

Attention-based for Multiscale Fusion Underwater Image Enhancement

Zhixiong Huang¹, Jinjiang Li^{2*}, and Zhen Hua¹

¹ College of Electronic and Communications Engineering, Shandong Technology and Business University
Yantai, Shandong 264005, China

² Co-innovation Center of Shandong Colleges and Universities: Future Intelligent Computing,
Shandong Technology and Business University, Yantai, Shandong 264005, China
[e-mail: lijjiang@gmail.com]

*Corresponding author: Jinjiang Li

*Received October 22, 2021; revised January 31, 2022; accepted February 17, 2022;
published February 28, 2022*

Abstract

Underwater images often suffer from color distortion, blurring and low contrast, which is caused by the propagation of light in the underwater environment being affected by the two processes: absorption and scattering. To cope with the poor quality of underwater images, this paper proposes a multiscale fusion underwater image enhancement method based on channel attention mechanism and local binary pattern (LBP). The network consists of three modules: feature aggregation, image reconstruction and LBP enhancement. The feature aggregation module aggregates feature information at different scales of the image, and the image reconstruction module restores the output features to high-quality underwater images. The network also introduces channel attention mechanism to make the network pay more attention to the channels containing important information. The detail information is protected by real-time superposition with feature information. Experimental results demonstrate that the method in this paper produces results with correct colors and complete details, and outperforms existing methods in quantitative metrics.

Keywords: Underwater image enhancement, Multiscale fusion, Convolutional neural network, Attention mechanism, Local binary pattern.

1. Introduction

In recent years, with the continuous development of underwater detection technology, the number of underwater-related applications has increased, such as subsea oil exploration, submarine biological assessment, archaeology, etc. And a lot of underwater work has been carried out by vision sensors carried by autonomous underwater vehicles and remotely operated vehicles [1,2,3]. However, the actual rendering of underwater images captured by vision sensors is of poor quality, often exhibiting degradation effects such as color deviation, scattering blur, and low contrast [4]. As an important carrier of underwater information, the quality of underwater images is undoubtedly important in displaying and extracting underwater information. Therefore, underwater image enhancement is a meaningful and challenging task.

The propagation of light is affected by the water medium, undergoes two exponential attenuation processes of absorption and scattering. The absorption of light in the underwater environment is related to wavelength and depth. Light with longer wavelengths has a higher degree of attenuation at deeper depths, as blue and green light can reach deeper depths than red and yellow light, so the colors of underwater images tend to show blue and green color distortion. The scattering effect is produced by suspended particles and plankton that deflect the direction of light propagation, which makes underwater images have low contrast and blurring problems. Fig. 1 shows the three-color histogram of the underwater image and the enhanced image. As can be seen, the original underwater image had the severe green color deviation, so the green channel distribution was concentrated in the higher grayscale in the three-color histogram, which resulted in uneven color and low contrast of the image. However, as the colors are well corrected, the histogram has a more uniform gray distribution and the red, green and blue histogram trends are more similar, resulting in improved image contrast and sharpness, and thus improved image quality.

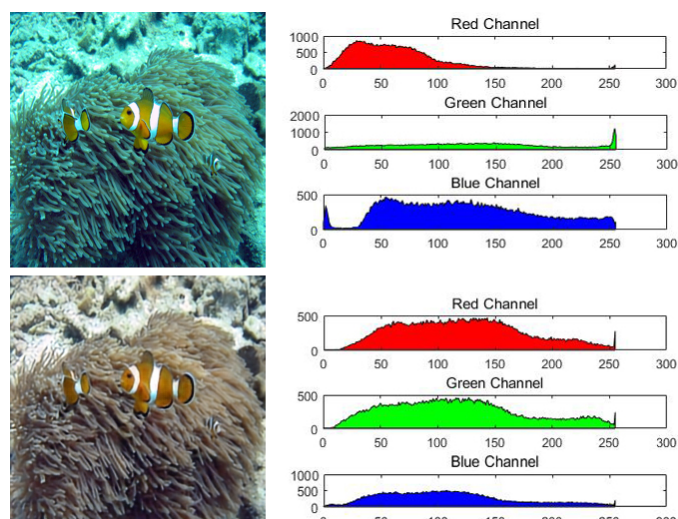


Fig. 1. Comparison before and after color correction. The top are the original underwater image and the corresponding three-color histogram, and the bottom are the image enhanced by the proposed method and the corresponding three-color histogram.

In order to solve the problems of underwater images, researchers have proposed many effective methods to recover and enhance underwater images, but there are still some problems to a greater or lesser extent. The first is the introduction of other color deviations and artifacts. The results produced by some methods are affected by other colors, so the color deviation still exists. In addition, excessive compensation for attenuated channel information, especially red channel information, will produce unexpected color artifacts, which will further affect image clarity and contrast. The second is that the color correction is not thorough enough. When some methods restore the overall color of the picture, there are still uncorrected color deviations in some areas of the scene. The third is to ignore the loss of details. Some methods will change the scale when dealing with the problem, and the scale change is likely to lead to the loss of details, and the loss of these details has not received the attention it deserves.

This paper proposes a multiscale fusion underwater image enhancement network combining channel attention and LBP. This method presents better performance on both referenced and non-referenced datasets, and is obviously superior to the advanced method in qualitative and quantitative aspects. The contributions of this paper are as following:

- Color correction of underwater images: an encoding-decoding network structure is used to restore the original color of underwater images and generate high quality underwater images.

- Introduction of channel attention: The channel attention is added between the feature aggregation module and the image reconstruction module to make the network more focused on important information.

- Effective protection of image details: LBP images are aggregated with features in real time when the image resolution changes, thus protecting details that may be corrupted.

The rest of the paper is organized as follows. Section 2 describes the related work. Section 3 describes the underwater enhancement network of this paper in detail, including the LBP mechanism and the principle of channel attention mechanism. Section 4 gives a comparison with existing methods on reference and no-reference datasets, and ablation study on the relevant components of the network. Section 5 describes conclusion and future work direction.

