International Journal for Housing Science and Its Applications

Publish September 9, 2025. Volume 47, Issue 1 Pages 811-826

https://doi.org/10.70517/ijhsa47171

Underwater litter detection network

Xinya Lu[1,*]

[1] School of Computer Science and Artificial Intelligence, Shandong University of Finance and Economics, Jinan, Shandong, 250000, China

Corresponding authors: (e-mail: 18854706666@163.com).

Abstract The accuracy of underwater garbage detection and identification plays a very important role in improving the garbage cleaning work carried out by underwater robots. Based on this, this paper proposes an improved underwater trash detection network model based on YOLOv5s. In order to improve the recognition performance of underwater garbage images, this paper also proposes a weighted fusion-based underwater image enhancement algorithm, which fuses the CLAHE algorithm and Retinex algorithm on the basis of weighted logarithmic transformation and adaptive Gamma correction, so as to improve the quality of underwater garbage images. For underwater garbage detection, GhostNet is introduced to improve the backbone network of YOLOv5s to enhance the feature extraction capability, and combined with the ECA attention mechanism and CARAFE up-sampling mechanism to further realize the model lightweighting and enrich the features and semantic features. The results show that the YOLOv5s-G-E-C model improves the detection average accuracy from 60.92% to 86.19% and the model computation reduces the model computation from 18.42 GLOPs to 15.78 GLOPs compared to the YOLOv5s model.It is feasible to apply the improved YOLOv5s model to underwater garbage detection with better detection performance.

Index Terms gamma correction, CLAHE algorithm, GhostNet, ECA attention mechanism, YOLOv5s-G-E-C model, underwater trash detection

I. Introduction

Underwater litter pollution has become a worldwide environmental problem, posing a serious threat to underwater ecosystems [1], [2]. In order to solve this problem, a large number of garbage in waters have been treated all over the world, but the effect is not very good [3], [4]. In order to remove floating objects from the water, various types of water litter cleaning robots have been introduced [5]. Most of the robots, collect the garbage, in which case the efficiency is reduced and cleaning is troublesome [6], [7]. In order to effectively remove and improve the efficiency, the application of underwater trash detection network has become the focus of academic attention [8], [9].

Underwater litter detection is the process of recognizing and detecting underwater litter in an underwater environment by using different sensors and technologies [10], [11]. Underwater litter detection is one of the core technologies in the fields of marine resource exploitation, marine scientific research and military sector [12]. The underwater environment is very different from the land environment, and the absorption and scattering properties of water make the underwater images limited in quality and resolution [13], [14]. At the same time, there are many interfering factors in the underwater environment, such as water currents, air bubbles, suspended objects, etc., which further degrade the quality of underwater images. Therefore, underwater trash detection is a challenging task [15]-[17]. In order to solve the problem of underwater target detection, researchers have proposed different methods, among which sonar imaging, optical imaging and LiDAR are the most commonly used techniques [18], [19].

The underwater debris generated due to human activities, industrial production and other methods can have a large impact on the underwater ecosystem and pose a serious threat to human and water quality health. In this paper, an underwater trash detection model based on YOLOv5s-G-E-C model is proposed. Aiming at the problem of low quality of underwater garbage images, the article combines CLAHE algorithm, Retinex algorithm, and Gamma correction, and establishes an underwater garbage image enhancement algorithm using weighted fusion. The enhanced underwater garbage image is used as input, and the improved YOLOv5s-G-E-C model with GhostNet, ECA and CARAFE is combined to realize the real-time monitoring of underwater garbage. The experiments show that the model has high detection accuracy for underwater garbage and has practical applications for underwater garbage cleaning.

II. Underwater image enhancement algorithm based on weighted fusion

Studies have shown that removing plastic debris from underwater would benefit underwater ecosystems exponentially. However, underwater lighting environments are complex, with bright light affecting waters of good quality and turbid waters

making it difficult to observe objects. Therefore, the quality of underwater images and the accuracy and efficiency of underwater trash identification can be improved by using appropriate underwater image enhancement algorithms. Aiming at underwater environmental imaging, where images suffer from insufficient light, image fogging and low contrast, this chapter proposes a weighted fusion-based image enhancement algorithm for underwater garbage, which aims to provide reliable underwater garbage image support for accurate detection and recognition of underwater garbage.

II. A. Underwater image enhancement algorithm

II. A. 1) CLAHE algorithm

The Constrained Contrast Adaptive Histogram Equalization (CLAHE) algorithm is an improvement of the Adaptive Histogram Equalization (AHE) algorithm, which introduces the concept of contrast limitation, which restricts the enhancement of the contrast while equalizing the histogram, which avoids the introduction of too much noise while preserving the local details of the image. In the CLAHE algorithm, if the number of some pixels in a sub-block exceeds a predefined limit, the exceeding portion is pruned, and the pruned portion is reallocated to other gray levels,. Subsequently, the algorithm calculates the cumulative distribution function (CDF) based on the histogram of the pruned reassignments, and finally maps the new gray values based on the position of the pixel points [20].The specific procedure of the CLAHE algorithm is as follows:

(1) Split the original image into equal-sized, non-overlapping sub-blocks, the number of sub-blocks is selected based on the actual situation.

(2) For each sub-block, count its grayscale histogram separately.

(3) Set a pruning limiting factor β, the value of β is between 0 and 1, the default is 0.01, the closer the value is to 1, the greater the contrast of the enhanced image. The formula for the trimming limiting value is:

$$T = \frac{xy}{L} + \left(\beta \times \left(xy - \frac{xy}{L} \times T \right) \right) \tag{1}$$

where x and y denote the number of pixels in the rows and columns of the sub-block image, respectively, and L is the number of gray levels of the sub-block.

(4) Crop and allocate the histograms of all sub-blocks according to the pruning limit T, firstly, the pixels exceeding T in the histogram are cropped down, and here the number of pixels cropped down is assumed to be N_{clip}, and then N_{clip} is evenly allocated to each gray level, and the number of pixels allocated to each gray level is N_{ave} as shown in equation (2). If there are pixels remaining after the first allocation is completed, the allocation operation is continued in a loop until the cropped pixels are completely allocated, and the histogram of the allocated sub-blocks $H_d(i)$ is obtained, as shown in equation (4). Namely:

$$N_{ave} = \frac{N_{clip}}{L} \tag{2}$$

Among them:

$$N_{clip} = \sum_i (\max(H(i) - T, 0)) \tag{3}$$

$$H_d(i) = \begin{cases} T, & H(i) > T \\ T, & H(i) + N_{ave} \geq T \\ H(i) + N_{ave}, & H(i) + N_{ave} \leq T \end{cases} \tag{4}$$

(5) Do histogram equalization again for each sub-block after reallocation.

(6) Combine the equalized sub-blocks into the final enhanced image, which uses linear interpolation to find new pixel values to ensure that the output image is continuous.

The CLAHE algorithm is able to effectively control the local contrast enhancement when enhancing the image, thus improving the visual quality of the image.

II. A. 2) Retinex enhancement algorithm

The Retinex algorithm is a more commonly used algorithm in the field of image enhancement. The human visual system does not only perceive the color and brightness of an object by receiving the brightness of the light, but also by comparing the relative intensities of the light on the surface of the object and the light of the surrounding environment, which makes it possible to have a relative consistency of the color and brightness of the observed object. A certain degree of stability can be maintained even under different lighting conditions without the interference of light changes. Its mathematical expression is:

$$S(x,y) = R(x,y) \cdot I(x,y) \tag{5}$$

where x and y denote the coordinates of the pixel points, $S(x,y)$ denotes the initial image, $I(x,y)$ denotes the image irradiation component, and $R(x,y)$ denotes the image reflection component.

II. A. 3) Gamma correction algorithm

Gamma correction refers to the direct correction of the gray value of the image through the nonlinear function mapping, so as to achieve the effect of image contrast enhancement.Gamma correction has a key role in enhancing the brightness of low illumination images in the spatial domain, and its expression is:

$$O(x,y) = 255 \times \left(\frac{I(x,y)}{255} \right)^{\gamma} \tag{6}$$

where $I(x,y)$ is the pixel value of the input image, $O(x,y)$ is the pixel value of the output image after gamma correction; γ is the correction coefficient, which usually takes the value of 0.5–1 under the condition of low-lighting, which can make the pixel points with very low pixel value reach the effect of stretching obviously, the dynamic range is improved, and the image contour is clearer.

