RESEARCH ARTICLE

Underwater Image Enhancement Method Based on Feature Fusion Neural Network

YUAN TIAN[1], YUANG XU[1], AND JUN ZHOU[2,3]

[1]School of Mechanical, Electrical and Information Engineering, Shandong University, Weihai, Shandong 264209, China
[2]No.59 Institute of China Ordnance Industry, Chongqing 401331, China
[3]Chongqing Research Institute of Business Economics, Chongqing 401331, China

Corresponding author: Jun Zhou (hgzhou2008@163.com)

ABSTRACT Aiming at the problems of uneven illumination of underwater image caused by supplementary illumination in deep-sea and night waters, image noise, low contrast and color deviation caused by suspended particles in water, a new underwater image enhancement method under non-uniform illumination is proposed. The heterogeneous feature fusion module is designed to fuse different levels and different levels of features, so as to improve the overall perception ability of the network to detail information and semantic information. Secondly, a new feature attention mechanism is designed to improve the traditional channel attention mechanism, and the improved channel attention and pixel attention mechanism are added to the heterogeneous feature fusion process to strengthen the ability of the network to extract pixel features with different turbidity. Then, the dynamic feature enhancement module is designed to adaptively expand the receptive field to improve the adaptability of the network to the image distortion scene and the ability of model conversion, and strengthen the network's learning of the region of interest. Finally, the color loss function is designed, and the absolute error loss and structural similarity loss are jointly minimized to correct the color deviation on the basis of maintaining the image texture. A multi-scale feature extraction module is designed to extract different levels of features at the beginning of the network, and the output results are obtained through the convolution layer with jump connection and the attention module. The experimental results on several data sets show that this method can have good results in processing synthetic underwater images and real underwater images, and can better restore the image color and texture details compared with the existing methods. It conforms to the characteristics of human vision, and the visual effect is better than the existing underwater image enhancement algorithms.

INDEX TERMS Deep learning, neural network, attention mechanism, underwater image enhancement.

I. INTRODUCTION

The complex underwater environment will lead to the problems of color deviation, haze effect and low contrast of the image, which seriously hinders the robot's visual perception of the underwater scene [1]. Therefore, it is necessary to study the underwater image enhancement algorithm to improve the performance of underwater robot performing advanced

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

vision tasks. Underwater image enhancement can be roughly divided into traditional enhancement methods and enhancement methods based on deep learning. Traditional methods can be roughly divided into non physical model-based methods and physical model-based image enhancement methods. Methods based on non physical models, such as histogram equalization [2], gray world hypothesis [3], wavelet transform [4], automatic white balance [5], improve the quality of underwater image by directly adjusting the image pixel value. However, due to the lack of physical model, it is easy to

cause problems such as large image color deviation and image artifacts. The methods based on physical model include the restoration method based on image defogging model Jaffe mcglamery [6], [7], such as dark channel prior [8], [9], fuzzy prior [10], [12], etc. The method based on physical model depends on the assumption of a priori underwater physical formation image model, the process of calculating features is very complex, and the a priori assumption is ideal and has great limitations. It can not well realize the color correction of underwater image and remove artificial light spot.

The characteristics of color deviation and fuzzy boundary in marine biological image lead to the poor effect of existing image detection algorithms for marine biological image detection. Researchers have proposed a variety of improved algorithms. Ping *et al.* [13] added the trained underwater image enhancement GAN generation network to the YOLOv3 detection network as its enhancement network, so that the enhancement part of the network is more conducive to the improvement of marine biometric recognition accuracy. Kun *et al.* [14] applied the automatic color equalization algorithm to the image enhanced by the multi-scale retina with color restoration algorithm to optimize the image color and brightness. At the same time, they replaced the residual module in the YOLOv3 feature extraction network with dense blocks to effectively strengthen the feature propagation, But this method makes the detection accuracy not very high. Song *et al.* [14] proposed a method for detecting underwater. This method realizes high-precision detection of small sample data sets by combining image enhancement algorithm with mask RCNN, but the disadvantage is that the model detection speed is low and the practicability is poor, and this method is not suitable for long-distance underwater target detection. Wang *et al.* [15] proposed an improved image enhancement algorithm based on adaptive algorithm to clearly process the collected underwater degraded image and detect marine or GANisms using YOLOv3, but the detection accuracy of sea cucumber and scallop is not high. The existing algorithms still can not achieve high detection accuracy in marine biological image detection, and there is still room for improvement.

Common image displacement detection methods include template matching algorithm based on gray scale, such as Normalized Cross Correlation [16], sum of absolute difference algorithm, etc. But only whole pixel displacement can be detected; Matching algorithms based on feature points, such as accelerated up robust features (surf), scale invariant feature transform (SIFT), etc. [17] The above methods can be classified as spatial detection methods. This kind of algorithm not only has low detection accuracy, but also the gray information in spatial domain is vulnerable to noise, which affects the matching accuracy Methods based on phase information in frequency domain include phase correlation algorithm (PCA) [18], phase difference algorithm [19]. Frequency domain information has better anti-interference ability than spatial domain the underwater image enhancement algorithm does not consider the physical model in the imaging principle, and

TABLE 1. Summarizing the advantages/limitations of different models.

Method	advantages	limitations
Based on physical model	Fewer parameters	Low accuracy
Mask RCNN	Fewer parameters	Low accuracy
MSRCP	Fewer parameters	Low accuracy
Apriori principle	High accuracy	High complexity
Airspace detection	High accuracy	High complexity

achieves the purpose of image enhancement by processing some mathematical models in the spatial domain or frequency domain. He *et al.* [20] calculated the atmospheric light component A and transmittance T through a priori principle, and obtained the target value to realize the defogging processing of the image. The underwater image is similar to the fog image. This model is also often used for underwater image enhancement, but it does not perform well when processing the image with low exposure. After processing, the overall color of the image is black and the background is blurred. Ancuti *et al.* [21] processed the images with white balance and sharpening, and adopted the multi-scale fusion strategy to fuse the images with white balance and sharpening, so as to improve the global contrast and edge definition. However, it has limitations in processing deep scenes and failed to eliminate the blue-green background. Peng *et al.* [22] used the underwater scene depth estimation based on image blur and light absorption for the image formation model to restore and enhance the underwater image. However, this method only considers the scene depth and background light and ignores the noise in the water, which will aggravate the image noise in the restoration process. Song *et al.* [23] proposed a backlight database based on manual annotation, which processes the underwater image according to the relationship between the background light and the histogram distribution of the underwater image, but the background light is too compensated, the processed underwater image is overexposed, the background is blurred, and the details in the image are also covered by the white fog caused by the over compensated background light. The above algorithm is only improved for individual cases in underwater environment, and its practical application has limitations. The evolving fusion based visibility restoration model has outstanding advantages in the overall enhancement effect, but it does not meet the expectations in the details of individual underwater blurred images [24].