2. Related Work

In this section, related work is presented in three aspects: restoration-based methods, enhancement-based methods and deep learning-based methods.

2.1 Restoration-based Methods

The underwater image restoration method combines the theory of light propagation in the water medium and an underwater imaging model. First, the prior information is extracted from the underwater image, and then the parameters of the underwater imaging model are derived from this to reverse the degradation process of the underwater image, and finally, the underwater image with the correct color is obtained.

Li et al. [4] performed underwater image restoration into two stages, firstly restoring the natural appearance of underwater images based on the principle of minimum information loss, and then improving the contrast and brightness based on the histogram prior distribution. In view of the uneven light field in the underwater environment, Hu et al. [5] proposed an underwater image restoration method based on estimating the polarization degree and backscattering intensity of different positions of the image. Akkaynak et al. [6] took into

account underwater wavelength-dependent degradation, and proposed a revised underwater imaging model. In the paper [7], they proved that the method using the improved model is better than the old method using the atmospheric model. Sánchez-Ferreira et al. [8] proposed a method to estimate the model parameters with a bionic optimization metaheuristic, and at the same time constructed a dataset with different degrees of underwater degradation. Dai et al. [9] calculated the transmission map by decomposing the degradation curve on the RGB channel, and improved the visual effect of the underwater image by correcting the transmission map and color balance algorithm. Berman et al. [10] recovered the underwater image color by estimating blue-red and blue-green channel attenuation ratios for different types of underwater images, and collected a new underwater dataset with real reference images. Kar et al. [11] proposed a real image restoration model based on the image degradation characteristics derived from the Koschmieder model, which showed good results on multiple real image datasets.

Underwater image restoration methods derive model parameters from priori information, so they are sensitive to changes in the underwater imaging environment, and are likely to fail to present the desired results in challenging situations such as low-light regions or near-far regions. In addition, the imaging model has a large number of parameters, and it is likely to go through a large number of calculation processes before obtaining the restored image.

2.2 Enhancement-based Methods

Underwater image enhancement methods directly improve image quality by modifying pixel values, and do not require underwater imaging models or other optical models. Ancuti et al. [12] used white balance to correct underwater color deviation and image fusion to enhance edge details, which showed good enhancement performance on underwater images. Due to the feedback relationship between the color-sensitive horizontal cells and the cone cells of the teleost fish retina, Gao et al. [13] solved the problem of blur and color unevenness in underwater images by simulating the retina function and red channel compensation. Ancuti et al. [14] proposed a color channel compensation (3C) method to improve the image color. This method subtracts the local mean value from each opposing color pixel to return to the zero-mean value, thereby restoring the lost color. Hegde et al. [15] proposed an adaptive color curve correction method based on cubic spline interpolation. The color correction curve is estimated through the CIE $L^*a^*b^*$ color space. Song et al. [16] established the first database for underwater image background light estimation, and divided the underwater image enhancement process into two steps, first image restoration by the proposed background light estimation model and the optimal transmission map optimizer, and then color correction by an improved white balance algorithm. Considering the wavelength-dependent attenuation of different colors, Liang et al. [17] proposed to solve the color distortion problem with the attenuation map of each color channel, and then based on multi-scale decomposition to eliminate the fogging effect and compensate for the loss of details. Zhang et al. [18] used sub-interval linear transformation and dual-interval histograms to correct colors and enhance contrast, and obtained high-quality underwater images.

The enhancement-based method can effectively improve the contrast and brightness of the underwater image, resulting in improved quality of underwater images. But the result of the enhancement is prone to residual uncorrected color deviation, and may introduce other color artifacts. In the variable underwater environment, it does not always produce the expected results.

2.3 Deep Learning-based Methods

In recent years, deep learning has received a lot of attention in the field of computer vision, and methods to solve various vision tasks through deep learning have emerged in an endless stream [19,20,21,22,23]. Li et al [24] first applied GAN networks to underwater style image generation. The proposed WaterGAN generates an underwater dataset of degraded images paired with reference images through an unsupervised confrontation model, and these images then form the training dataset of an underwater monocular image color correction network. In the literature [25], they found that WaterGAN can't process the image correctly when the underwater conditions are unknown, so they proposed a weakly supervised adversarial network (UCycleGAN), which avoids the requirement of one-to-one correspondence between the original image and the reference standard. Hou et al. [26] proposed a residual CNN for joint learning in the transmission domain and the image domain, and combined the paired dataset with the NYU depth dataset [27] and the Jaffe-McGlamery underwater imaging formulation [28], the imaging formula is as follows:

$$I(x) = J(x)t(x) + B_\lambda(1 - t(x)), \#(1)$$

where $I(x)$ represents the low-quality image, $J(x)$ represents the scene brightness, and $t(x)$ is the transmittance of the propagation medium, which is defined as follows: $t(x) = e^{-\beta(\lambda)d(x)}$, $\beta(\lambda)$ is the attenuation coefficient of light in the water medium, λ represents the three-color channel, $d(x)$ represents the distance between the camera and the object in the picture, and B_λ represents the atmospheric ambient light in the scene. The residual network model trained by the synthetic dataset effectively removes the fogging effect. Jamadandi and Mudanagudi [29] proposed a new deep learning framework combined with wavelet transform, which turned the underwater image enhancement problem into a realistic style conversion problem. Desai et al. [30] synthesized underwater style images based on the revised underwater imaging model [6], and trained on the proposed conditional generation confrontation network.

Deep learning methods enable powerful color correction of underwater images, but this powerful correction ability may affect the color of the thing itself, resulting in over-enhancement or over-saturation effects, and the feature information of the image is not well protected after processing by the deep network, producing results with loss of details. Considering these problems, this paper introduces the channel attention mechanism in the end-to-end underwater enhancement network, so that the network pays more attention to important channels when generating aggregated results, and the resulting images generated have more accurate environmental colors. In addition, LBP images are introduced to protect the detail information of the underwater images after the scale change, and the LBP image changes and fuses in real time with the change of feature matrix resolution in the network, making the enhanced image texture features richer and more detailed.

