

Article

Real-Time Underwater Fish Detection and Recognition Based on CBAM-YOLO Network with Lightweight Design

Zheping Yan ^{1,2}, Lichao Hao ^{1,*}, Jianmin Yang ¹ and Jiajia Zhou ¹ ¹ College of Intelligent Systems Science and Engineering, Harbin Engineering University, Harbin 150001, China² Qingdao Innovation and Development Base, Harbin Engineering University, Qingdao 266404, China

* Correspondence: haolichao-hrbeu@outlook.com

Abstract: More and more underwater robots are deployed to investigate marine biodiversity autonomously, and tools are needed by underwater robots to discover and acknowledge marine life. This paper has proposed a convolutional neural network-based method for intelligent fish detection and recognition with a dataset used for training and testing generated and augmented from an open-source Fish Database regarding 6 different types. Firstly, to improve image quality, a hybrid image enhancement algorithm is used to preprocess underwater images with a weighted fusion strategy of multiple traditional methodologies and comparisons have been made to prove the effectiveness according to various indexes. Secondly, to increase detection and recognition accuracy, different attention modules are integrated into the YOLOv5m network structure and the convolutional block attention module (CBAM) has outperformed other modules in recall rate and mAP while maintaining the capability of real-time processing. Lastly, to meet real-time requirements, lightweight adjustments have been made to CBAM-YOLOv5m with the GSConv module and C3Ghost module and a nearly 25% reduction in network parameters and a 20% reduction in computational consumption are obtained. Besides, the lightweight network has realized better accuracy than YOLOv5m. In conclusion, the method proposed in this paper is effective in real-time fish detection and recognition with practical application prospects.

Keywords: fish detection and recognition; CBAM-YOLO; lightweight design; underwater image; image enhancement



Citation: Yan, Z.; Hao, L.; Yang, J.; Zhou, J. Real-Time Underwater Fish Detection and Recognition Based on CBAM-YOLO Network with Lightweight Design. *J. Mar. Sci. Eng.* **2024**, *12*, 1302. <https://doi.org/10.3390/jmse12081302>

Academic Editor: Rafael Morales

Received: 5 June 2024

Revised: 19 July 2024

Accepted: 31 July 2024

Published: 1 August 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Due to accelerating changes in the global environment, protections for various kinds of fish have received more and more attention, especially protections for deep sea fish. In order to detect and recognize fish, underwater robots, which are equipped with a single camera or several cameras, are commonly deployed to obtain images of the underwater environment. As one key research content of underwater vision, underwater image processing has lots of differences compared with air image processing because of its color cast effect, low contrast, noise and so on [1], and all these factors make most of the proven image processing methods for air images unsuitable for underwater image processing, so studies on underwater image processing are necessary and crucial.

Aqueous media's strong absorption and scattering of light make underwater images deteriorate badly in imaging distance and quality, contrast also drops dramatically, and blurriness becomes higher, and as a result, it is hard to extract useful image information [2]. Image enhancement methods are adopted to clarify underwater images. After years of research and development, plenty of enhancement methods are more targeted and have better effects. As it is known that underwater images always have low contrast and uneven colors, to solve these problems, Iqbal et al. [3] designed an Unsupervised Color Correction Method (UCM). Firstly, color equalization was adopted to improve color projection. Secondly, RGB format images' R channels were strengthened while B channels were weakened.

Lastly, strengthened every channel's contrast of HSI format images. The advantages and processing effects of UCM were proven through comparison with extant methods like Gray World(GW), Histogram Equalization(HE) and so on. Ancuti et al. [4] described one underwater image-enhancing strategy based on the fusion principles which derived the inputs and the weight measures only from the degraded version of the image. This strategy effectively improved underwater images' and videos' quality and enhanced the dark areas' brightness, as a result, the whole contrast of images and videos could be upgraded and edges could be seen clearly. A hybrid CLAHE algorithm was proposed by Hitam et al. [5], this algorithm took the contrast-limited adaptive histogram equalization(CLAHE) as the enhancement method and the Euclidean norm as the fusion method to process HSV and RGB images. Experimental results showed that this algorithm could enhance image quality significantly, improve contrast, and reduce noise and artifacts. To solve the spectral degradation effects, Ghani and Isa [6] designed one method that stretched and integrated histograms of different components in RGB and HSV spaces. This method had achieved satisfactory results in both blue-green effects reduction and image enhancement, its advantage was proved through the results of processing 300 images. Based on principles of underwater imaging, Fengyun Cao et al. [7] proposed an adaptive weighted enhancement algorithm, this algorithm took advantage of white balance and contrast enhancement to acquire double inputs images firstly, then combined illuminance with contrast to obtain three-weight images, an adaptive weighted fusion of all weighted images was conducted to obtain output images finally. This algorithm's application value was shown through its improvement of underwater images' quality while ensuring real-time property. Huang et al. [8] adopted an adaptive relative global histogram stretching method, which consisted of contrast correction and color correction, to deal with low-quality shallow water images. High quality, abundant information, normal color and low noise shallow images were obtained by using this method. To solve problems that existed in typical low-quality underwater images, Jia et al. [9] proposed an enhancement method that was based on the HSI model and adopted improved Retinex. It has been proved that this method could improve underwater images' contrast and highlight images' details both in qualitative and quantitative analysis. A weighted fusion-based multi-space transformation method for underwater image enhancement was proposed by Du et al. [10], Gamma correction, CLAHE, SSR(Single-Scale Retinex), and some other traditional algorithms were adopted to implement enhancement on corresponding channels of different color models, and then weighted the sum of all results. Through comparison with the three existing methods both in subjective visual and objective evaluation indicators, this method's advantage in improving color deviation and enhancing image quality was proved. Dong et al. [11] designed a processing method for every channel in underwater images' RGB and LAB color spaces, the specific coefficient based on the contrast between normal and abnormal channels in RGB color space was calculated to improve abnormal channels; histograms adopted fusion local enhancement and the exposure cut-off strategy was employed for L heft in LAB color space to enhance contrast. The beneficial equilibrium method was adopted to weigh the difference between A heft and B heft. This method acquired higher-quality underwater images.

Underwater target image enhancement is the basis for underwater target detection and recognition. The application of air target detection and recognition technology has been very mature, while the research on underwater target detection and recognition technology started relatively late. However, as countries gradually pay more attention to the development of the ocean, underwater target recognition technology has also made great progress. Wang Shilong et al. [12] designed a new underwater target recognition system, which perfectly integrated the boundary moment theory and the improved FCM clustering method. The recognition experiments of four types of underwater objects proved the system's effectiveness, stability, reliability and real-time properties. Shi Min et al. [13] used the wavelet-transformed underwater target noise signal as the input of a probabilistic neural network to classify targets. The probabilistic neural network did not require training and was very suitable for underwater target recognition. Shi Yang et al. [14]

proposed a recognition method suitable for forward-looking sonar images. The method included a series of steps such as image enhancement, image segmentation, and feature classification, and used the particle swarm algorithm to optimize the parameters of the least squares vector machine. After 110 rounds of training, the recognition accuracy of the test set stabilized. Zhu et al. [15] proposed a real-time underwater target recognition method based on a prior transformable template matching method. This method extracted features from the sonar video sequence to analyze and construct a template, then used fast saliency detection technology to find the target area in the image, and finally obtained the normalized gradient feature by performing an affine transformation on the template and calculating the template and target area's similarity between them. An underwater acoustic target recognition method based on multi-dimensional fusion features was put forward by Wang et al. [16]. This method used a Gaussian mixture model to optimize the structure of a deep neural network. This method could reduce redundant features and improve recognition accuracy. Experiments were conducted on a dataset with underwater background noise and weak targets, and the method achieved an accuracy of 94.3% after 800 iterations. Li Yu et al. [17] designed an autonomous underwater target recognition system. This system was based on the LeNet-5 network model, used a 13-layer network structure, and was trained in 100 iterations. The final recognition rate reached 99.18%. Chen et al. [18] proposed an underwater target recognition model based on the improved YOLOv4 network. The model used deconvolution to replace upsampling and fused depth-separable convolution, and the training set was preprocessed through the improved mosaic enhancement method. According to subjective and objective evaluations, the improved network model could effectively improve the accuracy of underwater target recognition while ensuring real-time performance and was no longer overly dependent on hardware performance. Wang et al. [19] proposed an underwater target recognition method based on a convolutional neural network that optimized the two processes of feature extraction and target classification, and completed related underwater simulation experiments. Experimental results showed that the underwater target recognition accuracy was improved after optimization, which fully proved the effectiveness and innovation of the optimization method.