To sum up, the deblurring effect of the traditional network model has been improved, but the training efficiency is low, as Table 1. When restoring the blurred image with low resolution, the image will still have the problems of artifacts and untrue image details, which seriously affects the quality of the restored image and cannot meet the subsequent detection requirements. In order to improve the illumination unevenness, blur and color distortion of underwater image,
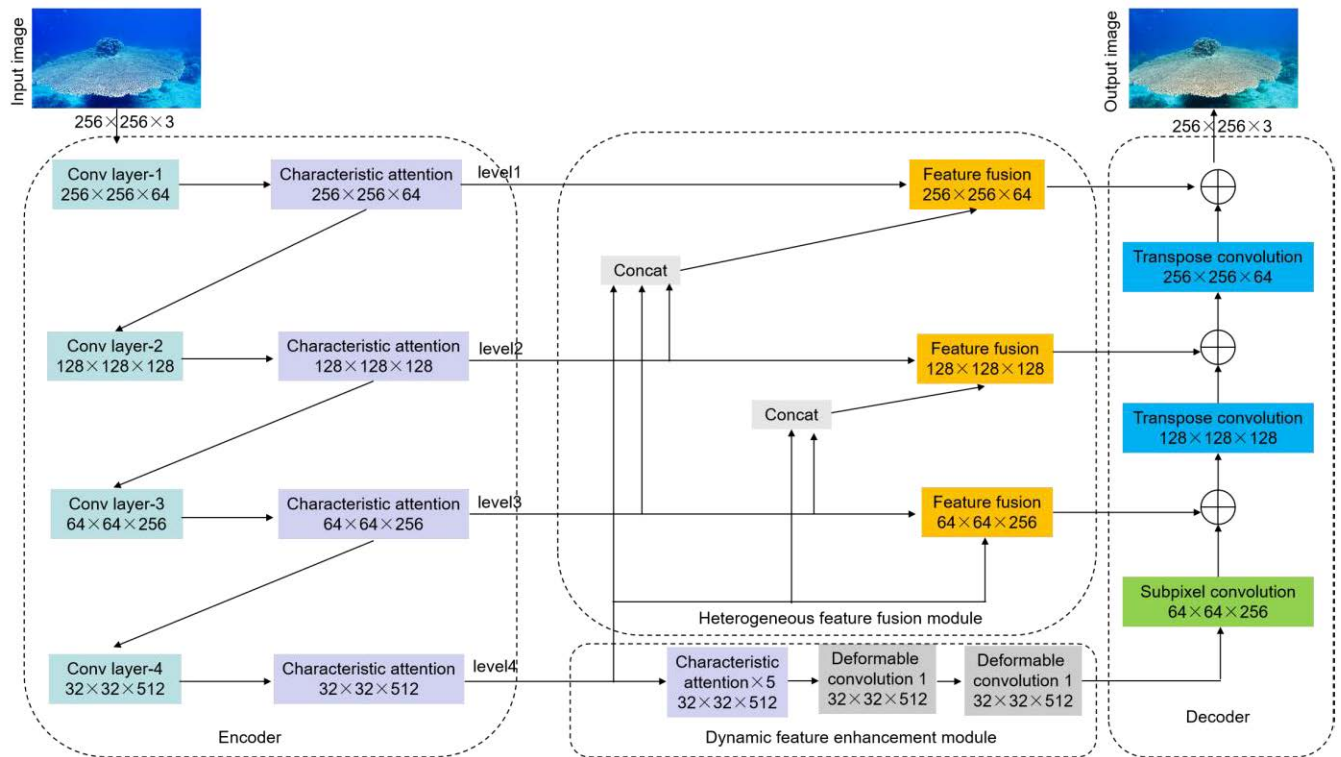
FIGURE 1. Underwater image enhancement network structure.

an underwater enhancement algorithm based on guided filter and adaptive operator is proposed. Firstly, according to the principle of rapid attenuation of visible light and red light in underwater environment, an adaptive color correction operator is designed for RGB three channels to restore the real color of underwater image, make the image more color contrast and enhance the appreciation of underwater image. Then put the image into the model combined with guided filtering and Retinex model, and use the characteristics of guided filtering to smooth the underwater image while retaining the image edge information and enhance the definition and detail information of the underwater image. Finally, the image weight is calculated and multi-scale fusion is carried out according to the image weight to avoid the artifact phenomenon of general image fusion. Experiments show that compared with the existing algorithms, this algorithm can better restore the color of the image in the atmosphere, has better numerical performance, and has more image detail information. Improve the low exposure of dark areas and enhance the overall contrast of the image. And the above GAN based image enhancement algorithm directly learns the degradation

II. METHOD
A. DYNAMIC HETEROGENEOUS FEATURE FUSION NETWORK

In order to solve the problems of uneven enhancement effect and loss of image details in existing methods,

an end-to-end heterogeneous feature fusion and dynamic feature enhancement underwater image restoration network is designed in this paper. The network structure is shown in Figure 1. Based on the encoder decoder structure, feature attention module, dynamic feature enhancement module and heterogeneous feature fusion module are integrated. The improved attention mechanism is added to the encoder structure to gradually extract different types of features from low level to high level; The dynamic feature enhancement module enhances feature extraction in low resolution space; The decoder uses the up sampling operation to reconstruct the feature vector and gradually restore it to the corresponding clear image. With the deepening of the network layer, the shallow features gradually degenerate, and the feature fusion module is used to integrate the low-level and high-level information.

Row fusion can improve the information transmission between layers and reduce the loss of detailed features. Among them, the encoder is composed of 8x down sampling operation and improved feature attention mechanism. 8x down sampling consists of a standard 7 with a step of 1×7 convolution and 3 standard convolutions with 2 steps, and the decoder uses transpose convolution as the corresponding up sampling method to gradually restore the image resolution.

Because most encoder decoder structures treat the whole image indiscriminately in the process of image processing, when the image information is large, more irrelevant features
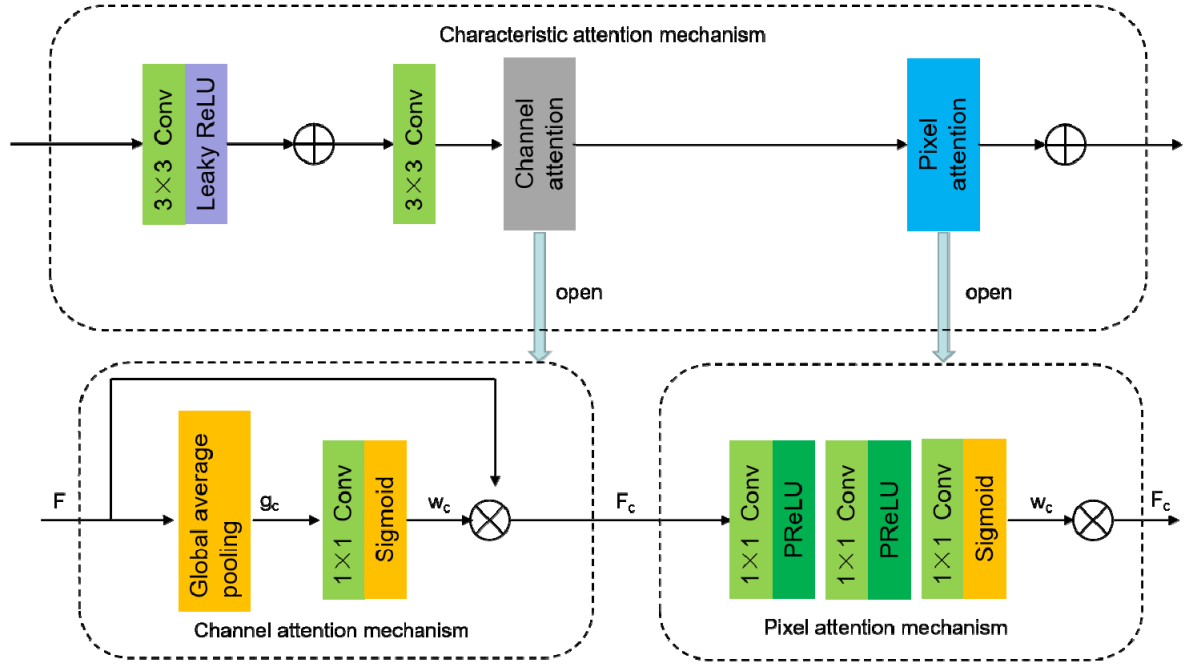
FIGURE 2. Feature fusion backbone module combined with attention mechanism.

are extracted, which is difficult to adapt to the complex and changeable underwater environment. Therefore, attention mechanism needs to be added to improve the accuracy of feature extraction of the model. In order to solve the above problems, this paper improves the channel attention mechanism and pixel attention mechanism, and designs the feature attention module to improve the adaptability of the network to different turbidity regions of underwater images. Fig. 2 shows the structure diagram of the feature attention module. The local residual structure is designed. The output features of the upper layer are first convoluted, and then converted into nonlinearity through the leaky relu activation function. The output results are superimposed with the input features, and the superimposed features are convoluted again. The obtained features are introduced into the improved channel attention mechanism and pixel attention mechanism to extract the more implicit mapping relationship.

Previous channel attention mechanisms usually use global average pooling for input feature f to convert channel related information into channel descriptors g_c.

$$g_c = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} F(i,j) \tag{1}$$

Under the effect of global average pooling, the shape of the characteristic graph changes from $C \times H \times W$ becomes $C \times 1 \times 1$. The full connection layer is used to obtain the weights of different channels, reduce the complexity of the model by reducing the dimension, project the channel features into the low-dimensional space, and then re map. This method will destroy the direct correspondence between the channel and the weight and affect the prediction effect of

channel attention. In order to improve the correlation between channels, this paper replaces the full connection layer in the channel attention module with a fast one-dimensional convolution with a kernel size of k, and all channels share learning parameters.

$$W_c = \sigma(Conv1D_k(g_c)) \tag{2}$$

Among them, σ Represents sigmoid activation function, conv D_1 represents 1-dimensional convolution. The mapping function between channel dimension C and kernel size k is designed as shown in formula (3), and the kernel size k is determined adaptively.

$$C = \phi(k) = 2^{(\gamma \times k - b)} \tag{3}$$

Given the channel dimension C, The k is expressed as:

$$k = \psi(C) = |\frac{\log_2(C)}{\gamma} + \frac{b}{\gamma}|_{odd} \tag{4}$$

Among them, will γ set to 2, b set to 1, $|\xi|_{odd}$ representative ξ recent odd number.