3. Methodology

To restore the normal color of the underwater image while retaining the original detailed features of the image, this paper proposes a network for multiscale feature extraction similar to an encoder-decoder. Because part of the information will be lost in the encoding process, the feature information is mapped to the channel attention region before down-sampling, and then aggregated with the features before up-sampling through a jump connection to make the final output result of the image more accurate. In addition, LBP images with rich image details are also introduced to aggregate them in real time as the image resolution changes, so as to avoid the loss of detailed information due to the scale change of features. The underwater image

enhancement learning process of our network is shown in Algorithm 1.

Algorithm 1. Attention-based for Underwater Image Enhancement
input: Paired synthetic dataset: d^s
output: Enhanced underwater image: I^e
1. Synthesize underwater style image I^s and true standard I^t by sampling from d^s
2. Generate LBP images based on I^s : $I^L(x_c, y_c) = \sum_{i=0}^{P-1} 2^i S(I_i^s - I_c^s)$
3. The aggregated results are obtained by the feature aggregation module: $I^a = F_a(I^s)$ When the scale of I^a changes during the aggregation: $I^a = I^a + F_a(I^L)$
4. Enhanced results are obtained by image reconstruction module: $I^e = F_r(I^a)$ When the scale of I^e changes during reconstruction: $I^e = I^e + F_r(I^L)$
5. Update network weights by MSE loss function: $L_{MSE}(I_e, I_t) = \frac{1}{mn} \sum_{i=1}^{m-1} \sum_{j=1}^{n-1} [I_e(i, j) - I_t(i, j)]^2$

The input of the network consists of the original underwater image with color deviation and its corresponding LBP image, and the final output is the underwater image without the influence of color deviation. The whole network consists of three modules, namely, the feature aggregation module, the image reconstruction module and the LBP enhancement module. The network structure can extract and encode the image feature information effectively at different scales, use the different characteristics existing in feature information, and then aggregate them with the features obtained after decoding to finally obtain the enhanced underwater images. Fig. 2 shows the overall framework of the proposed network in this paper. Next, we will introduce each part of the network structure in detail.

3.1 Feature Aggregation Module

When restoring the image color information, detailed information of the structural features is often lost, so it is necessary to use the multi-scale features of the image to perform the aggregation operation. In the network, the original image undergoes three channel expansion processes. Firstly, the channel of input feature is expanded from 3 to 64 through the convolution block, and then the features are down-sampled into two blocks, in which the number of channels changes while the scale of the features changes, resulting in three features with the shapes $H \times W \times 64$, $\frac{H}{2} \times \frac{W}{2} \times 128$ and $\frac{H}{4} \times \frac{W}{4} \times 256$. The structure of the convolution block and down-sampling block are explained below. In order to protect the detailed information lost during the scale change, the LBP image of original input is convolved to maintain the same number of channels as the result of channel expansion results, and then the two are superimposed and aggregated through convolution to obtain the aggregation results of 64 channels, 128 channels and 256 channels respectively.

To make the image information more complete, the 64-channel and 128-channel aggregation results are put into the corresponding channel attention module, and then skip

connected to the image reconstruction module. The 64-channel and 128-channel aggregation results are further aggregated with 128-channel and 256-channel aggregation results after the down-sampling process, and the 128-channel and 256-channel secondary aggregation results are obtained respectively. The 128-channel secondary aggregation result will be further aggregated with the 256-channel secondary aggregation result after another down-sampling process, and finally the aggregation result of the module is output.

① in Fig. 3 shows the specific structure of the convolutional block and the down-sampling block in detail. The latter layer of each convolutional layer in the convolutional block has a ReLU activation layer, and the down-sampling block is composed of a convolution layer, a maxpooling layer and a ReLU activation function. The down-sampling process eliminates the redundant information of the image while retaining important feature information. With the increase of the layer number, the gradient information will decay exponentially when the network updates the weight, resulting in the gradient disappearance problem. Due to the uniform gradient distribution of ReLU, it can solve the gradient disappearance problem of deep network, and has the characteristics of simple operation and fast convergence. Therefore, the network uses the nonlinear ReLU as the activation function to transform the linear model into the nonlinear model. The definition of the ReLU activation function is as following:

