# Design and implementation of Garbage Detection in Water Area Based on Yolov5 Algorithm

Jiajia Meng[a], Liuling Lang[b], Zhenzhan Lu[c], Xiling Tang[d], Huimin He[e*]

{E-mail: [a]486976596@qq.com, [b]651160968@qq.com, [c]2687564580@qq.com, [d]2308058522@qq.com, [e]hehuimin@gxmu.edu.cn;}

School of Information and Management, Guangxi Medical University, Nanning, China

**Abstract.** The target detection of water waste helps to timely discover and warn the floating garbage in the water area, realize the 24-hour uninterrupted monitoring of water waste, provide data support for the corresponding garbage cleaning, improve the cleaning efficiency, and optimize the management of water environment. In this paper, the YOLOv5 algorithm is combined with floating garbage in waters, and the detection interface designed by PyQt5 is used to realize the convenient detection of images, videos and cameras of water garbage. The YOLOv5 model used in this study is trained on a new dataset, including 4591 images and 6622 bounding boxes of three types of common garbage. Five models of YOLOv5 are trained, and the optimal model under the same experimental conditions is selected. The model achieves 99.5% and 82.5% average precision values (mAP@0.5 and mAP@0.5:0.95).

**Keywords:** object detection; YOLOv5; Water litter; Pyqt5

## 1.Introduction

With the continuous development of economy, urban water resources have been seriously polluted due to the accumulation of garbage. China has taken some effective measures to deal with the problem, including strengthening supervision and law enforcement, issuing relevant policies and regulations, promoting sustainable water resources management technology, and establishing and improving water treatment facilities and monitoring networks to strengthen the monitoring and control of water pollution. However, water pollution comes from a variety of pollutants, including plastic waste, bottles and other waste, posing a serious threat to Marine ecology. Plastic waste is piling up at an alarming rate, with more than 300 million tons produced each year and more than 8 million tons of plastic entering the Oceans, causing a devastating impact on Marine ecosystems and wildlife. Gary Stokes, Director of operations at oceanasia, said: "Plastic pollution is estimated to kill 100,000 Marine mammals and turtles, more than a million seabirds, and many more fish, invertebrates and other animals each year, and it has a negative impact on fishing and tourism. The cost to the global economy is estimated at $13 billion a year." Due to the COVID-19 pandemic, plastic consumption, which had been rising steadily for years, has increased significantly. Using the DPSIR framework, Song G et al.[1] found that COVID-19 led to a surge in the use of disposable face masks (DMs), and improper DMs entering the environment would be transformed into Marine

Correspondence: Huimin He(H.M.); Tel.: +86-189-7886-8509; Email:hehuimin@gxmu.edu.cn

PMBDA 2023, December 15-17, Nanjing, People's Republic of China

DOI 10.4108/eai.15-12-2023.2345400

microplastic pollution (MMP), endangering Marine ecosystems and human health, and posing a potential threat to the economy and society. Therefore, it is urgent to establish a sound supervision mechanism and use various advanced technologies to control water pollution and reduce the harm of plastic and other waste to the water ecological environment.

In recent years, target detection technology has been applied in the field of water, and many scholars have studied the target detection of water ecological environment. Lin et al.[2] proposed an improved YOLOv5s(FMA-YOLOv5s) algorithm and data expansion method, which proved to meet the requirements of target detection standards. Sannigrahi et al.[3] adopted support vector machine (SVM) and Random forest (RF) ML models to classify and analyze remote sensing data, improved model performance, and developed an automatic floating plastic detection system. ZailanNA et al.[4] developed an automatic detection system for floating garbage based on the improved YOLO model, which supports the detection of 5 types of garbage. Boer G et al.[5] recorded underwater organisms through a fixed underwater camera system, used adaptive background estimation algorithm to pre-process the initial video, and used YOLOv5 model to detect and locate aquatic organisms in the video frame, achieving high accuracy and precision. Chen Renfei et al.[6] proposed a floating object detection method based on continuous unsupervised domain adaptation strategy, which improved the detection network generalization ability through comparative experiments and provided a new method for the application of target detection in the field of water. Ye Zhiyang et al.[7] improved YOLOv5s model for underwater target detection. Cao Jianrong et al.[8] introduced underwater dark channel priority algorithm and efficient correlation channel to improve the YOLOv5 network and improve the detection accuracy.