Finally, the input feature F is multiplied by the learned channel weight W_c pixel by pixel, and the output channel feature F_c is expressed as:

$$F_c = F \otimes W_c \tag{5}$$

Due to the simple encoder decoder network structure, the image content will be processed indiscriminately, and the turbidity produced by different underwater environments is not uniform. Therefore, as a supplement to the channel attention mechanism, this paper designs a pixel by pixel attention mechanism to enable the network to focus on the image area

with higher turbidity. Similar to the channel attention mechanism, the feature F_c passing through the channel attention mechanism is introduced into the pixel attention layer as the input of pixel by pixel attention. The pixel attention layer is three convolution layers composed of two prelu activation functions and one sigmoid activation function.

$$W_p = \sigma(Conv(\delta(Conv(\delta(Conv(F_c)))))) \qquad (6)$$

where conv stands for 1×1 convolution, δ Represents the prelu activation function and sigmoid activation function. After a series of convolution operations, the characteristic shape is determined by $C \times H \times W$ becomes $1 \times H \times W$. Finally, the input feature F_c and the obtained weight WP are multiplied pixel by pixel to obtain the output feature FSC, which is expressed as:

$$F_{sc} = F_c \otimes W_p \qquad (7)$$

B. LINEAR MOTOR POSITION DETECTION PLATFORM

The ideal motion state of the linear motor mover is one-dimensional rigid body translation, but in the actual working conditions, the longitudinal vibration of the mover is difficult to avoid. At this time, the camera fixed on the mover will produce the offset of the lens and affect the collected signal sequence Therefore, the aperiodic fence stripe image shown in Figure 3 is designed. This kind of image has two characteristics:

1) The information in the horizontal direction is aperiodic, which avoids the information overlap in the process of signal acquisition

2) The gray consistency in the vertical direction can ensure that the signal collected during lens offset is consistent with the signal information under ideal conditions Moreover, if the displacement of the whole image is calculated, because the gray of the image is consistent in the vertical direction, it will cause a lot of redundant calculation In this paper, the linear array camera is used to obtain the one-dimensional information in the horizontal direction of the image to improve the calculation speed of the algorithm For dimension. The expression of M non periodic stripes can be constructed through the following pixels.

$$f(x, y) = \sum_{i=1}^{n} G_i \cdot \mathbb{R}_i(x, y) \qquad (8)$$

$$\mathbb{R}_i(x, y) = \begin{cases} 1 & (\omega_i - 1 \le x \le \omega_i) \\ 0 & other \end{cases} \qquad (9)$$

Among G_i is the gray value of the i-th stripe, and w_i is the width of the i-th stripe. The x, y are the width and depths for the i-th pixel, respectively.

C. MULTISCALE FEATURE EXTRACTION MODULE (MFE) AND LOSS FUNCTION

In this network, in order to obtain different levels of features [25]. Among them, multi-scale refers to the hole convolution with different size and sampling rate $= 1, 2, 4$ to extract
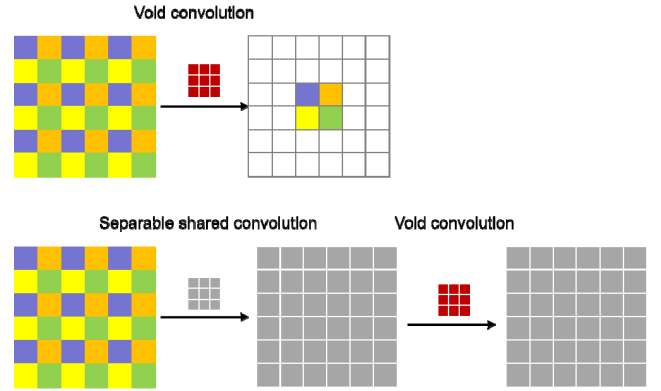

FIGURE 3. Smooth cavity convolution.

the characteristic information of different scale range. Due to the different expansion rate of hole convolution and different receptive fields, the size of characteristic area information obtained is different. In the 1-D case, given the 1-D input, the output of hole convolution is

$$o(i) = \sum_{j=1}^{k} f[i + r \times j]\omega[j] \qquad (10)$$

where, r is the expansion rate and k is the size of the receptive field. At that time, the expanded convolution corresponds to the standard convolution. As shown in Fig. 3, Four different points of the next layer are indicated by different colors. It can be seen that they are obtained from completely unrelated points of the previous layer, which may lead to mesh artifacts. It adds an additional separable shared convolution layer before expanding convolution, which increases the dependence between input units. The features obtained by smooth hole convolution with different sampling rates are weighted by the attention mechanism module, and then the features are connected together. After a layer of convolution, the features of different scales are extracted adaptively, and the obtained features are added with the input and input to the next stage

III. NETWORK IMPROVEMENT

A. RETINEX MODEL

The absorption of light by the water body makes the deepwater world dark. It is necessary to take a clear underwater image through auxiliary light. However, due to the scattering of light caused by granular substances in the water, the underwater imaging contrast is low, the image details are blurred, and there are fewer observable things. Aiming at the influence of water scattering on underwater images, this paper improves the guided filtering of Gaussian low-pass convolution function in single-scale Retinex [28]. Therefore, in order to avoid the halo phenomenon, this paper selects the single-scale Retinex model. The edge details of underwater image processed by single-scale Retinex model are incomplete. In order to solve this phenomenon, guided filtering is introduced. The formula of single scale Retinex model is as

follows:

$$I(x, y) = R(x, y) * L(x, y) \tag{11}$$

Retinex theory holds that the color, brightness, shape and details of the object by the human eye depend on the light irradiation and the reflection of the light on the surface of the object. That is, the image $I(x, y)$ received by the human body's own visual sensor is composed of incident information $L(x, y)$ and reflected information $R(x, y)$.

At present, single scale Retinex algorithm (SCR) mostly adopts Gaussian low-pass convolution [10] to obtain ideal illumination estimation. The specific calculation formula is as follows:

$$\hat{L}(x, y) = I(x, y) * F(x, y) \tag{12}$$

where: $\hat{L}(x, y)$ For ideal illumination estimation, it means convolution of the original image and Gaussian function [11]. The specific formula is as follows:

$$F(x, y) = K \exp[-(x^2 + y^2)\backslash\sigma^2] \tag{13}$$

where: σ Is the size of Gaussian function, and the coefficient K should meet:

$$\int\int F(x, y)dxdy = 1 \tag{14}$$

The specific formula of the ideal reflection image characteristic $R(x, y)$ of the target object is as follows:

$$R(x, y) = \ln I(x, y) - \ln[G(x, y) * I(x, y)] \tag{15}$$

The traditional simple weighted fusion has the advantages of easy implementation, simple principle, fast running speed, improving the peak signal-to-noise ratio of the image and reducing noise. However, this algorithm not only optimizes the image, but also reduces the detail information in the image to a certain extent, which makes the image edge incomplete, the image contrast low, artifacts and so on. In order to solve the above problems, this algorithm calculates Laplace weight, photometric weight and saliency weight, and introduces pyramid multi-scale fusion [30] to get a better image.

The significant weight focuses on the significant objects that lose significance in the underwater image. Firstly, convert RGB image into lab image, calculate the mean value of three channels, and calculate the sum of the squares of the mean values of three channels. The specific formula is as follows:

$$W_x = (L - L_{mean})^2 + (a - a_{mean})^2 + (b - b_{mean})^2 \tag{16}$$

where: W_x is the significant weight, L is the brightness, Which a and b are two colors respectively; L_{mean} is the average brightness, a_{mean} and b_{mean} is the color mean. Fig. 4 is a significant weight diagram after color compensation coefficient compensation and a significant diagram after improved guidance filtering. The significant weight tends to highlight the area with high brightness value. Therefore, the photometric weight based on reducing the saturation of the bright area is introduced. Firstly, RGB is converted to lab to obtain the bright channel L, and then the deviation between RGB tee

TABLE 2. The number of classes in training set.

Category	Quantity
Sea cucumber	6480
Sea urchin	25988
Scallop in Shell	9614
Starfish	11077

and 1 bright channel is calculated. The specific formula is as follows:

$$W_L = \sqrt{\frac{1}{3}[(R - l)^2 + (G - l)^2 + (B - l)^2]} \tag{17}$$

where: WL is the photometric weight, R, G, B are the pixel values of RGB three channels respectively, and l is the value of L bright channel.