$$ReLU(x) = \begin{cases} x, & x > 0 \\ 0, & x \leq 0 \end{cases} \quad (2)$$

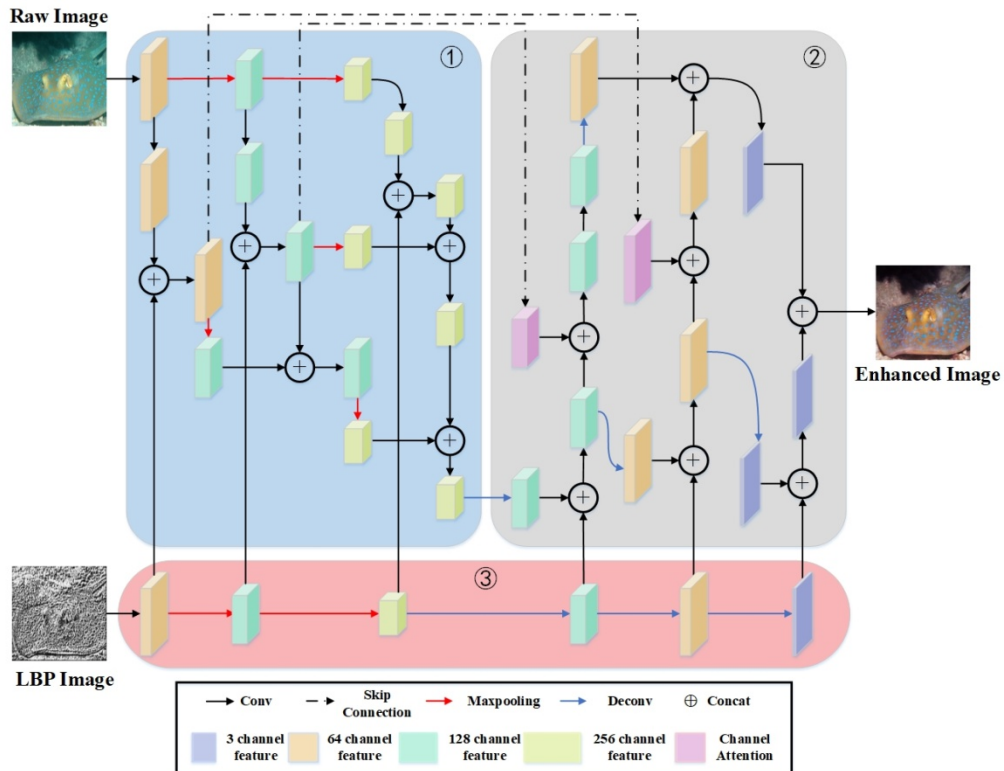


Fig. 2. The overall framework figure of the network proposed in this paper. ① represents the feature aggregation module, ② represents the image reconstruction module, and ③ represents the LBP enhancement module. The specific structure of each module is represented in Fig. 3, and the specific structure of the channel attention mechanism is shown in Fig. 5.

3.2 Image Reconstruction Module

The input of this module is the feature aggregation result of the previous module, from which the enhanced image result with the original scale size needs to be recovered. After three channel compression processes, the compression results of 128-, 64-and 3-channels are obtained respectively. The feature scale is also enlarged with the first two channel compression, and the resolution is restored to the original input. Consistent with the previous module, the LBP feature is superimposed with the features after the three channel changes, and then the two are aggregated by convolution. In addition, after the compression results of 128-and 64-channels are aggregated with the LBP features, the following skip connection is made:

$$F_{sc}(X) = \text{Conv}[F_{ca}(X_{ds}) \oplus X_{us}], \quad (3)$$

where $F_{sc}(X)$ represents the skip-connection result, $F_{ca}(X_{ds})$ represents the down-sampling module mapping through the channel attention module, X_{us} represents the up-sampling module feature, \oplus represents the stacking of feature channels, and Conv represents the convolution operation after the stacking to continue to maintain the original channel. Here, the network gets 128-channel and 64-channel skip-connection results. The former result will undergo an up-sampling process after a layer of convolution, and the result after aggregation with the latter result will undergo another convolution process to obtain the convolution result of 3-channel, and then perform the aggregation operation with the 3-channel aggregation result mentioned before, and output the final result of 3-channel by convolution.

② in Fig. 3 shows the components of this module in detail, which are 8 convolutional blocks and 4 up-sampling blocks, the structure of the convolutional block is consistent with the feature aggregation. The up-sampling block consists of a deconvolution layer and a ReLU activation function, and the convolution kernel number of deconvolution layer is set to 2, and the step size is 2. After multiple up-sampling operations, the feature map is restored to the size of the input image, and finally an enhanced underwater image with the correct color is obtained.

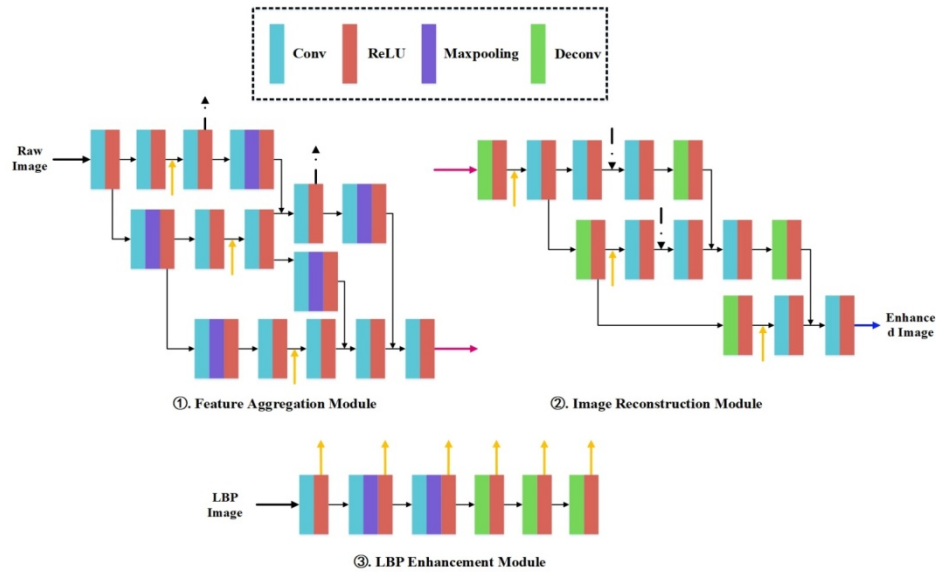


Fig. 3. The specific structure figure of the three main modules. The inputs of feature aggregation module are the raw underwater image and the convolved LBP feature image. The input of image reconstruction module is the aggregation result, the convolved LBP feature image and the channel attention output. The input of LBP enhancement module is the raw image corresponding LBP image.

3.3 LBP Enhancement Module

LBP is an operator that describes the texture features of an image, which has the characteristics of gray invariance. Therefore, the change of gray value in the region caused by illumination does not affect the description of the features, and the underwater environment often presents uneven illumination, so the light insensitive characteristic of LBP is applied to the whole process of network enhancement. Fig. 4 shows the LBP images corresponding to the original underwater image and the enhanced underwater image respectively, and the LBP features before and after are almost identical, and the proposed network takes advantage of this feature to make the underwater image enhancement more effective.