In summary, surface target detection technology has been widely studied and developed, but there are still many challenges and interference factors. Traditional floater detection algorithms usually carry out pre-processing such as smoothing and de-noising to reduce the influence of factors such as water surface ripple and light change. There have been many researches on the application of YOLO algorithm to surface target detection, which shows good detection efficiency and quality. YOLOv5 algorithm has faster inference speed and higher frame rate, and is implemented by PyTorch framework. Lightweight and stability are important considerations for practical application. Therefore, this study proposes the design and implementation of garbage detection in waters based on YOLOv5 algorithm. The main contributions of this study are as follows:

The optimal model in this study is used to design the interface of water waste detection, and the detection and marking of the data set can be completed with a small number of clicks.

The remainder of this paper is organized as follows: Section 2 discusses the relevant literature. Section 3 provides an overview of the experimental environment, data set acquisition and processing, and the evaluation metrics used in this study. Section 4 provides the experimental process and results. Section 5 summarizes the conclusions and potential future improvements.

## 2 Related Work

### 2.1 Target Detection

The development of object detection algorithm originated from the early stage of computer vision, which mainly relied on traditional image processing and computer vision technology in the early stage, but the effect was not good in complex scenes and changeable lighting conditions. With the rise of deep learning technology, object detection algorithms have been greatly developed, including Two-Stage and One-Stage object detection algorithms. One-Stage target detection has high efficiency and relatively good real-time performance, which is typically represented by YOLO and SSD series algorithms.[9] The YOLO algorithm treats object detection as a regression problem, using a single CNN to process the entire image and segment it into a grid to predict bounding boxes and class probabilities. The SSD algorithm extracts features through pre-trained CNN and combines multi-scale features for target detection. Its advantage is that it fuses feature maps of different levels to improve detection speed. Compared with the YOLO algorithm, the advantage of the SSD algorithm is that the feature maps of different levels are fused, which improves the detection speed.[10]Lu Tan et al.[11] compared the performance of RetinaNet, SSD and YOLO v3 recognition tablets, and the results showed that SSD performed poorly in MAP and FPS indicators, while YOLO v3 had advantages in detection speed and better performance for hard sample detection.

In actual applications, appropriate algorithms should be selected according to the requirements and application scenarios. SSD emphasizes high detection speed and YOLO emphasizes high detection accuracy. In addition to mainstream target detection algorithms, some other algorithms are also being developed and optimized, such as RCNN, Faster R-CNN, etc., which improve detection speed while maintaining high accuracy, making target detection technology more and more popular in real-time applications.

### 2.2 Overview of YOLOv5

YOLOv5 is an object detection algorithm that, similar to previous generations of YOLO algorithms, adopts the concept of a grid, dividing an image into multiple grids, each of which is responsible for predicting one or more objects. In the training process, the anchor in the grid where the middle point of the real manual marked box falls will "grow" or "shrink" madly towards the real box, and the confidence level is set to 1, indicating that the grid where the anchor is located has objects, while the confidence level of other anchors without prediction boxes is 0.

YOLOv5s network structure mainly consists of Input, Backbone, Neck, Head and so on. First of all, the Input terminal mainly completes image preprocessing, which mainly includes Mosaic enhancement, adaptive adjustment of image size and adaptive calculation of anchor frame [12]..Secondly, Backbone network focuses on input images, which mainly includes CSP, Focus and SPP modules [13].Next, the Neck structure fuses high-resolution features from lower levels with low-resolution features from higher levels to produce a series of multi-resolution feature maps, which are then fused to improve the accuracy and robustness of detection.Finally, the Head structure refers to the output layer, which is the last layer in the target detection network and is used to generate the final result of target detection based on the feature map of the middle layer.

