IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 23 September 2024, accepted 30 October 2024, date of publication 4 November 2024, date of current version 13 November 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3490799

RESEARCH ARTICLE

MGC-YOLO: Underwater Biomimetic Shrimp With Foreign Object Recognition

YING TANG, HONGLI YIN, CHANGHAO QUAN, QIAN MU, AND XIAOYUN ZHOU
School of Mechanical and Electrical Engineering, Chengdu University of Technology, Chengdu 610000, China

Corresponding author: Hongli Yin (yinhongli@stu.cdut.edu.cn)

This work was supported by Sichuan Provincial Science and Technology Plan Project under Grant 2022YFS0531.

ABSTRACT We propose a low-cost, small-volume, and easy-to-use underwater bionic robot shrimp to solve the problem of complex underwater scenes and the limited ability of target feature extraction. Specifically, an improved YOLOv8s underwater bionic robot shrimp model with foreign body recognition is proposed by introducing the improved Multi-scale Ghost (MGHost) convolution into the CSPDarknet53 to 2-StageFPN (C2F) module and finally adding an improved attention mechanism Convolutional priority multi space attention module (CPMS) after the Spatial Pyramid Pooling-Fast (SPPF), the MGC-YOLO algorithm model was proposed. At the same time, the bionic robot shrimp has a detachable manipulator design and can then achieve underwater movement, lifting, and hovering functions so that the underwater robot shrimp can carry out underwater maintenance, sampling, and other operations. Secondly, the image transmission module can cooperate with the APP side to control the machine shrimp, and the APP side has a deep waveform display function. The experimental results show that compared with the original model, the improved model reduces the number of parameters and the amount of calculation by 0.94 and 3.6, respectively, and improves its accuracy and Mean Average Precision (mAP) by 5.7% and 8.9%, respectively. After deploying the equipment, the lifting speed reaches 50mm/s, the time for the manipulator to identify the foreign body is less than 0.2s, and the delay for the return of underwater information is less than 0.2s, which verifies the feasibility and effectiveness of the model in underwater foreign body recognition.

INDEX TERMS Bionic shrimp, YOLOv8s, CPMS, detachable manipulator design.

I. INTRODUCTION

At present, according to national data, the earth's total area is approximately 510 million square kilometers, the total area of the ocean is approximately 361 million square kilometers, and the ocean's surface area accounts for approximately 70.5% of the earth's surface area [1]. However, with the continuous increase in the country's population and increasing demand for industrial and agricultural development, the water resources in nature are decreasing exponentially. At the same time, according to the Annual report of China's Ecological Environment Statistics, about 1.15 million to 2.41 million plastic fragments are discharged into the ocean through rivers every year [2], equivalent to about a truck of garbage dumped into the sea every minute [3]. However, the degradation of most plastic waste takes hundreds of years, and the long-term accumulation will drift to every part of the water with the ocean system. The water system will eventually absorb the garbage into the food chain, which the human body will absorb. It poses a potential threat to the biosphere and seriously affects human health. Therefore, it is urgent to protect the marine system's environment.

With the rapid development of artificial intelligence, artificial intelligence has gradually come to all fields of life. Traditional manual garbage salvage has low efficiency, security risks, and poor human-computer interaction [4]. In this paper, we design an underwater bionic machine shrimp with foreign body recognition against this background. The

The associate editor coordinating the review of this manuscript and approving it for publication was Shaohua Wan.

underwater bionic robot shrimp can identify and detect the target through the algorithm of the host computer. Then, the host computer sends instructions to bionic shrimp to solve the dredging function in the narrow water area and complete complex underwater tasks. Bionic robotic shrimp can improve work efficiency, reduce labor, maintain safety, and enhance reliability. [5]In response to this issue, a biomimetic robot underwater intelligent detection framework has been proposed, which can achieve intelligent motion planning, yaw control, and obstacle avoidance, providing a valuable solution for underwater operations in the ocean. However, the rapid development of computer vision technology provides new opportunities for improving the visual capabilities of underwater robots [6].

Underwater robots are being applied in increasing fields, including environmental monitoring, search and rescue, and industrial inspections. The typical workflow design for underwater object detection utilizes robots for complex underwater environments. This article uses a biomimetic machine shrimp to detect foreign objects underwater, approach them, and finally pick them up. At present, a relatively novel approach is autonomous mobile robots, whose primary purpose is to achieve agile path tracking and safe collision avoidance in chaotic environments [7]. The rise in robotics technology has led to increased interest in three-wheeled mobile robots (TWMRs) due to their agility and adaptability across various applications [8]. However, based on the technology of three-wheeled mobile robots, intelligent models based on neural networks are gradually emerging for real-time control of quadcopter drones [9]. When operating autonomously, underwater systems tend to use a different set of specialized sensors, including (3D) sonar, acoustic sensors, Doppler velocity logging devices (DVL), and inertial measurement units (IMU) but primarily rely on two-dimensional imaging sonar measurements for underwater positioning and navigation. [10]Describes feature-based underwater positioning and navigation using two-dimensional imaging sonar measurements but fails to address potential limitations and requires improvement in areas of use, charging systems, and other vital aspects such as control schemes. However, in the bionic robotic shrimp, the main task is underwater foreign object detection, so target detection algorithms are essential to complete subsequent tasks after successful recognition. Service-oriented systems are typically environment-aware [11] and may generate multidimensional quality data. Recently, some CNN-based methods have improved models' performance with different magnifications, which is different from various segmentation methods based on MS images [12]. However, there are still challenges in using information for images with different magnifications. Traditional recognition methods involve manually designing features and using classifiers to achieve recognition. Deep learning has gradually emerged with the rapid development of neural networks [13]. Deep learning methods train network models end-to-end through big data, effectively overcoming the limitations of traditional methods [14]. In object detection, deep learning mainly falls into two categories: two-stage and one-stage [15]. The two-stage algorithm first uses the Region Proposal Network (RPN) to generate candidate boxes that may contain targets in the input image. Then, it sends the candidate boxes to the classification network for classification operations. Representative algorithms include R-CNN [16], Fast-RCNN [17], and Faster-RCNN [18]. The first stage algorithms mainly include SSD and YOLO series. Although existing object detection methods have shown excellent performance in general images, underwater object detection still poses challenges.

According to domestic and international research, Chen et al. [19] proposed SWIPENet in 2020 and validated it with robot selection data URPC2017 and URPC2018, obtaining 46.3% of mAP. However, its recognition accuracy could be higher, and there are false positives. In 2018, Fulton et al. [20] used the Faster RCNN algorithm to achieve a detection accuracy of 83.3% for plastic recognition in garbage recognition. However, the detection speed could have been faster and met the real-time detection accuracy of garbage categories underwater. Zeng et al. combined the Faster RCNN network with an adversarial occlusion network to construct a new network called the Faster RCNN-AON network [21]. This network performs well suppressing overfitting during model training, but its detection accuracy is low. Song et al. improved the Mask RCNN network, which can perform object detection in complex underwater environments [22]. Although the proposed region proposal-based object detection algorithm has high detection accuracy, it is time-consuming and not suitable for real-time underwater object detection tasks. In 2024, Gao et al. [23] proposed a path-enhanced Transformer detection framework to solve the detailed problem of small-scale underwater targets. They also set up a way to embed local path detection information to achieve real-time detection tasks, but the detection accuracy is not high. The main focus of the first stage of algorithm research is that in 2021, Dinakaran et al. [24] proposed an algorithm model combining DCGAN+SSD to improve the model's performance in various aspects of object detection. In the same year, Lei et al. [25] improved the path aggregation network (PANet) of the YOLOv5 model. They introduced the Swin Transformer as the backbone network of YOLOv5, making it more practical for underwater images with blurred targets. The improved accuracy can reach 87.2%. In 2022, Tian et al. [26] proposed an improved YOLOv4 model to deploy and implement a garbage cleaning robot model, mainly transforming YOLOv4 into a four-scale detection method to improve detection speed and the detection speed can reach 66.67 frames per second. In 2023, Wen et al. [27] embedded SE and CA modules based on the YOLOv5s model to improve detection accuracy. Then, the number of necks in the first C3 was increased from 1 to 3, improving attention to the target. They conducted experiments on data from the China Underwater Robot Competition, and the

results showed an average accuracy improvement of 2.4%. In 2024, Wen et al. [28] proposed the YOLOv8-C2f Faster EMA algorithm, which mainly focuses on optimizing the backbone layer, neck network, and C2f module, improving the accuracy of underwater garbage detection.

However, no matter the two-stage or one-stage algorithm model mentioned above, the number of parameters must be more significant. In which the two-stage detection takes a long time and has low accuracy, it can not meet the real-time detection of underwater foreign bodies. Given the above research, compared with YOLOv8, YOLOv8 has significant advantages in precision and speed. Its model is more easily deployed than other algorithms. Therefore, we choose the YOLOv8s model as the primary detection algorithm model.

Therefore, our research aims to achieve two objectives. Firstly, our goal is to improve the recognition accuracy and precision of the YOLOv8s model. Secondly, we will deploy the identification number model to embedded devices to realize the underwater bionic machine shrimp model for foreign object recognition. This article provides four different contributions:

In this paper, the contributions of the underwater bionic robot shrimp on foreign body recognition are as follows:

1. We introduced the Ghost module and improved it to obtain a multi-scale Ghost module, namely the MGhost module, which is embedded into the C2f module. The C2f module selected is the C2f module in the backbone network, called the C2f MGConv model, making the model lightweight and more conducive to deploying embedded devices. Through experimental analysis, the introduction of a multi-scale Ghost model not only reduces the number of model parameters but also improves the accuracy.

2. We use the CPCA model to replace the channel attention mechanism and spatial attention mechanism module in Convolution Block Attention Module (CBAM) with the Multi-Scale Convolutional Attention (MSCA) module. The improved model is the CPMS model, placed after SFFP, and adds an enhanced attention mechanism CPMS layer. We propose the MGC-YOLO algorithm model. When improving the attention mechanism, the mAP for underwater foreign object detection has also been enhanced, surpassing previous algorithm models.

3. After algorithm improvements, the STM32H7 chip used in the hardware has an RT-Thread operating system. The PID algorithm controls the depth stability of the bionic shrimp in the water, achieving the hovering function, and enables the robot arm to capture the identified target.

4. Use the mobile phone interface to observe the underwater depth waveform in real time and control the bionic shrimp capture target to provide more accurate information for job decision-making.

Through rigorous experiments, we have verified that the model generated by the proposed method is effective for foreign object recognition underwater.

The rest of this article is organized as follows. Section II outlines several related works. Section III introduces the method of optimizing the model in detail. Then Section IV presents the data set, the experimental environment, and the analysis of the experimental results, compares it with the previous research, and discusses the limitations. Finally Section V summarizes the main conclusions of this study.

II. RELATED WORK

A. DESIGN OF BIONIC SHRIMP

1) DIMENSIONAL COMPOSITION

Foreign bodies picked up by mechanical arms play an important role in underwater autonomous detection. It can not only prevent pollution with underwater organisms but also provide healthy seafood for human beings. By integrating advanced sensing technology, the foreign body recognition system can obtain accurate data about the underwater environment. Therefore, this paper proposes an underwater bionic shrimp with foreign body recognition. Studying the independent exploration method using the bionic robot fish as the carrier [29], we conduct the design of the bionic shrimp, illustrating the conceptual design of the bionic robot shrimp as depicted in FIGURE 1. The system of foreign body recognition bionic shrimp is mainly composed of the primary control device, mechanical arm, and power device, which includes two manipulators, which are respectively composed of four waterproof steering gears, which are used to control up and down, left and right movement and end rotation. The grip at the end of the manipulator realizes the picking function. Additionally, we can divide the conceptual calculation module of the bionic shrimp into the upper decision layer and the bottom control layer. Specifically, the top-level decision-making module applied to deep reinforcement learning planning uses the NVIDIA Xavier NX edge computing platform to provide high-performance CPU, GPU, and AI computing capabilities for robots and other applications. The bottom layer adopts the STM32H7B0VBT6 microcontroller and ARM Cortex-M7 core, which can meet the requirements of high computing performance and low power consumption.

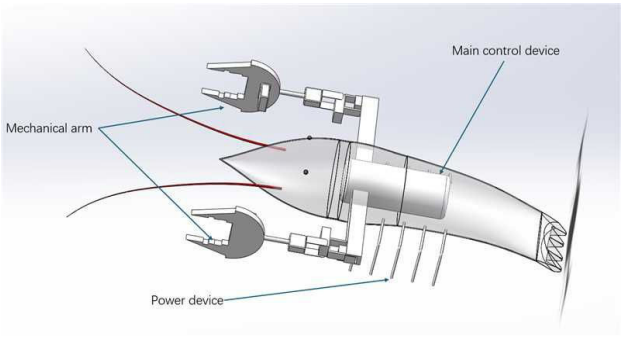


FIGURE 1. The structure diagram of bionic shrimp.