When the Gamma correction coefficient is less than 1, the low gray value increases greatly, the high gray value increases slowly, and the overall gray value is improved to varying degrees, thus making the dynamic range of the image larger. When the Gamma correction factor is equal to 1, the grayscale value remains unchanged, and the output image and the input image remain consistent. When the Gamma correction coefficient is greater than 1, the pixel points in the low and high gray value areas continue to decrease, which improves the contrast of the brighter areas, but the contrast of the darker areas decreases more seriously. Therefore, when selecting the correction coefficient, the appropriate parameter should be selected in the range of less than 1, so as to realize the effective improvement of brightness.

In this paper, we will design the Gamma coefficients to realize the adaptive Gamma correction function, which can take into account the overall brightness change of the image and realize the brightness information more in line with the human eye's vision.

II. A. 4) Underwater dark channel a priori algorithm

Considering the underwater environment, some researchers have proposed an underwater dark channel prior (UDCP) algorithm.The UDCP algorithm is proposed based on the assumption of the dark channel prior (DCP), which is based on the principle that there exists at least one object with the least intensity of the color channel as well as the dark object in the underwater image, and therefore the intensity of the pixels in these color channels can be almost considered as zero [21].

The expression for calculating the underwater dark channel using the UDCP algorithm is:

$$J_{UDCP} = \min_{y \in \pi(x)} \left(\min_{c \in (g,b)} J_c(y) \right) \tag{7}$$

where $\pi(x)$ is a local block centered at x in the underwater image and $J_c(y)$ is the scene image for a particular channel.

Neglecting the forward scattering component, the mathematical description of the underwater imaging model can be approximated as the sum of the backscattering component and the direct reflection component, i.e:

$$I(x) = J(x)t(x) + B(1 - t(x)) \tag{8}$$

where x is a point in the underwater scene, $I(x)$ is the image captured by the camera, $J(x)$ is the scene brightness at the point, $t(x)$ is the ratio of the remaining energy of the point that reaches the camera after reflections in the underwater scene, and B is the background light of the water source.

Applying the minimum operator on both sides of Equation (8) and using the concept of an underwater dark channel yields:

$$\min_{y \in \pi(x)} \left(\min_{c \in (g,b)} I_c(y) \right) = \min_{y \in \pi(x)} \left(\min_{c \in (g,b)} (J_c(y)t(x) + B_c(1 - t(x))) \right) \tag{9}$$

where $I_c(y)$ denotes the channel-specific input intensity and B_c denotes atmospheric light.

Normalizing the above equation with respect to B_c gives:

$$\min_{y \in \pi(x)} \left(\min_{c \in (g,b)} \frac{I_c(y)}{B_c} \right) = t(x) \min_{y \in \pi(x)} \left(\min_{c \in (g,b)} (J_c(y) + B_c(1 - t(x))) \right) \tag{10}$$

The atmospheric light is calculated by selecting the brightest pixel in the underwater dark channel. Substituting the calculated

values for the underwater dark channel and atmospheric light, the expression for the transmittance t is calculated as:

$$t = 1 - \min_{y \in \pi(x)} \left(\min_{c \in (g,b)} \frac{I_c(y)}{B_c} \right) \tag{11}$$

Using the UDCP algorithm to calculate the dark channel, it can be seen that for the images of the underwater garbage dataset, compared to the UDCP algorithm that only calculates the blue and green channels the dark channel shows richer image information, and its effect is better than the DCP algorithm that calculates the three channels.

II. B. Weighted fusion for underwater image enhancement

II. B. 1) Enhancement algorithm based on weighted fusion

The CLAHE algorithm improves the visual effect by adjusting the distribution of image pixel values, and the Retinex algorithm improves the image quality by inverting the imaging model and eliminating the illumination component. These two algorithms are simple and easy to understand in theory, and the computational overhead is small.The CLAHE algorithm can effectively enhance the details in the image by setting a threshold to limit the excessive enhancement of contrast and avoiding the introduction of unwanted noise.The Retinex algorithm enhances the results with serious edge dispersion, and the overall visual effect is not very good. In this subsection, a weighted fusion-based underwater image enhancement algorithm is proposed to address the above problems by fusing the weighted logarithmic transformation, adaptive Gamma correction, improved multi-scale Retinex algorithm and CLAHE algorithm to enhance the image quality while controlling the computational complexity.

(1) RGB to HSV color space conversion

In RGB color space, the color of each pixel is represented as a combination of three colors: red (R), green (G), and blue (B).HSV color space is a method of representing colors using three dimensions, namely hue, saturation, and lightness.The formula for converting from RGB to HSV is:

$$h = \begin{cases} 0^\circ, & if \ \max = \min \\ 60^\circ \times \dfrac{g-b}{\max - \min} + 0^\circ, & if \ \max = r \ and \ g \geq b \\ 60^\circ \times \dfrac{g-b}{\max - \min} + 360^\circ, & if \ \max = r \ and \ g < b \\ 60^\circ \times \dfrac{b-r}{\max - \min} + 120^\circ, & if \ \max = g \\ 60^\circ \times \dfrac{r-g}{\max - \min} + 240^\circ, & if \ \max = b \end{cases}$$

$$s = \begin{cases} 0^\circ, & if \ \max = 0 \\ \dfrac{\max - \min}{\max} = 1 - \dfrac{\min}{\max}, & otherwise \end{cases} \tag{12}$$

$$v = \max\{r, g, b\}$$

(2) Multi-scale Retinex algorithm based on bilateral filtering

In the traditional Retinex algorithm, a Gaussian filter is used to estimate the illumination component, but it only considers the spatial relationship of pixels and does not consider the correlation between pixels, so the enhanced image will have a loss of details and more serious edge dispersion. To address this problem, this subsection uses bilateral filter instead of Gaussian filter, and bilateral filter is used to estimate the illuminance component instead of Gaussian filter in the multi-scale Retinex algorithm, in order to achieve better enhancement effect [22].

(3) Enhancement transform based on weighted logarithmic transformation and adaptive Gamma correction

After the image has been logarithmically transformed, the contrast of the dark region will be enhanced to achieve the purpose of enhancing the dark details of the image. The logarithmic transformation formula is:

$$s = c \log(1 + r) \tag{13}$$

where s is the result of the transformation, c is an adjustment constant to control the effect of the transformation, and r is each pixel point of the image.

The dark region of the image after logarithmic transformation is not obvious, so in this paper we use weighted logarithmic transformation for brightness enhancement. A coefficient θ is added to the logarithmic transformation for local brightness enhancement, and the transformation formula is:

$$s = \frac{\sum_{x=0}^{m}\sum_{y=0}^{n} e\lg((x,y)+\varepsilon)\theta(\nabla(x,y),\tau)}{\sum_{x=0}^{m}\sum_{y=0}^{n} e\lg((x,y)+\varepsilon)}$$
$$\theta = \begin{cases} 1, & x = y \\ 0, & otherwise \end{cases} \tag{14}$$

where s is the output of the corresponding pixel point (x,y) after weighted logarithmic transformation, m and n are the length and width of the image, respectively, e is the weighted logarithmic transformation coefficient, ε is the correction coefficient, which is generally taken to be 1, ∇ is the third-order Laplacian operator, and τ is the luminance level, $\tau \in [0,255]$.

(4) Image Fusion

The above work can get two underwater images after the enhancement of the original underwater image, which need to be weighted fusion of the two.

First extract the R, G, B three-channel values of the two enhanced images, and calculate the weights W_{pta}, i.e.:

$$W_{pta} = \sqrt{\frac{1}{3}\left[(R_i-\sigma)^2+(G_i-\sigma)^2+(B_i-\sigma)^2\right]} \tag{15}$$

where R_i, G_i, B_i are the image red, green and blue channel values, and σ is the weight calculation parameter.