Compared with previous generations of YOLO algorithm, YOLOv5 is optimized in inference speed and adopts Focus structure and CSP structure. Among them, CSP structure designs two structures in the Backbone network. Taking Yolov5s network as an example, one is CSP1_X structure applied to the backbone network, and the other is CSP2_X structure applied to the Neck. In addition, YOLOv5 Neck adopts FPN+PAN structure. Therefore, YOLOv5 has improved detection accuracy and speed, and can adapt to target detection requirements in different scenarios.

## 3.Materials and Methods

### 3.1 Data acquisition and preprocessing

#### 3.1.1 Image capture

The objects of this study are plastic bags, plastic bottles and cans, and masks. Some images of plastic bottles and cans are obtained in the public data set.And then, the basic images of plastic bags, masks and plastic bottles are obtained through network keyword retrieval, and they are classified and processed, unified format modification and renaming, etc., which lays a foundation for subsequent dataset annotation and other operations.

#### 3.1.2 Dataset annotation and augmentation

In order to solve the problem of insufficient original data set and unbalanced samples, image processing technology is used to increase the amount of data in the initial image.The random enhancement method was used, including rotation 90 degrees, 180 degrees and 270 degrees. Brightness and dark adjustment horizontal and vertical flip; Zooming in and out; Add Gaussian noise and salt and pepper noise.The resulting modifications to the overall scene lighting, enhance the resilience and accuracy of the model.After random image processing, the total number of images in the dataset is 4591, including 1534 images for bottles, 1608 images for masks, and 1449 images for plastic bags.And the dataset is randomly divided according to the ratio of 7:2:1. The number of images and sample sizes in the dataset are shown in Figure 1.

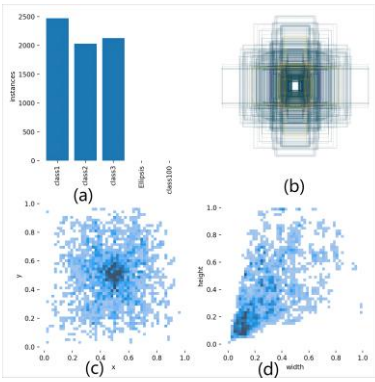


**Figure 1.** Visualization results of the attributes of the datasets in this study :(a) the number of labels in the dataset,(b) the proportion of labels in the dataset,(c) the location of labels in the dataset, and (d) the size of labels in the dataset.

Process the collected dataset and use LabelImg labeling software to annotate the positions of plastic bags, masks, bottles, and cans in the dataset. The annotated data for each image obtained is stored uniformly in YOLO format. Among them, a total of 6622 annotation boxes were marked, with 2468, 2029, and 2125 annotation boxes for bottles, masks, and plastic bags, respectively. Figure 2 shows an example of the annotation process, which includes coordinate information on the image.