The method proposed in this paper for real-time fish detection and recognition contains two main parts, underwater image enhancement and fish detection and recognition. The following contents are organized as follows. In Section 2 the image enhancement method is stated, and a weighted fusion algorithm is obtained by analyzing and improving several classic algorithms. The fish detection and recognition method is illustrated in Section 3, the model is established by adding attention modules and lightweight improvements to the YOLOv5m model. Finally, Section 4 introduces and analyzes the results of underwater simulation experiments.

2. Underwater Image Enhancement

The main problems encountered in underwater image processing include insufficient detail, low contrast, and color distortion, etc. Lots of image enhancement algorithms have achieved great effects in dealing with the problems described above. The Gray World algorithm [20] aims to weaken or completely remove the color deviation caused by water absorption in underwater images. It is based on the assumption that the average pixel values of the underwater image's R, G, and B color channels are all approximately the same value. Underwater images processed by the traditional Gray World algorithm are greatly improved in terms of the blue-green cast but will produce a certain degree of red artifacts. The CLAHE algorithm [21] can not only make up for the block effect of the AHE algorithm but also has a small amount of calculation. Details of underwater images processed in HSV space using the CLAHE algorithm are more prominent, and at the same time, noise is less. Gamma Correction [22] could improve images' contrast by performing nonlinear operations on their gamma curves. In order to obtain information from images more easily, variable parameters are applied to dynamically adjust images' brightness. However, when

processing images in batches, different images need different parameters, and it is obviously unrealistic to change parameters every time an image is processed. Retinex [23] theory aims to minimize the impact of light on the object's irradiation component on the appearance of the image and retain the essential attributes of the object to the greatest extent.

The underwater image enhancement algorithm proposed in this paper is based on traditional classic algorithms and improvements are implemented to meet fish detection and recognition needs.

An improved Gray World Algorithm is adopted to acquire better color correction results by compensating the red channel [24], that is, first compensate the R channel of the image and then calculate the gain of each channel according to the traditional algorithm.

$$R_c(x, y) = R(x, y) + \alpha \cdot (G_{av} - R_{av}) \cdot [1 - R(x, y)] \cdot G(x, y) \quad (1)$$

$R_c(x, y)$ represents the pixel corresponding to the compensated R channel, $R(x, y)$ and $G(x, y)$ represent the pixel values corresponding to the R component and G component of the original image, respectively, α is a variadic parameter, which should be dynamically adjusted based on the different degree of color deviation of underwater images.

Adaptive Gamma transformation is adopted to solve parameter adjustment issues while processing images in batches. This method is implemented through adaptive parameters, that is, each image determines its own Gamma transformation parameters through its own brightness. First of all, the underwater image is converted to LAB space. After channel separation, the L channel is separately Gamma transformed, and then the channels are merged and converted back to RGB space. This method significantly improves the contrast and clarity of the image while ensuring that the image color is not distorted.

In order to enhance the details of the underwater image and ensure the information content, this article first decomposes the underwater image into a lighting layer and a detail layer based on the Retinex theory and then enhances the detail layer based on the Sigmoid function. In order to ensure that the color of the color image is not affected by this algorithm, the original image needs to be converted to HSV space, and only the V component is decomposed and enhanced.

Image fusion can perfectly integrate the characteristics of multiple images into the output image, and make the output image more informative and detailed. Proper image fusion can increase the stability and reliability of the image enhancement system. In view of the characteristics of underwater images and based on the above analysis and improvement of these classic enhancement methods, this paper proposes an underwater image enhancement algorithm based on weighted fusion. The workflow is shown in Figure 1, and the contents in the dotted boxes are the algorithm proposed or improved by this paper.

The steps of the underwater image enhancement algorithm in this paper are as follows:

- (1) The Gray World algorithm that compensates for the red channel is used to process the original underwater image, which solves the problem that underwater images usually appear blue-green due to water absorption, making it close to the real color.
- (2) Convert the color-corrected image from RGB space to HSV space and LAB space, respectively, and separate the three channels of the HSV image and LAB image.
- (3) The CLAHE algorithm is used to process the V component to enhance the contrast of the underwater image. The color enhancement algorithm is used to process the S component to enhance the color saturation of the underwater image. The original H component is combined with the processed S and V components and converted to RGB space to obtain the first output image I_1 . The expression of the color enhancement algorithm is as follows:

$$S_o = \begin{cases} \alpha \ln(S_i + 1), & 0 \leq S_i \leq 0.2 \\ e^{S_i} - 1, & 0.2 < S_i \leq 0.7 \\ \beta(e^{S_i} - 1), & 0.7 < S_i \leq 1.0 \end{cases} \quad (2)$$

S_o is the output image of the color enhancement algorithm, S_i is the input image of the color enhancement algorithm, α and β are two variable parameters to change the degree of color enhancement and their best value could be determined experimentally.

- (4) The brightness of the image after color correction is often too high and the image brightness needs to be reduced. Therefore, the L component is processed through Adaptive Gamma Correction, then merged with the original A and B components and the LAB image to RGB space to obtain the second output image I_2 .
- (5) The V component is processed using the detail enhancement algorithm based on the Sigmoid function to enrich the details of the image. The original H and S components are merged with the enhanced V component and then convert the HSV image to RGB space to obtain the third output image I_3 .
- (6) The three images I_1 , I_2 , and I_3 are weighted and fused to achieve complementary advantages and obtain underwater images of higher quality and more in line with research needs. The expression of the weighted fusion output image I_o is as follows:

$$I_o = pI_1 + qI_2 + rI_3 \quad (3)$$

Weighting coefficients p , q , and r need to be determined experimentally.

- (7) The multi-scale detail enhancement method based on Gaussian difference [25] is used to improve the clarity and solve the blur problem caused by weighted fusion. The expression of the multi-scale detail enhancement method based on Gaussian difference is as follows:

$$B_1 = G_1 \otimes I_o \quad (4)$$

$$B_2 = G_2 \otimes I_o \quad (5)$$

$$B_3 = G_3 \otimes I_o \quad (6)$$

$$D_1 = I_o - B_1 \quad (7)$$

$$D_2 = B_1 - B_2 \quad (8)$$

$$D_3 = B_2 - B_3 \quad (9)$$

$$D_O = (1 - \omega_1 \times \text{sgn}(D_1)) \times D_1 + \omega_2 \times D_2 + \omega_3 \times D_3 \quad (10)$$

G_1 , G_2 , and G_3 represent Gaussian kernels with standard deviations of 1.0, 2.0, and 4.0, respectively. ω_1 , ω_2 and ω_3 value 0.5, 0.5, 0.25, respectively. This method first performs three-scale Gaussian filtering on the original image and then obtains different levels of detail information by subtracting from the original image. Finally, the obtained detailed information is integrated into the image through a specific algorithm to improve the clarity of the image.

Comparative experiments were conducted to prove the effectiveness of the image enhancement algorithm proposed in this paper, and experimental results are shown in Figure 2 and Table 1.