The whole structure of the self-deformed bionic robot shrimp adopts a pressure-free structure, and the internal equipment and electronic equipment are installed in

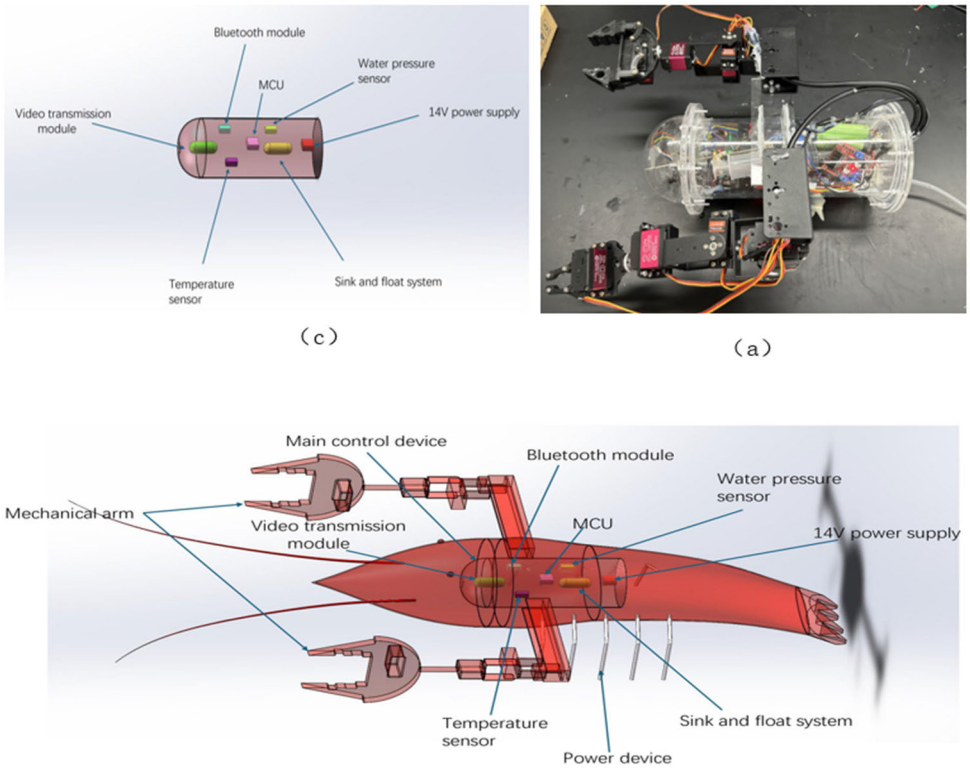


FIGURE 2. The prototype of the bionic robot shrimp displayed; (b) The three-dimensional overall structure model of the bionic robot shrimp; (c) The three-dimensional model of the internal power plant.

an independent pressure-sealed chamber. As shown in FIGURE 2, we divide the overall design into the primary device, the mechanical arm, and the power plant. The central part includes, The main chip; Water pressure sensor; Power supply; Temperature sensor; Picture transmission module; Bluetooth module; Ups and downs system. The left and right solid wings include two mechanical arms, each composed of four waterproof steering gear.

2) OVERALL PICTURE OF UNDERWATER ROBOT SHRIM

The underwater intelligent operation robot integrates several modules, and the central control based on the STM32H7B0VBT6 chip realizes close coordination and efficient interaction so as to complete the underwater intelligent operation task.

First, as the central control part, the STM32H7 chip has an RT-Thread operating system and robust computing and control capabilities. Then, the underwater motion control, lifting control, and hovering control modules realize intelligent motion and stable hover through the STM32H7 chip so that the robot can accurately perform tasks underwater. The manipulator control module controls two detachable manipulators to achieve object grasping and handling operations. Then, the picture transmission module transmits the underwater picture to the mobile phone APP in real-time, allowing the operator to observe and remotely monitor. Secondly, the communication module realizes the two-way data transmission between the robot and the mobile phone APP so that the operator can remotely control the robot and receive its feedback information. Finally, the depth sensing module measures the underwater depth and transmits the data to the mobile phone APP to provide depth information.

Overall, through the central control of the STM32H7 chip, the modules realize data exchange and coordination, cooperate closely, and work together so that the robot can have an intelligent underwater operation and improve the efficiency and safety of underwater resource exploration, scientific research, and environmental monitoring. The detailed process of this article is as follows: after booting up, the system initializes, and the bionic machine shrimp is placed in water to enter hover mode. When the bionic shrimp patrols in the water, the detection camera on the top of the bionic shrimp sends real-time data to the image transmission module through a USB data cable. The image processor then recognizes the underwater image. The visual recognition system sends a signal to the controller to execute the salvage program, driving the motor to rotate the bionic shrimp's hand and pick up foreign objects. Its physical structure is shown in FIGURE 3.

B. CONTROL OF BIONIC SHRIMP

1) MODULE PARSIN

The data exchange and coordination among the modules are realized through the central control of the STM32H7B0VBT6 chip. As shown in FIGURE 4, it is the

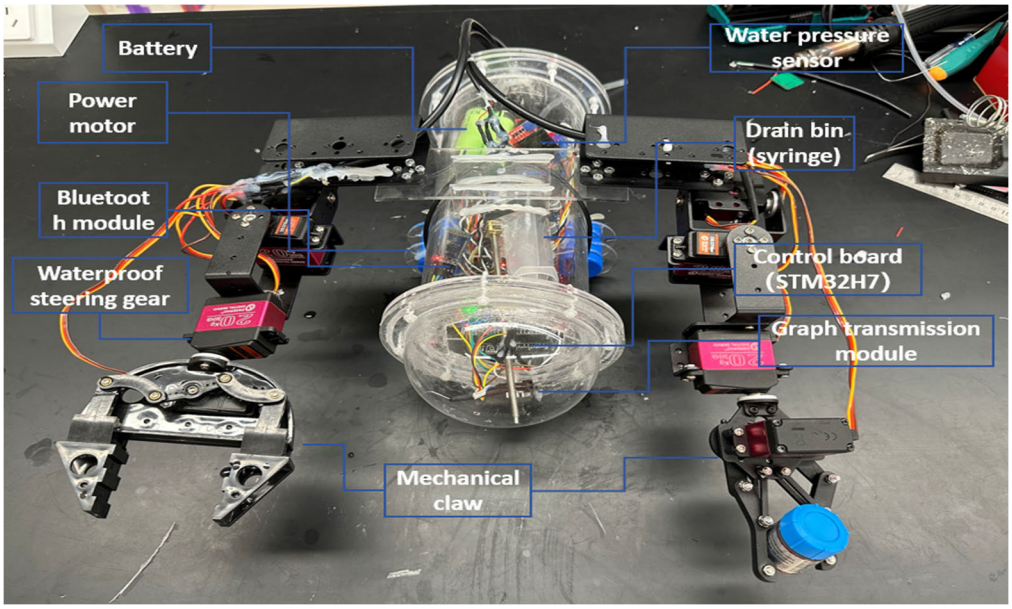


FIGURE 3. A real effect picture.

most miniature system board connecting the modules. Each module cooperates closely and works together to give the robot the ability to perform intelligent underwater operations and improve the efficiency and safety of underwater resource exploration, scientific research, and environmental monitoring. The introduction of each module is as follows, as shown in FIGURE 5, which is the overall module display diagram of the bionic shrimp.

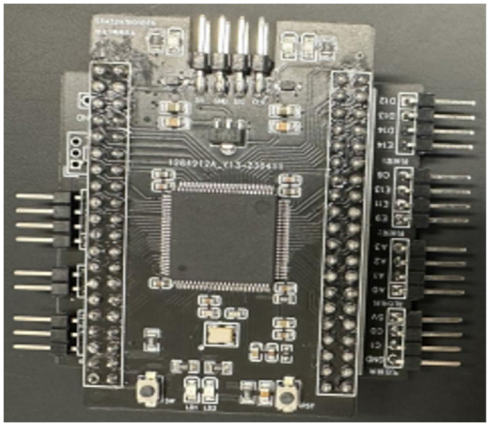


FIGURE 4. Minimum system board.

We will introduce the functions of each module in the modules.

a: GRAPHIC TRANSMISSION MODULE

The image transmission module is installed on the head of the underwater intelligent operation robot and is directly connected to the mobile phone through WIFI. The WIFI module has become the primary medium for communication between mobile phones and machines. The image transmission

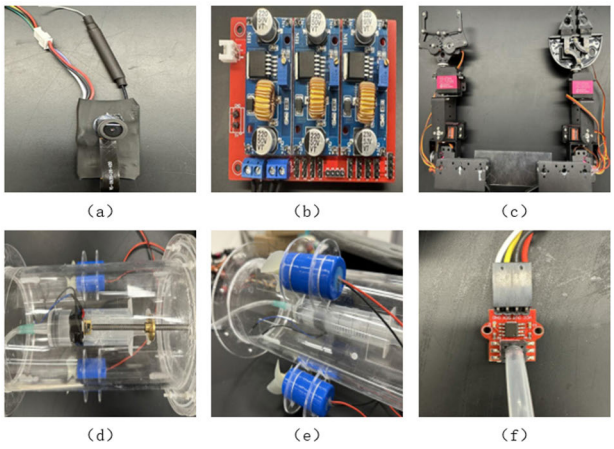


FIGURE 5. (a) Graphic transmission module; (b) Power management board; (c) Mechanical arm; (d) Floating system; (e) Power motor; (f) Water pressure sensor.

module can transmit underwater images to the mobile phone, allowing users to observe the underwater environment in real-time. Ensure efficient and stable image transmission through appropriate image transmission equipment and protocols. Users can easily monitor and control the robotic shrimp through a mobile APP, enhancing the maneuverability of the robotic shrimp and the efficiency of underwater operations, providing a convenient and reliable solution for intelligent underwater operations.

b: POWER MANAGEMENT BOARD

Due to the high power consumption of various modules of the underwater robot, as well as the sizeable total power consumption of the servo module of the robotic arm and the motor controlling the water storage tank, we use a power

supply consisting of four Panasonic NCR18650B batteries connected in series, which has an output voltage of 14V and a working current of up to 10A. At the same time, due to the independence of each functional module, different power supply voltages are required. Therefore, we designed and produced a power management board, which divides the 14V battery power supply into 12V and 5V through a switch power module and supplies power to the motor and servo, respectively. Then, the 5V is divided into 3V3 through a linear regulator module to supply power to the central control chip and sensors.

c: MECHANICAL ARM

The control module of the robotic arm is the core part of the underwater intelligent operation machine shrimp. It consists of two robotic arms, each consisting of four waterproof servos, used to control up and down, left and right movements, and end rotation. An independent servo controls the gripper at the robotic arm's end to achieve grasping and releasing functions. Using the PWM output port of the STM32H7B0VBT6 chip, the pulse signal of each servo is controlled to achieve precise motion control.

The control module of the robotic arm communicates with the central control system. It can flexibly execute underwater tasks according to instructions, improving the multifunctionality and adaptability of the robotic shrimp. At the same time, the design of the robotic arm structure and servo drive circuit considers waterproof performance, using waterproof servos to ensure the reliable operation of the robotic shrimp in underwater environments. The design of this module enables the robotic shrimp to complete complex tasks such as underwater maintenance and sampling, providing an efficient and reliable solution for underwater resource exploration and operations.

d: FLOATING SYSTEM

The sinking and floating system of underwater robotic shrimp use a syringe as a storage tank and adjusts the drainage volume by controlling the suction or drainage process of the syringe through the motor and screw inside the robotic shrimp, achieving underwater lifting function.

The motor and screw system inside the machine shrimp control the movement of the flange, which does not come into contact with water, ensuring that the components inside the machine shrimp are not affected by water. Through the commands of the control system and feedback from the water pressure sensor, the robotic shrimp can achieve stable lifting and lowering movements underwater, adapt to underwater environments at different depths, and improve the flexibility and efficiency of underwater operations. The overall design is simple and efficient, making the sinking and floating system of the robotic shrimp an essential component for achieving intelligent underwater operations.

e: POWER MOTOR

The power motors of the underwater machine shrimp are two waterproof motors installed on both sides of the machine shrimp. The L298N drive module controls them and can achieve movements such as forward, backward, and steering of the robotic shrimp underwater. Ensure the reliability and stability of the motor in underwater environments through appropriate power supply and waterproof design. The installation position of the power motor and the L298N control module enables the robotic shrimp to adapt to the needs of different underwater tasks flexibly, providing strong power support for the entire robotic shrimp. This power motor design allows the underwater intelligent operation machine shrimp to move efficiently, achieve diverse underwater operation tasks, and improve the performance and adaptability of machine shrimp.

f: WATER PRESSURE SENSOR

The water pressure sensor of this work is the HD710 model sensor, installed inside the machine shrimp and connected to the outside of the work through a hose. It is used to measure the current water pressure and thus determine the depth of the machine shrimp underwater. The sensor is connected to the STM32H7B0VBT6 chip to convert the measured pressure value into depth information. Deep information is transmitted in real-time to the mobile app through a Bluetooth module for operators to monitor the underwater environment. At the same time, depth information is also used to provide feedback for the buoyancy system. By controlling the motor screw to adjust the displacement, the machine shrimp can stably hover at the designated depth position, improving the accuracy and efficiency of underwater operations. This design ensures that the sensor does not come into contact with water during underwater operations, making the sensor's measurements more accurate and reliable.