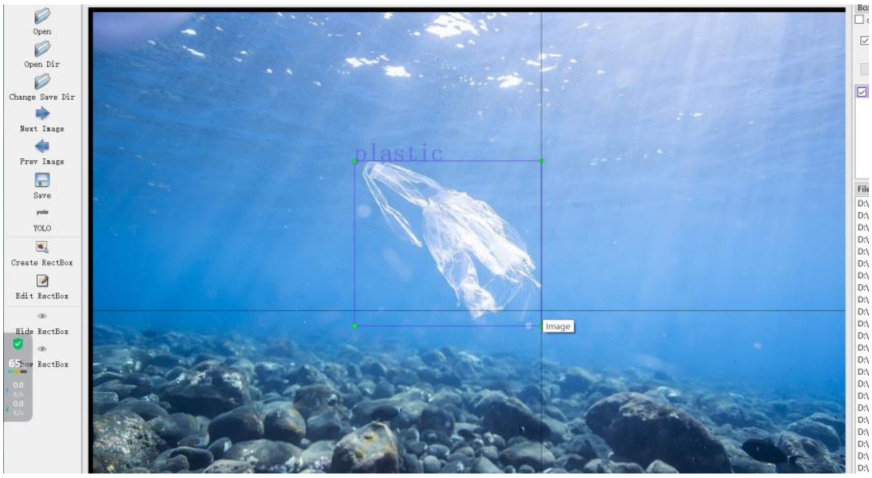


**Figure 2.** LabelImg dataset annotation interface was used.

Figure 1 presents the statistical results of the shape, distribution, and location of the samples in the visualization dataset.The upper left panel in Figure 1 shows the distribution of the number of samples, indicating that the number of samples is relatively sufficient and the distribution is relatively uniform.The upper right panel in Figure 1 shows the aspect ratio of the sample box in the original image. The bottom right figure of Figure 1 shows the sample center point of the whole image, each box represents the occurrence of a sample, and the color depth reflects the number of occurrences.The darker the color, the more frequent it appears. The results show that the distribution of sample locations in the image is concentrated in the center of the image.The bottom right plot in Figure 1 depicts the ratio of the width and height of the sample in the whole picture. Each point indicates that the sample appears in this horizontal and vertical coordinate scale.It can be observed that the sample concentration is near the origin, the sample is concentrated on the ray with a 45° slant Angle,Small target samples dominate the dataset. In general, the total amount, distribution and composition of the data set are relatively uniform and reasonable.

### 3.2 Model evaluation indicators

Precision P, recall R, and average recognition accuracy (mAP@0.5 and mAP@0.5:0.95) were used to evaluate the performance of the model.mAP is the average of the average accuracy AP computed for all classes, where N is the number of categories.mAP@0.5 indicates that the mAP threshold calculated at the IoU intersection is 0.5. MAP@0.5:0.95 indicates that the average mAP of the loU threshold ranges from 0.5 to 0.95 with 0.05 interval.

## 4.Experimental procedure and results

### 4.1 YOLOv5 model parameter Settings

The hyp hyperparameters used in the experiment of this study are the initial parameters of the model, the initial learning rate lr0 is 0.01, the weight decay coefficient weight_decay is 0.0005, and the learning rate momentum is 0.937. The training parameters are shown in Table 1.

YOLOv5's loss function consists of several parts, including bbox regression loss, target confidence loss, and class loss. The loss weights and IOU thresholds of these parts are designed to control the training process of the loss function. The loss function of the model in this study is the default value of the official code of YOLOv5, where the weight of bbox regression loss is 1, the threshold of IOU is 0.02, and the maximum value of IoU between the predicted box and the real box is 0.2. The weight of class loss is 1, the weight of class is 0.2, and the confidence threshold is 4.0. The weight of the target loss is 1, the positive sample threshold is 0.2, and the negative sample threshold is 4.0.

**Table 1.** Training parameter Settings

| Model configuration file cfg | argument | Numerical value |
|---|---|---|
| YOLOv5n | Epoch/Batch-size | 100/64 |
| YOLOv5s | Epoch/Batch-size | 100/64 |
| YOLOv5m | Epoch/Batch-size | 100/32 |
| YOLOv5l | Epoch/Batch-size | 100/32 |
| YOLOv5x | Epoch/Batch-size | 100/16 |

### 4.2 Model training and results

Firstly, YOLOv5s model is trained, and no parameters are modified except the data configuration required for training. After 300 epochs of training, it is found that the model overfits prematurely by observing the training process and results, and its training indicators are shown in Figure 3. Measures taken to correct the overfitting problem include:(1) Expand the amount of data. The data set studied in this paper was expanded from 3863 images to the current 4591 images, with little effect.(2) The early stop method. The training rounds were reduced to 100 epochs, the effect was not significant enough.(3) L2 regularization is added to the training code. This can be done by changing the model definition file or training script. By default, the weight_decay parameter is set to 0.001, and increasing its value will increase the regularization strength.Therefore, from the original 0.001 to 0.005, 0.01, 0.05, 0.5, 1, 2. In the progressive training process of weight_decay parameter, for the consideration of training efficiency, this study chooses YOLOv5n model for training, each increment is trained for 100epochs. The model training results of each time are collected, as shown in Table 2.