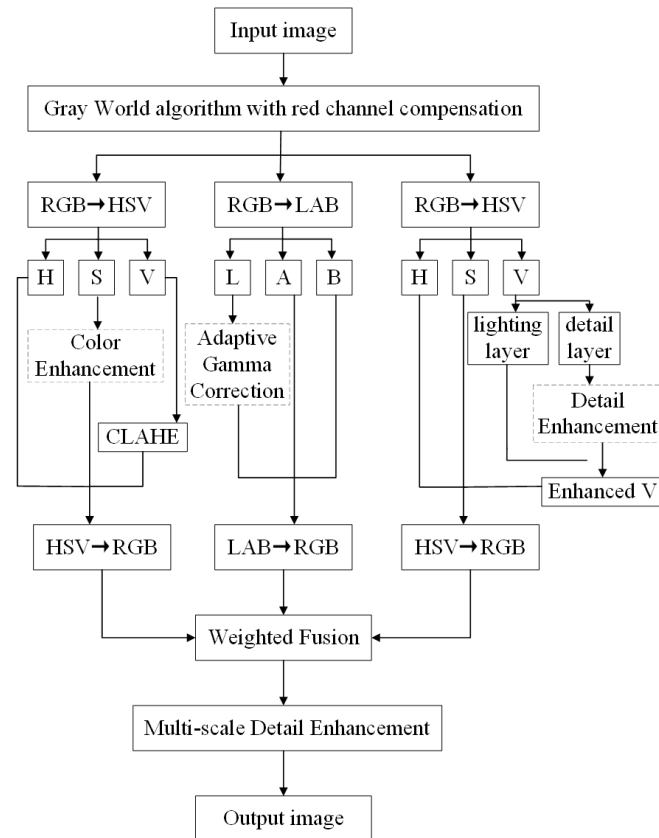


Figure 1. Workflow of the Proposed Image Enhancement Algorithm. The enhancement procedure consists of gray world compensation, color enhancement, gamma correction, and detail enhancement. Outputs are weighted fusion.

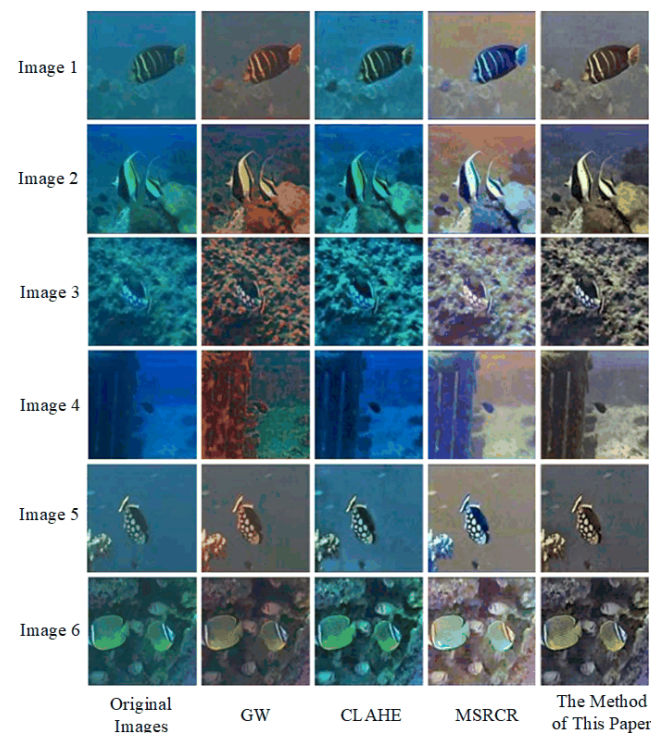


Figure 2. Comparative Experiment Results of Different Image Enhancement Methods. 6 typical images are used to present the image enhancement results of the GW algorithm, the CLAHE algorithm, the MSRCR algorithm, and the method proposed in this paper.

Table 1. Quantitative Measures of Different Image Enhancement Methods.

Image	Indicator	OI	GW	CLAHE	MSRCR	Method of This Paper
Image 1	Information Entropy	4.816	4.906 (+1.9%)	5.613 (+16.5%)	5.817 (+20.8%)	6.047 (+25.6%)
	PSNR	—	15.134	28.101	11.891	14.449
	UIQM	0.417	0.438 (+5.0%)	0.427 (+2.4%)	0.560 (+34.3%)	0.541 (+29.7%)
	UCIQE	0.344	0.382 (+11.0%)	0.382 (+11.0%)	0.411 (+19.5%)	0.433 (+25.9%)
Image 2	Information Entropy	5.970	6.511 (+9.1%)	6.628 (+11.0%)	7.014 (+17.5%)	7.278 (+21.9%)
	PSNR	—	13.147	25.969	10.258	13.493
	UIQM	0.566	0.834 (+47.3%)	0.689 (+21.7%)	0.685 (+21.0%)	0.698 (+23.3%)
	UCIQE	0.476	0.491 (+3.2%)	0.488 (+2.5%)	0.481 (+1.1%)	0.527 (+10.7%)
Image 3	Information Entropy	6.529	6.871 (+5.2%)	7.059 (+8.1%)	7.370 (+12.9%)	7.442 (+14.0%)
	PSNR	—	13.169	23.297	10.837	15.136
	UIQM	0.971	1.326 (+36.6%)	1.199 (+23.5%)	0.991 (+2.1%)	1.268 (+30.6%)
	UCIQE	0.442	0.512 (+15.8%)	0.468 (+5.9%)	0.446 (+0.9%)	0.499 (+12.9%)
Image 4	Information Entropy	5.760	5.837 (+1.3%)	6.066 (+5.3%)	7.056 (+22.5%)	7.163 (+24.4%)
	PSNR	—	12.646	28.959	12.389	13.345
	UIQM	0.564	0.701 (+24.3%)	0.602 (+6.7%)	0.658 (+16.7%)	0.837 (+48.4%)
	UCIQE	0.449	0.505 (+12.5%)	0.469 (+4.5%)	0.484 (+7.8%)	0.523 (+16.5%)
Image 5	Information Entropy	4.467	4.618 (+3.4%)	5.491 (+22.9%)	5.499 (+23.1%)	5.469 (+22.4%)
	PSNR	—	17.386	27.381	13.117	17.032
	UIQM	0.024	0.416 (+1633.3%)	0.112 (+366.7%)	0.537 (+2137.5%)	0.629 (+2520.8%)
	UCIQE	0.382	0.375 (−1.8%)	0.399 (+4.5%)	0.432 (+13.1%)	0.447 (+17.0%)
Image 6	Information Entropy	6.337	6.348 (+0.2%)	7.010 (+10.6%)	7.347 (+15.9%)	7.451 (+17.8%)
	PSNR	—	19.768	23.643	11.636	18.118
	UIQM	0.289	0.637 (+120.4%)	0.524 (+81.3%)	0.749 (+159.2%)	0.808 (+179.6%)
	UCIQE	0.397	0.394 (−0.8%)	0.419 (+5.5%)	0.470 (+18.4%)	0.486 (+22.4%)

In Figure 2, images of the first column are original underwater images, images of the second column are underwater images processed by the GW algorithm, images of the third column are underwater images processed by the CLAHE algorithm, images of the fourth column are underwater images processed by MSRCR algorithm, and images of the fifth column are underwater images processed by the image enhancement method proposed in this paper. It can be seen that the clarity and contrast of the underwater images processed by the GW algorithm are slightly improved; however, while correcting the blue-green bias of underwater images, it causes severe red artifacts or even significant color distortion in some images. The CLAHE algorithm does not correct the color of underwater images but significantly improves the contrast. The MSRCR algorithm can noticeably enhance the details, clarity, and contrast of underwater images and correct colors to some extent, but it may lead to color distortion. The underwater image enhancement method proposed in this paper can not only effectively improve the quality of underwater images in terms of details, clarity, and contrast but also decrease color distortion and make underwater images closer to real scenarios.