B. DETECTION MODULE
In YOLOv4, the loss function loss is composed of CIO loss, confidence loss and classification loss, and BCE is used in classification loss Loss, the formula is:

$$BCEloss = -\omega * [m * \ln(n) + (1 - m) * \ln(1 - n)] \tag{18}$$

Among them, ω is the weight, m is the real value, and n is the predicted value. In the training set used in this paper, the number of four categories is unbalanced, as shown in Table 2, which will lead to the decline of precision measurement accuracy. Therefore, a focal Loss is used to replace BCE loss [31] in YOLOv4.

Focal loss is to use an appropriate function to measure the contribution of difficult and easy to classify samples to the total loss $\alpha_t = 0.25$ not only adjust the weight of positive and negative samples, but also control the weight of difficult and easy to classify samples. among $\alpha_t = 0.25$, $\gamma = 2$. The value is from the literature [12]. The model structure of YOLOv4 is shown in Fig.4.

This paper expands the idea of literature, improves the red channel compensation strategy, and solves the problem caused by the direct application of gray world algorithm to partial green underwater images in the improved compensation strategy, the attenuation of the significant area of the red channel is increased. Finally, the compensation formula of the red channel in this paper is:

$$I_{rc} = I_r + G_{com} \cdot (\bar{I}_g - \bar{I}_r) \cdot I_g \tag{19}$$

Among them, G_{com} is the improved red channel compensation rate in this paper, which is defined as:

$$G_{com} = \frac{1}{(1 + e^{\beta I_r - \gamma})} \tag{20}$$

Among them, β And γ Is a constant, which is obtained from the experiment $\beta = 13.15$, $\gamma = 3.79$. After using the improved color correction algorithm to preliminarily correct the color deviation of underwater image, a simple feature extraction network is designed to obtain the feature F1 of
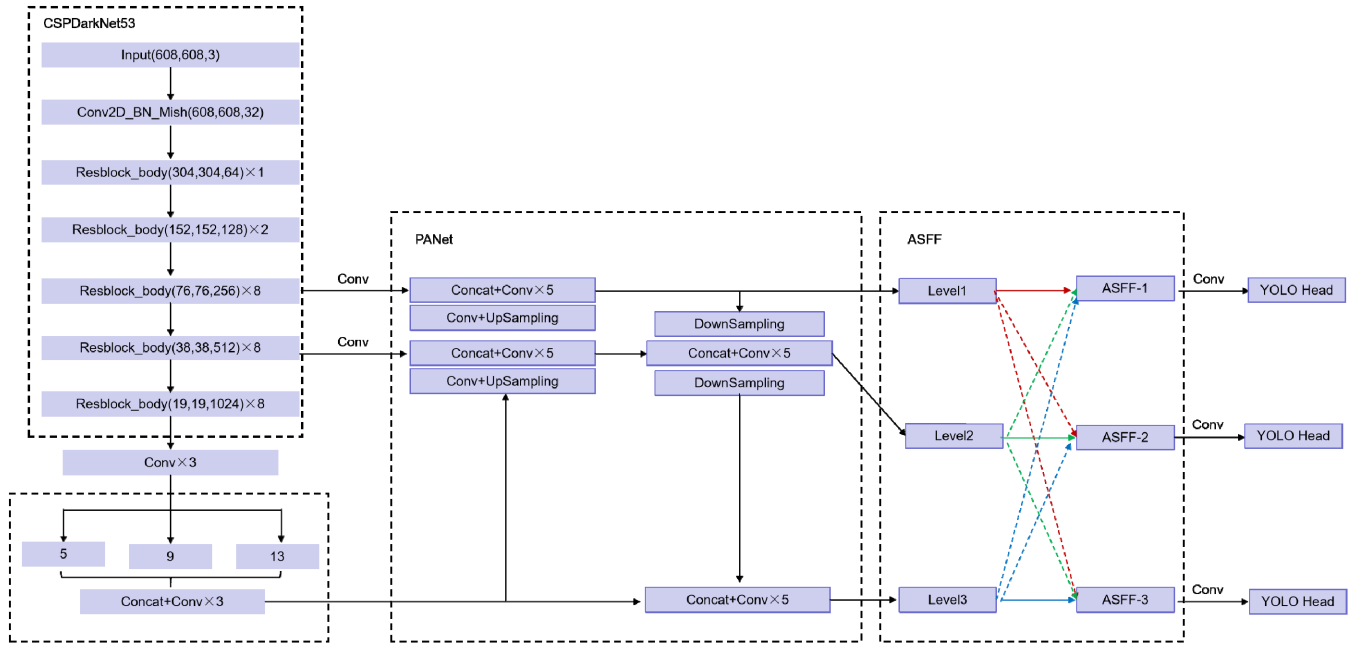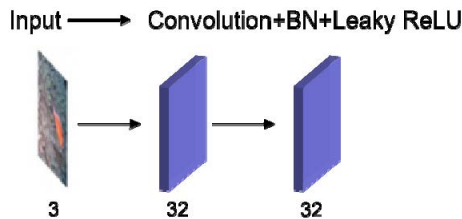
FIGURE 4. Network detected structure.

FIGURE 5. Color correction feature extraction module.

FIGURE 6. U-net feature extraction network structure.

color corrected image. Its structure is shown in Fig. 5 The structure is composed of two convolution layers. A convolution core with a size of 3 and a convolution step of 1 is used, and a batch normalization (BN) layer and a leaky relu activation layer are added after the convolution layer. The number directly below the blue block represents the number of characteristic channels.

In this paper, the u-net feature extraction module is used to extract the features of underwater degraded images. The symmetrical codec structure is adopted, and a layer hopping connection is added between the corresponding codec layers. Its structure is shown in Fig.6, The feature extraction module is mainly composed of a lower sampling unit and an upper sampling unit. The lower sampling unit includes a convolution layer with convolution step of 1 and a convolution layer with convolution step of 2, while the upper sampling unit includes a convolution layer with convolution step of 1 and a deconvolution layer with convolution step of 2 In the feature extraction module, all convolution layers adopt convolution kernel with size of 3, and BN layer and leaky relu activation layer are added behind the convolution layer
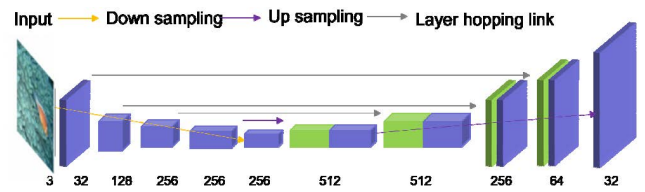
The underwater image is mapped into 32 channel feature F2 through the feature extraction network.

At the end of G, the features of the color correction image and the original underwater image are fused by multiplying the corresponding elements, and the fused features are mapped into the enhanced underwater image through the convolution layer Among them, the convolution layer uses a convolution core with a size of 1, and a sigmoid activation layer is added after the convolution layer.

C. LOSS FUNCTION DESIGN

There are some problems in underwater image, such as low brightness, image blur, color deviation and so on. The previous underwater image enhancement methods may not achieve good restoration effect only by using a single loss function. This paper comprehensively considers the minimization of absolute error loss L_1, structural similarity loss lssim and color loss lcolor, and designs a loss function suitable for underwater images [33]. L_2 loss is often used in previous image restoration tasks, but due to its own defects, it will over punish large errors and reduce the quality of image restoration. In order to avoid this problem, this paper uses the

minimization of absolute error loss, such as formula (14):

$$L_{I1}(U) = \frac{1}{N} \sum_{p \in U} |x(p) - y(p)| \quad (21)$$

where p is the index of the pixel and U is the pixel block; $x(p)$ and $y(p)$ are the values of pixel blocks in the enhanced image and pixels in the reference image, respectively. L_{I1} loss calculates the absolute value by comparing pixel by pixel to keep the brightness and color of the image unchanged.