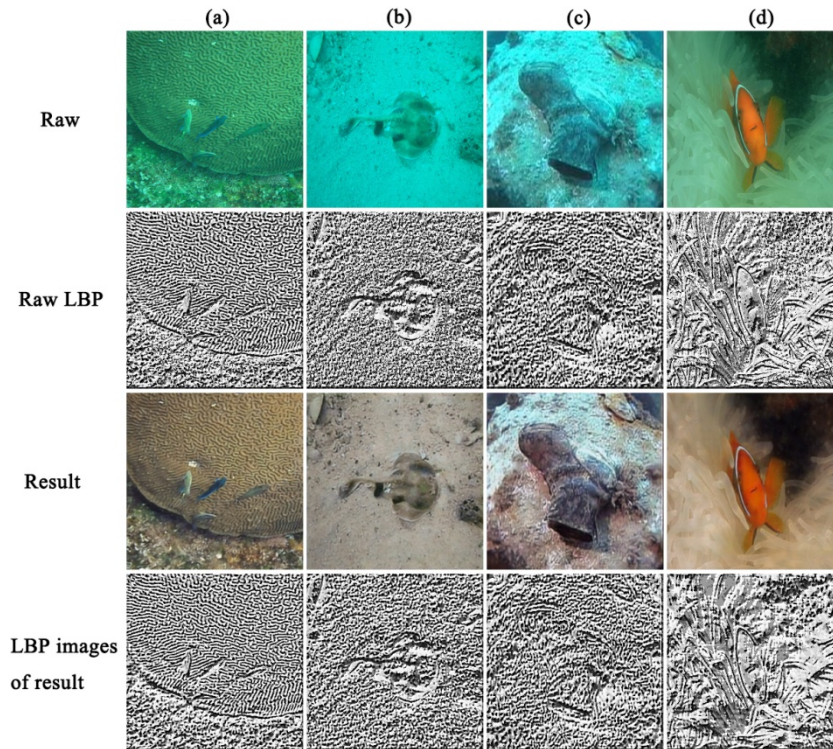


Fig. 4. Comparison of the raw underwater images with the result images of this paper and the corresponding LBP images

The early LBP extracts the neighborhood relationship by setting the center gray value as the threshold, and comparing the gray value of the neighboring pixels with the threshold to obtain the eight-bit binary number, which corresponds to $2^8=256$ LBP codes. The binary number is converted into a decimal number to represent the texture information of the region. In this paper, the LBP with circular neighborhood is used to remove the limitations on the number of samples and the size of the area. The formula is as following:

$$LBP_{P,R}(x_c, y_c) = \sum_{i=0}^{P-1} 2^i S(g_i - g_c), \#(4)$$

$$S(x) = \begin{cases} 1, & \text{if } x \geq 0 \\ 0, & \text{otherwise} \end{cases}, \#(5)$$

where (x_c, y_c) represents the center point coordinates, P is the number of all sampled points, R represents the neighborhood radius, g_i represents the gray value of the i -th neighbor point on

the circle with radius R , and g_c represents the gray value of the center point.

The LBP image is a single-channel gray scale image, which is first stacked into a three-channel image as the original input in order to enable normal aggregation. The feature information in the network undergoes three down-sampling and three up-sampling processes, so the LBP images also undergo the same process and aggregate with them when the image resolution changes. The aggregation formula is as following:

$$F_a(X) = \text{Conv}[X_U \oplus F_L(X_L)], \#(6)$$

where $F_a(x)$ represents the aggregation result, X_U represents the underwater image feature extracted from the network, and $F_L(X_L)$ is the LBP image feature after the scale change. The aggregation result still maintains the original feature shape through convolution operation. The introduction of LBP protects image detail information when the network changes scale, and the enhanced image output by the network has better visual performance.

3.4 Channel Attention Mechanism

To deal with the loss of feature detail when the image scale changes, the feature information before down-sampling is mapped to the image reconstruction module to make skip connection with the feature map before up-sampling, and the feature details of the underwater image enhancement results are more complete. Since the feature maps in the network are multi-channel, and the information contained in different channels contributes to the key information differently. Therefore, the channel attention mechanism is introduced in the network to add weight to the information on each channel. Fig. 5 shows the detailed structure of the channel attention mechanism.

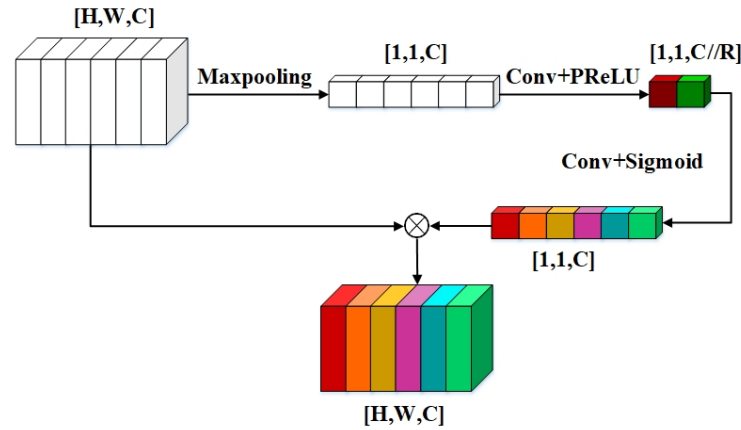


Fig. 5. Structure of channel attention mechanism. H and W are the height and width of input and output feature images, the number of channels and compression factor are C and R respectively, "⊗" indicates matrix multiplication, and the different colors of each channel represent their different weights.

The channel attention mechanism focuses on the overall information of each feature channel, while ignoring the local information in the channel, thus discriminating the overall importance of each channel. Therefore, the feature information first goes through the adaptive average pooling process, which compresses the local information in the channel into real numbers. The size of the feature map changes from $H \times W$ to 1×1 , while the number of channels does not change. Next, the learning weight information is trained by several processes of convolution layer, PReLU, and Sigmoid activation function, and the real number within the channel is adjusted according to the learning results. The formula is as following:

$$W_c = \sigma(W_2 \left(\gamma \left(W_1(W_{avg}) \right) \right)), \quad (7)$$

$$f_c = W_c \cdot f_0, \#(8)$$

where W_c represents the weight of each channel obtained by learning, σ and γ represent the sigmoid and ReLU activation function, W_1 and W_2 represent the weight information in the learning process, W_{avg} represents the initial weight processed by adaptive averaging pooling, and the final result f_c is obtained by multiplying the learning weight W_c with the input feature f_0 , which is essentially multiplying different weights on different channels, so that the channels with more key information can play a more important role.