OI is short for Original Images in Table 1. As can be seen from Table 1, information entropy values of underwater images processed by the GW algorithm are all smaller than other algorithms but part of UIQM and UCIQE values are larger due to color distortion. Overall, the GW algorithm has less improvement in the quality of underwater images. Information entropy values of underwater images processed by the CLAHE algorithm have increased significantly, UIQM and UCIQE values have also increased to a certain extent, and PSNR values are larger since they have no color correction effect, so the CLAHE algorithm can effectively improve most underwater images' contrast and clarity. Information entropy, UIQM and UCIQE values of underwater images processed by the MSRCR algorithm are significantly increased, but PSNR values are generally small, which means this algorithm can effectively improve the overall quality of underwater images but may cause distortion problems. Information entropy, UIQM and UCIQE values of underwater images processed by the method of this paper are almost the largest or close to the maximum compared with the other three algorithms. PSNR values are also within an acceptable range. Hence, the method of this article is relatively comprehensive and can effectively enrich images' information, and increase contrast and clarity while ensuring that processed underwater images' colors are close to real scenes.

3. Fish Detection and Recognition

With enhanced underwater images obtained from the above section, fish detection and recognition methods can be presented in the following.

3.1. Underwater Target Detection and Recognition Models

Due to the complex underwater environment, and various types and shapes of fishes, traditional underwater target detection and recognition methods have disadvantages of low accuracy and poor versatility. With the rapid rise of deep learning in various fields, it has also been widely used in underwater target detection and recognition. Compared with traditional methods, deep learning methods have higher accuracy and robustness. Among all kinds of deep-learning-based models, a highly representative series of YOLO (You Only Look Once) algorithms [26–30] have shown advantages in target detection and classification performed together. The overall performance of YOLOv5m has been proved in many fields in both detection speed and accuracy, making it more suitable for application scenarios with high real-time performance and low computation cost.

In addition, since it is difficult to photograph underwater images, especially different fish pictures, only a small sample data set of underwater images is available. Due to the wide variety of fish, if a shallow convolutional neural network is only trained on a small sample data set, there will be a problem of incomplete learning target features, that is, under-fitting; while training a deep convolutional neural network on a small sample data set, an over-fitting problem will occur. In order to avoid these problems and improve the

detection and recognition effect of underwater targets, this article will train an underwater target detection and recognition network model based on the pre-trained backbone network of YOLOv5m and integrate transfer learning, image enhancement, and data enhancement methods to improve the effectiveness, as shown in Figure 3.

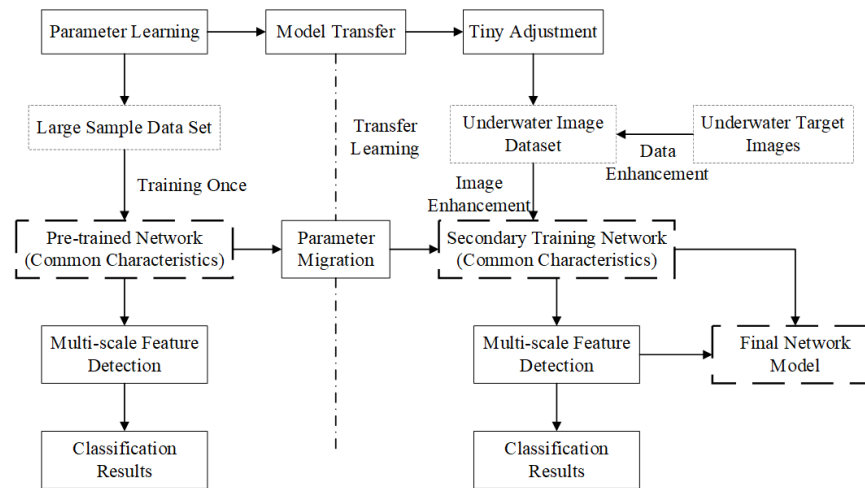


Figure 3. Model Training Method.

Firstly, select an appropriate deep convolutional neural network and conduct training on a large sample data set once to obtain a pre-trained network model. Secondly, use data enhancement technology to expand underwater target images to generate the underwater image data set, and then the underwater image enhancement algorithm proposed in this article will be implemented to pre-process the data set. Finally, the parameters of the feature extraction network in the pre-trained model are frozen, and then the pre-trained network is trained the second time on the enhanced underwater image data set, and the parameters of the feature detection network and classify network should be renewed, and the final underwater target detection and recognition network will be generated.

3.2. Model Evaluation

For the task of detecting and identifying fishes, the trained model must be able to accurately identify fishes and determine the location and range of fishes. Then, on this basis, it should ensure the recognition rate of fish and at the same time reduce the probability of misidentification or missed identification as much as possible. In order to prove the performance of the basic YOLOv5m model, a simple comparative experiment is carried out between the SSD (Single-Shot Detector) model, YOLOv4 model, YOLOv5s model, and YOLOv5m model on the self-built underwater image data set based on transfer learning and image enhancement. The training and testing of all models are under the same experimental environment.

According to results shown in Table 2 the underwater target detection and recognition model trained through YOLOv5m based on transfer learning and image enhancement has the best overall performance. Therefore, this article will adopt YOLOv5m as the basic model to conduct subsequent research on underwater target detection and recognition.

Table 2. Results of Models' Comparative Experiments

Model	SSD	YOLOv4	YOLOv5s	YOLOv5m
Precision/%	86.8	91.6	98.4	91.4
Recall/%	77.4	82.4	45.6	84.7
mAP@0.5/%	77.5	82.9	45.9	84.1
FPS/(frame/s)	48.07	42.62	72.46	54.95
Size/MB	122.1	244.5	14.4	42.5

3.3. Model Improvements

Although the original YOLOv5m model has achieved proper results in underwater target detection and recognition, there is still room for improvement. On the one hand, the attention mechanism can be introduced to further raise the ability to extract target features and reduce the negative-true and false-negative rates. On the other hand, the convolutional layer structure can be improved to make the network more lightweight, and reduce model complexity and computing costs.

3.3.1. Improvement of Feature Extraction Layer Based on Attention Mechanism

Since some fish can be pretty small and easily blend into the environment background, this article introduces an attention mechanism into the YOLOv5m model to improve the feature extraction layer's learning ability for target features and improve target feature extraction ability to achieve better detection results.

Introducing the attention mechanism is to tell the model where to put the “attention” without paying attention to the image background or other parts. Commonly used attention mechanisms include SENet [31], CBAM [32], ECA [33] and CA [34] with features as follows.

- (1) SENet puts “attention” on channel information and models the inherent relationships of each channel to make each channel more sensitive to target features.
- (2) CBAM not only takes into account channel information but also takes spatial information into account; therefore, the introduction of the CBAM attention mechanism can obtain richer feature information.
- (3) The ECA module does not involve dimensionality reduction operations, ensuring that the underlying features will not be lost, and the information interaction operation effectively reduces the calculation.
- (4) CA module encodes channel information along the horizontal and vertical directions of the coordinate axis. The CA module can be added to any feature map in the network model to process it.

Usually, the attention mechanism will be added to the backbone part of the YOLOv5m network, because this part plays the role of feature extraction in the entire YOLOv5m network, including all feature information of the detected fishes. Adding the attention mechanism will correct the importance of this information and remove some useless information, thereby improving the network's feature extraction capability and detection accuracy. As a result, there are two ways to add the attention module to the backbone part: one is to put the attention module at the end of the part; the other is to integrate the attention mechanism into each C3 module to obtain the C3-attention module.