When using self encoder like structure for image reconstruction, the structure texture of the original image is easy to be distorted. Structural similarity [34] can measure image similarity by calculating the combined differences of brightness, contrast and structure between the original image and the label image. In underwater image enhancement and other low-level visual tasks, it is very necessary to ensure the consistency of image brightness, contrast and texture structure. The mean value is used to estimate the brightness, the standard deviation to estimate the contrast and the covariance to estimate the structural similarity, as shown in formula (22)

$$S(p) = \frac{2u_x u_y + C_1}{u_x^2 + u_y^2 + C_1} \cdot \frac{2\sigma_{xy} + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \quad (22)$$

among, u_x and u_y Represents the average of pixel x and pixel y; σ_{xy} Represents the covariance of x and y; σ_x^2 and σ_y^2 represents the variance of x and y respectively; To prevent the denominator from being zero, set $C_1 = 0.01$, $C_2 = 0.02$. The expression of similarity loss of design structure is:

$$L_{ssim} = 1 - S(p) = 1 - \frac{2u_x u_y + C_1}{u_x^2 + u_y^2 + C_1}$$
$$\cdot \frac{2\sigma_{xy} + C_2}{\sigma}{}_x^2 + \sigma_y^2 + C_2 \quad (23)$$

Although the $L1$ loss has the function of calculating the color error, because it only measures the color difference between the generated image and the real image by comparing the pixel values, it can not guarantee that the color vector has the same direction, and this measurement method may still lead to obvious color difference. Therefore, this paper designs a loss function with color angle information. The color is further corrected by calculating the angle between the pixel color of the generated image and the real image pixel by pixel and summing it [35]. The expression of color correction loss is as follows:

$$L_c = \sum_p \angle((\tilde{x}_i)_p, (\tilde{y}_i)_p) \quad (24)$$

Including $(\cdot)$ P represents one pixel; $\angle$ $(\cdot)$ Is an operator that calculates the angle between two colors and converts RGB colors into three-dimensional vectors. Formula (24) pairs $\tilde{x}_i$ and $\tilde{y}_i$ Sum the angles between the color vectors of each pixel pair in the. To sum up, the total loss function L designed in this paper consists of three components. As shown in formula (25):

$$L = \sum_{i=1}^N \omega_1 L_{l1} + \omega_2 L_s + \omega_3 L_c \quad (25)$$

IV. EXPERIMENT AND ANALYSIS

The compensation amount of the red channel is proportional to the difference between the normalized mean of the green and red channels In order to avoid the supersaturation of the red channel in some areas when the gray world algorithm is used after the red loss compensation, the red channel compensation follows the principle of enhancing the area with small pixel intensity of the red channel and reducing the compensation of the area with large pixel value of the red channel However, when the algorithm is directly applied to the greenish underwater image, there may still be excessive red compensation, especially when the color deviation problem is serious, as shown in Fig. 7.

The essence of Gaussian low-pass filtering is to set different weights according to the distance between each point in the neighborhood and the central point for smoothing, so as to ensure the overall gray value characteristics of the image at the same time. However, after Gaussian low-pass filtering, the edge of the image is excessively smoothed, resulting in the lack of edge details in the ideal illumination image, and the resulting ideal reflection image inevitably lacks a large number of edge details. General filtering cannot distinguish between noise and edge. The pixel gradient around image noise is large, resulting in gradient fault at the edge [29]. The advantage of this paper is to use the gradient filter instead of the Gaussian filter to obtain the maximum illumination of the image. The edge details are preserved to the greatest extent, the essence of the image is preserved, and the purpose of noise removal is achieved.

As shown in Fig.8, the underwater image processed by Retinex with improved guidance filtering has higher definition and accuracy than single channel Retinex (SCR), multi-channel Retinex (MSRCR) [30] and the original image Reduction degree, which is closer to the image under natural light. In Fig. 9, after the image is locally enlarged by 300 times, the algorithm in this paper can smooth the image and remove noise better than SCR and MSRCR, while retaining the edge details of the image to make the overall details more prominent. The dark area of the enhanced image is brighter, the color is real, and the edge details remain intact.

A. OBJECTIVE INDEX AND ABLATION EXPERIMENT

In order to analyze and evaluate the effect of the algorithm more objectively, in addition to subjective feelings, full reference index evaluation and no reference index evaluation are also carried out. The test sets test A and test B containing label images were evaluated by Peak Signal to Noise Ratio (PSNR) and structural similarity index (SSIM). The higher the score of PSNR, the closer the enhancement result to the label image in image content, and the higher SSIM score, the closer the enhancement result to the label image in structure and texture. For the test set test C without label image, the enhanced image is evaluated by using the non reference underwater image quality indexes Underwater Image Quality Measure (UIQM) [33]. UIQM evaluates the degree of underwater image degradation by measuring the weighted sum of underwater image
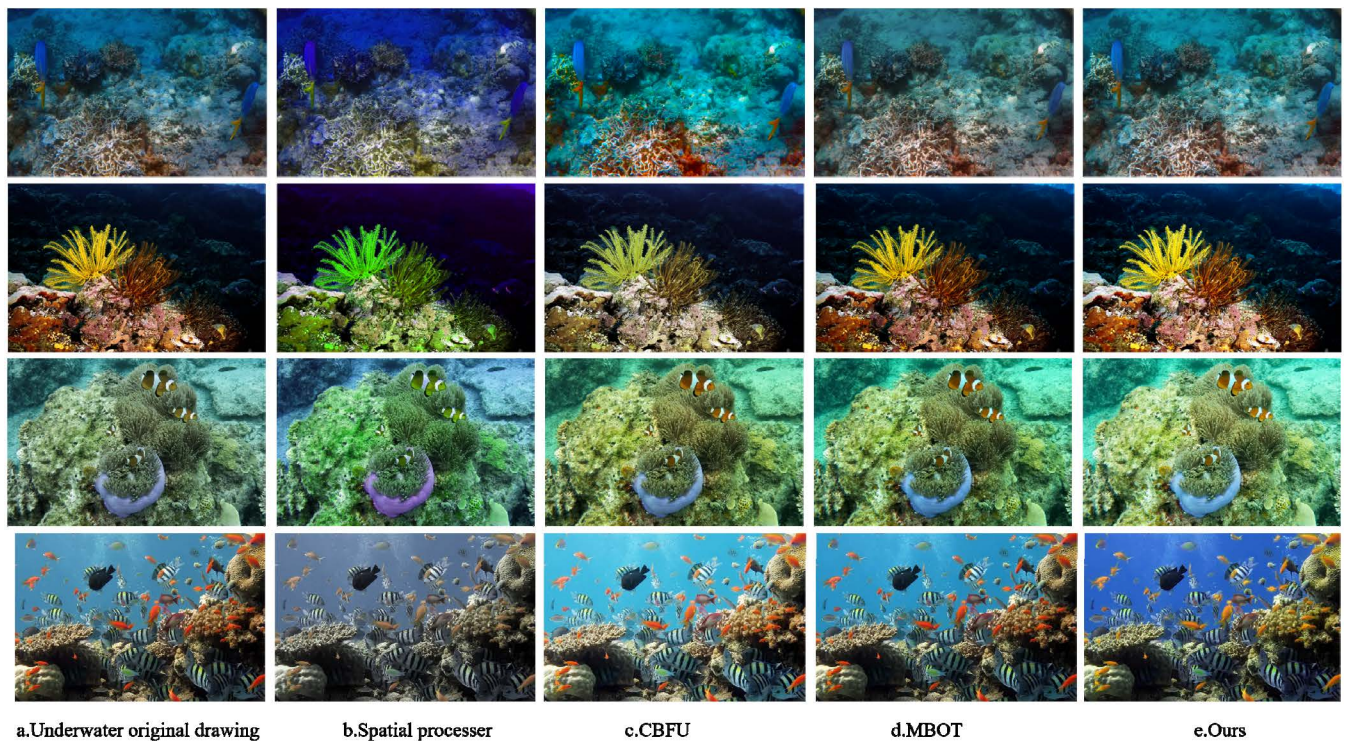
| a.Underwater original drawing | b.Spatial processer | c.CBFU | d.MBOT | e.Ours |

FIGURE 7. The evaluation results of underwater image visualization by this method and other representative methods.
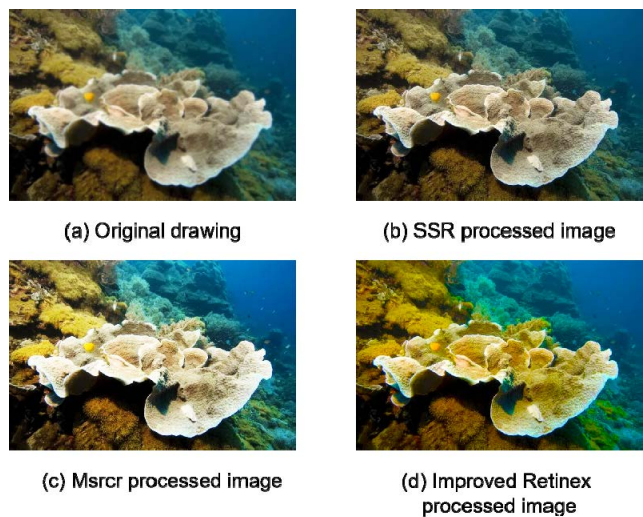

(a) Original drawing

(b) SSR processed image

(c) Msrcr processed image

(d) Improved Retinex processed image

FIGURE 8. comparison diagram after Retinex processing with improved guided filtering.

TABLE 3. Comparison of full reference indexes of different methods.

