ISSN: 3007-9640 DOI: https://doi.org/10.70114/acmsr.2026.6.1.P230

Research on Underwater Debris Detection Technology Based on YOLO Algorithm

Chaolong Liu [1, a], Xiaoting Bu [2, b], Shaobo Yang [1, c] *, Aicheng Xiong [3, d]

[1] College of Ocean Science and Engineering, Shandong University of Science and Technology, 579 Qianwan Port Road, Huangdao District, Qingdao, 266590, Shandong Province, China;

[2] China Resources Recycling Group Equipment Assets Co., Ltd.

[3] Tianjin Luhai Surveying and Mapping Co., Ltd.

[a] 18266107327@163.com, [b] 2207838973@qq.com, [c] yangskyle@tju.edu.cn,

[d] 2917273859@qq.com

Abstract. Marine debris not only severely pollutes the health and habitat of marine life but also indirectly threatens human life. Deep learning-based target detection algorithms have achieved significant results in underwater debris detection, with the YOLO algorithm being widely used by researchers due to its fast detection speed and high accuracy. However, the underwater environment is complex and variable, and underwater debris detection faces challenges such as low detection accuracy, false negatives and missed detections of small targets, and a large number of parameters. This paper mainly focuses on three aspects: experimental environment configuration, dataset construction, and platform design. The experimental environment setup primarily included Anaconda environment debugging and Python programming language installation. Dataset construction involved initial collection of underwater debris images, followed by image labeling and classification. The dataset was divided into two main categories: biological (bio) and plastic (plastic) debris. Finally, the platform was designed using PyQt5. The platform interface is divided into three areas: settings, detection, and data display. The settings area allows users to select images and videos for detection. The detection area displays the detected underwater debris in the currently detected images and videos. The data display shows the time taken to detect the images, the number of detected targets, their type, and the confidence level.

Keywords: YOLOv8; underwater debris; plastic debris; detection platform

1. Introduction

1.1 Research Background and Significance

Water pollution has become a major environmental problem threatening the stability and balance of the global ecosystem and the sustainable development of humankind. According to statistics, about 11 million tons of plastic waste enter the ocean every year. Among them, materials such as polyethylene have a natural degradation cycle of up to 470 years in seawater. These pollutants not only cause the death of more than 1 million seabirds and 100,000 mammals every year.

Figure. 1 Pollution caused by underwater debris

ISSN: 3007-9640 DOI: https://doi.org/10.70114/acmsr.2026.6.1.P230

With the development of technology, waste detection methods are constantly improving, gradually transforming from the most primitive manual detection to intelligent and automated detection. In recent years, marine ecological environment protection issues have become increasingly prominent. Underwater waste detection technology, as a crucial link in marine environmental monitoring, has seen even more significant technological iterations. In particular, advancements in deep learning technology have greatly propelled the development of detection technology, achieving remarkable results in target detection and providing new solutions for waste identification in complex environments. Among these, the YOLO model (You Only Look Once) is the most representative detection technology, boasting high accuracy and real-time performance, demonstrating enormous potential in many applications and widely used in fields such as industrial quality inspection and autonomous driving. However, in complex underwater environments, the basic YOLO algorithm suffers from low detection accuracy, high computational load, and slow detection speed. To address these issues, some researchers have improved the YOLO model and achieved corresponding results. While the improved model performs excellently in experimental environments, the real marine environment is much harsher and more complex, leading to a decrease in detection accuracy compared to experimental settings.

1.2 Current Status of Research at Home and Abroad

Underwater debris detection is a crucial step in underwater debris management, and many researchers both domestically and internationally have conducted numerous experiments in this field, achieving preliminary research results. However, given the complex underwater environment and the limitations of various detection technologies in underwater debris detection, despite numerous modifications made by researchers to improve detection accuracy and efficiency, further optimization of current detection techniques is still necessary. This paper investigates and studies two aspects: traditional object detection techniques and deep learning-based object detection algorithms.