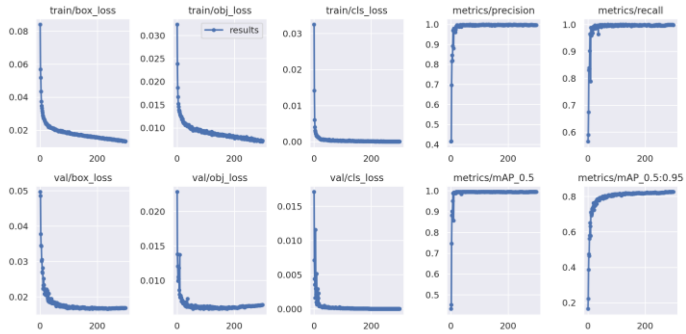


**Figure 3.** Results of YOLOv5s Training for 300 epochs

**Table 2.** Training results with only the weight_decay parameter tuned

| Cfg | Epoch | Weight_decay | Training duration | Training result file size |
|---|---|---|---|---|
| YOLOv5n.pt | 100 | 0.001 | 0.898hours | 3.9MB |
| YOLOv5n.pt | 100 | 0.005 | 0.920hours | 3.9MB |
| YOLOv5n.pt | 100 | 0.01 | 0.896hours | 3.9MB |
| YOLOv5n.pt | 100 | 0.5 | 0.915hours | 3.9MB |
| YOLOv5n.pt | 100 | 1 | 0.893hours | 3.9MB |
| YOLOv5n.pt | 100 | 2 | 0.881hours | 3.9MB |

With the increase of the value of the weight_decay parameter, the overfitting problem of the model was gradually alleviated, the loss function gradually decreased, and the result curve gradually stabilized. However, if weight_decay is too large, it can cause the model to be too conservative during training to fully learn the features of the data. Therefore, under the conditions of this study, the weight_decay value of 1 is relatively good. In this process, the performance of the verification set is gradually improved, and the overfitting situation is reduced. Therefore, by comparing the loss function and verification set performance under different weight_decay parameters, we can select an optimal parameter value that can prevent the model from overfitting while maintaining the model's generalization ability. By synthesizing FIG. 3 and Table 2 and based on the data set identity of this study, we determined the value of weight_decay parameter to be 1 for further model training. Then, the training of YOLOv5s.pt, YOLOv5m.pt, YOLOv5l.pt and YOLOv5x.pt on the same data set was continued on the same server, and the trained model was verified using the verification set. The results were shown in Table 3 below.

**Table 3.** Training results of each model for YOLOv5

| cfg | epoch | Weight _decay | Training duration | Training result file size | P | R | mAP @0.5 | mAP@0.5 :0.95 |
|---|---|---|---|---|---|---|---|---|
| YOLOv5n.pt | 100 | 1 | 0.893hours | 3.9MB | 0.985 | 0.982 | 0.993 | 0.779 |
| YOLOv5s.pt | 100 | 1 | 0.903hours | 14.4MB | 0.997 | 0.996 | 0.995 | 0.825 |
| YOLOv5m.pt | 100 | 1 | 1.015hours | 42.2MB | 0.996 | 0.993 | 0.995 | 0.816 |
| YOLOv5l.pt | 100 | 1 | 1.318hours | 92.8MB | 0.995 | 0.996 | 0.995 | 0.816 |
| YOLOv5x.pt | 100 | 1 | 2.822hours | 173.1MB | 0.995 | 0.991 | 0.994 | 0.822 |