To improve the original network model, an attention module is integrated at different positions and with different structures to explore its effectiveness in detection and recognition. A total of five cases are discussed in this paper, that is, YOLOv5m+SE(pos1), YOLOv5m+SE(pos2), YOLOv5m+ECA, YOLOv5m+CBAM, YOLOv5m+CA. The positions of the backbone to add ECA, CBAM, and CA are the same as SE(pos2). The training data set uses the image-enhanced underwater image data set. The settings of hyperparameters are shown in Table 3. The improved models are trained for 300 rounds, respectively.

Table 3. Hyperparameter settings.

Name	Image Size	Batch Size	Optimizer	Lr0	Lrf	Momentum	Weight Decay
Numerical value	$640 \times 640 \times 3$	16	SGD	0.01	0.1	0.937	0.0005

The comparison results are shown below in Section 3.3.2. Although the size of the underwater target detection and recognition model trained by YOLOv5m with the SENet module added to each C3 module has been reduced, the detection speed has also been

significantly reduced, and the detection accuracy has not been significantly improved, the mean average precision is only 0.8% higher than the original YOLOv5m model. The underwater target detection and recognition model developed by YOLOv5m by adding the SENet module at the end of the backbone is not different in size, computing cost, and detection speed compared with the original YOLOv5m model, but the precision rate, recall rate and mean average precision have increased by 4%, 4.5% and 5.4%, respectively. The detection accuracy of the model has been significantly improved, and the number of false detection and missed detection is effectively reduced. Then, based on the analysis of the above experimental results, adding the attention module at the end of the backbone is more suitable for underwater fish detection and recognition tasks.

After determining the location to add the attention module, the very attention module should be selected to achieve better performance. This article adds SENet, CBAM, ECA and CA at the end of the YOLOv5m model's backbone, respectively, and then trains the network model. Table 4, compares the training results and the performance of the models on the test set to analyze which attention mechanism is more suitable for underwater fish detection and recognition tasks. Although the YOLOv5m model with the SENet module added has a great improvement in detection accuracy compared with the original YOLOv5m model, its overall performance is worse than the YOLOv5m model with other modules added; the size and parameter quantity of the YOLOv5m model with the ECA module added are basically unchanged compared with the original YOLOv5m model, but its detection accuracy is not as good as the YOLOv5m models with the CBAM module and CA module added; compared with adding the CA module, although the precision of the YOLOv5m model with the CBAM module added is slightly smaller, its recall and mean average precision both are larger, and it is also more advantageous in terms of size, parameter quantity and detection speed. In summary, the overall performance of the underwater target detection and recognition model trained by the YOLOv5m model with the CBAM module added is better, because it effectively improves the precision and reduces the probability of false detection and missed detection of targets while ensuring that the model is not complicated and the detection speed remains basically unchanged compared with the original YOLOv5m model.

Table 4. Comparison of Adding Different Attention Modules.

Model	Size/MB	Precision/%	Recall/%	mAP@0.5/%	FPS/ (Frame/s)
YOLOv5m	42.5	91.4	84.7	84.1	54.95
YOLOv5m +SENet(pos1)	36.9	95.2(+3.8)	84.7	84.9(+0.8)	45.87
YOLOv5m +SENet(pos2)	43.1	95.4(+4.0)	89.2(+4.5)	89.5(+5.4)	52.91
YOLOv5m +ECA	42.5	97.6(+6.2)	89.5(+4.8)	89.7(+5.6)	52.35
YOLOv5m +CBAM	43.1	97.2(+5.8)	90.4(+5.7)	90.7(+6.6)	53.19
YOLOv5m +CA	43.4	97.6(+6.2)	90.1(+5.4)	90.5(+6.4)	51.55

In order to show the effects of adding a CBAM module in identifying small targets or targets that are easily confused with the background, simulation experiments are conducted here with the original YOLOv5m model and the CBAM-YOLOv5m model to execute target detection and recognition on the same test set. As shown in Figure 4 with the typical part of the results, after adding the CBAM attention mechanism, small targets that cannot

detected by the original YOLOv5m could be detected and accurately identified by the CBAM-YOLOv5m model in different scenarios.

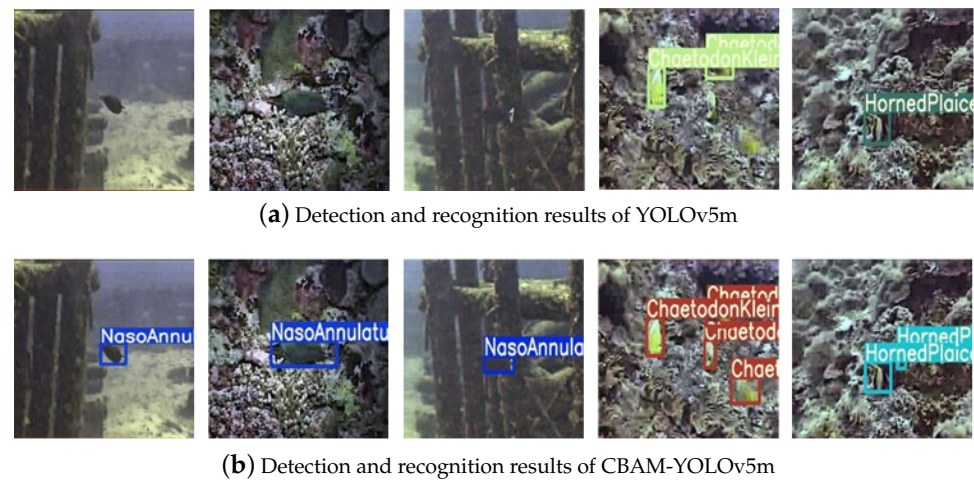


Figure 4. Detection and Recognition results of YOLOv5m and CBAM-YOLOv5m. For images containing small or blurred targets, five typical images are presented to show the effectiveness.

3.3.2. Lightweight Improvements of CBAM-YOLOv5m Model

To further reduce the model size, parameter quantity, and computation burden of CBAM-YOLO while ensuring precision, a lightweight improvement strategy is implemented in this section. Lightweight convolution modules are adopted to replace standard convolution modules to achieve a lightweight network. Commonly used lightweight convolution modules mainly include DSC (Depth-wise Separable Convolution) [35], Ghost-Conv [36] and GSConv [37], etc. Since convolution modules and C3 modules are almost distributed throughout the entire network model, and each module has a large number of parameters, a lightweight design may improve the whole network efficiently.

(1) Replacement location selections

The backbone part plays the role of feature extraction, and dense convolution operations will retain the connections between channels, while sparse convolution operations will ignore all these connections; therefore, although replacing the backbone part will reduce the model size and computing cost, loss of semantic features caused by changes in the number of feature map channels will lead to a reduction in precision. The feature map of the neck part is small in size and has a large number of channels, it is generally no longer compressed or expanded. Therefore, lightweight processing of convolution modules and C3 modules of the neck part can not only reduce the model size and computational cost but also ensure that precision is not reduced.

(2) Choice of lightweight modules

The DSC module may reduce both the detection and recognition accuracy and detection speed of the model, so it is not considered a lightweight module for replacement. The GhostConv module and the C3Ghost module can effectively reduce the number of parameters and size of the model, but they require more calculation costs. The GSConv module and VoVGSCSP module are not as effective as the GhostConv module and C3Ghost module in reducing the number of parameters and size of the model, but they occupy less computing resources. Therefore, this article weighs the advantages and disadvantages of each lightweight module, in the neck part, the C3Ghost module is selected to replace C3 modules, whose number of parameters is large, to effectively reduce the model size and parameter quantity, at the same time, the GSConv module is selected to replace convolution modules to ensure that the computing cost is not too high. The lightweight CBAM-YOLOv5m network structure is shown in Figure 5.