1.2.1 Traditional underwater detection technology

Traditional underwater target detection technology has long been a focus of attention for scholars at home and abroad, and has now become relatively mature. These detection technologies are mainly used for the localization and identification of underwater targets. As early as 2006, Liu Jing et al. studied the preprocessing of underwater images and the identification of underwater targets[1] . They processed and analyzed underwater images of real environments, and then preprocessed the images to make full use of the entropy information of underwater images to enhance underwater images and improve the accuracy of underwater target detection. However, although the algorithm can improve the accuracy of inspection through image enhancement, it still has the problem of large workload. Later, in 2011, Mao Dun et al. proposed an underwater target detection algorithm based on frogman detection sonar sequence images[2] . It mainly improves the model's ability to detect targets by using an adaptive double frame difference method based on image binarization and region growing. However, its accuracy in detecting targets in harsh environments still needs to be improved. In the same year, Liu Dandan et al. proposed an underwater target detection algorithm based on sonar image resolution[3] . This algorithm can accurately detect the movement of targets in relatively complex environments, but the computational load of this algorithm is large and its application scenarios are limited. In 2015, Ma Weiying et al. proposed an underwater target detection algorithm based on polarization features[4] . This is a detection model that combines polarization information with the Itti model. Compared with the traditional Itti model, it analyzes and extracts polarization features and edge features that are more suitable for underwater target detection, which is especially important for the underwater environment. This makes it more effective for underwater target detection. In 2018, Li Xiaofang et al. proposed an underwater target saliency detection algorithm based on polarization information[5] . This algorithm addresses the problem that traditional saliency detection methods rely too much on spectral and light intensity features and cannot be well applied to complex and variable scenarios. It proposes an underwater

ISSN: 3007-9640 DOI: https://doi.org/10.70114/acmsr.2026.6.1.P230
target detection method that combines polarization features with a saliency model, which can greatly improve the accuracy of detection. In 2022, Li Xiang et al. proposed an underwater identification code detection algorithm based on spatial curved surfaces[6] . This algorithm can significantly improve the positioning and recognition speed. In the same year, Zeng Teng et al. proposed an underwater target detection algorithm based on combined features. This algorithm integrates multiple information such as acoustic intensity and distance, improving the underwater detection performance of the model. Traditional underwater target detection algorithms have significant limitations in application scenarios and are difficult to adapt to complex and diverse underwater targets. Therefore, to address these limitations, it is necessary to further improve and develop traditional underwater target detection algorithms to enhance their detection capabilities in complex underwater environments.

1.2.2 Deep learning-based object detection algorithms

In 2014, Ross and his collaborators published the R-CNN object detection algorithm (region-based convolutional neural network), which is considered the first representative work of the successful application of deep convolutional neural networks in the field of object detection. This groundbreaking advancement in deep learning-based object detection research marked the beginning of two-stage detectors. Despite the significant achievements of R-CNN, problems such as low computational efficiency, excessive storage space, and the complexity of the training process remain.

The key technology of spatial pyramid pooling was introduced in SPP-Net[7] proposed by Kaiming He's research team . The fixed limitation of input size in traditional convolutional neural networks was effectively overcome by this technology, and the efficiency of object detection tasks was significantly improved, and the accuracy of image classification was also improved. In 2015, the Fast R-CNN algorithm was further proposed by Ross et al[8] . As an important improved version of R-CNN. The problem of low computational efficiency was solved in this algorithm, and the defect of complex training process was also corrected. Fast R-CNN is an important milestone in the development of two-stage detectors. Its core innovation lies in the convolutional sharing mechanism of feature maps and the end-to-end training method. Examples show that the speed and accuracy of object detection are significantly improved as a result.

In 2015, Joseph Redmon and other researchers proposed YOLO (You Only Look Once)[9] , a single-stage object detection algorithm that has become a classic. Unlike the dual-end detection system R-CNN, as mentioned above, its real-time performance and efficiency are particularly outstanding. The core idea of the YOLO algorithm is to transform the object detection task into an end-to-end regression problem. Through a single forward propagation process, the category information of all objects in the image and their bounding boxes can be directly predicted. Examples show that this design idea brings a significant improvement in inference speed.