Method	Test A		Test B		Test C	
	PSNR	SSIM	PSNR	SSIM	UIQM	UCIQE
Fusion	17.90	0.73	17.41	0.87	4.71	0.77
UDCP	16.28	0.70	13.65	0.46	2.13	0.69
FIEGAN	25.66	0.88	17.95	0.75	6.82	0.62
UWCNN	23.62	0.82	14.82	0.61	6.36	0.57
Our method	27.58	0.94	20.58	0.78	6.73	0.88

color difference, underwater image sharpness and underwater image contrast. The larger the value, the better the image enhancement effect. As shown in Table 3, the algorithm in this paper achieves the best in test set test a and test B, which shows that the difference between the enhanced image and the label image is the smallest, and has achieved remarkable results in improving image brightness, contrast and maintaining texture structure. Compared with test a data set, the

restoration effect of different methods on the real underwater image data set test B is relatively poor. Compared with the comparison method, the restoration effect of this algorithm on the real underwater image is still significantly improved. In addition, the UIQM index of this algorithm on the test set test C is slightly weak, and the FIEGAN method achieves the best Underwater Color Image Quality Evaluation (UCIQE), which shows that this method has achieved good restoration effect in improving image clarity, color saturation and color saturation, and is more in line with human visual perception.

In order to illustrate the effectiveness of each component of the network proposed in this paper, ablation experiments are carried out on the feature attention module
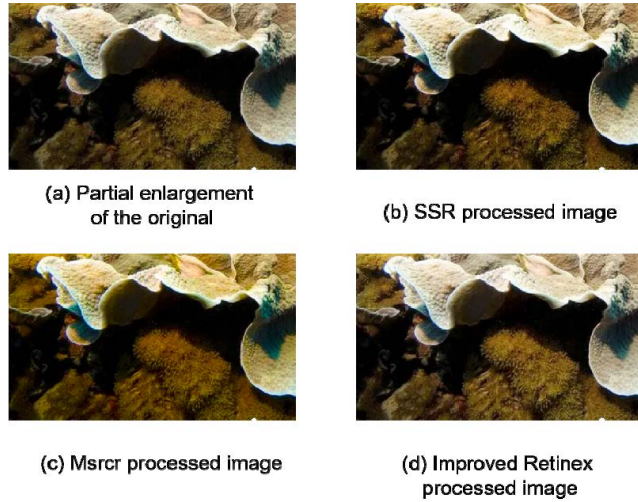
(a) Partial enlargement
of the original

(b) SSR processed image

(c) Msrcr processed image

(d) Improved Retinex
processed image

FIGURE 9. Visual result evaluation of detail enhancement after magnification.

FIGURE 10. Aperiodic fence fringe image and its single line sampling.

CP (channel pixel), dynamic feature enhancement module DFE (dynamic feature enhancement), heterogeneous feature fusion module HFF (heterogeneous feature fusion), and loss function under the same experimental conditions. The basic network is composed of encoder, decoder and hop connection, and different modules are gradually added to the basic network. In the ablation experiments of different modules, only L1 loss is used to train the network, and euvp image data set is used for training and testing. The number of experimental steps is set to 52 10. On the basis of using only L1 loss, the structural similarity loss and color loss are gradually added to train the network, which proves the effectiveness of the designed joint loss function.

According to the one-dimensional rigid body translation motion characteristics of linear motor mover, a linear motor mover position detection platform based on linear array camera and aperiodic fence stripe image is designed As shown in Fig.10, the linear array camera is installed on the mover and remains relatively stationary with the mover, and the aperiodic fence image is fixed in front of the camera lens During the movement of the mover, the linear array camera collects the signal sequence in real time, calculates the sub-pixel displacement between adjacent signal sequences through this algorithm, and then obtains the actual displacement of the mover combined with the calibration coefficient The mathematical model is: the signal sequences of two adjacent frames collected by the linear array camera are I_1 (x) and I_2 (x) respectively, and there is a one-dimensional rigid body translation conversion relationship between them as shown in formula (26).

$$I_2(x) = I_1(x - \Delta x) \qquad (26)$$

Among them Δ X is the sub-pixel displacement between two frame signal sequences, which is accurately calculated by the algorithm in this paper Δ X to obtain the mover displacement. Due to the low intensity of light and the gradient information
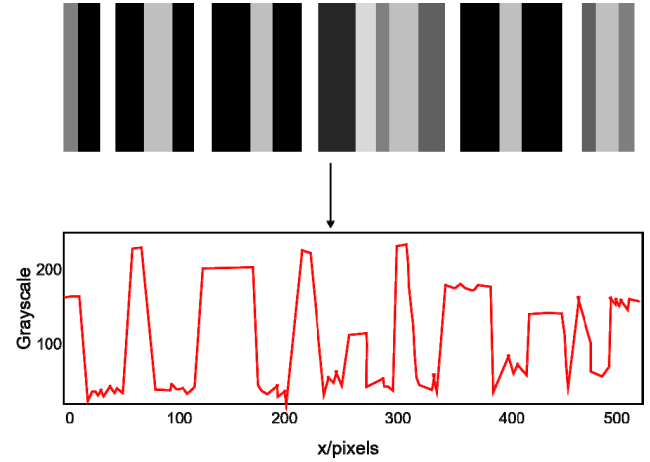
of the CCD, the overall intensity of the image decreases, and the image is sensitive to the intensity of light in reference [18], Taylor series expansion of the two signal sequences to be displaced proves that when the gray gradient and sum of the signal sequence are reduced, the detection accuracy will also be reduced. Therefore, it is necessary to preprocess the collected signal sequence to improve the detection accuracy and anti-interference ability to the light The image enhancement method based on gray linear transformation method is simple in calculation and will not add too much burden to the calculation time of the algorithm, so it is introduced into the preprocessing process of signal sequence If the collected signal sequence is I (x), the signal sequence I_e (x) enhanced by gray linear transformation is::

$$I_e(x) = \frac{n - m}{b - a}[I(x) - a] + m \qquad (27)$$

where [*a*, *b*] and [*m*, *n*] are the gray range of *I* (x) and transformed I_e (x), respectively, [*m*, *n*] takes the maximum range of gray range, i.e. [0255] As shown in Fig. 11, Fig.11 (a) and Fig.11 (b) are two-dimensional projection and gray distribution of reference signal sequence and signal sequence under low light intensity respectively It can be seen that the gray value of the signal sequence is generally low under low light intensity, and it can be seen from table 4 that the gray gradient and sum of the signal sequence collected under low light intensity are significantly reduced for the same target image Fig.11(c) and Fig.11(d) respectively show the signal sequence after image enhancement in Fig. 11(b) through histogram equalization and linear transformation method. The histogram equalization image has more distortion than the reference signal after enhancement, while the linear transformation method better retains the gray change characteristics of the signal.