Then the weights of the image under HSV color space are calculated, i.e.:

$$W_{ptb} = \sqrt{[(H_i-\overline{H})^2+(S_i-\overline{S})^2+(V_i-\overline{V})^2]} \tag{16}$$

where H_i, S_i, and V_i are the values of the three channels H, S, and V, respectively, and $\overline{H}$, $\overline{S}$, and $\overline{V}$ are the average values of the three channels H, S, and V, respectively.

The weights are normalized, i.e:

$$W_1 = (W_{pta1}+W_{ptb1})/(W_{pta1}+W_{pta2}+W_{ptb1}+W_{ptb2})$$
$$W_2 = (W_{pta2}+W_{ptb2})/(W_{pta1}+W_{pta2}+W_{ptb1}+W_{ptb2}) \tag{17}$$

Finally, the weighted fusion of the two enhanced images I_1 with I_2 yields the final enhancement result I_{res}, viz:

$$I_{res} = W_1 I_1 + W_2 I_2 \tag{18}$$

II. B. 2) Underwater Image Enhancement Algorithm Flow

In order to improve the visual effect of garbage images collected under underwater conditions, a weighted fusion-based underwater garbage image enhancement algorithm is proposed, and its specific flow is shown in Fig. 1. The algorithm well improves the contrast as well as the brightness of the enhanced underwater garbage image, removes the noise, and makes the color information of the enhanced underwater garbage image richer and more natural.

The specific steps are as follows:

Step1 Transfer the underwater garbage image from RGB color space to HSV color space to obtain the luminance component V of the image.

Step2 Process the V component using bilateral filter to estimate the light component of the underwater garbage image.

Step3 Derive three lighting inputs, input 1 is the original lighting component, which contains the original lighting information of the image to prevent image distortion. The weighted logarithmic transformation and adaptive Gamma correction algorithm mentioned in the previous section are used to perform the light correction process on the original illumination component, which is labeled as light input 2, in order to effectively correct the effect of illumination on the brightness of the image. The CLAHE algorithm is used to process the original illumination component to improve the overall contrast of the image.

Step4 A detailed feature-weighted fusion strategy is used to fuse the input 1 obtained in Step3 with input 2 to obtain the corrected light component.

Step5 Morphological operations are performed on the reflection component to reduce the effect of noise on the image.

Step6 Finally, multiply the reflection component with the illumination component to obtain the enhanced luminance component, merge it with the hue and saturation, and convert the enhanced result to the RGB color space to obtain the enhanced image.

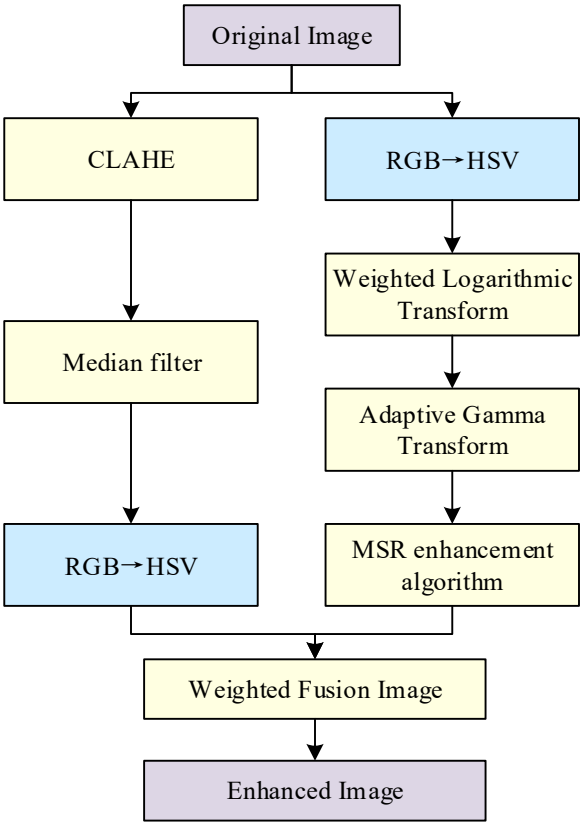

Figure 1: The process of underwater image enhancement algorithm

III. Underwater garbage detection model based on improved YOLOv5s

With the continuous development and utilization of water resources by human beings, the problem of underwater garbage pollution is becoming more and more serious, which not only poses a great threat to the water ecosystem, but also affects the sustainable utilization of water resources by human beings. Therefore, the development of an efficient and accurate underwater trash detection method has become an important research topic in the field of environmental protection and marine science and technology. The aim of this study is to design and implement a detection network model that can autonomously recognize underwater garbage through deep learning techniques, in order to improve the efficiency and accuracy of underwater garbage cleaning.

III. A. Relevant Technical Basis

III. A. 1) GhostNet network

In order to achieve efficient deployment of the model on the devices in this paper, the number of parameters and size of the model need to be reduced and adapted to mobile and embedded devices, the GhostNet network structure is introduced in this paper to reduce the model complexity. Compared with the traditional convolutional neural network, GhostNet has fewer parameters and lower computational effort without adjusting the output feature maps.The GhostNet network mainly consists of the Ghost module.The core idea of the Ghost module is to utilize simple linear transformations to generate "phantom" feature maps, thus reducing the computational effort of convolutional operations. The core idea of the Ghost module is to utilize simple linear transformations to generate "phantom" feature maps, thus reducing the computational effort of convolution operations.

The convolution process of the Ghost module is divided into three steps. The first step is to perform a standard convolution operation on the input feature maps to generate a part of the main feature maps. The number of main feature maps is usually smaller than the number of feature maps generated by conventional convolution. The second step is to generate more phantom feature maps by applying a series of simple linear transformations to each main feature map. The third step is to merge the main and phantom feature maps to form the final output feature map [23].

Assuming that the size of the convolution kernel is $k \times k$, the number of input feature channels is P, and the size of the output feature map is $H \times W \times N$, the formula for the amount of computation utilizing conventional convolution floating-point is as follows:

$$F = P \times k \times k \times N \times H \times W \tag{19}$$

Let the number of output feature channels of the first step of the Ghost module be m, and apply linear operations to each original feature map to generate s feature maps, where $m = N / s$, then there are $(s-1) \times m$ linear transformations in the second step, and the size of the convolution kernel is likewise $k \times k$, and the amount of floating-point computation required to use the Ghost module for feature extraction is:

$$Q = P \times k \times k \times m \times H \times W + (s-1) \times k \times k \times m \times H \times W \tag{20}$$

The resulting ratio r_b of ordinary convolution to Ghost computation and the ratio r_c of model parametric quantities are as follows:

$$r_b = \frac{P \times k \times k \times N \times H \times W}{k \times k \times m \times H \times W (P + s - 1)} = \frac{P \times s}{P + s - 1} \approx s \tag{21}$$

$$r_c = \frac{P \times k \times k \times N}{k \times k \times m (P + s - 1)} = \frac{P \times s}{P + s - 1} \approx s \tag{22}$$

From the above calculations, it can be seen that the Ghost module obtains the same number of feature maps as the traditional convolution module through a simple linear transformation, which compresses the computational and model parameter counts by a factor of s while ensuring that the model can learn enough features. In this paper, the original convolution is replaced by the Ghost module convolution in the YOLOv5s model. In order not to affect the feature extraction ability of the backbone network, the convolution operation in the Bottleneck structure of the C3 module on the backbone network is replaced by the Ghost convolution to obtain the C3Ghost module, and the rest of the convolution operations in the module are kept unchanged.

III. A. 2) ECA Attention Mechanism

The ECA attention mechanism is a lightweight and efficient channel attention mechanism designed to improve the performance of convolutional neural networks (CNNs) by rationally assigning channel weights.The design concept of ECA is to achieve the effect of the attention mechanism through simpler operations, avoiding the introduction of excessive parameters and computational complexity.The core idea of ECA is to model the interdependence between the channels by applying a one-dimensional pooling result of the global average of each channel's convolution to model the interdependence between channels. This process eliminates the need for a fully connected layer, thus reducing the introduction of parameters while preserving the interaction of global information [24].The specific computational flow of the ECA attention mechanism is as follows:

(1) Global average pooling. Global average pooling (GAP) is performed for each channel of the input feature map to obtain a scalar value for each channel, which represents the global information of that channel.