In recent years, many researchers have proposed a series of underwater target detection algorithms based on the improved YOLO algorithm. In 2017, Sun Jing et al. proposed a biological target detection algorithm based on the YOLOv2 algorithm[10] . This algorithm uses the obtained regression box as the region of interest for binocular feature matching, which improves the localization and recognition speed of the model. In 2022, Tian Manjun et al. proposed a visual detection algorithm for underwater garbage cleaning robots based on the improved YOLOv4[11] . This algorithm improves the traditional YOLOv4 into a four-scale detection algorithm and combines it with the robot. Experimental results show that the improved algorithm performs well in terms of accuracy and detection speed. In 2024, Wang Yannian, Li Ying et al. proposed a lightweight underwater detection method of YOLOv5[12] , namely YOLOv5s-MCS. To address the issue that YOLOv5 algorithm is prone to detection loss when identifying small underwater debris, this algorithm proposes to use MobileNetv3-Small to optimize the backbone network in YOLOv5s in order to improve speed and reduce the number of model parameters. Finally, the loss function is optimized to improve detection accuracy. In 2024, Yuan Hongchun et al. proposed the FDC-YOLOv8 underwater biological target detection method based on improved deformable

ISSN: 3007-9640 DOI: https://doi.org/10.70114/acmsr.2026.6.1.P230
convolution[13] , which is mainly used to solve the problems of target overlap, occlusion and small target size in underwater biological target detection tasks, resulting in false detection and missed detection. In 2024, Han Zhigen et al. further improved the YOLOv9 algorithm[14] . In order to deal with the problem that it is difficult to detect and identify due to the influence of dark light and seawater color, color enhancement and image enhancement processing are performed in the preprocessing, which can significantly improve the image recognition. The Squeeze and Excitation attention mechanism is introduced to improve the detection efficiency.

2. Experimental Process and Results Analysis

2.1 YOLOv8 Architecture

YOLOv8 (You Only Look Once, version 8) is an open-source real-time object detection and image segmentation model released by Ultralytics in 2023, belonging to the eighth generation of the YOLO series. The YOLOv8 algorithm not only inherits the advantages of previous generations but also improves upon them, enhancing both accuracy and efficiency, making it the most widely used vision model currently. YOLOv8 can be mainly categorized into several types based on detection speed: YOLOv8n, YOLOv8s, YOLOv8m, YOLOv8l, and YOLOv8x. Its architecture primarily consists of four core parts: the backbone network, the neck network, the detection head, and the loss function.

Backbone: The backbone is the core function of the YOLO algorithm, primarily responsible for extracting multi-scale features. The addition of the SPPF module significantly enhances its feature extraction capabilities.

Neck Network: In deep learning object detection tasks, the neck network is typically located between the backbone network and the output. The main purpose of the neck network is to fuse features from different levels, aiming to reduce the feature dimensionality of the target and thus extract more useful information for object detection, ultimately enhancing the algorithm's feature representation capabilities.

Detection Head: The detection head in YOLOv8 is primarily used for target prediction, detecting, classifying, and locating objects based on features . This detection head also introduces a decoupling head structure, separating the target classification and detection tasks to improve model performance and accuracy. Using a decoupling head not only enhances feature extraction capabilities but also adapts to multi-class detection.

Loss Function : The loss function is mainly used to calculate the difference between the predicted result and the actual situation, so as to facilitate the improvement and optimization of the model. It mainly consists of four parts: classification loss, regression loss, mask loss, and pose loss.

2.2 Dataset

The dataset prepared for this experiment contains a total of 1900 images. This dataset has two categories: 1426 images in the training set and 474 images in the validation set. After preparing and classifying the dataset, the professional image annotation tool LabelImg was used to annotate each image . Bounding boxes were used to annotate the debris in each image, and the category of each type of debris was labeled next to them. There are two categories of underwater debris in this study: bio and plastic. LabelImg was used to annotate each image. The following are the images after annotation using LabelImg .

ISSN: 3007-9640 DOI: https://doi.org/10.70114/acmsr.2026.6.1.P230

Fig. 2 Image of underwater plastic waste after labeling.

Fig. 3 Image of underwater organisms after labeling.

2.3 Evaluation Indicators

Object detection metrics primarily focus on the accuracy of the model in detecting objects. These metrics include: Precision (P), Recall (R), Mean Average Precision (MAP), Floating-point Operations Per Second (FLOPs) , and Intersection over Union (IoU). The formulas for Precision and Recall are shown below:

$$P=\frac{TP}{FP+TP} \tag{1}$$

$$R=\frac{TP}{FP+TP} \tag{2}$$

Where TP represents the number of correctly detected positive samples, FP represents the number of targets that incorrectly detect negative samples as positive samples, and FN represents the number of targets that incorrectly detect positive samples.