B. COMPARISON BETWEEN iYOLOv4 AND YOLOv4 ALGORITHMS

There are 100 epochs in the whole training process. The first 50 epochs are trained by freezing network. The initial
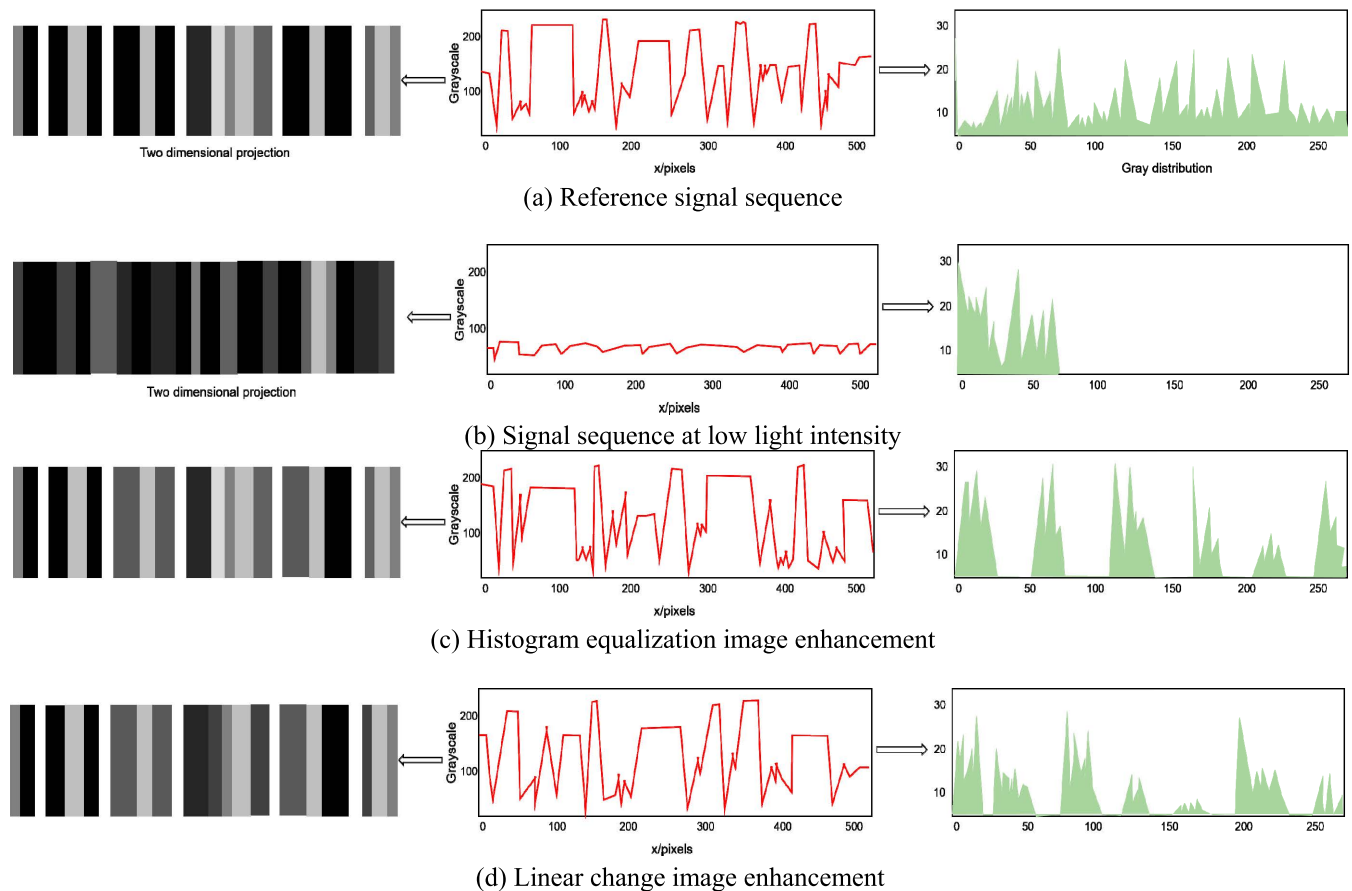
(a) Reference signal sequence

(b) Signal sequence at low light intensity

(c) Histogram equalization image enhancement

(d) Linear change image enhancement

FIGURE 11. One dimensional signal sequence and its two-dimensional projection and gray distribution.
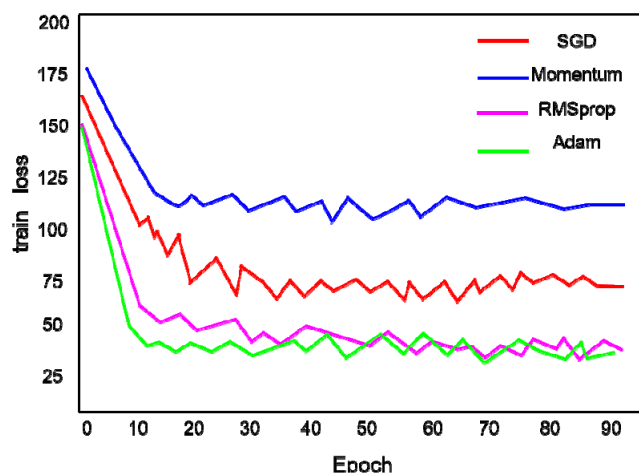
FIGURE 12. Comparison of four classic optimizers.

TABLE 4. Normalized gray gradient.

	Normalized gray gradient sum
Reference signal sequence	18.084
Signal sequence at low light intensity	1.803
Histogram equalization image enhancement	16.438
Linear transformation image enhancement	18.085

reached a stable state after 100 times of training. Using Adam optimizer for parameter learning, it can iteratively update the weight of neural network based on training data, which is suitable for large-scale data and parameter problems. Take out 1000 images in the data set and use four classical optimizers: SGD, momentum, RMSprop and Adam to train 100 epochs in the same network. Since the loss values are relatively large at the beginning of training, the loss values of the first 10 epochs are discarded. The results are shown in Figure 12. It can be seen that the effect of SGD and momentum optimizer is general, and the effect of RMSprop and Adam is better. However, after training 40 epochs, the loss value of Adam

learning rate is 0.001, and the cosine annealing is used to attenuate the learning rate, batch_ When the size is set to 64, the last 50 epoch thaw network training, the initial learning rate is 0.0001, and the cosine annealing is used to attenuate the learning rate, batchsize Set to 6. Using the same environment to train YOLOv4 and YOLOv4 respectively, the loss has

TABLE 5. Comparison of FLOPs and params.

Experience group	Method	FLOPs/G	Params/M
G1	YOLOv4	63.848	63.954
G2	YOLOv4+ASFF	70.785	73.181
G3	YOLOv4+ASFF+Focal Loss	70.785	73.181

optimizer is lower. Therefore, Adam optimizer is selected for training in this paper.

In deep learning, the complexity analysis mainly includes two indicators: the amount of computation and the amount of parameters. Flops refers to the number of floating-point operations per second, and flops corresponds to the time complexity; Params refers to the total number of parameters to be trained in the network model, and the parameter quantity corresponds to the spatial complexity. In this paper, the complexity of each method is compared by using the control variable method, which is divided into three groups for comparison, as shown Table 5. The first group (G1) is the original YOLOv4 algorithm; The second group (G2) is to add ASFF structure on the basis of YOLOv4; The third group (G3) is to replace bceloss with focaloss on the basis of G2. The results are shown in Table 5. Compared with G1 and G2, lops and params only increase by 9.6% and 14.2%, indicating that they have little impact on the time and space complexity of the network; Compared with G2 and G3, the replacement loss function does not affect flops and params.

V. CONCLUSION

This paper proposes an underwater image enhancement network based on heterogeneous feature fusion and dynamic feature enhancement. The network to improve the encoder and decoder structure, design and add features module, dynamic enhancement module and heterogeneous characteristics of fusion module with the method of heterogeneous feature fusion converged network to extract the low-level features and advanced features, improve communication between various modules, retain more image details. Aiming at the problems of uneven illumination, low contrast, large amount of noise and color deviation of underwater images taken in deep sea and night waters, the underwater image imaging model and water parameters are used to synthesize more real data sets for training. The trained network not only achieves good results on the synthetic data sets, but also achieves good results on the collected in reality data sets. For real underwater images, the color and detail of the scene are also well restored. The experimental results show that the proposed method is better than many existing methods, and can effectively remove the discoloration caused by water, and the enhanced image can restore the original features of the object well. It has better robustness for underwater images with severe color decay. In the future, we will consider further optimizing the network structure, reducing the number

of network parameters and single frame operation time; At the same time, although the algorithm in this paper has a good effect of image enhancement and restoration, it only focuses on the enhancement of static images. In the future, the network will be considered to be applied to continuous video scenes, and the enhancement of low resolution fuzzy video can be achieved by introducing the extraction of video multi frame information into the network structure.

REFERENCES

[1] C. Wenyi, Y. Chengxun, and Y. Hui, "Multiscale Retinex infrared image enhancement based on guided filtering and logarithmic transformation algorithm," *Infr. Technol.*, vol. 44, no. 4, pp. 397–403, 2022.

[2] M. Lu, "Low illumination image enhancement based on multi-scale wavelet U-shaped network," *Infr. Technol.*, vol. 44, no. 4, pp. 410–420, 2022.

[3] H. Yan, M. Zongfang, and W. Hao, "Multi scale image enhancement and flame segmentation based on ant colony algorithm in strong light," *Fire Sci. Technol.*, vol. 41, no. 4, pp. 435–439, 2022.

[4] W. Xingrui, P. Yan, and W. Yumo, "Research on composite residual network in low illumination image enhancement," *Liquid Cryst. Display*, vol. 37, no. 4, pp. 508–518, 2022.

[5] Q. Xin and W. Yi, "Design and implementation of adaptive night scene enhancement system based on FPGA," *Liquid Cryst. Display*, vol. 37, no. 4, pp. 501–507, 2022.

[6] W. Nan, L. Yiming, Z. Rui, Z. Jiao, and X. Yi, "Material orGA-Nization image enhancement algorithm based on generated countermeasure network," *Inf. Technol.*, vol. 12, no. 5, pp. 1–7, 2021, doi: 10.13274/j.cnki.hdzj.2021.12.001.

[7] H. Shouwu and K. Yongqiang, "Research on illumination interference image enhancement based on visual communication technology," *Laser J.*, vol. 42, no. 12, pp. 136–139, 2012, doi: 10.14016/j.cnki.jgzz.2022.12.136.