2) PID ALGORITHM

This design must maintain its depth stability in underwater operation, so it combines a depth sensor and PID algorithm to control the water tank system. When the design reaches the underwater target position, it can maintain depth stability in the face of external factors. The design goal of the PID algorithm is to quickly recover to the target position error of less than one centimeter in case of interference. And the oscillation phenomenon is not prominent.

The PID algorithm operates based on the combined operation of three controllers: proportional, integral, and differential. The proportional controller adjusts according to the difference between the controlled quantity and the expected value; the integral controller adjusts according to the cumulative error between the controlled quantity and the expected value, which is used to remove the static error; the differential controller adjusts according to the rate of change of the controlled quantity, It restrains overshoot and oscillation in the dynamic process.

The PID algorithm adjustment program built in MCU automatically adjusts the depth in real-time. After importing the data of the depth sensor into the parameters of the PID algorithm, the system can calculate the direction and power

of the motor in real time and import the estimated parameters into the DC motor driver to drive the motor to change the volume of the water tank. As a result, the submarine is close to the depth of the target. In the case of reasonable parameter setting, the farther away the submarine is from the target, the faster the adjustment is, and the closer it is to the target, the slower the adjustment is to achieve fast and accurate depth adjustment. The PID algorithm uses ultrasonic and steering gear speed sensors to convert the collected environmental parameters into electrical signals when the underwater vehicle is working and then sends them to the STM32 controller. Then, the STM32 controller processes and analyzes the electrical signals and finally sends the processed signals to the motor and other executive components, thus realizing the intelligent operation of the underwater vehicle. FIGURE 6 shows its working principle.

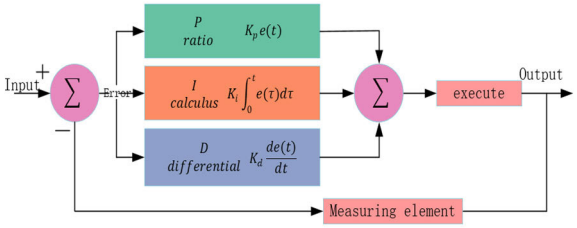


FIGURE 6. How PID works.

The relation of the transfer function of the PID mathematical model in the control system, as shown in equations 1 and 2.

$$e(k) = r(k) - c(k) \tag{1}$$

$$u(k) = K_p \left\{ e(k) + \frac{T}{T_I} \sum_{j=0}^{k} e(j) + \frac{T_D}{T} [e(k) - e(k-1)] \right\} \tag{2}$$

where: index k represents sequential sampling instances, incremented to 0, 1, 2, etc., to mark discrete evaluation or adjustment times. r(k)represents the target value at each k instance and is the system's output target. c(k) captures the actual input value of the same instance, reflecting the received real-world input. The control output u(k) corresponds to the system and aims to solve the difference between the control output and the target value. The deviation e(k) quantifies the gap between the target r(k) and the actual input c(k) at each instance, indicating the error that the system is minimizing. The previous deviation e(k-1) provides a basis for comparing the variation of errors with time. The proportional coefficients P and K affect the immediate response of the system to the current error, the integral time constants I and T affect the cumulative error correction with time, and the differential time constant DT affects the predictive response of the system to error changes. Finally, T represents the sampling period and determines the interval between online observation and adjustment, thus indicating the response rhythm of the system.

3) INTODUCTION TO HOVER FUNCTION

Implementing the hovering function in underwater environments is challenging, and achieving underwater hovering is crucial for precise operations. The robotic shrimp must hover stably at a designated position for foreign object recognition and sampling tasks.

The machine shrimp achieved stable hovering underwater through a well-designed drainage bin control system and PID control algorithm. This innovation enables the robotic shrimp to be more accurately positioned and controlled during operations.

The sinking and floating system of underwater robotic shrimp uses a syringe as a water storage tank. Through the motor and screw inside the robotic shrimp, the suction or drainage process of the syringe is controlled to adjust the drainage volume, achieving underwater lifting and hovering functions. As shown in FIGURE 7

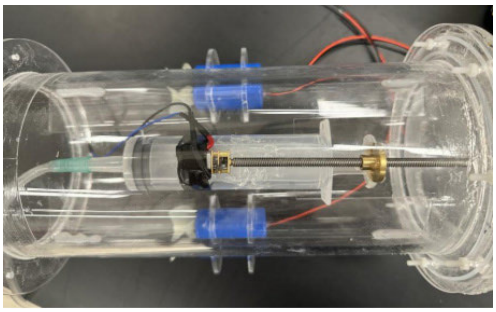


FIGURE 7. Water storage tank.

The hardware functions mainly consist of three parts: a DC motor and drainage chamber, an air pressure sensor, and a microcontroller.

1. DC motor and drainage chamber

The DC motor is mounted on the drainage chamber (syringe), and the volume of discharge or suction of the drainage chamber (syringe) is controlled by controlling the movement of the motor, thereby controlling the sinking and floating of the submarine.

2. Air pressure sensor

Used as a water level sensor to read the current depth of the submarine in real-time.

3. Single-chip microcontroller

Processing the data transmitted by the air pressure sensor, applying specific algorithms to process the data from the air pressure sensor, and sending control signals to the DC motor based on the processed data to control the motor is the control core.

The detailed introduction of its software program is divided into four parts: the HD710 pressure sensor, DC motor driver program, PID algorithm, and operating system.

1. HD710 pressure sensor

Controls the pressure module's real-time measurement accuracy and frequency and correctly returns real-time depth to the microcontroller.

2. DC motor

By adjusting the duty cycle of the PWM control signal, high-precision motor power control can be achieved over an extensive range, improving motor stability and operating efficiency.

3. PID algorithm

The MCU's built-in PID algorithm adjustment program is the basis for real-time automatic depth adjustment. After importing the data from the depth sensor into the parameters of the PID algorithm, the system can calculate the direction and power of the motor in real-time and import the estimated parameters into the DC motor drive program to drive the motor to change the volume of the water tank. As a result, the submarine also approached the target depth. Under reasonable parameter settings, the further the submarine is from the target, the faster the adjustment, and the closer it is to the target, the slower the adjustment, thus achieving fast and accurate depth adjustment. PID regulation is the communication bridge between the motor and the depth sensor.

4. Operating System

Embedded operating systems can integrate the above three through thread synchronization, running synchronously at a macro level, which can control submarines to quickly and accurately dive to the target depth while also resisting external interference when suspended, maintaining hovering at the target depth as much as possible and reducing oscillation.

The system diagram of its implementation is shown in FIGURE 8.

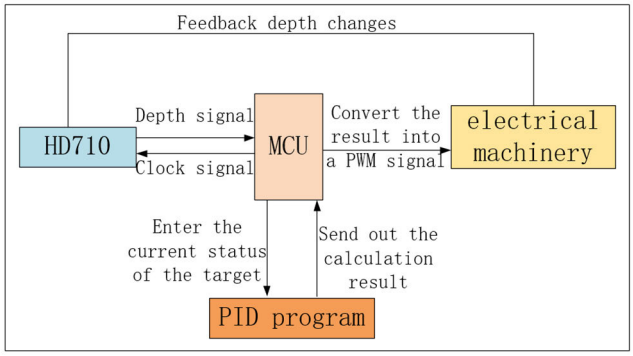


FIGURE 8. System block diagram.

C. CONTROL OF BIONIC SHRIMP

1) YOLOV8 MODEL FRAMEWORK

YOLOv8 is a state-of-the-art object detection system known for its accuracy and speed, tailored to meet the unique needs of aerial surveillance in marine environments. It retains YOLOv5's CSP concept, integrated feature fusion technology (PAN FPN), and SPPF module, ultimately achieving a series of enhancements. This includes introducing state-of-the-art models for object detection and instance segmentation at various resolutions, a coefficient scalable model based on YOLOv5, and a novel C2f module inspired by the ELAN structure in YOLOv7. In addition, YOLOv8 has innovated in the detection head by separating classification and detection functions, using binary cross entropy (BCE) for classification loss, and introducing a complex form of regression loss (CIOU loss+DFL and VFL) to improve detection accuracy. A key feature of YOLOv8 is its scalability and compatibility with previous YOLO versions, which helps with performance comparison analysis. This feature, coupled with its advancements in accuracy and the introduction of anchor-free dynamics and dynamic task alignment allocators, makes YOLOv8 the most accurate detector to date. This algorithm's adaptability and improved computational and detection efficiency make deploying embedded devices more convenient, so the latest model was chosen for foreign object recognition in underwater biomimetic machine shrimp. Finally, the feature extraction network in YOLOv8 is designed to extract single-scale features from images processed by SPPF and C2f modules. The C2f module is a simplified version of the original C3 module, which combines convolutional layers to reduce the burden on the model while integrating the advantages of YOLOv7's ELAN structure. This enhancement broadens the gradient flow information using bottleneck modules in the gradient branches. The SPPF module reduces the number of layers in the standard SPP (Spatial Pyramid Pooling), reduces unnecessary operations, and accelerates feature fusion.

In real-time target detection, the YOLO algorithm is a model of innovation and efficiency, and many widely praise its pioneering function. The popularity of this algorithm stems from its lightweight network architecture, effective feature fusion methods, and more accurate detection results. In its iterative process, YOLOv5 and YOLOv7 have become prominent versions, using deep learning to achieve efficient and real-time target detection [30]. With the introduction of YOLOv8 in 2023, its structure retains the CSP module of YOLOv5, integrates feature fusion technology and SPPF module, and finally realizes a series of enhanced functions. Due to the influence of YOLOv7, C2f innovates to replace the original structure in the Efficient Layer Attention (ELAN) module, YOLOv8 adopts a new classification loss function, Varifocal Loss (VFL Loss), which can better handle the category imbalance problem and enhance classification accuracy. Second, the regression loss function of YOLOv8 adopts the combination of Complete Intersection Over Union Distribution Focal Loss (CIOU DFL Loss), which can handle the bounding box regression problem more effectively and improve detection accuracy.

The main structure of YOLOv8 comprises a backbone network, a neck network, and a detection head. The backbone network is used to extract the features of the image, and the C2f module is a simplified version of the original C3 module, which integrates a convolution layer to reduce the model and integrates the ELAN advantages of YOLOv7. SPPF module minimizes the number of layers of standard SPP, reduces unnecessary operations, and speeds up feature fusion. The neck network is for feature fusion, and the detection head part is to output the final detection results. It uses features sampled from 8, 16, and 32 scales to perform multi-scale

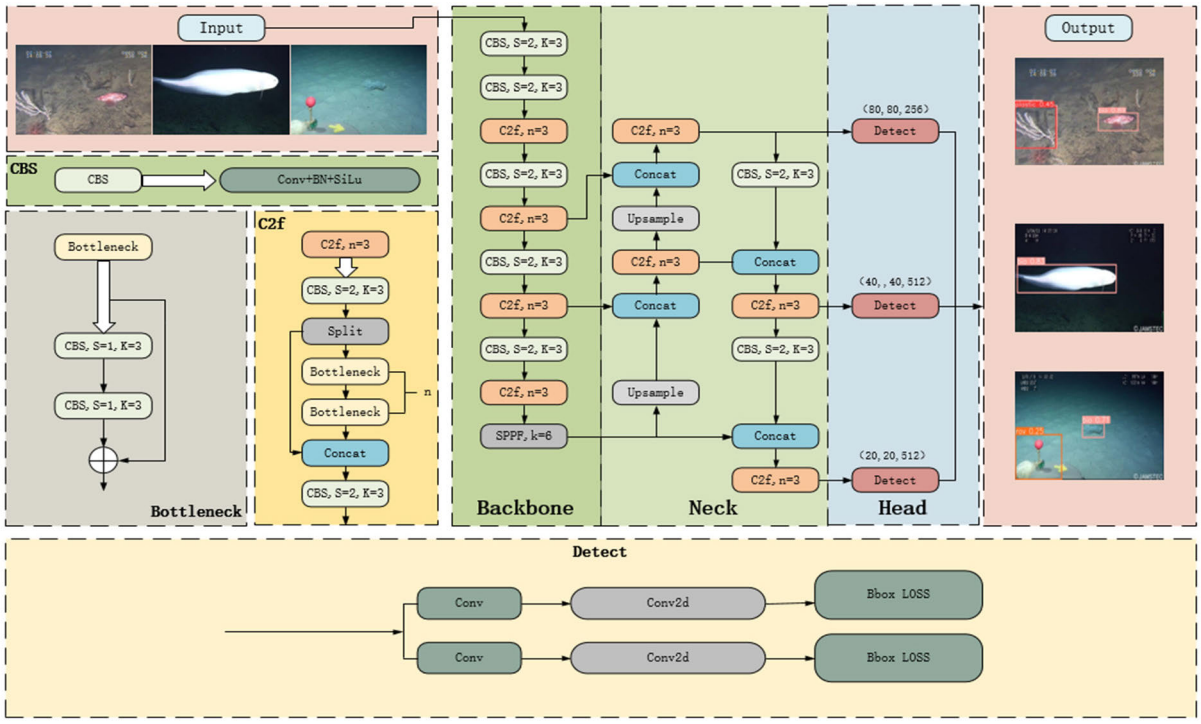


FIGURE 9. System block diagram.

prediction to fine-tune the prediction of small, medium, and large targets,and its overall model block diagram is shown in FIGURE 9.