4. Experiments

4.1 Experimental Details

The experimental environment is a computer with NVIDIA GeForce GTX 2060, 16GB RAM and Intel i7-10700 CPU. And the method is built in Python based on the Pytorch framework. The dataset in this paper is obtained from the underwater dataset provided by Fabbri et al [31], which contains 6128 real color images and the corresponding underwater style images generated by the CyCleGAN network. The first 90%, a total of 5515 pairs of images, are used as training data and validation data, where the training data contains 5000 pairs of image sets and the validation data contains 515 pairs of images. The test datasets are dataset with reference standard and dataset without reference standard. The reference dataset comes from the last 10% of Fabbri underwater dataset, totaling 613 pairs of images. The non-reference dataset from Imagenet [32] totaling 1813 real underwater color skewed images. The mean square error (MSE) was chosen as the training loss function, its formula is as follows:

$$L_{MSE}(I_e, I_t) = \frac{1}{mn} \sum_{i=1}^{m-1} \sum_{j=1}^{n-1} [I_e(i, j) - I_t(i, j)]^2 \#(9)$$

where I_e is the enhancement result of the network, I' is the reference standard, and L_{MSE} is the cumulative square of the difference between the enhanced image and the corresponding reference image, according to which the network weights will be updated. And the validation part uses the index SSIM to measure the structure similarity between the enhanced image and the reference image. When the SSIM gets closer to 1, the similarity of the two images is higher.

4.2 Batch Selection

To determine the best training batches for real underwater images, a total of 200 batches of models were trained, of which every 10 batches were saved and their enhancements were compared on the test set. In this paper, two non-reference indexes, UIQM [33] and UCIQE [34] were used to objectively evaluate each batch models.

UIQM is an underwater image quality evaluation index, which is widely used in this field. A linear combination of three image quality measures is used to obtain an evaluation score. The higher UIQM score, the image is more comfortable for human vision. The formula is as follows:

$$UIQM = c1 * UICM + c2 * UISM + c3 * UIConM, \#(10)$$

where $UICM$ represents the image color degree, $UISM$ represents the image clarity, and $UIConM$ represents the image contrast. The three scale factors maintain the original suggested values: $c1 = 0.02821$, $c2 = 0.29532$, and $c3 = 3.57533$.

UCIQE is also a widely used underwater image quality evaluation index that measures chromaticity, saturation, and contrast. A higher score indicates a better balance between the three. Its formula is as follows:

$$UCIQE = c_1 \times \sigma_c + c_2 \times con_l + c_3 \times \mu_s, \#(11)$$

Fig. 6 shows the change process of UIQM and UCIQE values of each batch model on the test set. In the early training that has not converged, the enhancement effect of the network is not comprehensive, so there may be fluctuations in two indicators. We have extended the total batch to 200 to ensure that the network is fully trained. It can be seen that in the later batches, the two indicators have shown an overall downward trend, which indicates that the network has converged. The 100 batch obtained the best value of UIQM and the second-best value of UCIQE, while the 140 batch obtained the second-best value of UIQM. The 130 batch obtained the best value of UCIQE, but did not perform well in UIQM. Therefore, this paper decides to adopt 100 batches as the best batch, and compare this batch model with other underwater image restoration or enhancement methods.



Fig. 6. A line graph of the UIQM and UCIQE values changes for each batch model with respect to the test set, where the red points indicate the best and the blue points indicate the second best.

4.3 Comparison on Reference Dataset

On a validation set containing 613 images with reference standards, this paper presents a qualitative and quantitative comparison with seven different methods, including four underwater image enhancement/ restoration methods and three deep learning methods. These methods are: UIBLA [35] based on image blur and light absorption, RGHS [36] based on relative global histogram stretching, DMHP [4] based on histogram distribution prior and minimum information loss, L2UWE [37] using local contrast with multiscale fusion, UDIM [38] based on the underwater imaging model, MLFcGAN [39] based on conditional generation adversarial network, and AIODA [40] based on domain adversarial learning.

The enhanced results of each method are shown in **Fig. 7**. UIBLA, DMHP, and L²UWE all improve the brightness and contrast, but UIBLA and L²UWE deepen the color deviation phenomenon, and DMHP shows excellent color correction ability among the traditional methods, but the color deviation problem of the underwater image has not been completely resolved. RGHS deepens the black and white areas in the image by histogram stretching, although the image contrast has been improved, the color deviation is still more obvious. The color deviation situation is alleviated in the results produced by UDIM, but it can be seen in (h) that UDIM destroys the green color of the leaves and makes the leaves appear yellowish. The results produced by AIODA are darker overall, and the color deviation still has a non-negligible effect. MLFcGAN and the proposed method not only solve the color deviation problem, but also retain the colors of other things in the scene well, so that the brightness and

contrast can be improved. However, the MLFcGAN results have the effect of yellow color, and the colors are not as bright as the results of this paper, which shows that the method of this paper is closer to the reference standard.

We further evaluated the quantitative evaluation of all the results, and two full-reference image quality evaluation indexes, PSNR and SSIM, were used to compare the performance of all methods. The average indicators and running time of all methods on the dataset with reference are shown in [Table 1](#). The proposed method in this paper achieved the best score in both metrics, and MLFcGAN achieved the second-best score in both metrics. In terms of running time, UDIM adopts a smaller number of convolutional kernels and thus exhibits the fastest speed, while AIODA achieves the second fastest speed by encoding to reduce the generator operations. Due to the introduction of more complex model parameters, the operation speed of UIBLA and L^2UWE is affected. Our network focuses on the compensation of information loss in the encoding-decoding process, so the speed of our method is not outstanding.