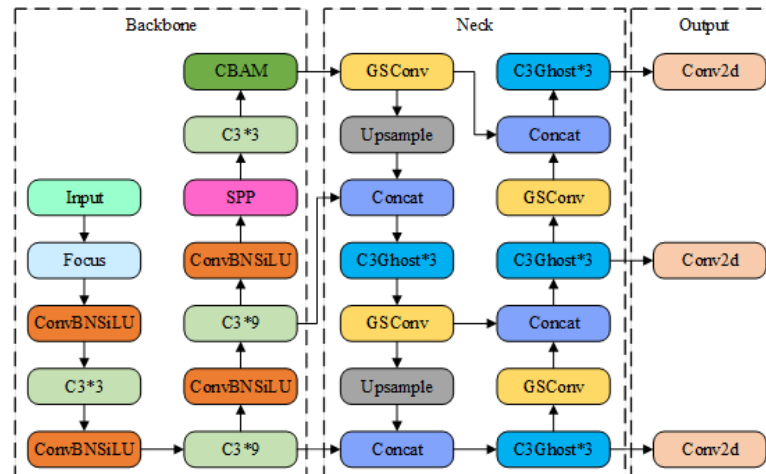


Figure 5. Light-weighted CBAM-YOLOv5m.

Then experiments are conducted to prove the correctness of the analysis of replacement location selections and choice of lightweight modules.

(1) Changes in parameter quantity

Each module's parameter quantity of the neck structure before and after lightweight improvement is shown in Table 5. Obviously, after the lightweight improvement, the parameter quantity of each module in the neck structure is significantly reduced, effectively reducing the size of the entire model.

Table 5. Comparison of Modules' Parameter Quantity before and after Lightweight Improvement.

Original Neck		Lightweight Neck	
Module	Parameter Quantity	Module	Parameter Quantity
Conv+ Upsample	295,680	GSConv+ Upsample	153,024 (−48.2%)
C3 × 3	1,182,720	C3Ghost × 3	489,120 (−58.6%)
Conv+ Upsample	74,112	GSConv+ Upsample	39,648 (−46.5%)
C3 × 3	296,448	C3Ghost × 3	124,752 (−57.9%)
Conv	332,160	GSConv	168,672 (−49.2%)
C3 × 3	1,035,264	C3Ghost × 3	341,664 (−67.0%)
Conv	1,327,872	GSConv	669,120 (−49.6%)
C3 × 3	4,134,912	C3Ghost × 3	1,346,880 (−67.4%)

(2) Changes in performance

In this paper, under the same experimental conditions, the lightweight model was trained for 300 rounds on a self-built underwater image data set, and the training results were compared with the CBAM-YOLOv5m model obtained in the above section, as shown in Figure 6. It can be seen that the lightweight improvement has almost no impact on the convergence speed and precision of the CABM-YOLOv5m model. Change trends in the total loss, recall and mean average precision are almost the same. This illustrates that the lightweight CABM-YOLOv5m model designed in this article has almost no impact on the model training speed and detection accuracy.

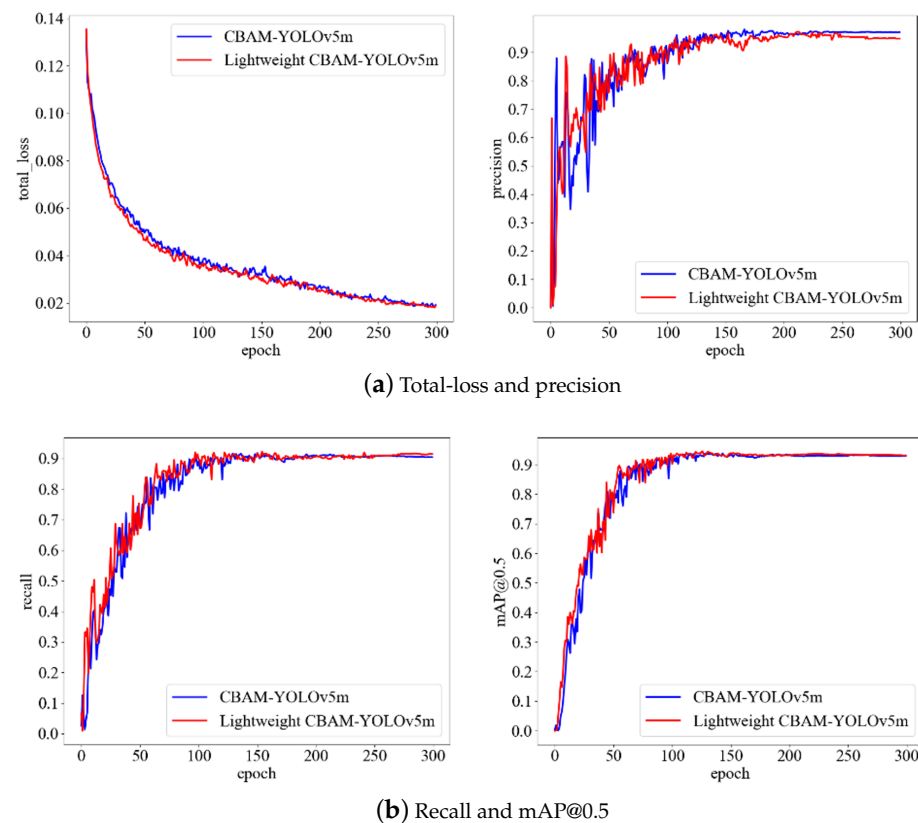


Figure 6. Comparison of Training Results Before and After Lightweight Improvement.

(3) Performance comparison of different lightweight methods

In addition, to further demonstrate the advantage of the proposed method, comparisons between the CBAM-YOLOv5m mode, the lightweight model designed, and models improved by selecting different replacement positions or other modules are carried out on the same test set. Experiment results are shown in Table 6, serial number 1 is the proposed CBAM-YOLOv5m here, and serial number 7 is the lightweight improved model adopted in this article.

Table 6. Comparison of Models Improved by Different Lightweight Methods.

Model Structure				No.	Size /MB	GFL OPs	Recall /%	mAP @0.5/%	FPS/ (Frame/s)
Backbone		Neck							
Conv	C3	Conv	C3	1	43.1	50.8	90.4	90.7	53.19
GS Conv	C3 Ghost	GS Conv	C3 Ghost	2	18.2	18.9	79.1	79.0	42.19
Conv	C3	DSCConv	DSC3	3	32.8	41.4	89.1	90.4	52.45
Conv	C3	Ghost Conv	C3 Ghost	4	32.4	40.6	89.4	90.1	56.78
Conv	C3	GS Conv	VoVG SCSP	5	35.6	39.2	89.7	90.0	56.54
Conv	C3	Ghost Conv	VoVG SCSP	6	35.6	39.2	89.2	89.7	56.08
Conv	C3	GSConv	C3Ghost	7	32.4	40.6	91.5	91.4	57.56

Comparing model No.7 with model No.1, it can be seen that the lightweight method proposed in this article reduces the size and computing cost of the CBAM-YOLOv5m model by 25% and 20%, respectively, and the detection speed is also improved. Not only does the precision not decline, but the recall rate and mean average precision increase by 1.1% and 0.7%, respectively. Comparing model No.7 with model No.2, although the size and computing cost of the model have been significantly reduced, the precision becomes very low, and the detection speed is greatly reduced. Comparing model No.7 with models No.3, No.4, No.5, and No.6, the improved model in this article has more advantages in model size, detection speed, and precision. Although the computing cost is slightly higher than that of model No.5 and model No.6, the difference is small.