2) LOSS FUNCTION OF YOLOV8 MODEL

IOU is a way to describe the degree of coincidence between boxes. In the regression task, the regression degree of the box can be measured by the ratio of the "goal box" to the "prediction box." The loss function of YOLOv8 consists of IOU is a way to describe the degree of coincidence between boxes. In the regression task, the regression degree of the box can be measured by the ratio of the "goal box" to the "prediction box." The loss function of YOLOv8 consists of several parts, classified as loss (VFL Loss) and regression loss (CIOU Loss + DFL). Among them, CIOU and YOLOv5 are consistent, and the main improvement of VFL is the asymmetric weighting operation. Both FL and QFL are symmetrical; the definition of Varifocal Loss is shown in the equation 3:

$$VFL(p, q) = \begin{cases} -q\left(q\left(\log p\right) + (1-q)\log\left(1-p\right)\right) & q > 0 \\ -\alpha p^{\gamma}\log\left(1-p\right) & q = 0 \end{cases} \tag{3}$$

q is the IOU of bounding box (bbox) and real box, IOU is the intersection of the prediction box and the real box divided by the union of the two boxes, and then p is the score, namely probability. When q is greater than 0, it is a positive sample, that is, if two boxes intersect, so that q is negative, then the two boxes do not intersect.

For DFL, the position of the box is modeled as a general distribution, so that the network can quickly focus on the distribution of the position close to the target location, as shown in FIGURE 10.

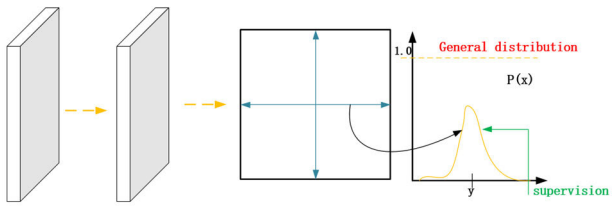


FIGURE 10. DFL general distribution.

The meaning of DFL expressed in the figure is to optimize the probability of the two positions closest to the label y in the form of cross entropy, so as to make the network focus on the distribution of the adjacent area of the target position more quickly; that is to say, the learned distribution is theoretically near the real floating-point coordinates, and the weight of the integer coordinates around the distance is obtained by the mode of linear interpolation. As shown in equation 4.

$$\begin{aligned} &DFL\left(S_i, S_{i+1}\right) \\ &= -\left(\left(y_{i+1} - y\right)\log S_i + \left(y - y_i\right)\log\left(S_i + 1\right)\right) \end{aligned} \tag{4}$$

In the formula, S_i and S_{i+1} are the "predicted value" and "approaching predicted value" of the network output. y, y_i and y_{i+1} are the "actual value", "label integral value" and "near label integral value" of the label.

III. METHODS

A. IMPROVE THE MODEL OF YOLOV8

1) LIGHTWEIGHT MODEL

In the model of YOLOv8, we can see that YOLOv8n has the smallest number of model parameters and the least amount of calculation. Still, its speed is relatively slow, and YOLOv8s is relatively moderate compared with other models, so this paper chooses YOLOv8s as the basic model Details of the parameters are shown in TABLE 1.

Using the C2f feature extraction backbone network in YOLOv8s is very large and needs a lot of computing resources, so this paper introduces an algorithm to improve the GhostNet model. GhostNet is a lightweight end-to-end neural network with significant advantages over other lightweight networks. Its design focuses on improving the computational efficiency of the model, maintaining accuracy while reducing the number of model parameters, making it very suitable for deployment on embedded devices and mobile devices [31]. Because the target detection object of this paper is underwater target detection and deployed to the mobile phone, this paper selects the lightweight backbone network of improved GhostNet to be used in the basic model of YOLOv8s to achieve efficient and accurate detection.

Han et al. based on the fact that one feature graph in the redundant feature graph can be transformed into another feature graph by cheap operation, but one of the feature graphs can be seen as a "Ghost" of another feature layer, and then based on this concept, a GhostNet network model [32] is proposed. The model GhostNet is divided into two parts: first, conventional convolution is used to generate feature graphs; second, additional graphs are generated through the linear transformation of the initial test set; and third, these mappings are combined to produce the final output. The core Ghost model structure diagram in GhostNet is shown in FIGURE 11: The calculation formula of Ghost model is as follows:

$$Y' = X \times F' \tag{5}$$

$$Y_{ij} = \phi_{ij}\left(Y_i'\right), i \in [1, M], j \in [1, G] \tag{6}$$

$$f_1 = M \times H' \times W' \times C_{in} \times K^2 \tag{7}$$

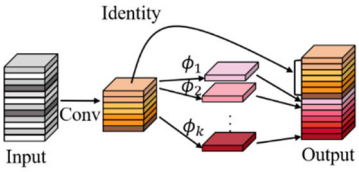


FIGURE 11. The Ghost module.

In the formula, ϕ is a cheap operation, which can be either depth wise convolution or other convolution, where c is the number of channels, h is the height and w is the width. The specific meaning is to generate m original features Y' after the input image X passes the convolution kernel Y_i' Then, the two main structures of GhostNet are shown in FIGURE 12, in which the step size of the left image in the depth separable convolution is 1, and it is used as an extension layer through the Ghost module, including two Ghost modules in the model to increase the corresponding number of channels. The step size of the figure on the right is 2, which also contains two Ghost modules, mainly to reduce the number of channels. The difference between the left and the right is that the right does not use ReLU after the last Ghost module.

According to the Ghost module, the multi-scale Ghost is improved. First, the multi-scale convolution layer does not change the size of the original feature image. Still, it enriches the features of the image through the convolution operation of different convolution kernels. The improved model structure is shown in FIGURE 13.

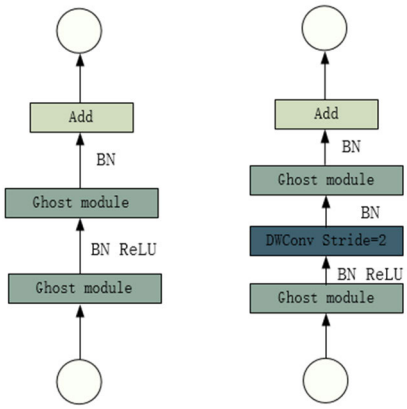


FIGURE 12. G-Bneck network structure.

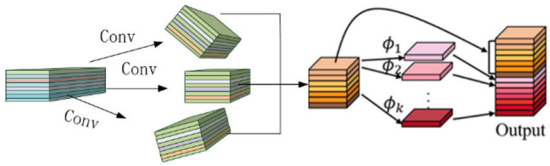


FIGURE 13. Improved Ghost model diagram.

2) LIGHTWEIGHT MODEL

Because the classical attention mechanism Squeeze-and–Excitation-Networks (SENet) [33] only considers the information of the internal channel and does not pay attention to the position information, in the visual effects, the spatial structure is equally important, which promotes the emergence of the attention mechanism of CBAM. Still, because CBAM ignores the critical details in underwater detection, this paper proposes an improved CBAM, called CPMS, as shown in FIGURE 14. CPMS includes not only a spatial attention mechanism module but also a multi-scale channel attention mechanism module, which can achieve better results than the SENet attention mechanism with only an attention channel.

The CPMS model proposed in this paper is inspired by channel-first convolution attention, and the overall structure

TABLE 1. YOLOv8 model parameters.

Model	Params (M)	FLOPs (B)	Speed TensorRT	mAP @0.5
YOLOv8n	3.2	8.7	0.99	37.3
YOLOv8s	11.2	28.6	1.20	44.9
YOLOv8m	25.9	78.9	1.83	50.2
YOLOv8l	43.7	165.2	2.39	52.9
YOLOv8x	68.2	257.8	3.53	53.9

of CPCA model still follows the structure of CBAM. First channel (Channel Attention Module) CA, and then (Spatial Attention Module) SA, the equation is as follows.

$$F_c = CA(F) \otimes F \tag{8}$$

$$\hat{F} = SA(F_c) \otimes F_c \tag{9}$$

The channel attention mechanism is improved to a multi-scale channel attention mechanism on the original channel. The formula of the original channel and the improved attention mechanism is shown in formula 5 and 11.

$$CA(F) = \sigma(MLP\,(AvgPool\,(F)) + MLP(MaxPool(F))) \tag{10}$$

$$CAM(F) = Conv_{1\times1}(\sum_{i=0}^{3} Input_i(DwConv(F))) \tag{11}$$

Then the expression of the spatial attention mechanism is formula 8, and the improved equation of the spatial attention mechanism is shown in 13.

$$SA(F) = \sigma(f^{7\times7}\,([AvgPool\,(F)) + MLP(MaxPool\,(F)])) \tag{12}$$

$$SAM(F) = Conv_{1\times1}(\sum_{i=0}^{3} Branch_i\,(DwConv\,(F))) \tag{13}$$

In the above formula, F_c is the feature graph, CA and SA represent channel attention and spatial attention respectively, $\otimes$ represents element-by-element multiplication, and F_c and $\hat{F}$ represent the feature output graph after channel attention and spatial attention, respectively. Where σ is signmoid, where Dwconv represents the depth convolution, $Input_i$, $i\in\{0, 1, 2, 3\}$ represents i the branch, $Input_0$ represents the characteristic graph of the input image, $f^{7\times7}$ represents a convolution operation with filter size 7×7, $Branch_i$, $i\in\{0, 1, 2, 3\}$ represents the i branch, $Branch_0$ represents the characteristic graph after the output of the channel attention mechanism, and finally CAM (F) and SAM (F) represent the improved multi-scale channel and multi-stew spatial attention module.

B. IMPROVE THE MODEL OF YOLOV8

First of all, due to the large number of parameters of the model, the improved Ghost lightweight model is introduced and added to the C2f module of the backbone network to generate the C2f-MGhost model, which can extract more and more detailed features and lay the foundation for subsequent feature fusion. Then, by adding the attention mechanism CPMS after SPPF, we can increase the ability to pay attention to features, obtain refined features as output, enhance the model to pay more attention to the information of small targets and improve the accuracy of small target recognition and location. The blue box details the enhanced part, and the improved overall model framework is shown in FIGURE 15.

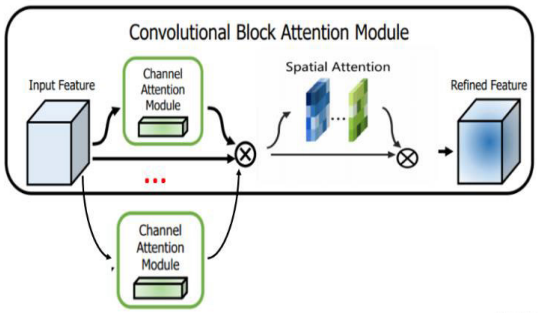


FIGURE 14. CPMS model.

IV. EXPERIMENT

A. INTRODUCTION TO THE EXPERIMENT

1) DESCRIPTION OF THE DATASET

The data set used in this paper is Trash_ICRA19 [20], which is currently widely used. The data set consists of three categories: plastic, biological species, and others. Four thousand pictures are selected according to Trash_ICRA19 and then collected according to the open underwater photos of the ocean on the Internet and integrated into the ROBAT-Date data set needed in this paper. It contains a total of 5800 pictures. The data set is divided according to 6:2:2, and the detailed data set classification and label data location size distribution show that the sample data set has the characteristics of multi-scale distribution, as shown in FIGURE 16.

There are three main categories of data sets, namely, plastic (plastic fragments), bio (biology), and rov (other categories). The recognition effect is shown in FIGURE 17.