Precision is the ratio of correctly detected images to the total number of detected targets. Recall is the ratio of the number of correctly detected targets to the total number of true targets. Intersection over Union (IoU) is the accuracy of target prediction in an object detection algorithm; it is the ratio of the overlap area between the predicted bounding box and the ground truth bounding box to the sum of the areas of those two boxes. The specific calculation formula is as follows:

$$IOU=\frac{Ao}{Au} \tag{3}$$

In the above formula, Ao is the overlapping area of the ground truth bounding box and the target bounding box, which is the intersection of the two frames, while Au is the sum of the areas of the detection box and the target bounding box, representing the union of the two frames.

2.4 Model Training Experimental Environment

All experiments in this study used the same equipment, and the specific equipment configuration is shown in the table below.

Table 1. Three Scheme comparing

ISSN: 3007-9640 DOI: https://doi.org/10.70114/acmsr.2026.6.1.P230

Experimental Platform	Configuration Environment
CPU	Intel i7-11800H
GPU	NVIDIA RTX 3050 Ti
Operating System	Windows 10
Memory	32MB/s
Deep Learning Framework	PyTorch 2.20
Programming Language	Python 3.9

2.5 Model Training

The training dataset contains 1426 images. The training parameters are set as follows: 100 iterations and 8 batch sizes. In order to ensure the fairness of the experiment, no pre-training was used during the training process.

Fig. 4 Training process

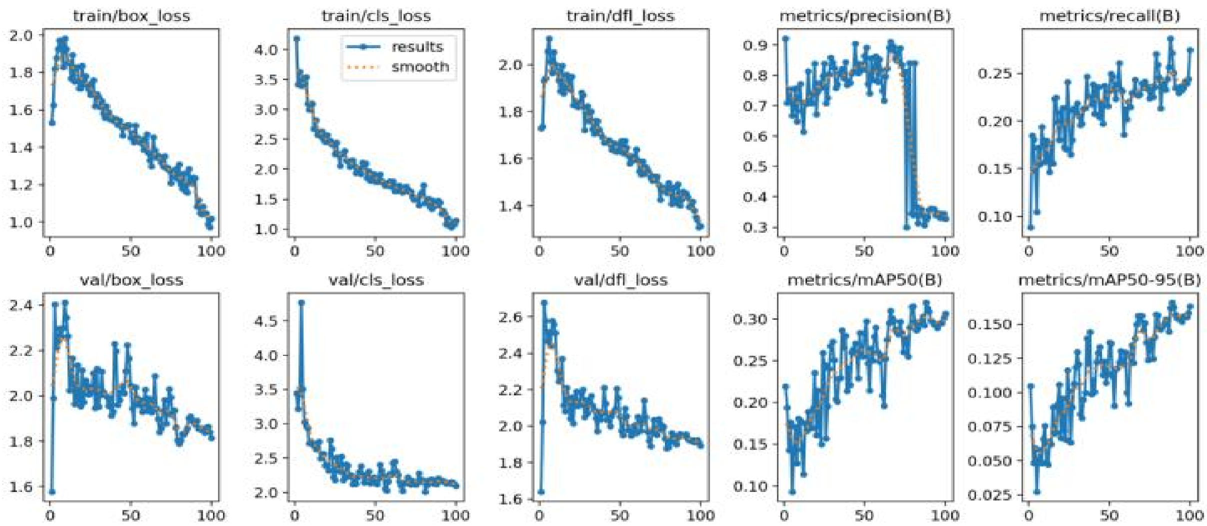


Figure. 5 Training Results

The main parameters of each table are train/box_loss, val/box_loss, train/obj_loss, val/obj_loss, cls_loss, and val/cls_loss. The horizontal axis of each table represents the number of training iterations, while the vertical axis has different meanings. The following is an explanation of the meaning of the vertical axis for each chart.

train/box_loss: represents the mean of the loss function for regression of bounding boxes in the training set;

val/box_loss: represents the mean of the loss function for bounding box regression on the validation set;

train/obj_loss: represents the mean of the confidence loss function for the training set. The smaller the value, the more accurate the object detection.