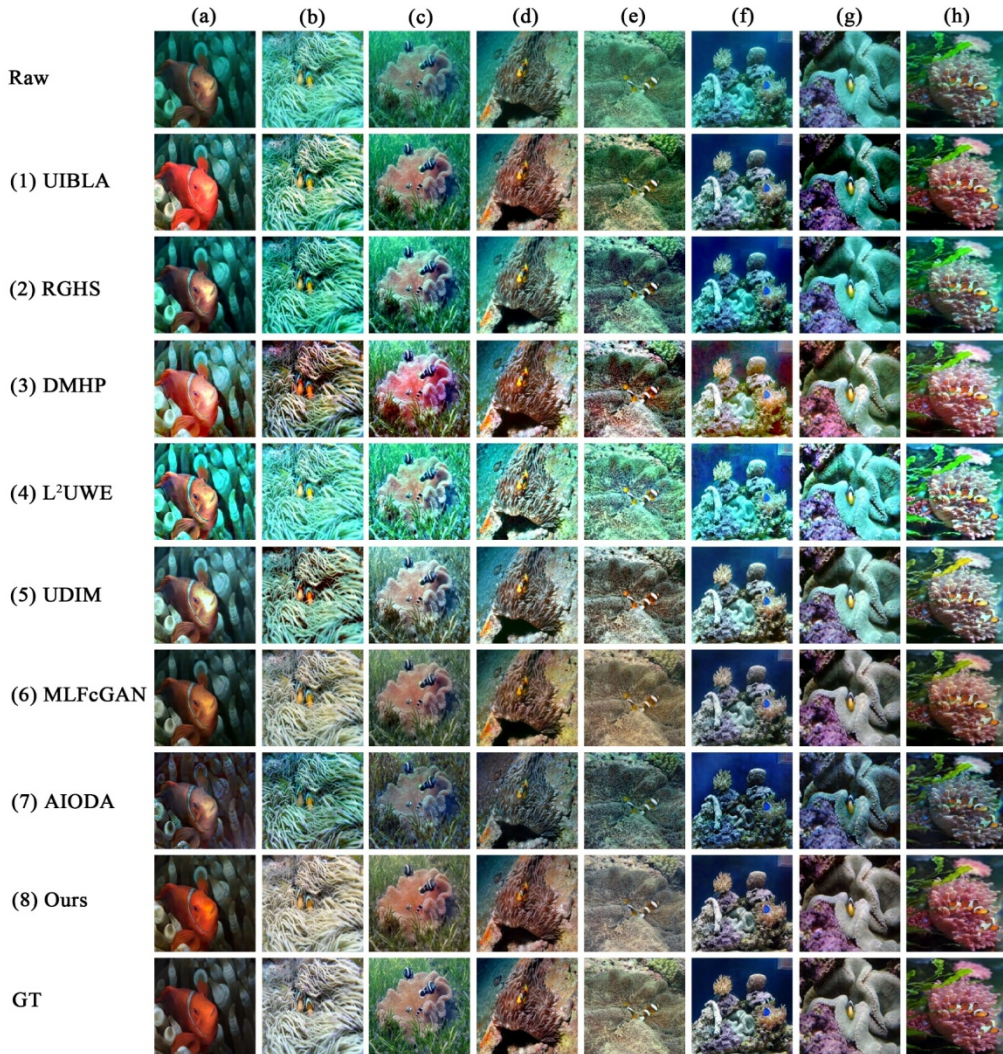


Fig. 7. Subjective comparisons on the reference dataset. The images from top to bottom are raw images, the results of UIBLA[35], RGHS[36], DMHP[4], L^2UWE [37], UDIM[38], MLFcGAN[39], AIODA[40], the proposed method and ground truth.

Table 1. The average results of PSNR and SSIM and running time for all method on the reference validation dataset. The red values represent the best results, the blue values represent the second-best.

Methods	PSNR	SSIM	Running time(s)
UIBLA	20.0612	0.7161	3062
RGHS	22.9738	0.8263	474
DMHP	17.9850	0.7100	144
L ² UWE	14.2027	0.6304	1917
UDIM	20.1607	0.7809	12
MLFcGAN	25.3435	0.8482	200
AIODA	17.6478	0.7341	24
Ours	27.0611	0.8822	236

4.4 Comparison on Non-reference Dataset

This paper also conducts a comparative experiment on the test set of non-reference images to demonstrate the performance of the method in real underwater images. Fig. 8 shows 8 underwater original images and the corresponding results of all methods. The results produced by UIBLA, RGHS, and L²UWE still have a serious blue-green color influence, and DMHP has a very obvious red color deviation. The results produced by UDIM and MLFcGAN have excellent performance in terms of color and brightness, but there is still the influence of blue-green color deviation that cannot be eliminated. AIODA enhances the performance of dark colors in the picture, resulting in improved contrast, but the image becomes darker due to this, and the result is severely affected by the red color. The method in this paper produces results that not only correctly corrected the blue-green color deviation, but also did not introduce the influence of other colors, with excellent clarity and visibility. As can be seen from the comparison of the scene object details in Fig. 11, the method in this paper is the best at restoring the colors of the scene objects themselves, and the detailed information of the image is also well preserved.

Table 2. The average results of UIQM and UCIQE and running time for all method on the no-reference dataset. The red values represent the best results, the blue values represent the second-best.

Methods	UIQM	UCIQE	Running time(s)
UIBLA	1.9519	0.5228	8900
RGHS	2.3628	0.5709	1486
DMHP	2.7271	0.5181	367
L ² UWE	2.2078	0.5252	2727
UDIM	2.8086	0.6396	21
MLFcGAN	2.8732	0.6257	305
AIODA	2.7983	0.5958	39
Ours	3.0322	0.6748	378

Fig. 9 and Fig. 10 show the UIQM and UCIQE evaluation values for all the results of Fig. 8. The method in this paper achieves the best results on UIQM in most cases, and MLFcGAN achieves the second best UIQM values. Regarding the UCIQE values, UDIM, MLFcGAN and the proposed method all perform well. The mean UIQM and UCIQE and the running time of all methods on the no-reference dataset are shown in Table 2, and the proposed method of this paper has the best scores in both indexes, which indicates that the proposed method can

generate results with better color, saturation, and contrast. In terms of running time, UDIM and AIODA are still the shortest and second shortest methods.