The actual image data of the seafloor is shown in FIGURE 18, indicating white plastic garbage on the seafloor and biological and other prominent types of waste.

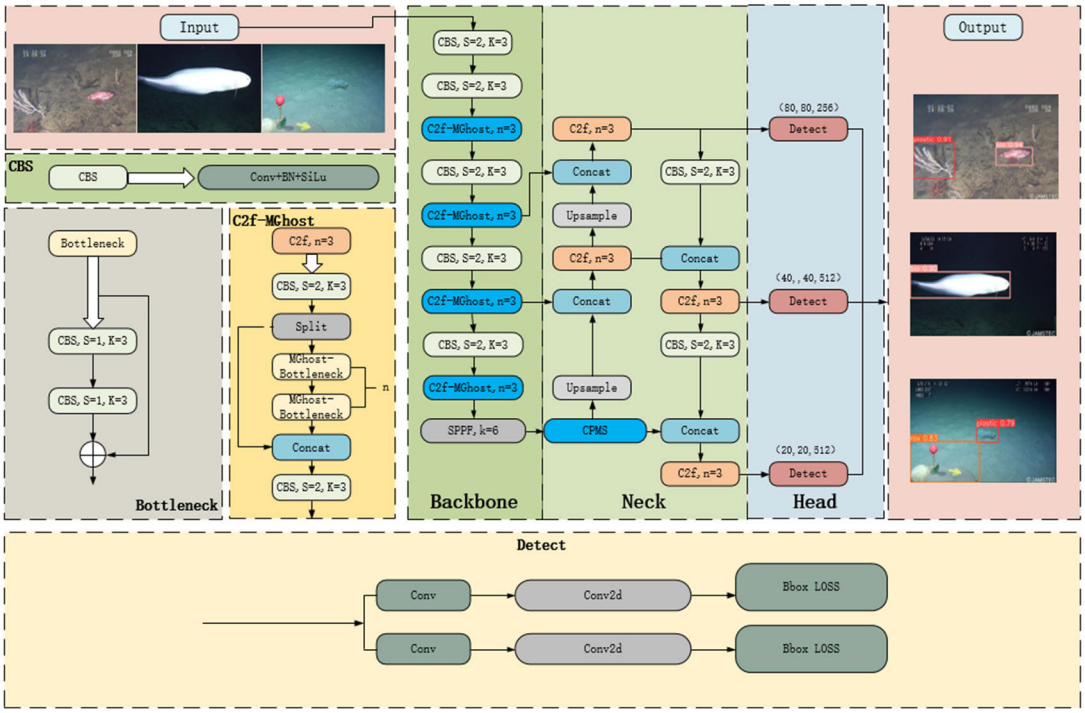


FIGURE 15. MGC-YOLO model.

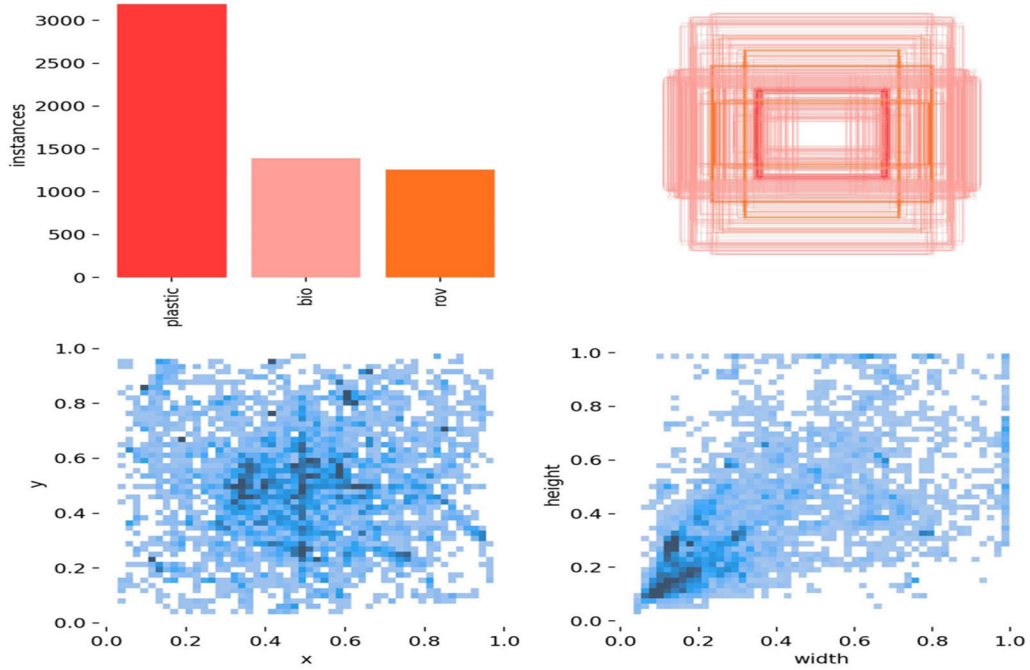


FIGURE 16. Tag information distribution.

2) EXPERIMENTAL ENVIRONMENT AND PARAMETER DESIG

The settings of the environment are shown in TABLE 2.

During the experiment, the training set of the data set is 6, and the training set of the dataset consists of 6 units, with a proportion of 2 units each allocated to the validation set and the test set, used for training, validating, and testing the improved algorithm model. The size of the obtained images is 640-640, the Batch-Size size is 16, the number of iterations for training (epoch) is 100, and the initial learning rate is 0.01. The model selects the SGD optimizer for its calculation speed

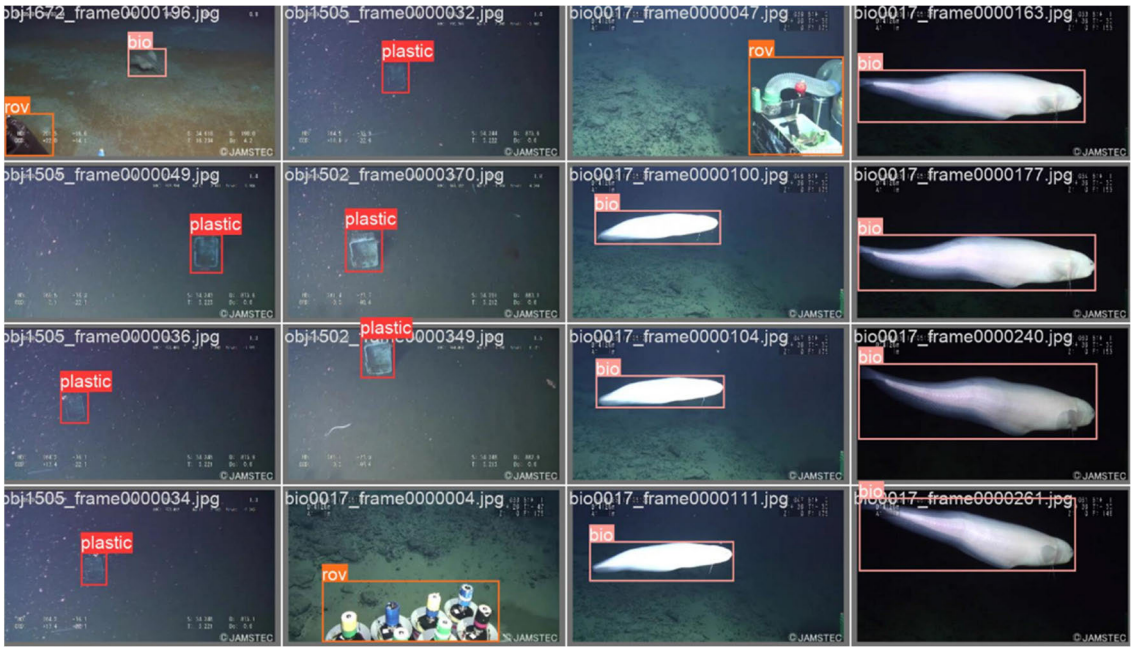


FIGURE 17. Visual example of a dataset.

and convergence speed; the momentum is set to 0.93, and the weight decay is set to 0.0005.

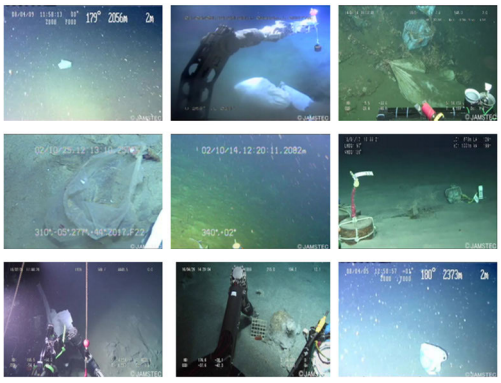


FIGURE 18. Real sea bottom image.

TABLE 2. Detailed data of experimental environment configuration.

Configure	Version
Frame	Pytorch 1.11.0
Python	3.8
CUDA	11.3
GPU	NVIDIA RTX A2000
Operating system	Ubuntu18.0.4

3) ASSESSMENTOF INDICATORS

To objectively evaluate the model's performance, the indicators selected in this experiment are accuracy (Precision, P) [34] and recall rate (Recall, R), which can comprehensively evaluate the detection effect of the algorithm model. Then, the average accuracy (mean Average Precision, mAP), the number of parameters (Parameters), and the number of floating-point operations (floating-point operations, FLOPs) are selected to evaluate the performance of the algorithm. According to the categories of the selected marine data sets, the average accuracy average mAP is used as the evaluation index, which comprehensively considers the detection accuracy and recall rate of marine target categories. Calculating the average accuracy of different categories reflects the model's overall performance in seafloor foreign body recognition, making the evaluation more comprehensive and objective. The number of parameters can measure the complexity of the algorithm model. In contrast, FPS can measure the real-time detection speed of the model, and FPS can be obtained by calculating the predicted consumption time. Each index expresses its formula as follows:

$$P = \frac{TP}{TP + FP} * 100\% \tag{14}$$

$$R = \frac{TP}{TP + FN} * 100\% \tag{15}$$

$$AP = \int_0^1 P(R)dR \tag{16}$$

$$\text{mAP} = \frac{1}{N}\sum_{i=0}^{n} AP_i \tag{17}$$

In the formula: if the sample is correctly predicted as a positive sample, it is called TP (True Positive), that is, a positive sample; the sample is correctly predicted as a negative sample, called TN (True Negative), that is, a negative sample;

the number of negative samples predicted as a positive sample is the quantity, then it is called FP (False Positive), that is, a negative sample; the number of positive samples predicted to be negative is called FN (False Negative), that is, a positive sample. Among them, negative samples represent wrong predictions, while positive samples are marked with correct predictions. Then P represents precision and R represents recall rate in equation (16).

B. EXPERIMENTAL RESULTS

1) THE INFLUENCE OF MGHOST

To verify the effectiveness of the lightweight module MGhost proposed in this paper, this paper uses Partial Convolution (PConv) and Ghost, and the improved Ghost algorithm proposed in this paper as the backbone network of YOLOv8s, and the ablation results on the test set are shown in TABLE 3.

According to the data in TABLE 3, it can be seen that the parameters of the improved MGhost are compared with the other two lightweight models; the number of parameters of MGhost is 0.17M smaller than that of PConv, and then 0.19m higher than that of Ghost; according to its calculation amount FLOPS, MGhost is 0.4 and 0.5 slower than PConv and Ghost respectively, and then PConv is the fastest per second and Ghost the slowest; in terms of accuracy, PConv is obviously much lower than Ghost, followed by little difference between MGhost and Ghost. According to the recall rate, Ghost is the weakest of the three, and MGhost is 2% and 3.8% higher than PConv and Ghost, respectively; the average accuracy of PConv is the smallest, and there is little difference between Ghost and MGhost. By comparing the improved model with the original C2f model, it can be found that all the indexes after improvement are better than those without improvement. To sum up, the improved MGhost is efficient and lightweight in the backbone network of YOLOv8s. The enhanced model pays more attention to the information of small targets, improves the recognition and location accuracy of small targets, and names the improved model MGYOLOv8s. The visualization result analysis is shown in FIGURE 19.

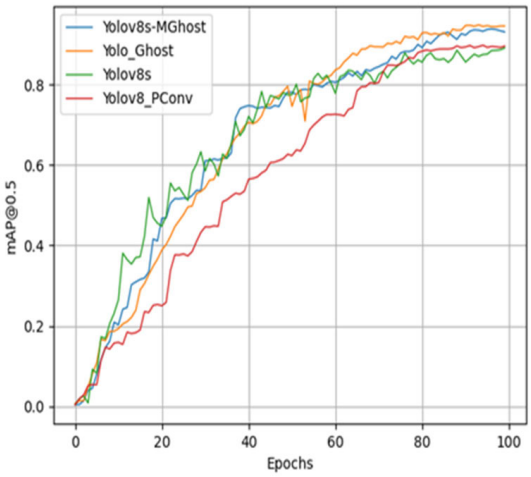


FIGURE 19. Lightweight model comparison diagram.