(4) Fish detection and recognition simulation experiment

Using this article's model and the original YOLOv5m model to conduct underwater fish detection and recognition simulation experiments on the same test set, part of the detection and recognition results are shown in Figure 7. From the first and fifth sets of images in Figure 7, the lighted-weighted model in this article has better detection and recognition capabilities of small fish, and the number of missed fish detections has been reduced. It can also be seen from the second set of images that the confidence of the fish detected by this improved model is higher than that of the original model. Additionally, the number of false detections of the improved model has been reduced according to the third and fourth sets of images.

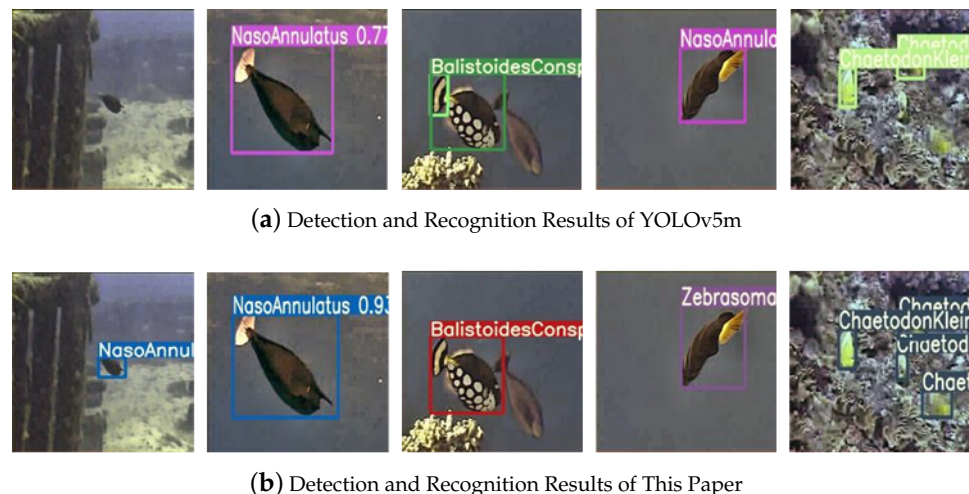


Figure 7. Comparison of Detection and Recognition Results Before and After Improvement.

Detection and recognition results of the two models for each type of fish in the test set are also statistically calculated, and statistical results are shown in Figure 8. It can be seen from Figure 8 that the improved model has better detection and recognition indicators of various fish with only the precision of *Zebrasoma scopas* slightly decreasing. Recalls and average precision of detection for *Zebrasoma scopas*, *Naso annulatus*, *Zanclus cornutus*, and *Chaetodon kleinii* in the original model are all not high mainly due to the existence of small targets or targets that are easily confused with the background or multiple objects in one image in the data sets. The improved model's detection and recognition capabilities for these targets have been significantly improved.

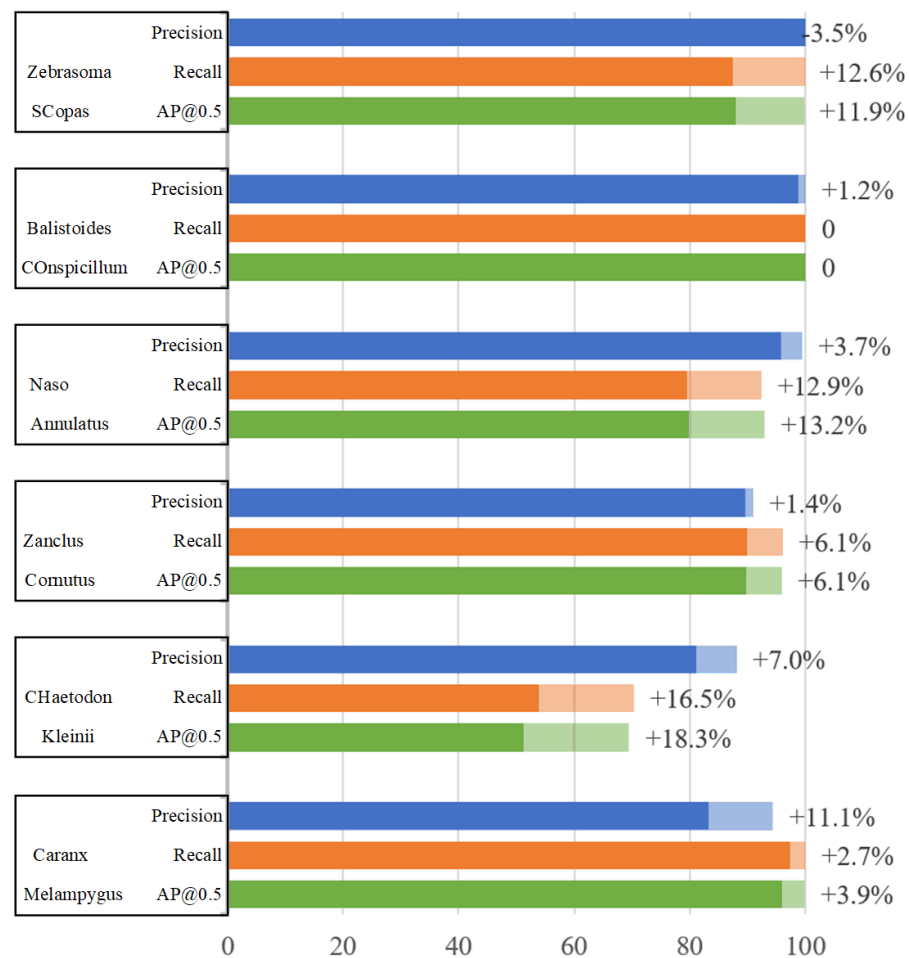


Figure 8. Statistics of Detection and Recognition Results Before and After Improvement.

4. Conclusions

This paper conducts in-depth research on underwater image enhancement and fish detection and recognition. Underwater image processing methods mainly include color enhancement algorithms to improve image color saturation, detail enhancement algorithms based on Sigmoid functions to improve image details, and adaptive Gamma correction algorithms to adjust image brightness. Based on the above algorithms, an underwater image enhancement algorithm based on weighted fusion is proposed. The detection and recognition model of this article is improved based on YOLOv5m. Based on deep learning and transfer learning, and integrating image enhancement algorithms, the underwater target detection and recognition network training method of this article is designed. In order to solve the problem that some small fish and fish that are easily confused with the background cannot be recognized in the underwater image data set, this paper effectively improves the detection and recognition accuracy and reduces the probability of miss-detection by adding a variety of attention modules while ensuring that the size and detection speed remain basically unchanged. In order to reduce the size and computing cost of the model, this paper designs a lightweight CBAM-YOLOv5m model by replacing ordinary convolutions with lightweight convolutions. The effectiveness of the proposed methods in this paper was demonstrated through experiments. Although the research contents of this article have made progress, there is still room for improvement, such as enriching the underwater image data set and further reducing the model size and computing cost while ensuring precision, so as to facilitate applications in mobile devices and embedded devices. In addition, experiments will be carried out in an environment of dirty or contaminated water in the future to verify whether the method described in this article can be applied to more scenarios.