ISSN: 3007-9640 DOI: https://doi.org/10.70114/acmsr.2026.6.1.P230

val/obj_loss: represents the mean of the confidence loss function for the validation set;

cls_loss: represents the mean of the classification loss function on the training set. The smaller the loss value, the more accurate the classification.

val/cls_loss: represents the mean of the classification loss function on the validation set;

metrics/precisiong: This represents the precision rate; the higher the value, the more accurate the detection.

metrics/recall: represents the recall rate;

2.6 Experimental Results

2.6.1 Test Data

The partial test data used in this experiment are shown below:

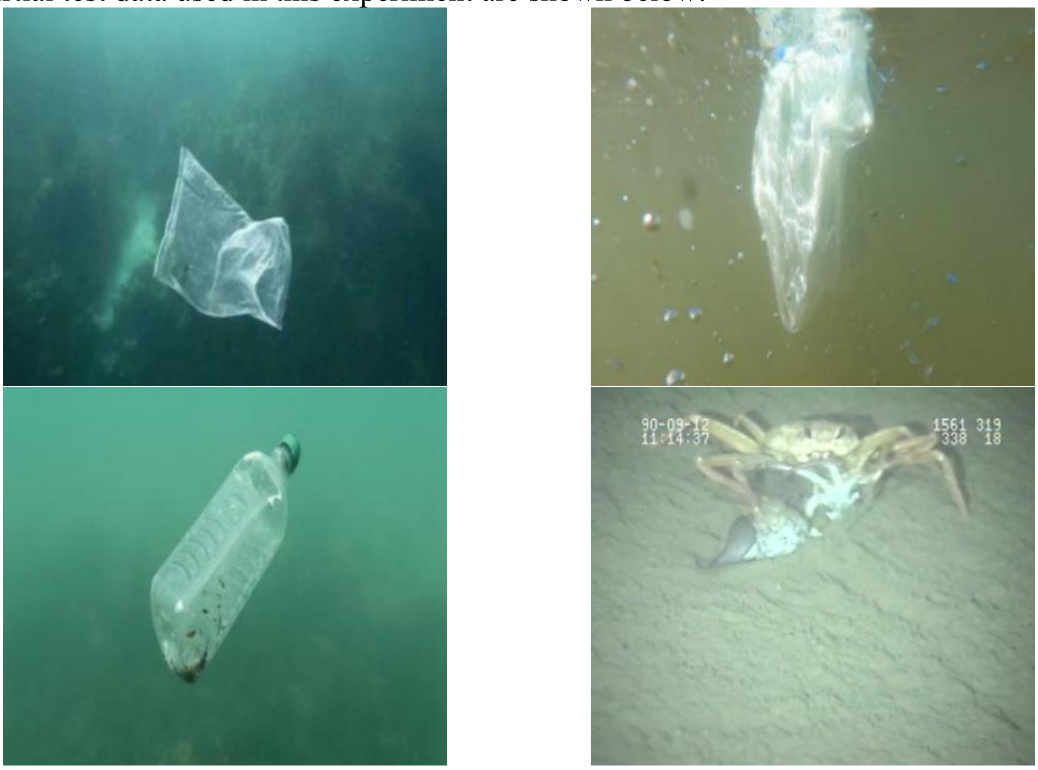
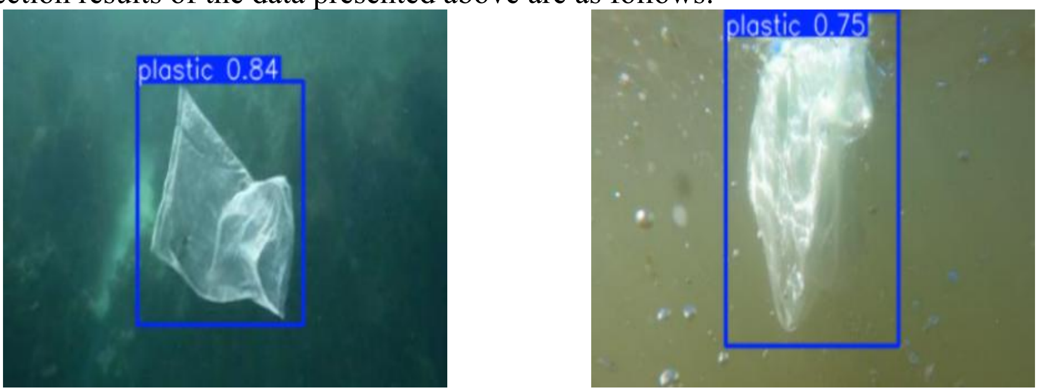