2) THE INFLUENCE OF CPMS

To evaluate the model more comprehensively, the researchers conducted the ablation experiment on the model in the test set.

As shown in TABLE 4 according to the number of parameters, the amount of calculation, and accuracy, the experimental results constructed MGYOLOv8-CBAM and MGYOLOv8- CPMS models, in which experiment 1 was compared with the basic model YOLOv8s and experiment 2 was compared after introducing MGhost convolution into the YOLOv8s model. The first experiment is the YOLOv8s model before improvement, and the second experiment adds the improved MGhost module and CBAM attention mechanism based on the YOLOv8s model.

The experimental results in TABLE 4 show that the primary benchmark is constructed based on the YOLOv8s model, embedding the CBAM attention mechanism module into the network. The precision P increases by 2.8%, the recall rate R rises by 3.4%, and the number of parameters and GFLOPs decreases by 3.8%, respectively. When embedding the CPMS attention mechanism in the network, the precision P increases by 5.7%, the recall rate R increases by 7.3%, the map rises by 8.9%, and the frame rate per second of the improved model is much higher than that of the original model.

In summary, the average accuracy of the improved algorithm is higher than that of other attention mechanism modules introduced and shows better index characteristics than the YOLOv8 model. The detection speed of this algorithm is much higher than that of the YOLOv8 model, and it can detect 256 frames of images per second, which can meet the real target of underwater garbage recognition. Through table analysis, the changed model is named MGC-YOLO.

To observe the change of index parameters more intuitively, this paper combines the improved network model with the comparison results of the above two models, such as FIGURE 20, mAP@0.5 in the left represents the average accuracy when the IoU value is 0.5, and mAP@0.5:0.95 in the right represents the average accuracy when the IoU value is 0.5-0.95. From the figure, it is evident that the average accuracy of the improved algorithm surpasses that of YOLOv8 significantly. When the algorithm iterates to the 80th round, the mAP@0.5 rises to about 0.861 and finally stabilizes to about 0.941, while when the YOLOv8 algorithm iterates to the 80th round, the value of mAP@0.5 is about 0.822 and finally stabilizes around 0.853. The model with the CBAM attention mechanism is also quite different from the improved CPMS model.

3) HORIZONTAL CONTRAST EXPERIMENT

Comparing the experiments to mainstream object detection algorithms benchmarks the performance of the enhanced

TABLE 3. Comparative analysis of lightweight models.

Model	Params(M)	GFLOPs	FPS(f/s)	P(%)	R (%)	mAP@0.5(%)
C2f	10.62	28.4	188.6	87.8	76.9	85.4
PConv	9.24	24.2	217.3	84.5	79.1	86.2
Ghost	8.88	23.3	101.1	91.2	77.3	90.5
MGhost	9.07	23.8	175.4	90.2	81.1	89.9

TABLE 4. Comparison of ablation experiments.

Model	MGhost	CBAM	CPMS	Params(M)	GFLOs	FPS	P (%)	R(%)	mAP@0.5(%)
One	×	×	×	10.62	28.4	188.6	87.8	76.9	85.4
Two	√	√	×	9.32	24.2	172.4	90.6	80.3	89.2
Ours	√	×	√	**9.68**	**24.8**	**256.4**	**93.5**	**84.2**	**94.3**

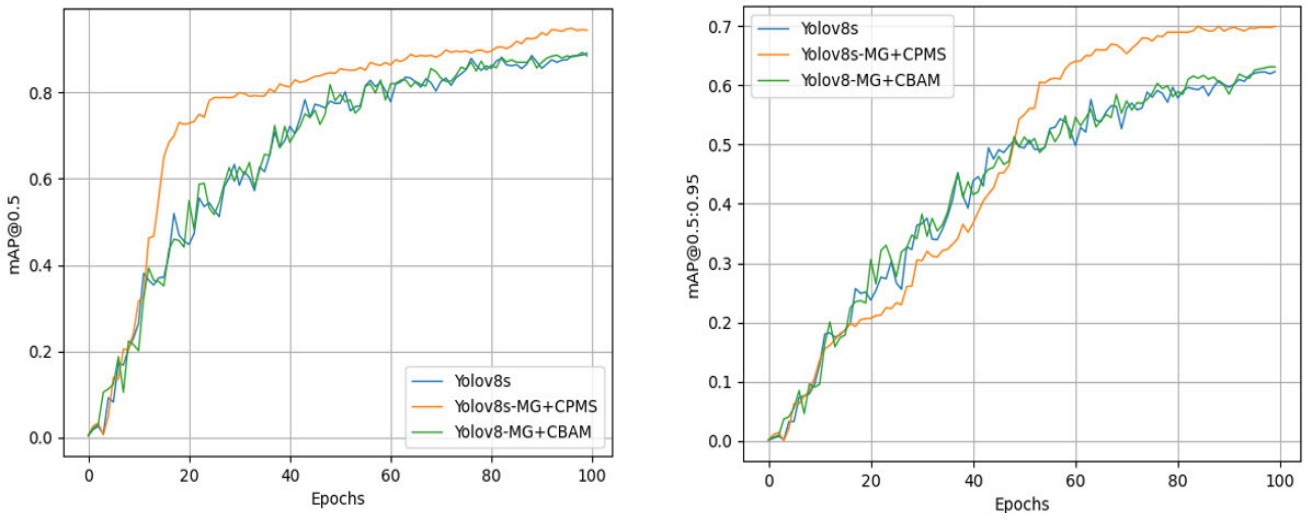


FIGURE 20. MAP curve analysis.

YOLOv8 model and the alternative model proposed in this study. These include highly accurate two-stage detectors, faster R-CNN, and single-stage detectors such as SSD and YOLOv5. We are presenting the results in TABLE 5.

The improved YOLOv8 model proposed in this paper achieves an average accuracy of 94.3%, reducing the number of model parameters to 9.68m, and its floating-point operation is 24.8, 256.4 seconds per frame. Compared with the SSD model, the number of parameters decreased by 16.6 and 37.9, and the mAP_0.5 was also much higher than that of the SSD model, while the mAP_0.5 increased by 5%. Compared with the Faster R-CNN model, the number of parameters and GFLOs of the improved model were much lower than those of the Faster R-CNN model, the frame rate of FPS was higher than that of the model, and the FPS was 4.6 higher. Compared with the YOLO series, the parameter number of v5 is also much higher than that of the improved model, but the performance of FPS and mAP_0.5 is not good. Although there is little difference in the number of parameters between v8 and v8, its average accuracy mAP_0.5 is lower than that of the improved model by 8.4%. In summary, the large GFLOP and larger model size make it less suitable for actual scenario deployment. The improved model has obvious advantages over other single-stage detectors regarding mAP_0.5 and GFLOs indexes. Display the horizontal target detection and recognition map as shown in FIGURE 21.

In short, the improved model described in this paper maintains high detection accuracy and achieves reduced model size and optimized computational complexity. They are proving that the enhanced model has better recognition results than YOLOv5, YOLOv8, Faster R-CNN, and SSD in the mainstream models. This paper studies the problems of large model size, recognition accuracy, and deployment difficulties in underwater foreign body recognition and detection. The lightweight GhostNet is integrated into the backbone network to reduce the dimension of the model. In addition, we are introducing the attention mechanism CPMS for the improved CBAM. The experimental results show that the enhanced

TABLE 5. Algorithm comparison test.

Model	Params(M)	GFLOs	FPS(f/s)	mAP@0.5(%)
SSD	26.28	62.7	80.47	89.3
YOLOv5	44.01	107.7	57.47	84.7
YOLOv8	10.62	28.4	188.6	85.4
Faster R-CNN	137.09	370.21	19.74	89.9
Ours	**9.68**	**24.8**	**256.4**	**94.3**

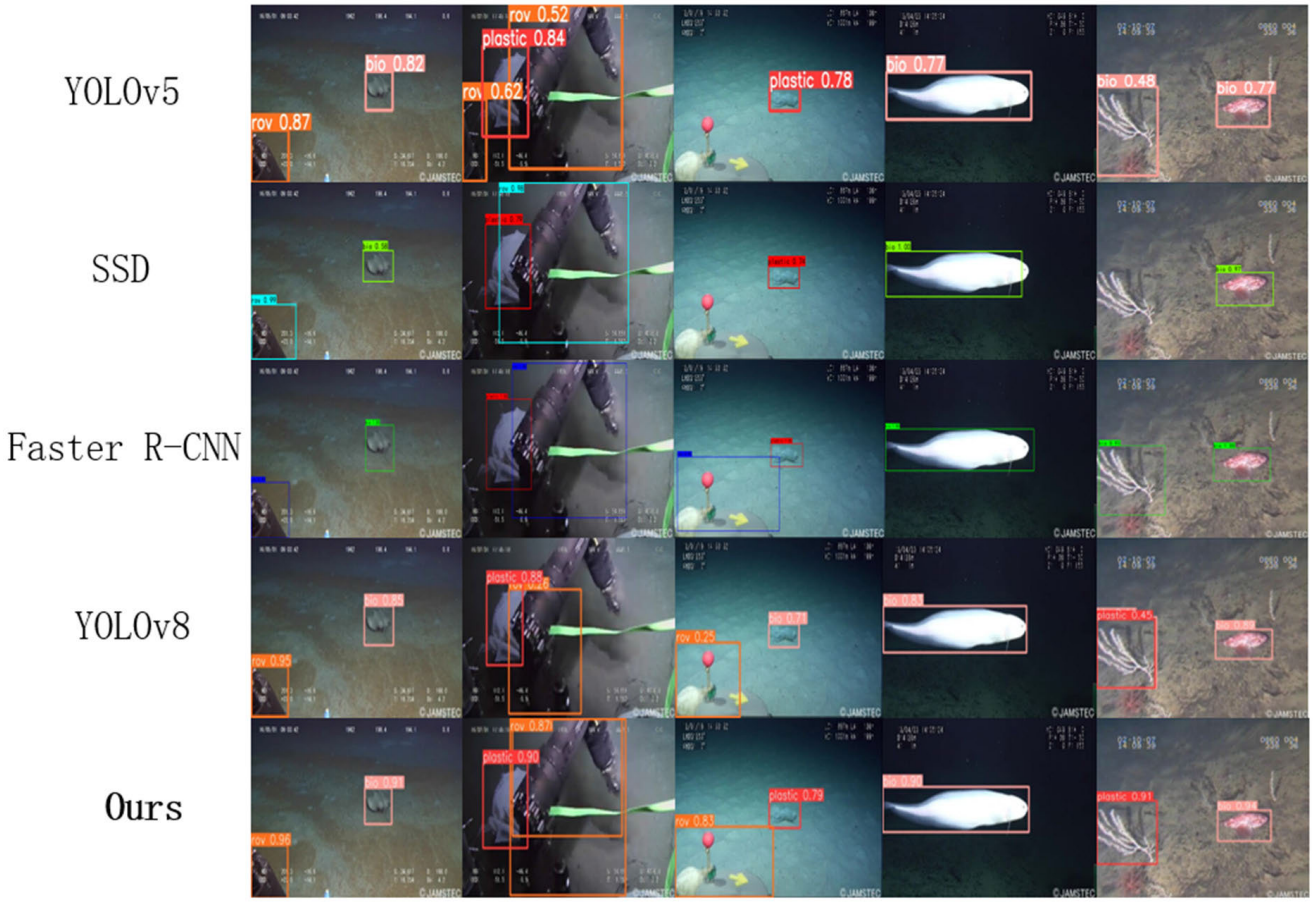


FIGURE 21. Different model inspection diagrams.

model exceeds the baseline in most evaluation indicators. It reduces the model's size and enhances its generalization ability without affecting detection accuracy.

4) TARGET DETECTION RESULTS OF DATASET

As shown in FIGURE 22, the PR curve of the model is displayed. The rate of change in accuracy increases with the increase in recall. The PR curve of the proposed model is close to the upper right corner, indicating that the proposed framework has high recall and accuracy. The area under the PR curve is relatively large, indicating the model's good performance. In addition, the PR curve is smooth, suggesting a relatively stable relationship between the PR curve and the relatively large one, indicating that the model's performance is good. In addition, the PR curve is also smooth, suggesting that there is a relatively stable relationship between recall and accuracy in the improved model.

To directly see the recall and accuracy of the model when dealing with different behavior characteristics, the experimental results tested by the test set are shown in TABLE 6.