Author Contributions: Conceptualization, Z.Y. and J.Z.; methodology, L.H. and J.Y.; software, L.H. and J.Y.; validation, L.H. and Z.Y.; formal analysis, Z.Y. and J.Z.; investigation, L.H. and J.Y.; writing—original draft preparation, Z.Y. and L.H.; writing—review and editing, Z.Y. and L.H.; visualization, J.Y.; supervision, Z.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This work is partially supported by National Natural Science Foundation of China under grants 52071102, 51909044.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are contained within the article.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Liu, Z. *Research on Target Detection and Recognition Method Based on Underwater Image Enhancement*; ShanDong University: Jinan, China, 2021.
2. Zhang, H.; Xu, Y.; Wan, L.; Tang, X.; Cai, H. Processing Method for Underwater Degenerative Image. *J. Tianjin Univ.* **2010**, *43*, 827–833.
3. Iqbal, K.; Odetayo, M.; James, A.; Salam, R.A.; Talib, A.Z.H. Enhancing the Low Quality Images Using Unsupervised Colour Correction Method. In Proceedings of the 2010 IEEE International Conference on System, Man and Cybernetics, Istanbul, Turkey, 10–13 October 2010; pp. 1703–1709.
4. Ancuti, C.; Ancuti, C.O.; Haber, T.; Bekaert, P. Enhancing underwater images and videos by fusion. In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012; pp. 16–21.
5. Hitam, M.S.; Awalludin, E.A.; Yussof, W.N.J.H.W.; Bachok, Z. Mixture contrast limited adaptive histogram equalization for underwater image enhancement. In Proceedings of the 2013 International Conference on Computer Applications Technology (ICCAT), Sousse, Tunisia, 20–22 January 2013; pp. 20–22.
6. Ghani, A.S.A.; Isa, N.A.M. Underwater image quality enhancement through integrated color model with Rayleigh distribution. *Appl. Soft Comput.* **2015**, *27*, 219–230. [\[CrossRef\]](#)
7. Cao, F.; Zhao, K.; Wang, X.; Qian, Y.; Wang, H. Adaptive underwater color image enhancement algorithm. *J. Electron. Meas. Instrum.* **2016**, *30*, 772–778.
8. Huang, D.; Wang, Y.; Song, W.; Sequeira, J.; Mavromatis, S. Shallow-Water Image Enhancement Using Relative Global Histogram Stretching Based on Adaptive Parameter Acquisition. In *MultiMedia Modeling*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 453–465.
9. Jia, P.; Li, B.; Zhao, X. Improved Retinex Underwater Image Enhancement Algorithm Based on HSI Model. *Res. Explor. Lab.* **2020**, *12*, 1–4.
10. Du, X.; Sun, X.; Li, C.; Wang, C.; Wang, N.; Feng, Z. Weighted Fusion Based Multi-Space Transformation Method for Underwater Image Enhancement. *Electron. Opt. Control* **2021**, *28*, 101–105.
11. Dong, L.L.; Zhang, W.D.; Xu, W.H. Underwater image enhancement via integrated RGB and LAB color models. *Signal Process. Image Commun.* **2022**, *104*, 10–16. [\[CrossRef\]](#)
12. Wang, S.; Xu, Y.; Wan, L.; Tang, X. Underwater targets recognition based on contour moment and modified FCM algorithm. *Syst. Eng.-Theory Pract.* **2012**, *32*, 2809–2815.
13. Shi, M.; Xu, X.; Yue, J. Underwater target recognition based on wavelet transform and probabilistic neural network. *Ship Sci. Technol.* **2012**, *34*, 85–87.
14. Shi, Y.; Hu, C. Forward-looking sonar target recognition based on particle swarm and least squares support vector machine. *Tech. Acoust.* **2018**, *37*, 122–128.
15. Zhu, J.; Yu, S.; Han, Z.; Tang, Y.; Wu, C. Underwater object recognition using transformable template matching based on prior knowledge. *Math. Probl. Eng.* **2019**, *10*, 1888. [\[CrossRef\]](#)
16. Wang, X.; Liu, A.; Zhang, Y.; Xue, F. Underwater acoustic target recognition: A combination of multi-dimensional fusion features and modified deep neural network. *Remote Sens.* **2019**, *11*, 72–79. [\[CrossRef\]](#)
17. Li, Y.; Wang, J.; Huang, W. A Research on AUV Underwater Recognition System Based on Convolutional Neural Network. *Nav. Archit. Ocean Eng.* **2021**, *37*, 20–25.
18. Chen, L.; Zheng, M.; Duan, S.; Luo, W.; Yao, L. Underwater Target Recognition Based on Improved YOLOv4 Neural Network. *Electronics* **2021**, *10*, 634. [\[CrossRef\]](#)
19. Wang, H.; Wang, P.; Deng, S.; Gu, Z. MRFCNN: The optimisation method of convolutional neural network for underwater target recognition. *Int. J. Model. Identif. Control* **2022**, *40*, 36–43. [\[CrossRef\]](#)
20. Forsyth, D.A. A novel algorithm for color constancy. *Int. J. Comput. Vis.* **1990**, *5*, 5–35. [\[CrossRef\]](#)

21. Zuiderveld, K. *Contrast Limited Adaptive Histogram Equalization*; Graphics Gems IV; Academic Press: Cambridge, MA, USA, 1994; pp. 474–485.
22. Liu, Z.; Wang, D.; Liu, Y.; Liu, X. Adaptive Adjustment Algorithm for Non-Uniform Illumination Images Based on 2D Gamma Function. *Trans. Beijing Inst. Technol.* **2016**, *36*, 191–196. 214.
23. Kimmel, R.; Elad, M.; Shaked, D.; Keshet, R.; Sobel, I. A Variational Framework for Retinex. *Int. J. Comput. Vis.* **2003**, *52*, 7–23. [[CrossRef](#)]
24. Ancuti, C.O.; Ancuti, C.; De Vleeschouwer, C.; Bekaert, P. Color balance and fusion for underwater image enhancement. *IEEE Trans. Image Process.* **2017**, *27*, 379–393. [[CrossRef](#)] [[PubMed](#)]
25. Kim, Y.; Koh, Y.J.; Lee, C.; Kim, S.; Kim, C.S. Dark image enhancement based on pairwise target contrast and multi-scale detail boosting. In Proceedings of the IEEE International Conference on Image Processing, Quebec City, QC, Canada, 27–30 September 2015; pp. 1404–1408.
26. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.
27. Redmon, J.; Farhadi, A. YOLO9000: Better, faster, stronger. In Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition CVPR 2017, Honolulu, HI, USA, 21–26 July 2016; pp. 6517–6525.
28. Redmon, J.; Farhadi, A. YOLOv3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767.
29. Bochkovskiy, A.; Wang, C.-Y.; Liao, H.-Y.M. YOLOv4: Optimal Speed and Accuracy of Object Detection. *arXiv* **2020**, arXiv:2004.10934.
30. Wu, D.; Lv, S.; Jiang, M.; Song, H. Using channel pruning-based YOLOv4 deep learning algorithm for the real-time and accurate detection of apple flowers in natural environments. *Comput. Electron. Agric.* **2020**, *178*, 105742. [[CrossRef](#)]
31. Zhong, X.; Gong, O.; Huang, W.; Li, L.; Xia, H. Squeeze-and-Excitation Wide Residual Networks in Image Classification. In Proceedings of the 2019 IEEE International Conference on Image Processing, Taipei, Taiwan, 22–25 September 2019; pp. 395–399.
32. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. CBAM: Convolutional block attention module. In Proceedings of the 15th European Conference on Computer Vision, Munich, Germany, 8–14 September 2018; pp. 3–19.
33. Wang, Q.; Wu, B.; Zhu, P.; Li, P.; Zuo, W.; Hu, Q. ECA-Net: Efficient channel attention for deep convolutional neural networks. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 11531–11539.
34. Hou, Q.; Zhou, D.; Feng, J. Coordinate attention for efficient mobile network design. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 13708–13717.
35. Chollet, F. Xception: Deep learning with depthwise separable convolutions. In Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition CVPR 2017, Honolulu, HI, USA, 21–26 July 2016; pp. 1800–1807.
36. Han, K.; Wang, Y.; Tian, Q.; Guo, J.; Xu, C.; Xu, C. GhostNet: More features from cheap operations. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 1577–1586.
37. Li, H.; Li, J.; Wei, H.; Liu, Z.; Zhan, Z.; Ren, Q. Slim-neck by GSConv: A better design paradigm of detector architectures for autonomous vehicles. *arXiv* **2022**, arXiv:2206.02424.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.