Figure. 6 Detection data display

2.6.2 Test Results

The detection results of the data presented above are as follows:



ISSN: 3007-9640 DOI: https://doi.org/10.70114/acmsr.2026.6.1.P230

Figure. 7 Data detection results

3. Platform Interface Implementation

3.1 Platform Function Introduction

The platform is designed to facilitate image and video detection for users. It is primarily divided into three areas: settings, detection, and data display. The settings area includes functions such as opening images, opening videos, saving, and exiting. The detection area displays the targets detected in the images and videos already examined. The data display area shows the detected data, including target type, confidence level, number of targets, etc. The image below shows the platform interface.

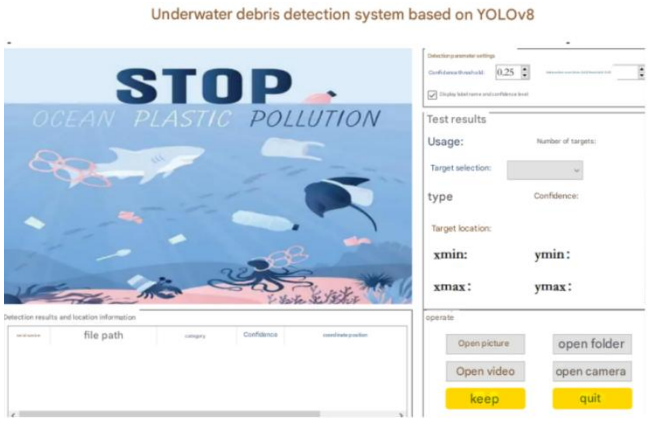


Figure. 8 Platform Interface Display

3.2 Selecting an image

After the user opens the platform, two parameters need to be set first: the Intersection over Union (IoU) threshold and the confidence threshold (conf). In this example, the IoU threshold is set to 0.35 and the confidence threshold to 0.40. The "Image" button can be found in the lower right corner of the detection interface. Clicking the button will bring up a file selection window. In this window, the user needs to select the image to be detected from locally stored images. After selecting the image, the platform will automatically begin detecting it.

3.3 Selecting a Video

In the operation area of the detection interface, you can find the "Open Video" button. After clicking the button, a new window will pop up. The user needs to select the video file to be detected. After selecting the video file, clicking the "Open" button will perform the video detection.