Suppose the input feature map is X and the dimension of the input feature map is $[C, H, W]$, where C is the number of channels, and H and W are the height and width of the feature map, respectively. After global average pooling, X_{GAP} is obtained and each of its elements is denoted as:

$$X_{GAP}(c) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} X(c, i, j) \tag{23}$$

(2) One-dimensional convolution. A one-dimensional convolution operation is performed on X_{GAP} to capture the interrelationships between channels.The ECA mechanism employs a convolution kernel of size $k = 5$, where the size of k is proportional to the number of channels to ensure the interaction effect across multiple channels. After the convolution operation, a new channel weight w is obtained, i.e.,:

$$w = Conv1D\left(X_{GAP}\right) \tag{24}$$

(3) Channel weight reassignment. The computed weights w are subjected to a channel-by-channel weighting operation with the original input feature map to obtain the final output feature map X', i.e:

$$X'(c, i, j) = w(c) \cdot X(c, i, j) \tag{25}$$

where c is the channel index, and i and j are the height and width indexes of the feature map, respectively.

By avoiding fully connected layers and adopting 1D convolution, ECA reduces the number of parameters and computational complexity, and is able to efficiently enhance the performance of various convolutional neural networks, especially in resource-limited scenarios. In addition, ECA has obvious advantages in enhancing the feature representation of important channels and improving the detection speed of YOLO models, which is especially suitable for real-time detection tasks.

III. A. 3) Sampling on CARAFE

The most widely used upsampling methods for YOLOv5s are nearest neighbor interpolation and bilinear interpolation. However, the two methods only consider the information in the neighborhood of the old image elements, which cannot capture the rich semantic information in the feature map, while the sensory field is also limited to 1×1 range.

Based on the above shortcomings, this section introduces the lightweight up-sampling operator CARAFE.This method uses a larger receptive field, which allows for a better extraction of the input feature map features without introducing excessive parameter counts and computation.CARAFE consists of two modules, namely, the feature reorganization module and the up-sampling kernel prediction module.

Assuming a given upsampling multiplicity of σ (let σ be an integer) and an input feature map χ of size $H \times W \times C$, CARAFE will produce a new feature map of size $\sigma H \times \sigma W \times \sigma C$. For any target location $I' = (i', j')$ in χ', there is a source location $I = (i, j)$ corresponding to it in the input feature map χ, where i and j can be expressed as:

$$i = \left\lfloor \frac{i'}{\sigma} \right\rfloor \tag{26}$$

$$j = \left\lfloor \frac{j'}{\sigma} \right\rfloor \tag{27}$$

$N = (\chi_1, k)$ denotes the $k \times k$ range of neighbors centered on I in χ.

The first step of the CARAFE operator process is the up-sampling kernel prediction module Ψ predicts the recombination feature $W_{I'}$ for each location based on the neighborhood of χ_1, and the recombination feature $W_{I'}$ is denoted as:

$$W_{1'} = \Psi(N(\chi_1, k_{encoder})) \tag{28}$$

The second step is to reorganize the features using the feature reorganization module Φ using the kernel $W_{I'}$ predicted in the previous step to output the feature map $\chi'_{1'}$. The output feature map $\chi'_{1'}$ is denoted as:

$$\chi'_{1'} = \Phi(N(\chi_1, k_{up}), w_{I'}) \tag{29}$$

where k_{up} represents the size of the kernel and k_{encoder} denotes the size of the kernel of the convolutional layer in the content encoder.

The up-sampling kernel prediction module consists of three modules: the channel compression module, the content encoder, and the normalization module. Firstly, there is the channel compression module, which performs 1×1 convolution operation on the input feature map with size $H \times W \times C$, and compresses the channel of the input feature map with size $H \times W \times C_m$, and then there is the content encoder, which uses the kernel size $k_{encoder}$ convolutional layer for the encoding operation, assuming that the number of channels of the input feature map is C_m, then the number of channels of the output feature map is $\sigma^2 k_{\text{up}}^2$. The output channel number is expanded in the spatial dimension to obtain the upsampled kernel of $\sigma H \times \sigma W \times k_{up}^2$, and finally the normalization module normalizes the upsampled kernel to make the convolutional weight 1.

In the feature reorganization module, the pixels of the output feature map are mapped back to the input feature map, and the corresponding region $N = (\chi_1, k_{\text{up}})$ region with $I = (i, j)$ as the centroid is taken out, and the part of the region is subjected to the dot-product operation with the up-sampling kernel in the up-sampling kernel prediction model. The operation formula is:

$$\chi_{1'} = \sum_{n=-r}^{r} \sum_{m=-r}^{r} W_{I'(n,m)} \cdot \chi_{(i+n, j+m)} \tag{30}$$

Finally a feature map χ' of size $\sigma H \times \sigma W \times C$ is obtained.

III. B. YOLOv5s-G-E-C Underwater Litter Detection Modeling

III. B. 1) YOLOv5s target detection network

YOLOv5s is the baseline model of underwater litter detection network proposed in this paper.YOLOv5s extends the model structure of previous YOLO series algorithms, which has a great improvement in performance and speed compared with previous YOLO series algorithms [25]. The model structure of YOLOv5s is shown in Fig. 2, which mainly consists of Input module, Backbone module, Neck Module, Head Module, whose functions are preprocessing the input image, feature extraction, feature fusion and prediction network are used to accomplish the target detection, respectively.

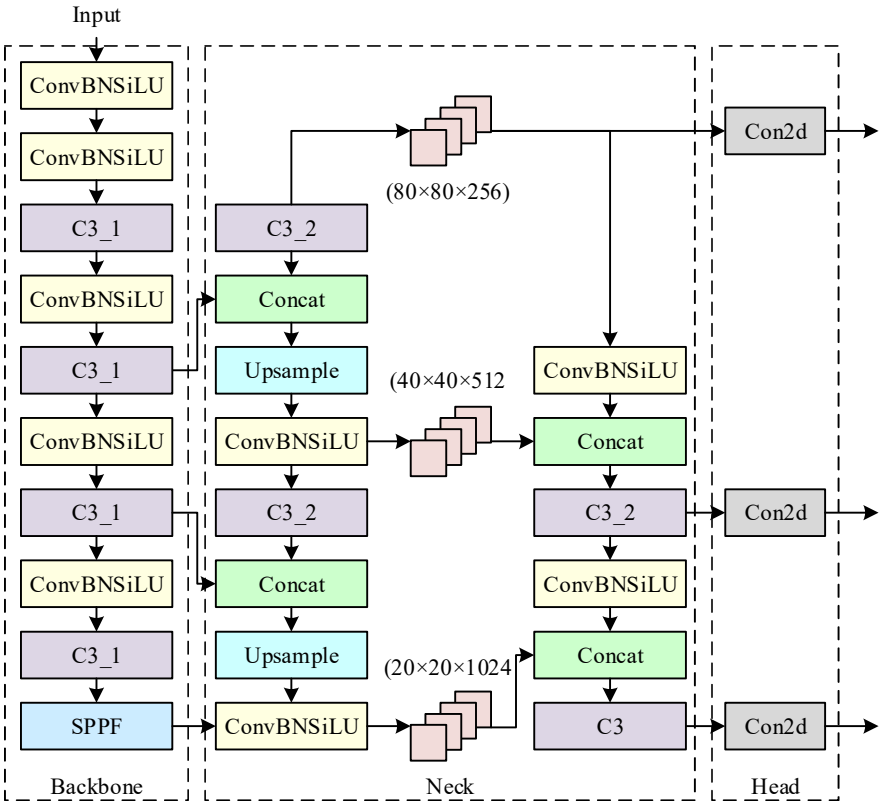

Figure 2: YOLOv5s model diagram

The composite loss function employed by the YOLOv5s model is disassembled into three key components, i.e., classification loss, confidence loss, and localization loss. The classification loss accurately evaluates the prediction accuracy of the object categories contained in each candidate box, and employs a binary cross-entropy (BCE) loss function mechanism to quantify the gap between the predicted probability distributions of the categories given by the model and the actual category labels. YOLOv5s supports multicategorical detection, and binary refers to the computation of a dichotomous classification problem independently for each category. The categorization loss is obtained by summing the binary cross-entropy loss over all categories and averaging or weighting the average according to the category weights to obtain the overall categorization loss.