[8] S. Vahidian, M. Joneidi, A. Esmaeili, S. Khodadadeh, S. Zehtabian, and B. Lin, "Spectrum pursuit with residual descent for column subset selection problem: Theoretical guarantees and applications in deep learning," *IEEE Access*, vol. 10, pp. 88164–88177, 2022.

[9] Z. Jie, L. Yipeng, D. Lu, and L. Xueyan, "NSCT enhancement of low brightness image based on quantum harmony search fuzzy set," *Adv. Laser Optoelectron.*, vol. 58, no. 24, pp. 371–382, 2021.

[10] W. Sili, S. Ying, W. Hui, and H. Wei, "SAR image enhancement based on Wavelet and Gaussian mixture model," *Shanghai Aerosp.*, vol. 38, no. S1, pp. 26–31, 2021.

[11] X. Minmin, Z. Feng, Z. Liqun, and W. Bo, "Application research of police low illumination image enhancement technology based on PCNN," *J. Armed Police College*, vol. 37, no. 12, pp. 57–62, 2021.

[12] A. maitusun, I. ilhamu, and T. abduklimu, "An image enhancement method based on wavelet domain," *Electron. Design Eng.*, vol. 29, no. 24, pp. 24–27, 2021, doi: 10.14022/j.issn1674-6236.2021.24.005.

[13] P. Tian, "Research on low quality image enhancement technology of airport monitoring based on *a priori* and depth neural network," *Sichuan Univ. Electron. Sci. Technol.*, vol. 15, pp. 177–189, Apr. 2021.

[14] L. Juan, C. Minhui, W. Yao, Z. Yu, Z. Shun, Z. Peijiang, and Z. Dequan, "Seedling image enhancement algorithm based on double gamma correction," *Jiangsu Agricult. J.*, vol. 36, no. 6, pp. 1411–1418, 2020.

[15] C. Xiao, D. Jicun, and L. Chuang, "Low illumination image enhancement algorithm based on improved BEMD decomposition," *Inf. Technol. Informatization*, vol. 12, no. 8, pp. 95–98, 2020.

[16] F. Lv, Y. Li, and F. Lu, "Attention guided low-light image enhancement with a large scale low-light simulation dataset," *Int. J. Comput. Vis.*, vol. 129, no. 7, pp. 2175–2193, 2021.

[17] B. Turan, "A image enhancement method for night-way images," *Balkan J. Elect. Comput. Eng.*, vol. 9, no. 1, pp. 8–16, 2021.

[18] E. H. Land, "The retinex theory of color vision," *Sci. Amer.*, vol. 237, no. 6, pp. 108–128, Dec. 1977.

[19] D. Xuwei and C. Dong, "Image enhancement and segmentation of small defects in low specific steel plate," *Machinery Electron.*, vol. 38, no. 12, pp. 65–69, 2020.

[20] Z. Chunli, D. Jingwei, X. Bo, and M. Xiaofeng, "Research on fog image enhancement algorithm integrating histogram equalization and homomorphic filtering," *J. Harbin Univ. Technol.*, vol. 24, no. 6, pp. 93–97, 2019, doi: 10.15938/j.jhust.2019.06.014.

[21] L. Yanan and S. Tao, "Research on palmar vein image enhancement method based on MATLAB," *J. Qilu Univ. Technol.*, vol. 33, no. 6, pp. 23–27, 2019, doi: 10.16442/j.cnki.qlgydxxb.2019.06.004.

[22] S. Yang and L. Zhu, "Night lane detection technology based on depth image enhancement," *Comput. Appl.*, vol. 39, no. S2, pp. 103–106, 2019.

[23] Z. Changyong, Z. Xin, M. Chao, and S. Jiumei, "Research on laser 3D image enhancement system based on virtual reality," *Laser J.*, vol. 40, no. 12, pp. 67–71, 2019, doi: 10.14016/j.cnki.jgzz.2019.12.067.

[24] D. Singh, M. Kaur, M. Y. Jabarulla, V. Kumar, and H.-N. Lee, "Evolving fusion-based visibility restoration model for hazy remote sensing images using dynamic differential evolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–14, 2022.

[25] Y. Yun, L. Yu, and Z. Quanhua, "A color remote sensing image enhancement method based on adaptive intensity transformation," *Remote Sens. Inf.*, vol. 34, no. 6, pp. 39–44, 2019.

[26] H. Zhen, T. Haijun, and W. Haifeng, "Low illumination image enhancement method based on conditional generation countermeasure network," *Chin. J. Image Graph.*, vol. 24, no. 12, pp. 2149–2158, 2019.

[27] X. Luyan, "Wang Hong power image enhancement algorithm based on CLAHE fusion wavelet denoising," *Data Commun.*, vol. 6, no. 12, pp. 31–33, 2018.

[28] Z. Jin, "Research on image enhancement algorithm based on multi-objective optimization," *J. Wuyi Univ.*, vol. 37, no. 12, pp. 54–57, 2018, doi: 10.14155/j.cnki.35-1293/g4.2018.12.011.

[29] L. Min, L. Chunmiao, and M. Jun, "Application of iPP image enhancement technology in the study of soil microstructure," *Coal Technol.*, vol. 4, no. 9, pp. 57–60, 2018, doi: 10.19896/j.cnki.mtkj.2018.04.018.

[30] S. Jingjing, "Research on Retinex image enhancement algorithm based on trilateral filtering in HSV color space," *Automat. Instrum.*, vol. 12, no. 7, pp. 12–14, 2018, doi: 10.14016/j.cnki.1001-9227.2018.12.012.

[31] B. Siwen and K. Yuxian, "Research on data enhancement algorithm of quantum remote sensing image," *J. Global Change Data Chin. English*, vol. 2, no. 4, pp. 367–376, 2018.

[32] Z. Cong, "A nonlinear image enhancement algorithm for underground coal mine," *Energy Energy Conservation*, vol. 12, no. 6, pp. 145–146, 2018, doi: 10.16643/j.cnki.14-1360/td.2018.12.059.

[33] J. Huihua, "Research on real-time low illumination color image enhancement method," *J. Anhui Electron. Inf. Vocational Tech. College*, vol. 17, no. 6, pp. 36–38, 2018.

[34] R. Mengjie and S. Shengbin, "High performance implementation of single scale Retinex image enhancement algorithm on Xilinx FPGA," *J. Gun Launch Control*, vol. 39, no. 4, pp. 44–49, 2018, doi: 10.19323/j.issn.1673-6524.2018.04.009.

[35] W. Linzhu and S. Qinglie, "Application and research of vehicle image enhancement algorithm based on human visual characteristics," *J. Guilin Inst. Aerosp. Technol.*, vol. 23, no. 4, pp. 435–438, 2018.

[36] L. Zhao, C. Jie, S. Hui, Y. Pu, and S. Liangliang, "Research on monitoring image enhancement algorithm of Underpass interchange based on improved MSRCR," *J. Anhui Jianzhu Univ.*, vol. 26, no. 6, pp. 75–79, 2018.

[37] H. Zhen and L. Xiaomeng, "Research and simulation of image enhancement algorithm based on wavelet tower decomposition," *Technol. Innov. Appl.*, vol. 35, no. 11, pp. 30–31, 2018.

[38] X. Pengfei and L. Guoen, "Application of image enhancement based on histogram specification in medicine," *Comput. Knowl. Technol.*, vol. 14, no. 34, pp. 192–195, 2018, doi: 10.14004/j.cnki.ckt.2018.4042.

[39] L. Xin, *Research on Low Illumination Image Enhancement Method Based on Depth Learning*. Tianjin, China: Tianjin Univ., 2018, doi: 10.27356/d.cnki.gtjdu.2018.001064.

[40] W. Xinyan, *Underwater Image Enhancement Based on Structure Texture Decomposition*. Tianjin, China: Tianjin Univ., 2018, doi: 10.27356/d.cnki.gtjdu.2018.001815.
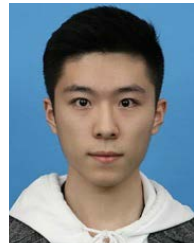
YUAN TIAN was born in Jining, Shandong, China, in 2000. He is currently pursuing the degree with the College of Mechanical, Electrical and Information Engineering, Shandong University. His research interests include computer science, software, and machine learning.

YUANG XU was born in Fushun, Liaoning, China, in 2000. He is currently pursuing the bachelor's degree in computer science with Shandong University. His research interests include NLP, program development, and java source code research.

JUN ZHOU was born in 1981. He received the Doctor's degree from Logistic Engineering University. His research interests include deep learning and big data analysis.

● ● ●