According to Table 3, the detection effects of all versions of YOLOv5 model are almost the same. The detected P value and R value are both higher than 0.980, the difference between the four groups of data with P value is only 0.001, the difference between the last four groups of data with R value is only 0.005 at most, and the difference between mAP@0.5 value is at most 0.002 in the five groups of data. The parameter mAP@0.5:0.95 has a maximum difference of 0.046 and a minimum difference of 0.037 among the 5 groups of data. Among them, the difference between 0.046 and 0.005 is 9.2 times. Therefore, the mAP@0.5:0.95 parameter has a large difference in the 5 groups of data, and it is more reasonable to use it as the selection basis. In this case, mAP@0.5:0.95 of YOLOv5s model has the highest value. By synthesizing the four parameter indicators, it can be seen that the data of YOLOv5s model is more stable. Combined with Table 2 and Table 3, when the depth and width of the model network increase, the batch-size of images read in each round of model training becomes smaller, the training time becomes longer, the hardware requirements become higher, and the weight file of the trained model also becomes larger. This study is based on YOLOv5 algorithm for surface garbage detection. On the premise of ensuring detection efficiency and quality, excessive weight files have higher requirements for realistic hardware factors, which restricts the deployment of mobile terminals. Therefore, YOLOv5s model was selected as the basic model for visualization of subsequent surface garbage detection results, and the detection effect of YOLOv5s model was shown in Figure 4.

**Figure 4.** Detection effect of YOLOv5s model

### 4.3 Program design of water garbage detection

In this study, PyQt5 framework was used to design the interface, and the model was called to realize the target detection function of three forms of pictures (JPEG format, PNG format, TIF format), video (AVI format, MP4 format) and camera connection for plastic bags, masks and bottles, and the results were visualized. Examples of interface and detection are shown in Figure 5.

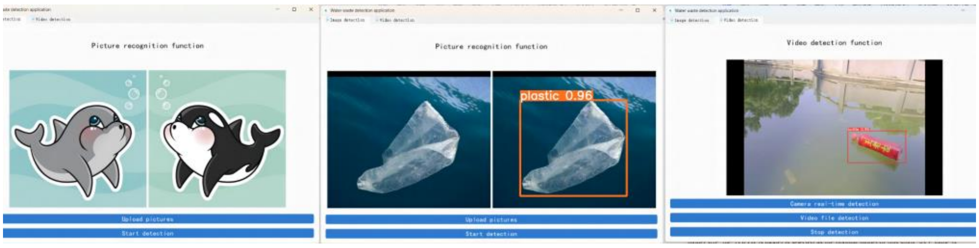


**Figure 5.** (a) Program home page interface;(b)Realizing the image recognition function; (c) Realize the video detection function

## 5.Conclusions and potential future improvements

In this study, the YOLOv5 algorithm was applied to water debris detection. Firstly, YOLOv5s model was trained for 300 epochs, and L2 regularization was performed on YOLOv5s after overfitting was found. The weight_decay parameter was adjusted to reduce overfitting, and the area of the mAP curve and the fluctuation of precision and recall were taken into account. The weight_decay parameter is chosen as 1 as the subsequent model training parameter. After training all models of YOLOv5 algorithm, indicators P, R, mAP@0.5, mAP@0.5: 0.95 and the model size, the YOLOv5s model is selected as the optimal model of this study, its P value is 0.997, R value is 0.996, mAP@0.5 value is 0.995, mAP@0.5:0.95 value is 0.825. pyqt5 was used to design the interface of water garbage detection, and the trained YOLOv5s model was called to realize the final target detection function. Comprehensive research process and research results, it is found that YOLOv5 algorithm takes into account both high efficiency and high quality target detection performance, which helps the in-depth development of target detection research.