The formula for L_{cls} is as follows:

$$BCE_{i,c} = -y_{i,c}\log(p_{i,c}) - (1-y_{i,c})\log(1-p_{i,c}) \tag{31}$$

$$L_{cls} = \frac{1}{N \cdot C}\sum_{i=1}^{N}\sum_{c=0}^{C} BCE_{i,c} \tag{32}$$

where i is the i th prediction frame, p_i represents the category prediction vector, $p_{i,c}$ represents the probability of belonging to category c within the model prediction frame i, C is the total number of categories, y_i represents the true labeling vector, and $y_{i,c}$ is a one-hot vector, for the true category c' with $y_{i,c'} = 1$ and the rest of the elements are 0. N is the total number of predictor frames involved in the computation of the loss, and $BCE_{i,c}$ is the binary cross-entropy loss of the i th predictor frame on category c.

The confidence loss is used to measure the confidence of the bounding box predicted by the model for the contained objects and the accuracy of the degree of matching between the bounding box and the real objects, and the binary cross entropy is also used in YOLOv5s, and the expression of the confidence loss is:

$$L_{obj} = -\sum_{i}\left[C_i \cdot \log\left(C_i\right) + \left(1-C_i\right)\cdot \log\left(1-C_i\right)\right] \tag{33}$$

where Σ_i denotes the summation of all prediction frames. For each prediction frame i, its confidence score C_i is the value predicted by the model and C_i is the corresponding true value.

The localization loss is responsible for calculating the error between the coordinates of the bounding box predicted by the

model and the actual labeled bounding box. The localization loss function used in YOLOv5s is CIoU Loss, and its specific function expression is:

$$L_{loc} = L_{CIoU} = 1 - IoU + \rho(IoU) + \alpha V \tag{34}$$

Among them, IoU is the standard intersection union ratio, $\rho(IoU)$ is the penalty item related to IoU, which is helpful to reduce the distance between the center points between the boxes and adjust the aspect ratio, V is the volume of the bounding box, that is, the sum of the area of the prediction box and the real box, and α is the parameter that controls the weight of the penalty items related to the aspect ratio.

Loss Loss consists of the above three components and the composition formula is:

$$Loss = \lambda_1 L_{cls} + \lambda_2 L_{obj} + \lambda_3 L_{loc} \tag{35}$$

where L_{cls} is the classification loss, L_{obj} is the confidence loss, and L_{loc} is the localization loss. λ_1, λ_2, and λ_3 are the weighting parameters used to balance the different loss terms.

III. B. 2) Improvements to the YOLOv5s network

The improved YOLOv5s network in this paper is called YOLOv5s-G-E-C. The network structure diagram of YOLOv5s-G-E-C is shown in Fig. 3. The aim of this paper is to reduce the computational and parametric quantities and to improve the model accuracy in order to make it more suitable for underwater embedded devices for underwater trash detection tasks.

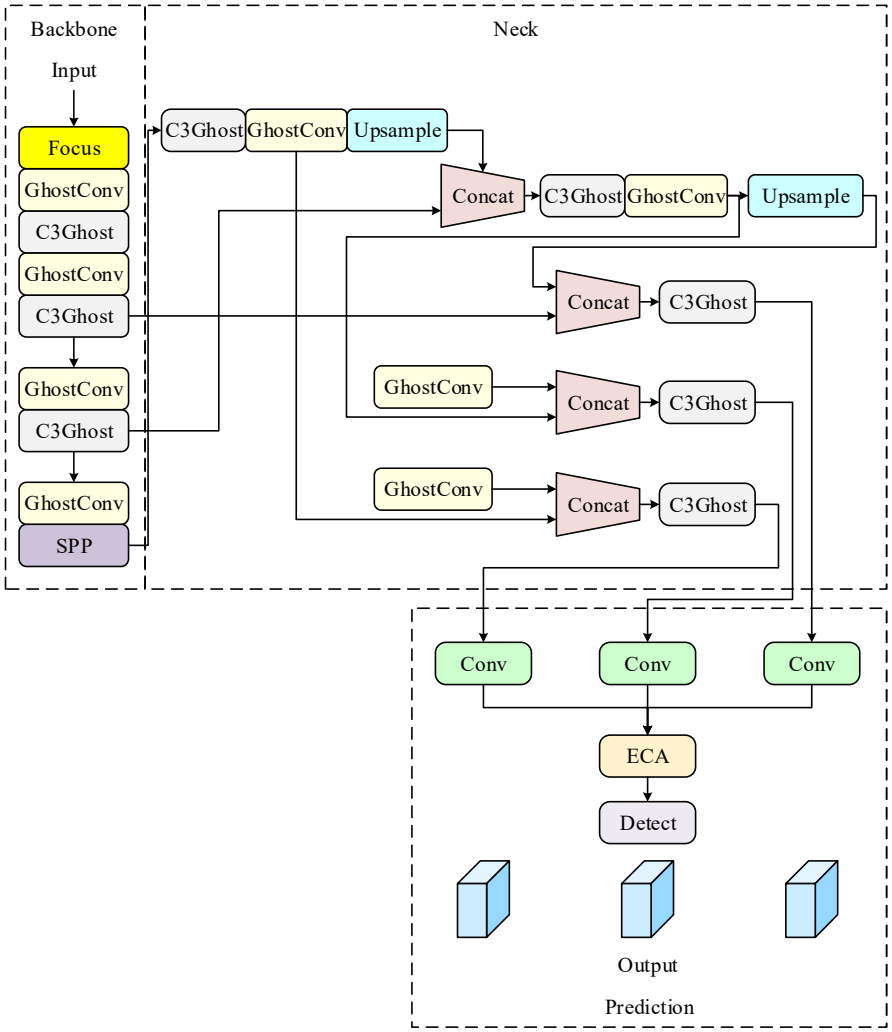

Figure 3: YOLOv5s-G-E-C network structure

First, the backbone and neck networks of YOLOv5s are lightened and improved using the Ghost network. Second, without

adding too much computational overhead, the ECA attention mechanism is added in front of the detection head of YOLOv5s to improve the network's ability to pay attention to important target regions, thus improving the accuracy of detection. Finally, the lightweight operator CARAFE is employed to improve the up-sampling method of the neck network in order to increase the model's perceptual ability to better utilize the global feature information and further improve the model's accuracy.

IV. Experimental results and analysis

Along with the country's overall rapid economic development, scientific and technological progress and the accelerating process of human development of water resources, the problem of underwater garbage pollution is becoming more and more serious. The flooding of underwater garbage will not only damage the underwater ecosystem, but also have an impact on the survival and health of human beings. Based on this, exploring and researching the construction of underwater garbage detection network is a necessary way to enhance the underwater garbage cleanup and protect the water resources environment and human health.

IV. A. Experiment on the effect of underwater image enhancement

IV. A. 1) Reference-free image quality assessment

In order to clarify the effectiveness of the underwater garbage image processing method, it is necessary to quantitatively assess the quality of the image results output by the algorithm, and the objective evaluation method will be used in this study to ensure the fairness and accuracy of the research results. The dataset of this paper comes from the marine debris dataset provided by JAMSTEC online, and the annotation is completed independently using free annotation tools, while using Google search to screen some of the marine plastic garbage images as the scene to expand the data, and ultimately obtain the marine underwater garbage image dataset UnderwaterTrash, that is, the experimental dataset of this paper.

When evaluating image quality, reference-free image quality assessment is a method that does not require the use of raw, unprocessed reference images, and the assessment is based on single or multiple image features, statistics, or models. Reference-free image quality assessment is often used to quantitatively assess image processing performance in situations where original image comparisons are not possible. Common reference-free image quality assessment methods include EAV point sharpness, Brenner gradient, Tenengrad gradient, Laplacian gradient, and energy gradient. These metrics do not require the original image as a reference, and can be combined with the analysis of the intrinsic properties of the image or the quantitative assessment of image quality using anthropomorphic visual perception models.