As shown in TABLE 6, the improved algorithm performs well in the detection data set, and its accuracy can reach 99.9% in biometric identification (bio), with an average accuracy of 94.3%. However, its recall rate reaches 75.4 in biometric identification (bio), but the overall recall rate is still reasonable.Compared with the unimproved YOLOv8 model, according to the accuracy analysis, the improved plastic the accuracy analysis, the improved plastic recognition is 8.8%

TABLE 6. Comparison of ablation experiments.

Class	YOLOv8 P (%)	YOLOv8 R (%)	Ours P (%)	Ours R (%)
plastic	88.7	84.3	97.5	92.9
bio	96.5	68.2	99.9	75.4
rov	78.1	78.1	82.9	84.4

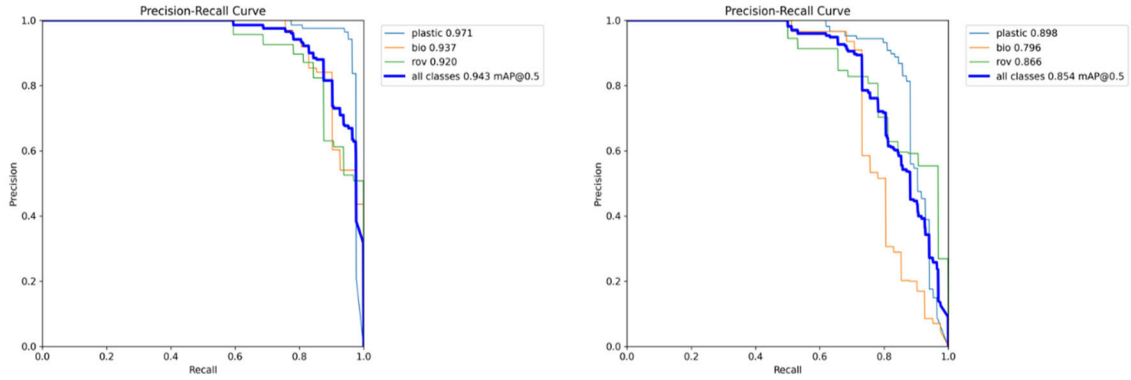


FIGURE 22. PR curve.

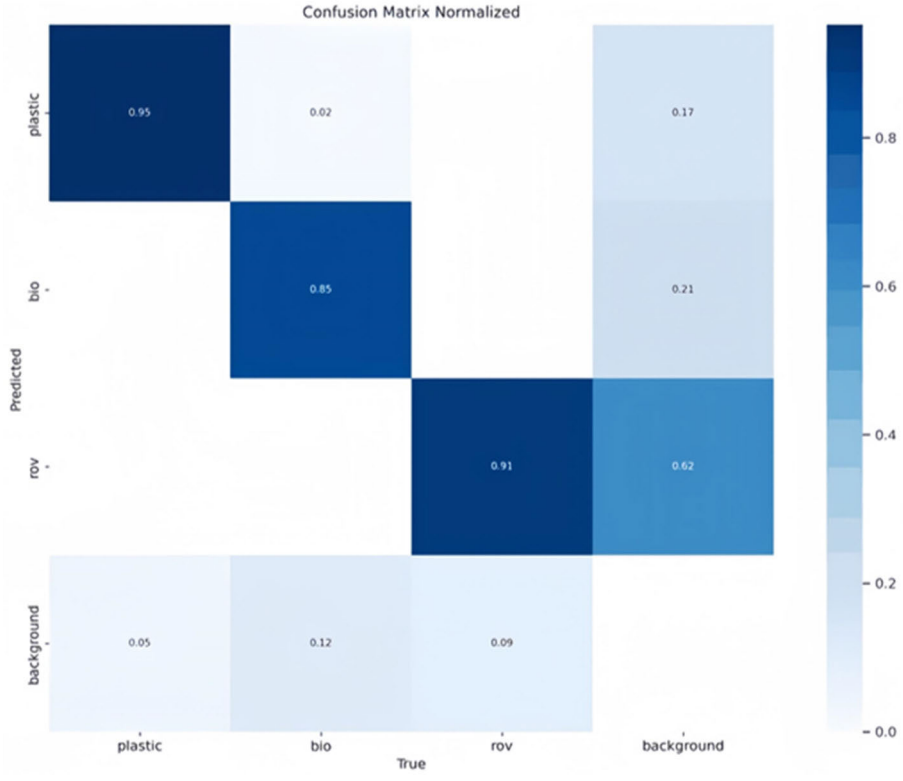


FIGURE 23. Confusion matrix.

higher than the unimproved. The biometric (bio) is 3.49% higher and 4.8% higher in other categories. The results show that the improved model can also achieve higher accuracy in marine foreign body recognition and more accurately identify object types. It displays the visualization results of two of these models in FIGURE 23 and FIGURE 24.

In addition, FIGURE 25 shows the heat map generated on the ROBAT-Date test set image using the GradCAM

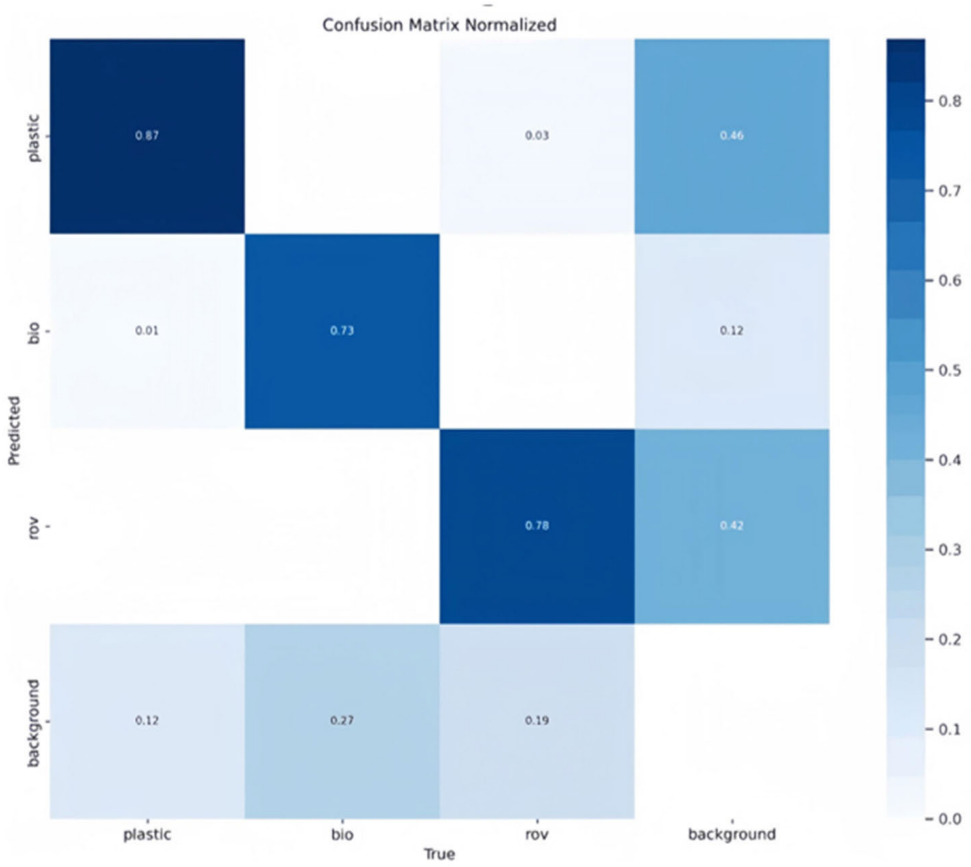


FIGURE 24. Confusion matrix.

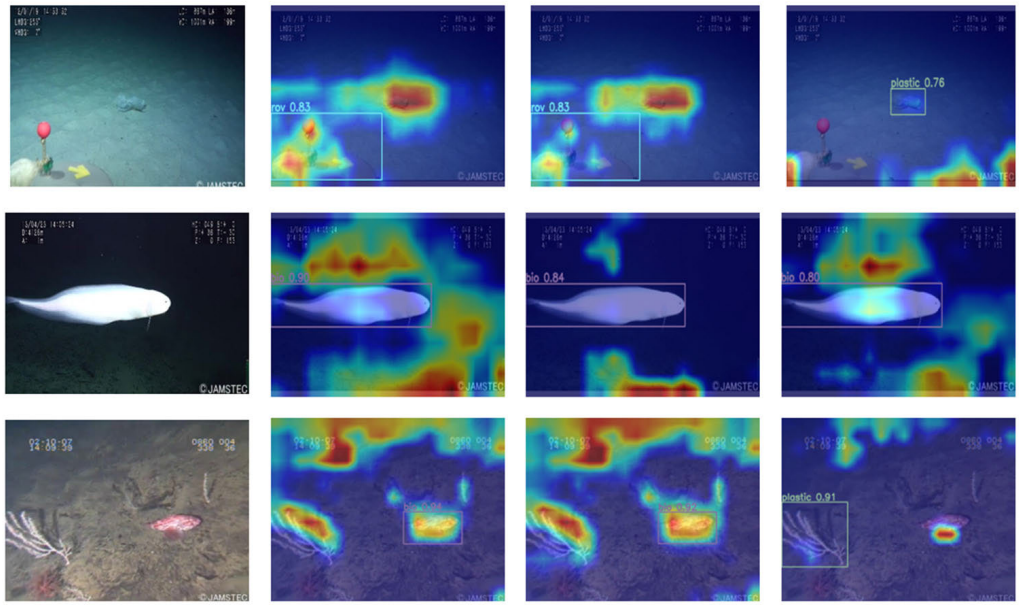


FIGURE 25. Thermal map.

visualization model highlighting vital behavioral characteristics. The prominent orange area in the heat map indicates the underwater target recognition network's successful location of the relevant image features. This visualization consolidates

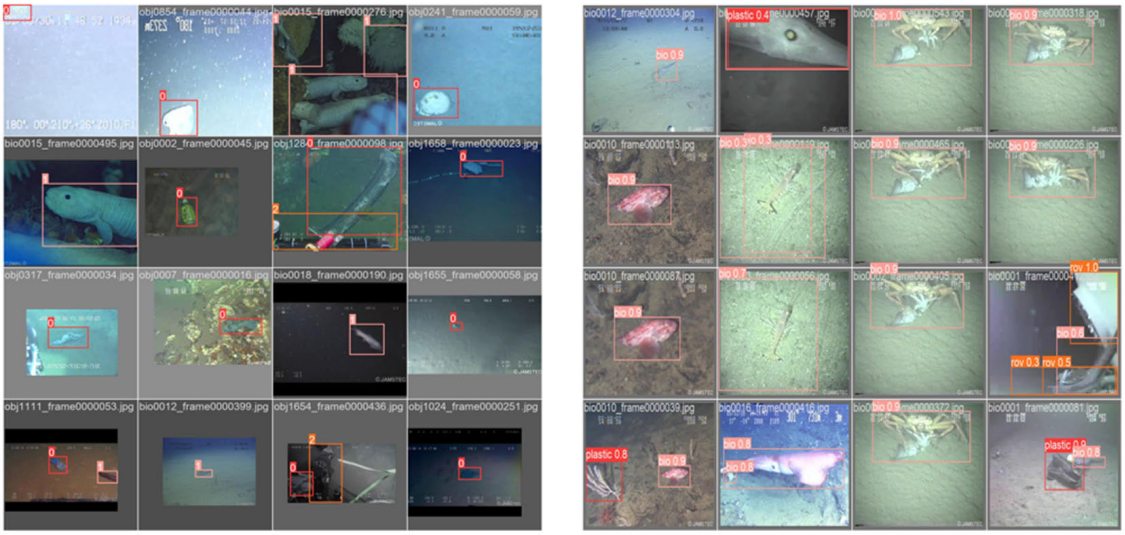


FIGURE 26. Test effect diagram.

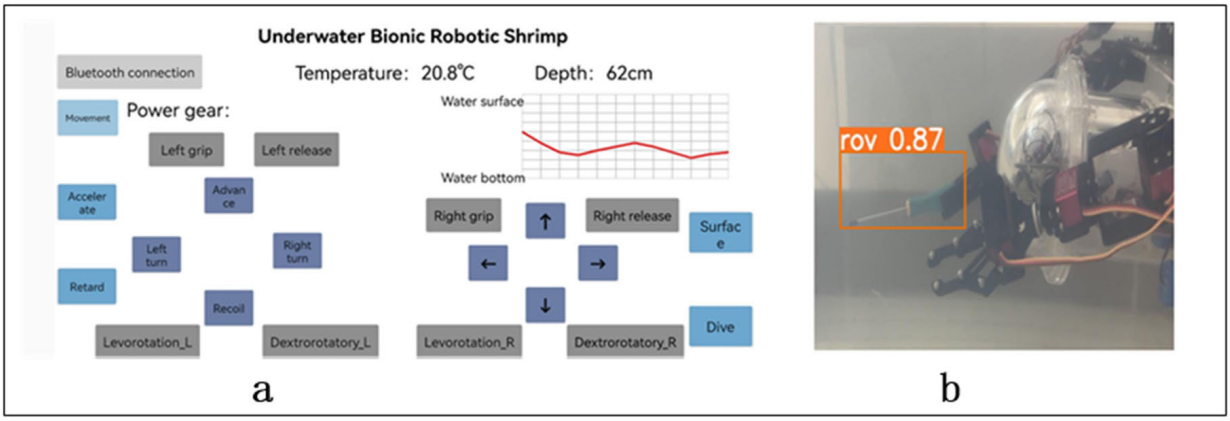


FIGURE 27. Mobile phone recognition interface diagram and UI function diagram.

the effectiveness of the proposed method in accurately identifying and highlighting the identification of underwater foreign bodies.