There are still some shortcomings in this study. The research object is only three categories: plastic bags, masks, bottles and cans, and the data types are few. The number of models obtained during the experiment is limited due to the few parameters adjusted. The application designed and implemented has few target detection functions, and at this stage, it has not been able to use PyCharm software. Future research will expand the data set, expand the scope of modified parameters, introduce camera detection resources, and move the application to the mobile terminal to realize more lightweight and fast water garbage detection and positioning.

**Fund:** Innovation Project of Guangxi Graduate Education(JGY2023070),Innovation Project of Guangxi Graduate Education(YCSW2022192).

## References

[1]Song G, Cao H, Liu L, Jin M. Analysis of Marine Microplastic Pollution of Disposable Masks under COVID-19 Epidemic-A DPSIR Framework. Int J Environ Res Public Health. 2022;19(23):16299. Published 2022 Dec 5. doi:10.3390/ijerph192316299

[2]Lin, Feng et al. "Improved YOLO Based Detection Algorithm for Floating Debris in Waterway." Entropy (Basel, Switzerland) vol. 23,9 1111. 27 Aug. 2021, doi:10.3390/e23091111

[3]Sannigrahi, Srikanta et al. “Development of automated marine floating plastic detection system using Sentinel-2 imagery and machine learning models.” Marine pollution bulletin vol. 178 (2022): 113527. doi:10.1016/j.marpolbul.2022.113527
[4]Zailan, Nur Athirah et al. “An automated solid waste detection using the optimized YOLO model for riverine management.” Frontiers in public health vol. 10 907280. 12 Aug. 2022, doi:10.3389/fpubh.2022.907280
[5]Böer G, Gröger JP, Badri-Höher S, et al. A Deep-Learning Based Pipeline for Estimating the Abundance and Size of Aquatic Organisms in an Unconstrained Underwater Environment from Continuously Captured Stereo Video. Sensors (Basel). 2023;23(6):3311. Published 2023 Mar 21. doi:10.3390/s23063311
[6]Chen Renfei, Peng Yong, Li Zhongwen. Surface floating object detection method based on continuous unsupervised domain adaptation strategy [J]. Systems Engineering and Electronics,2023,45(11):3391-3401. (in Chinese) DOI:10.12305/j.issn.1001-506X.2023.11.04.
[7]Ye Zhiyang, Liang Haolin, LAN Chengdong. Applied to YOLOv5s algorithm model of underwater target detection [J]. Journal of TV technology, 2023,47 (2) : 39-43. DOI: 10.16280 / j.v ideoe. 2023.02.010.
[8]Cao Jianrong, Manor, Wang Ming, et al. Based on ECA YOLOv5 fish underwater target detection [J]. Computer system application, 2023, 32 (6) : 204-211. The DOI: 10.15888 / j.carol carroll nki. Csa. 009149.
[9] Cai Jialei, Mao Zhihui, Li Jun, et al. Review of deep learning based object detection algorithms and applications [J]. Network Security Technology and Applications, 2023, (11): 41-45
[10]Wu Xue, Song Xiaoru, Gao Song, etc A review of deep learning based object detection algorithms [J] Sensors and Microsystems, 2021,40 (2): 4-7,18 DOI: 10.13873/J.1000-9787 (2021) 02-0004-04
[11]Tan L, Huangfu T, Wu L, Chen W. Comparison of RetinaNet, SSD, and YOLO v3 for real-time pill identification. BMC Med Inform Decis Mak. 2021;21(1):324. Published 2021 Nov 22. doi:10.1186/s12911-021-01691-8
[12] Lei Liang, Yin Yanwei, Liang Minghui, et al. Research on elderly fall recognition algorithm based on improved YOLOv5s [J] Journal of Chongqing University of Science and Technology (Natural Science Edition), 2023,25 (01): 85-90. DOI: 10.19406/j.cnki. cqkjxyxbzkb.2023.01.002
[13] Yang Jinhui, Li Hong, Du Yunyan, Mao Yao, Liu Qiong. Lightweight object detection algorithm based on improved YOLOv5s [J]. Electrooptic and Control, 2023,30 (2): 24-30