CLAHE, homomorphic filtering (HOMF), underwater dark channel a priori, and the algorithm of this paper are selected for underwater garbage image enhancement processing, and the results of reference-free image quality assessment are obtained as shown in Fig. 4. Analyzing the data changes in the figure, it can be seen that the method proposed in this paper has a significant effect on the enhancement of underwater garbage images, and the performance in EAV point sharpness, Brenner gradient, Tenengrad gradient and energy gradient presents a more stable advantage, but the performance in Laplacian gradient is slightly lower than that of the CLAHE algorithm.The CLAHE algorithm has a more significant effect on the de-fogging of the image with the The CLAHE algorithm has more obvious effect on image defogging and contrast improvement, but it is slightly inferior to other algorithms for the processing of dark details, and there exists a certain degree of color distortion phenomenon, so this paper proposes a weighted fusion of the image enhancement method for the application of underwater garbage image enhancement has practical value.

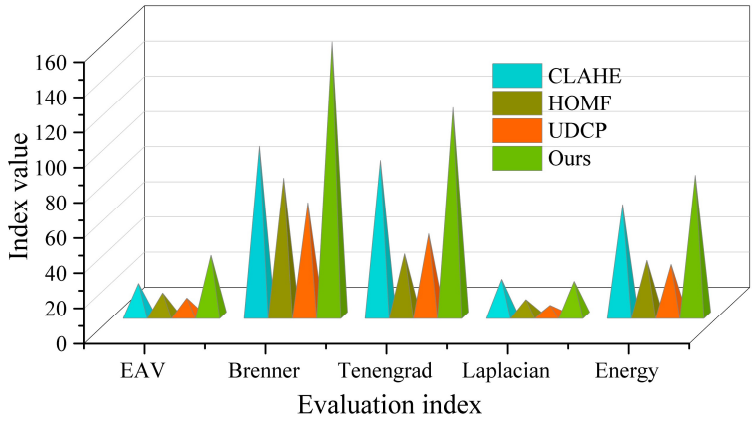

Figure 4: No-reference image quality evaluation results

IV. A. 2) Full reference image quality assessment

The full-reference image quality assessment method realizes the quantitative evaluation of image quality by analyzing the differences between the image to be assessed and the original image against each other, which is an objective, quantitative and widely used method that can meet the demand for accurate measurement of image quality. Peak signal-to-noise ratio (PSNR), structural similarity (SSIM), information entropy (IE), and average gradient (EG) are chosen as the evaluation indexes to evaluate the quality of the results of the underwater garbage image enhancement processing for CLAHE, MSR, MCMSR, LIME, MF, UDCP, and the algorithm in this paper.

The peak signal-to-noise ratio can reflect the ratio of signal to noise in the image, and the higher its value, the smaller the image distortion and the higher the quality of the enhanced image. Structural similarity is used to describe the similarity between the enhanced image and the source image, the larger the value, the smaller the deviation between the two. Information entropy describes the amount of information in the image, the higher the entropy value, the more information is in the image and the more rich and diverse the content is. The average gradient, on the other hand, can measure the detailed characteristics such as image edges and texture features, describing the clarity of the image. Table 1 shows the results of the full reference image quality evaluation, where the optimal values are bolded in the table and the sub-optimal values are underlined in the table.

As can be seen from the table, the weighted fusion-based underwater spam image enhancement algorithm in this paper performs optimally on both PSNR and SSIM values, with values of 17.218 dB and 0.619, respectively, which are 6.82% and 25.81% higher compared to the sub-optimal performance of the UDCP algorithm and MSR algorithm, respectively, and the performances of IE and MG are only lower than the values of the LIMF algorithm. Overall, the weighted fusion-based underwater trash image enhancement algorithm has less distortion and the enhanced underwater trash image is clear and natural.

Table 1: Full-reference image quality evaluation results

Algorithm	PSNR/dB	SSIM	IE	MG
CLAHE	11.365	0.318	7.152	8.834
MSR	15.948	0.492	6.938	8.316
MCMSR	14.492	0.431	7.073	9.252
LIME	15.483	0.453	**7.218**	**12.465**
MF	14.751	0.457	6.953	11.503
UDCP	16.119	0.474	7.065	11.217
Ours	**17.218**	**0.619**	7.209	12.379

In summary, compared with the classical image enhancement algorithms such as CLHAE, MSR and MSRCR, the algorithm proposed in this paper shows more excellent results. Even compared with the newly proposed LIME algorithm and MF algorithm in recent years, the algorithm in this paper also has its own advantages. However, it is worth noting that the LIME algorithm and MF algorithm have higher computational complexity, for example, the LIME algorithm uses a weighted least squares filter, which involves a large sparse matrix, if the resolution of an image is 1280×1024, the size of this matrix reaches a staggering 1310720×1310720, and an inverse operation is required for this matrix. For hardware platforms, this operation is cumbersome to implement. Similarly, the MF algorithm uses Laplace pyramid and Gaussian pyramid, which are still computationally intensive and difficult to implement on hardware platforms. In contrast, the weighted fusion-based underwater image enhancement algorithm in this paper not only enhances significantly, but also has low algorithmic complexity and is easier to implement in hardware. Overall, the algorithm in this paper does not show obvious color distortion in underwater garbage image enhancement, significantly improves the overall brightness and contrast of underwater garbage images, and is relatively easy to implement on hardware platforms such as FPGA.

IV. B. Validation of the underwater litter detection model

In the UnderwaterTrash experimental dataset established in this paper, it mainly consists of 6,840 underwater images in real environments, including all kinds of garbage, underwater robots, and a variety of undersea plants and animals. Before the experiment, the training set and test set, which contain 5800 and 1040 images respectively, are randomly divided by hand.

The experiments in this paper are built under Ubuntu LTS system using PyTorch deep learning framework. The computer hardware configuration includes NVIDIA GeForce RTX 4090Ti 32GB GPU, AMD Ryzen 95950X 16-Core CPU and 32 GB of running memory. The hyperparameters for training were set to 200 for Epoch, 64 for Batch size, 0.976 for Momentum, 0.0002 for Weight Decay Coefficient, and 0.0001 for Initial Learning Rate.

IV. B. 1) Comparative experiments before and after model improvement

In order to realize the effective detection of underwater garbage, this paper improves the original YOLOv5s network by

GhostNet network, ECA attention mechanism and CARAFE up-sampling mechanism. In order to verify the effectiveness of the improved YOLOv5s model, this paper conducts comparison experiments before and after the model improvement. The original YOLOv5s model is first trained on the UnderwaterTrash dataset using the original YOLOv5s model. Considering the hardware conditions and the small capacity of the dataset, the number of training layers is set to 300. The number of batches for data enhancement is set to 5, the learning rate is set to 0.001, and the input image size is 640×640.The YOLOv5s-G-E-C model is then trained with the hyper-parameters set in the previous section, and the loss value of the model is used as the evaluation index, and Fig. 5 shows the trend of the loss curve before and after the model improvement.

Due to the small capacity of the UnderwaterTrash dataset established in this paper, the number of training layers is set to 300, and it can be seen from the curve changes that the curves of the two models before and after the improvement tend to flatten out after 250 layers of training, which proves that the two models are close to convergence at this time, and neither one of them overfitted or underfitted for the UnderwaterTrash dataset. Both models do not over- or under-fit the UnderwaterTrash dataset. The loss curve of the YOLOv5s-G-E-C model decreases faster at the beginning of the training period, indicating that it converges faster. After more than 120 rounds of training the improved YOLOv5s-G-E-C model in this paper tends to converge, as can be seen from the curves, the loss function of the network for the YOLOv5s-G-E-C model is smaller. From the curve situation, it can be seen that for the UnderwaterTrash dataset established in this paper, the improved YOLOv5s network fits better.

