretation with manual interpretation and ImageJ interpretation.

## 5. Conclusion

In this paper, we analyzed the needs of immunohistochemistry positive cell counting and proposed a method to improve the YOLOv5s model, which achieved an average accuracy of 89.1%, an improvement of 3.8% over the original model. After analyzing and validating, The improved model is highly consistent with manual interpretation and was superior to ImageJ area interpretation, which can significantly improve the efficiency of the counting work, and solve the problem of time-consuming and poor consistency of manual counting. In subsequent research, other

lightweight models can be used for optimization, or pruning and distillation of the model to improve the response speed and efficiency of the system.

## Acknowledgments

National Natural Science Foundation of China (62001196);

Jiangsu Province “333 High-level Talent Training Project” project (No.2022-3-4-107);

Changzhou Science and Technology Planning Project (CM20223015);

Changzhou Applied Basic Research Project (CJ20220064, CJ20220059)

## References

- GORE J C. (2020) Artificial intelligence in medical imaging. Magnetic Resonance Imaging. Elsevier, A1-A4.
- Falk T, Mai D, Bensch R. (2019) U-Net: deep learning for cell counting, detection, and morphometry. Nature Methods, 16(1): 67-70.
- Xu X T. (2020) Research on cell counting based on deep target recognition. Hefei. Anhui University, 5-22.
- Igor Z, Nina Z, Gerald B. (2021) Deep Learning-Based Detection of Endothelial Tip Cells in the Oxygen-Induced Retinopathy Model. Toxicologic Pathology, 49(4): 862-871.
- Cui Z W. (2022) Research on blood cell counting based on deep learning. Guizhou: Guizhou University, 59-64.
- LI K Y, OU O, LIU G B. (2023) Target detection algorithm of remote sensing image based on improved YOLOv5. Computer Engineering and Applications, 59(9): 207-214.
- WANG J Q, CHEN K, XU R. (2019) CARAFE: content-aware reassembly of features [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Seoul: IEEE, 2019: 3007-3016. DOI: 10.1109/ ICCV. 2019. 00310.
- TONG Z J, CHEN Y H, XU Z W. (2023). Wise-IoU: Bounding box regression loss with dynamic focusing mechanism [EB/OL].