ISSN: 3007-9640 DOI: https://doi.org/10.70114/acmsr.2026.6.1.P230

3.4 Display of Test Results

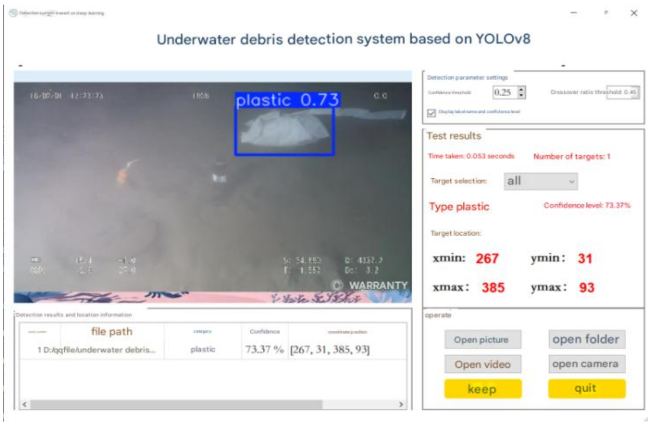


Figure. 9 Detection Results

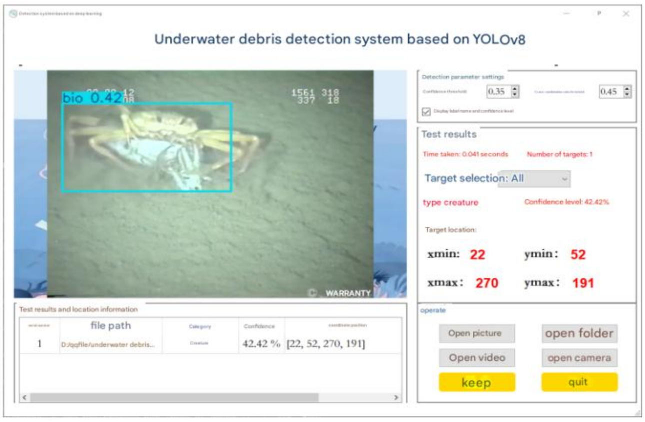


Figure. 10 Detection Results

4. Summary

In recent years, with the continuous advancement of computer technology, target detection algorithms have also been constantly improving, playing a vital role in many situations. However, due to the limitations of real-world detection environments, target detection algorithms need to be adjusted for different environments. In underwater debris detection, due to the complex and variable underwater environment and numerous influencing factors, underwater target detection is prone to problems such as missed or false detections of small targets, and difficulty in detecting targets that are similar to the environment, all of which affect the detection of underwater debris. Therefore, this study addresses these issues, and the main research results are as follows:

The design is based on YOLOv8. First, the environment was set up, then the dataset was constructed. Images of various types of underwater debris were found online and labeled using annotation software. After the images were labeled, underwater debris was classified, then the model was trained, and finally the detection was completed.

An underwater debris detection platform based on the Python programming language was built. This platform allows for a more intuitive and clear view of the detection results, enabling real-time monitoring of the type and quantity of underwater debris detected. The biggest advantage of this platform is its enhanced applicability to the model.

ISSN: 3007-9640 DOI: https://doi.org/10.70114/acmsr.2026.6.1.P230

Acknowledgements

This project was supported by the Tianjin Municipal Science and Technology Bureau under the project Research and Application of Intelligent Surveying and Mapping Unmanned Vessels for Complex Nearshore Sea Areas (24YFZCSN00090).

References

[1] Liu, Jing. Research on underwater image preprocessing and underwater target recognition methods based on entropy [D]. Harbin Engineering University, 2006.

[2] Mao Dun, Liu Zhong, Cheng Yuanguo. An underwater small target detection algorithm based on sonar sequence images detected by frogmen [J]. Journal of Sensor Technology, 2011, 24(07): 1027-1032.

[3] Liu Dandan, Tang Chunrui, Yuan Gannan. A block-based adaptive denoising method for underwater target acoustic images [J]. Acta Ordnance et al., 2010, 31(9): 1263-1267.

[4] Ma Weiying. Underwater target detection based on polarization characteristics [J]. Electronic Measurement Technology, 2014, 37(12): 90-92+98.

[5] Li Xiaofang, Xie Guangqian, Li Chunguang. Underwater target salience detection based on polarization information [J]. Journal of Changzhou Institute of Technology, 2018, 31(04):24-30.

[6] Li Xiang. Research on underwater target detection and pose estimation methods in shallow water [D]. Harbin Institute of Technology, 2019.

[7] Kaiming H, Xiangyu Z, Shaoqing R,et al.Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition.[J].IEEE transactions on pattern analysis and machine intelligence,2015,37(9):1904-16.

[8] Girshick BR .Fast R-CNN.[J].CoRR,2015,abs/1504.08083

[9] Redmon J ,Divvala KS ,Girshick BR , et al.You Only Look Once: Unified, Real-Time Object Detection.[J].CoRR,2015,abs/1506.02640

[10] Sun, Jing. Research on Deep Learning-Based Marine Biological Target Detection Technology [D]. Harbin Engineering University, 2019.

[11] Tian Manjun, Li Xiali, Kong Shihhan, et al. Visual detection algorithm for underwater garbage cleaning robot based on improved YOLOv4 (English) [J]. Frontiers of Information Technology & Electronic Engineering, 2022, 23(08): 1217-1229.

[12] Wang Yannian, Li Ying, Lian Jihong. Lightweight YOLOv5s underwater debris detection method [J]. Journal of Xi'an University of Engineering , 2025, 39(02):39-46.

[13] Yuan Hongchun, Zang Tianqi. Underwater debris target detection based on attention mechanism and Ghost-YOLOv5 [J]. Environmental Engineering, 2023, 41(07): 214-221.

[14] Han Zhiyin, Liu Xiaoqun, Hao Juan. Research on marine debris detection based on improved YOLOv9c [J]. Yangtze Information and Communication, 2024, 37(08):28-30.