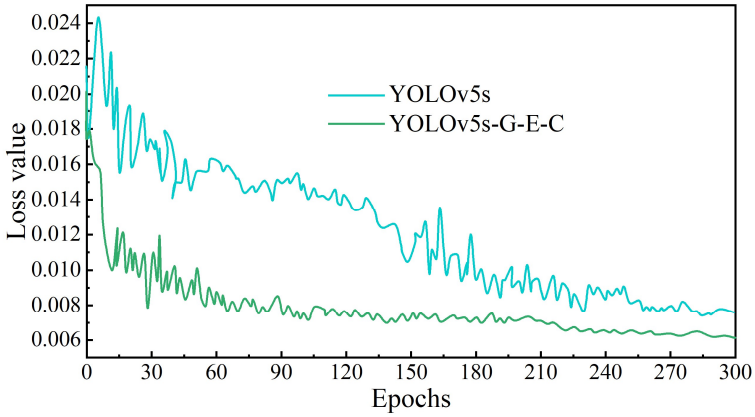

Figure 5: Comparison chart of loss value functions

IV. B. 2) Comparison results of the performance of different models

In order to further illustrate the performance of the YOLOv5s-G-E-C model proposed in this paper in performing underwater trash detection, YOLOv5s, YOLOv5s+CLAHE, YOLOv5s+SPPF, YOLOv5s+CLAHE+SPPF, YOLOv5s+SPPF+ST, YOLOv5s+CLAHE+SPPF +ST with the YOLOv5s-G-E-C model of this paper for comparison experiments, selecting precision (P), average precision (Pm), recall (R) and F1 as evaluation indexes, and obtaining the performance comparison results of different models as shown in Table 2.

It can be seen that the improved YOLOv5s-G-E-C model in this paper has improved the detection ability of each target species compared to the YOLOv5s algorithm. Among them, YOLOv5s+CLAHE improves the recognition ability of the algorithm by processing the input image with the CLAHE algorithm, which leads to an increase in the average precision by 8.13%, the recall by 0.19%, and the F1 value by 4.12%.YOLOv5s+SPPF enhances the model's texture feature extraction ability by introducing the SPPF module, which leads to an increase in the algorithm's average precision for the This dataset has improved the average precision by 3.87%, recall decreased by 1.27%, and F1 value increased by 1.49%.YOLOv5s+CLAHE+SPPF enhances the image features by adding the CLAHE algorithm after the introduction of the SPPF module, which makes the model improve the average precision and the F1 value for this dataset compared to YOLOv5s+CLAHE YOLOv5s+SPPF+ST enhances the generalization ability of the model and the detection ability of small targets by introducing the SPPF module while replacing the CSP2_X module with the ST module, which makes the algorithm's average precision and F1 value also improved compared to YOLOv5s+SPPF, and the recall reduced.YOLOv5s+CLAHE+ SPPF+ST achieves an increase in average precision, recall and F1 value for this dataset by 9.99%, 2.08%, and 5.53%, respectively, compared to YOLOv5s+SPPF+ST. The YOLOv5s-G-E-C model combines the above features with a modification of the neck network portion of the model, and achieves an increase in average precision, recall, and F1 value of 86.19%, 69.52%, and 76.96%, compared to the second best performing YOLOv5s+CLAHE+SPPF+ST model, for which the average precision, recall, and F1 value are all improved for this dataset, which fully demonstrates the feasibility of this paper's model for underwater garbage detection as well as its superior performance compared to existing related models.

Table 2: Improved YOLOv5s and YOLOv5s experimental results (%)

Model	P			Pm	R	*F*1
	Bottle	Plastic	Other			
YOLOv5s	61.45	61.02	60.28	60.92	68.24	64.37
YOLOv5s+CLAHE	73.28	73.45	72.41	73.05	68.43	70.66
YOLOv5s+SPPF	63.86	68.17	62.34	64.79	66.97	65.86
YOLOv5s+CLAHE+SPPF	74.37	76.31	72.19	74.29	67.15	70.54
YOLOv5s+SPPF+ST	69.52	74.68	71.65	71.95	66.08	68.89
YOLOv5s+CLAHE+SPPF+ST	82.03	84.26	79.52	81.94	68.16	74.42
YOLOv5s-G-E-C	85.79	88.83	83.96	86.19	69.52	76.96

On this basis, this paper further verifies the running effect of the model by Pm@0.5, Pm@0.5:0.95 and detection speed. When Pm@0.5 is the IoU threshold and 0.5 is taken, for one of the categories, there is a positive sample, and the accuracy of this sample is averaged. Pm@0.5:0.95 indicates that the IoU threshold of average accuracy is increased from 0.5 in steps of 0.05 to 0.95 and the average value is taken. The specific results are shown in Table 3.

As can be seen from the table, compared with the YOLOv5s model, the YOLOv5s-G-E-C model designed in this paper improves the two indicators of Pm@0.5 and Pm@0.5:0.95 by 14.89% and 9.45%, respectively, and reduces the detection speed by 17.86 frames/s. Compared with the second-best YOLOv5s CLAHE SPPF ST model, the proposed model improves the Pm@0.5 and Pm@0.5:0.95 by 0.88% and 1.78%, respectively, and the detection speed is reduced by 2.24 frames/s. Although the detection speed is reduced, it has little impact on the real-time detection performance of underwater garbage. Therefore, the detection of underwater garbage by using the enhanced image of underwater garbage after weighted fusion can significantly improve the problem of reducing the feature extraction ability caused by different color distortion, so as to effectively improve the detection accuracy of the model for underwater garbage.

Table 3: Improved YOLOv5s and YOLOv5s experimental results-2

Model	Pm@0.5/%	Pm@0.5:0.95/%	FPS/Frame·s^{-1}
YOLOv5s	60.18	32.18	166.65
YOLOv5s+CLAHE	70.26	36.04	138.42
YOLOv5s+SPPF	68.15	35.37	247.57
YOLOv5s+CLAHE+SPPF	71.24	38.42	239.31
YOLOv5s+SPPF+ST	68.93	37.96	156.84
YOLOv5s+CLAHE+SPPF+ST	74.19	39.85	151.03
YOLOv5s-G-E-C	75.07	41.63	148.79

IV. B. 3) Backbone Network Accuracy Impact Analysis

In the experimental part of this section, in order to verify the effect of the backbone network on the detection accuracy, by fixing the feature fusion part and the prediction part, selecting different backbone networks for feature extraction, and comparing the detection algorithms of different backbone networks in various aspects, such as the number of parameters of the algorithms (Param), the amount of computation (FLOPs), the speed of detection (FPS), and the mean of average accuracy (mAP), etc., to Verify the advantages of this paper's backbone network over other backbone networks in several aspects. Table 4 shows the comparison results of target detection models of different backbone networks.

From the table, it can be seen that YOLOv5s based on GhostNet achieves the best results in terms of the number of parameters and the amount of computation compared with YOLOv5s based on other backbone networks. In terms of computational speed, ResNet36 based on residual network structure has the fastest detection speed, and VGG18 has a more desirable detection speed due to the removal of the three fully-connected layers, but suffers from the problem of excessive number of parameters and computational volume.DenseNet120 achieves a more desirable result in terms of the number of parameters and computational volume due to the advantage of the densely-connected network structure, but DenseNet The increase in the number of explicit memory accesses due to the more feature multiplexed nature makes the algorithm the slowest in terms of detection speed. YOLOv5s based on DarkNet52 achieves more balanced results in terms of the number of parameters, computation volume, and detection speed. While the YOLOv5s network incorporating GhostNet is slower in detection speed compared to DarkNet52 and ResNet36 based on residual network, it also meets the demand of real-time detection.

In terms of detection accuracy, although GhostNet reduces the number of convolutional blocks in the network connection compared to DenseNet120, the improvement in target detection accuracy proves the advantage of using GhostNet network

with it, and also shows that the higher number of convolutional blocks of DenseNet has some feature redundancy for target detection in small datasets.

Taken together, GhostNet achieves the optimal results in terms of the number of parameters, computational volume, and detection accuracy when retaining the same feature fusion module and prediction module, and also proves that the feature extraction network designed in this paper is well suited for the classification and localization tasks of small sample datasets, and that it fully meets the real-time demands of underwater garbage detection, even though there is still the problem of slower detection speed.