The above comparative analysis proves that the model effectively recognizes underwater garbage and can provide critical environmental detection tasks for the current hardware equipment. As shown in FIGURE 26, the image on the left is the trained image, and based on verification, the picture on the right is displayed.

C. UI INTERFACE ANALYSIS

The bionic shrimp accurately controls its rise and fall underwater, ensuring operations are conducted at different depths and adapting to changes in varying water depths. At the same time, it can achieve underwater hovering, hovering stably in the designated position to carry out maintenance, sampling, and other tasks. The bionic shrimp can accurately measure the underwater depth and transmit the depth data to the mobile phone APP. The mobile phone APP can display the underwater depth change through a wave graph. This data visualization method enables the operator to understand the changing trend of the underwater environment more intuitively and provide more accurate information for operation decision-making. FIGURE 27 (b) below shows the control interface of the underwater bionic shrimp. It was deployed on embedded devices to verify the improved model's actual performance. Experiments demonstrated that the recognition accuracy of the enhanced model reached 87%, surpassing that of the comparison model. This indicates that the proposed improved model is suitable for deployment on edge computing devices, such as mobile terminals. FIGURE 27(a) shows the image of the underwater foreign body recognized by the mobile phone.

V. CONCLUSION

This paper introduces MGC-YOLO, an underwater bionic robotic shrimp model designed for accurate marine foreign body identification, addressing challenges such as

large memory requirements, object density, and occlusion. To enhance efficiency, the model replaces part of the standard convolution in the original feature extraction network with improved multi-scale ghost convolutions. A lightweight architecture is also adopted to reduce network complexity and increase computational speed for faster recognition. Furthermore, the CPMS attention mechanism is integrated into the network backbone to improve feature extraction capabilities for behavioral state and location information. Finally, the model is implemented in a mobile platform to enable the bionic shrimp to clean identified underwater foreign bodies autonomously. However, experimental evaluation has encountered limitations, mainly due to the dispersion of objects in the dataset, which poses challenges in achieving optimal detection on all targets. In addition, the scarcity of datasets tailored for underwater garbage limits further improvement and performance enhancement of the model. Future efforts to enhance system proficiency will require expanding datasets and continuously improving models. The subsequent experimental evaluation confirmed the enhanced ability of our model to recognize small objects, emphasizing its proficiency and universality in different target types.

The experimental results show that, compared with the original model, the recognition accuracy, recall rate, and average recognition accuracy of the improved algorithm are 93.5%, 84.2%, and 94.3%, respectively, and the number of parameters and GFLOPs are 9.68% and 24.8 respectively. A series of comparative experimental results prove the improved algorithm's real-time performance and effectiveness; the enhanced algorithm and the up-and-down speed of the improved algorithm can reach 50mm/s after deploying mobile devices. According to the corresponding time of the garbage identified by the manipulator, the corresponding time is less than 0.2s, and the delay of underwater information return is less than 0.2s, which verifies the feasibility and effectiveness of the model. It has practical application value for underwater foreign body recognition.

REFERENCES

[1] Q. L. Zhu, L. Liu, and Y. J. Li, "Analysis on comprehensive treatment of water environment of polluted lakes in cities," *J. North China Univ. Water Resour. Electr. Power*, vol. 40, no. 5, pp. 41–47, 2019.

[2] Ministry of Ecology and Environment of the People's Republic of China. *2022 China Marine Ecological Environment Status Bulletin*. Accessed: Dec. 29, 2023. [Online]. Available: https://www.mee.gov.cn/hjzl/sthjzk/sthjtjnb/202312/W020231229339540004481.pdf

[3] J. R. Jambeck, R. Geyer, C. Wilcox, T. R. Siegler, M. Perryman, A. Andrady, R. Narayan, and K. L. Law, "Plastic waste inputs from land into the ocean," *Science*, vol. 347, no. 6223, pp. 768–771, Feb. 2015.

[4] Y. H. Zuo, K Zhang, X. Y. Kong, J. W. Pu, Y. P. Zhao, and Y. Y. Hou, "Design and research of underwater pipeline cleaning robot," *Mech. Manage. Develop.*, vol. 38, no. 2, pp. 110–112, 2023.

[5] Y. Wang, J. Wang, L. Yu, S. Kong, and J. Yu, "Toward the intelligent, safe exploration of a biomimetic underwater robot: Modeling, planning, and control," *Biomimetics*, vol. 9, no. 3, p. 126, Feb. 2024.

[6] Z. Wu, Z. Wu, X. Chen, Y. Lu, and J. Yu, "Self-supervised underwater image generation for underwater domain pre-training," *IEEE Trans. Instrum. Meas.*, vol. 73, pp. 1–14, 2024.

[7] D.-H. Lee, S. Choi, and K.-I. Na, "ASAP: Agile and safe pursuit for local planning of autonomous mobile robots," *IEEE Access*, vol. 12, pp. 99600–99613, 2024.

[8] Y. Adgo, L. Negash, D. Lebsework, and Y. Adgo, "Enhancing trajectory tracking performance of wheeled mobile robot using backstepping fuzzy sliding mode control," *Eng. Res. Expressse*, Feb. 2024.

[9] M. M. Madebo, C. M. Abdissa, L. N. Lemma, and D. S. Negash, "Robust tracking control for quadrotor UAV with external disturbances and uncertainties using neural network based MRAC," *IEEE Access*, vol. 12, pp. 36183–36201, 2024.

[10] Y. Xu, R. Zheng, S. Zhang, and M. Liu, "Robust inertial-aided underwater localization based on imaging sonar keyframes," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–12, 2022.

[11] S. Ding, S. Qu, Y. Xi, and S. Wan, "Stimulus-driven and concept-driven analysis for image caption generation," *Neurocomputing*, vol. 398, pp. 520–530, Jul. 2020.

[12] S. Ding, H. Wang, H. Lu, M. Nappi, and S. Wan, "Two path gland segmentation algorithm of colon pathological image based on local semantic guidance," *IEEE J. Biomed. Health Informat.*, vol. 27, no. 4, pp. 1701–1708, Apr. 2023.

[13] J. Yang, L. Xin, H. Huang, and Q. He, "An improved algorithm for the detection of fastening targets based on machine vision," *Comput. Model. Eng. Sci.*, vol. 128, no. 2, pp. 779–802, 2021.

[14] X. Liang, R. Li, and H. Yu, "Improved underwater target detection algorithm for YOLOv7," *Comput. Eng. Appl.*, vol. 78, no. 2, pp. 1–4, 2024. [Online]. Available: http://www.cmp.net.cn/

[15] Z. Zou, K. Chen, Z. Shi, Y. Guo, and J. Ye, "Object detection in 20 years: A survey," *Proc. IEEE*, vol. 111, no. 3, pp. 257–276, Mar. 2023.

[16] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," 2013, *arXiv:1311.2524*.

[17] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1904–1916, Sep. 2015.

[18] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1440–1448.

[19] L. Chen, Z. Liu, L. Tong, Z. Jiang, S. Wang, J. Dong, and H. Zhou, "Underwater object detection using invert multi-class AdaBoost with deep learning," in *Proc. IJCNN*, Jul. 2020, pp. 1–8.

[20] M. Fulton, J. Hong, M. J. Islam, and J. Sattar, "Robotic detection of marine litter using deep visual detection models," in *Proc. Int. Conf. Robot. Autom. (ICRA)*, Montreal, QC, Canada, 2019, pp. 5752–5758.

[21] L. Zeng, B. Sun, and D. Zhu, "Underwater target detection based on faster R-CNN and adversarial occlusion network," *Eng. Appl. Artif. Intell.*, vol. 100, Apr. 2021, Art. no. 104190.

[22] S. Song, J. Zhu, X. Li, and Q. Huang, "Integrate MSRCR and mask R-CNN to recognize underwater creatures on small sample datasets," *IEEE Access*, vol. 8, pp. 172848–172858, 2020.

[23] J. Gao, Y. Zhang, X. Geng, H. Tang, and U. A. Bhatti, "PE-transformer: Path enhanced transformer for improving underwater object detection," *Expert Syst. Appl.*, vol. 246, Jul. 2024, Art. no. 123253.

[24] R. Dinakaran, L. Zhang, C.-T. Li, A. Bouridane, and R. Jiang, "Robust and fair undersea target detection with automated underwater vehicles for biodiversity data collection," *Remote Sens.*, vol. 14, no. 15, p. 3680, Aug. 2022.

[25] F. Lei, F. Tang, and S. Li, "Underwater target detection algorithm based on improved YOLOv5," *J. Mar. Sci. Eng.*, vol. 10, no. 3, p. 310, Feb. 2022.

[26] M. Tian, X. Li, S. Kong, L. Wu, and J. Yu, "A modified YOLOv4 detection method for a vision-based underwater garbage cleaning robot," *Frontiers Inf. Technol. Electron. Eng.*, vol. 23, no. 8, pp. 1217–1228, Aug. 2022.

[27] G. Wen, S. Li, F. Liu, X. Luo, M.-J. Er, M. Mahmud, and T. Wu, "YOLOv5s-CA: A modified YOLOv5s network with coordinate attention for underwater target detection," *Sensors*, vol. 23, no. 7, p. 3367, Mar. 2023.

[28] J. Zhu, T. Hu, L. Zheng, N. Zhou, H. Ge, and Z. Hong, "YOLOv8-C2f-Faster-EMA: An improved underwater trash detection model based on YOLOv8," *Sensors*, vol. 24, no. 8, p. 2483, Apr. 2024.

[29] S. Yan, Z. Wu, J. Wang, Y. Huang, M. Tan, and J. Yu, "Real-world learning control for autonomous exploration of a biomimetic robotic shark," *IEEE Trans. Ind. Electron.*, vol. 70, no. 4, pp. 3966–3974, Apr. 2023.

[30] D.-A. Pham and S.-H. Han, "Deploying a computer vision model based on YOLOv8 suitable for drones in the tuna fishing and aquaculture industry," *J. Mar. Sci. Eng.*, vol. 12, no. 5, p. 828, May 2024.

[31] D. D. G and S. Bo, "Traffic safety helmet wear detection based on improved YOLOv5 network," *Optoelectron. Lett.*, pp. 1–11, May 2024.

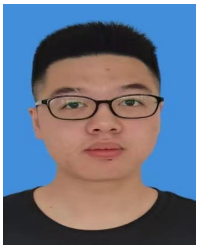
[32] F. G. Tan, Q. M. Liao, and C. Zhai, "Research on YOLOv5 lightweight object detection for urban rail transit vehicle vision," *J. Guangdong Vocational Tech. College Transp.*, vol. 23, no. 1, pp. 45–48, 2024.
[33] N. Silberman, D. Hoiem, P. Kohli, and R. Fergus, "Indoor segmentation and support inference from RGBD images," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2012, pp. 746–760.
[34] H. Wang, D. Zhang, J. Feng, L. Cascone, M. Nappi, and S. Wan, "A multi-objective segmentation method for chest X-rays based on collaborative learning from multiple partially annotated datasets," *Inf. Fusion*, vol. 102, Feb. 2024, Art. no. 102016.

YING TANG received the M.S. degree in electronic and communication engineering from the University of Electronic Science and Technology of China, China. She is currently an Associate Professor with Chengdu University of Technology. Her research interests include imagine processing, pattern recognition, and embedded technology.

HONGLI YIN was born in Sichuan, China. She received the bachelor's degree in 2022. She is currently pursuing the master's degree in electronic information with Chengdu University of Technology. Her research interests include object detection, pose recognition, and embedded systems.

CHANGHAO QUAN was born in Sichuan, China. He received the bachelor's degree in 2023. He is currently pursuing the master's degree in embedded systems with Chengdu University of Technology. His research interests include object detection, deep learning, and drug recognition.

QIAN MU received the bachelor's degree from Chengdu University of Technology, in 2024. He is currently a RF Hardware Engineer. His research interests include DC-DC converter and robotics.



XIAOYUN ZHOU received the bachelor's degree from Chengdu University of Technology, in 2024. He is currently an Automation Engineer. His research interests include robotics and control algorithms.