Table 4: Comparison of object detection models in Different backbone networks

Backbone	Param/M	FLOPs/G	FPS/Frame·s^{-1}	mAP/%
DarkNet52	61.45	155.14	33.25	90.75
DenseNet120	27.68	85.32	19.82	90.18
VGG18	40.37	356.78	35.63	89.27
ResNet36	46.52	97.85	**46.74**	90.53
GhostNet	**22.06**	**62.27**	32.18	**92.49**

IV. B. 4) Analysis of model ablation experiment results

In order to realize the accurate detection of underwater garbage, this paper introduces GhostNet, ECA attention mechanism and CARAFE up-sampling mechanism on YOLOv5s network as a way to improve the underwater garbage detection capability of YOLOv5s network. For the improvements made to the YOLOv5s network, a series of ablation experiments are conducted to verify the effectiveness of each part of the improvements, and the results are shown in Table 5.

The results show that for all the improvements used, the accuracy of the model is improved to different degrees, among which the GhostNet convolution method plays a very significant role in reducing the model parameters and computation. While the model using ECA attention mechanism has the highest accuracy without leading to an increase in model parameters and computation. The use of CARAFE upsampling mechanism also improves the model accuracy after preprocessing the data. After using the above three improvement methods, the final improved YOLOv5s-G-E-C model improved its accuracy and mAP by 6.29% and 6.26%, respectively, compared to the YOLOv5s model, and the size of the model weights was only 13.98 MB, which was reduced by 0.54 MB, and the computational complexity of the GFLOPs was reduced by 264 GFLOPs. The experimental results showed that the improved model has achieved good detection results.

Table 5: Improve the ablation experiment of the module

Model	P/%	R/%	mAP/%	GLOPs/G	Weight/MB
YOLOv5s	89.14	78.24	80.15	18.42	14.52
YOLOv5s+GhostNet	94.27	76.35	84.23	16.16	13.68
YOLOv5s+ECA	96.82	78.27	85.36	15.84	14.52
YOLOv5s+CARAFE	96.05	79.63	84.72	15.41	14.73
YOLOv5s-G-E-C	95.43	80.41	86.41	15.78	13.98

V. Conclusion

In this paper, we study an underwater garbage detection and recognition model based on YOLOv5s-G-E-C, and propose an underwater image enhancement algorithm based on weighted fusion for the problem of low underwater image quality. The experimental results show that the enhancement algorithm in this paper has a better overall adjustment effect on underwater garbage images, and the objective indexes are significantly higher than other comparison algorithms. For underwater garbage detection, the improved YOLOv5s-G-E-C model has a higher detection accuracy on the underwater garbage dataset, possesses better comprehensive performance, and meets the demand of real-time detection.

Future work will focus on further improving the detection speed of the YOLOv5s-G-E-C model, fully refining the comprehensive performance of the YOLOv5s-G-E-C model, and combining it with the practical application of underwater litter detection.

References

[1] Huang, C., Zhang, W., Zheng, B., Li, J., Xie, B., Nan, R., ... & Xiong, N. N. (2025). YOLO-MES: An Effective Lightweight Underwater Garbage Detection Scheme for Marine Ecosystems. IEEE Access.

[2] Long, Z. (2023). Begin ocean garbage cleanup immediately. Science, 381(6658), 612-613.

[3] Ozoh, A. N., Longe, B. T., Akpe, V., & Cock, I. E. (2021). Indiscriminate solid waste disposal and problems with water-polluted urban cities in Africa. Journal of Coastal Zone Management, 24(S5), 1000005.

[4] Goyal, V., & Dharwal, M. (2022). The puzzle of garbage disposal in India. Materials Today: Proceedings, 60, 926–929.
[5] Tian, M., Li, X., Kong, S., Wu, L., & Yu, J. (2022). A modified YOLOv4 detection method for a vision-based underwater garbage cleaning robot. Frontiers of Information Technology & Electronic Engineering, 23(8), 1217–1228.
[6] Chitra, L., Poornima, P., Veeramuneeswaran, P., Manobala, M. K., Karthikeyan, D., & Chauhan, S. K. (2024, February). Garbage Robot: A Solution for Waste Disposal and Sustainable Bio Product Management. In 2024 2nd International Conference on Computer, Communication and Control (IC4) (pp. 1–5). IEEE.
[7] Kong, S., Tian, M., Qiu, C., Wu, Z., & Yu, J. (2020). IWSCR: An intelligent water surface cleaner robot for collecting floating garbage. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 51(10), 6358–6368.
[8] Li, X., Tian, M., Kong, S., Wu, L., & Yu, J. (2020). A modified YOLOv3 detection method for vision-based water surface garbage capture robot. International Journal of Advanced Robotic Systems, 17(3), 1729881420932715.
[9] Deng, H., Ergu, D., Liu, F., Ma, B., & Cai, Y. (2021). An embeddable algorithm for automatic garbage detection based on complex marine environment. Sensors, 21(19), 6391.
[10] Chen, L., & Zhu, J. (2024). Water surface garbage detection based on lightweight YOLOv5. Scientific Reports, 14(1), 6133.
[11] Khriss, A., Elmiad, A. K., Badaoui, M., Barkaoui, A. E., & Zarhloule, Y. (2024). Exploring deep learning for underwater plastic debris detection and monitoring. Journal of Ecological Engineering, 25(7).
[12] Liu, X., Xiong, H., Gao, M., & Liu, W. (2025, January). Lightweight underwater garbage target detection algorithm based on improved YOLOv7-Tiny. In Fourth International Conference on Computer Vision, Application, and Algorithm (CVAA 2024) (Vol. 13486, pp. 447–457). SPIE.
[13] Jiang, L., Liu, F., Lv, J., Liu, B., & Wang, C. (2024). GST-YOLO: a lightweight visual detection algorithm for underwater garbage detection. Journal of Real-Time Image Processing, 21(4), 114.
[14] Guan, J., & Guo, B. (2024). An improved YOLOv5 algorithm for underwater garbage recognition. Journal of Computational Methods in Science and Engineering, 14727978241295566.
[15] Demir, K., & Yaman, O. (2024). Projector deep feature extraction-based garbage image classification model using underwater images. Multimedia Tools and Applications, 83(33), 79437–79451.
[16] Walia, J. S., & Seemakurthy, K. (2023, September). Optimized custom dataset for efficient detection of underwater trash. In Annual Conference Towards Autonomous Robotic Systems (pp. 292–303). Cham: Springer Nature Switzerland.
[17] Watanabe, J. I., Shao, Y., & Miura, N. (2019). Underwater and airborne monitoring of marine ecosystems and debris. Journal of Applied Remote Sensing, 13(4), 044509.
[18] Wu, G., Ge, Y., & Yang, Q. (2023). UTD-YOLO: underwater trash detection model based on improved YOLOv5. Journal of Electronic Imaging, 32(6), 063034-063034.
[19] Cheng, K., Yan, L., Ding, Y., Zhou, H., Li, M., & abdul Ghafoor, H. (2023). Sonar image garbage detection via global despeckling and dynamic attention graph optimization. Neurocomputing, 529, 152–165.
[20] Zong Tianyang & Yang Lei. (2024). Low illumination image enhancement based on improved CLAHE algorithm. Journal of Physics: Conference Series,2891(11),112029-112029.
[21] Sun Zhengping,Li Fubing,Chen Wenjian & Wu Mianxing. (2021). Underwater image processing method based on red channel prior and Retinex algorithm. OPTICAL ENGINEERING,60(9).
[22] Cao Ning,Lyu Shuqiang,Hou Miaole,Wang Wanfu,Gao Zhenhua,Shaker Ahmed & Dong Youqiang. (2021). Restoration method of sootiness mural images based on dark channel prior and Retinex by bilateral filter. Heritage Science,9(1).
[23] Tamminina Ammannamma & A S N Chakravarthy. (2025). A bio-inspired optimal feature with convolutional GhostNet based squeeze excited deep-scale capsule network for intrusion detection. Computers & Security,150,104221-104221.
[24] Xuhao Shi,Jie Zhang,Guoqiang Liu,Kun Yi & Muhammad Bilal. (2025). Self-attention based cloud top height retrieval for intelligent meteorological service recommendation. Information Sciences,713,122192-122192.
[25] Luo Liu,Jinxin Chen,Qi an Ding,Ruqian Zhao,Mingxia Shen & Longshen Liu. (2025). Detection and analysis of sow nursing behavior based on the number and location of piglets outside the suckling area using YOLOv5s. Computers and Electronics in Agriculture,235,110324-110324.