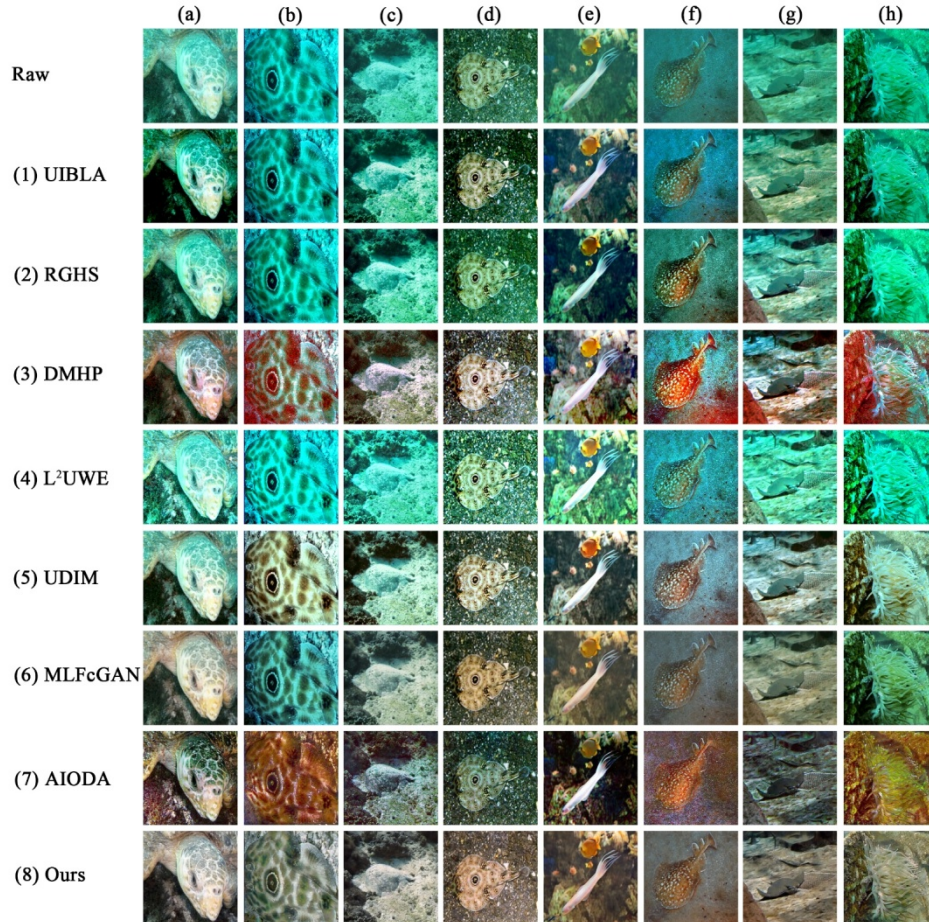


Fig. 8. Subjective comparisons on the no-reference dataset. The images from top to bottom are raw images, the results of UIBLA[35], RGHS[36], DMHP[4], L²UWE[37], UDIM[38], MLFcGAN[39], AIODA[40], the proposed method.

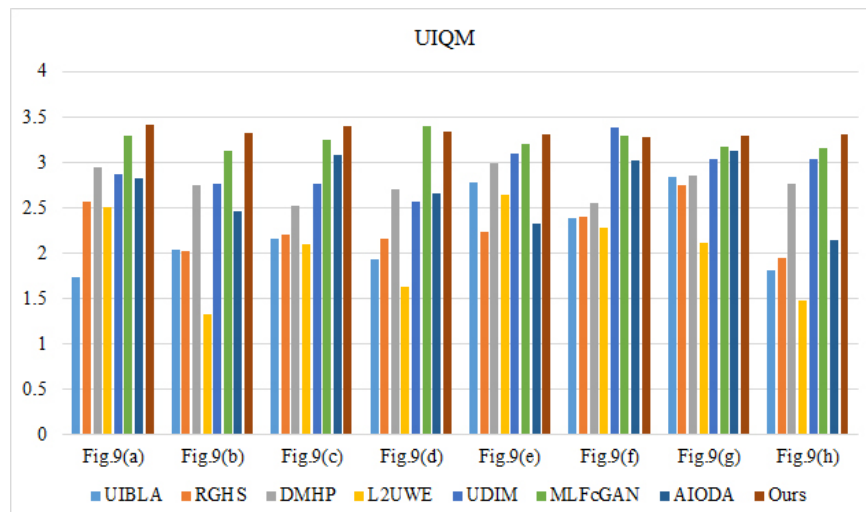


Fig. 9. The UIQM histogram of all results in Fig. 8

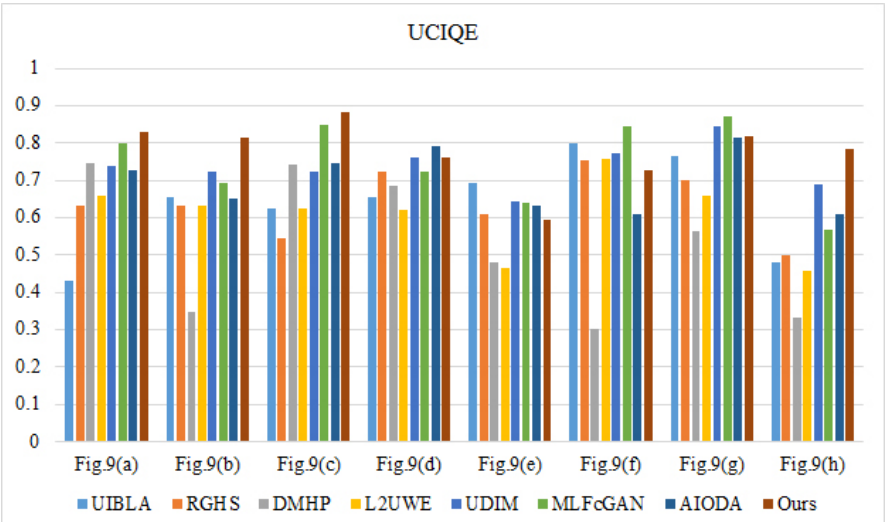


Fig. 10. The UCIQE histogram of all results in Fig. 8

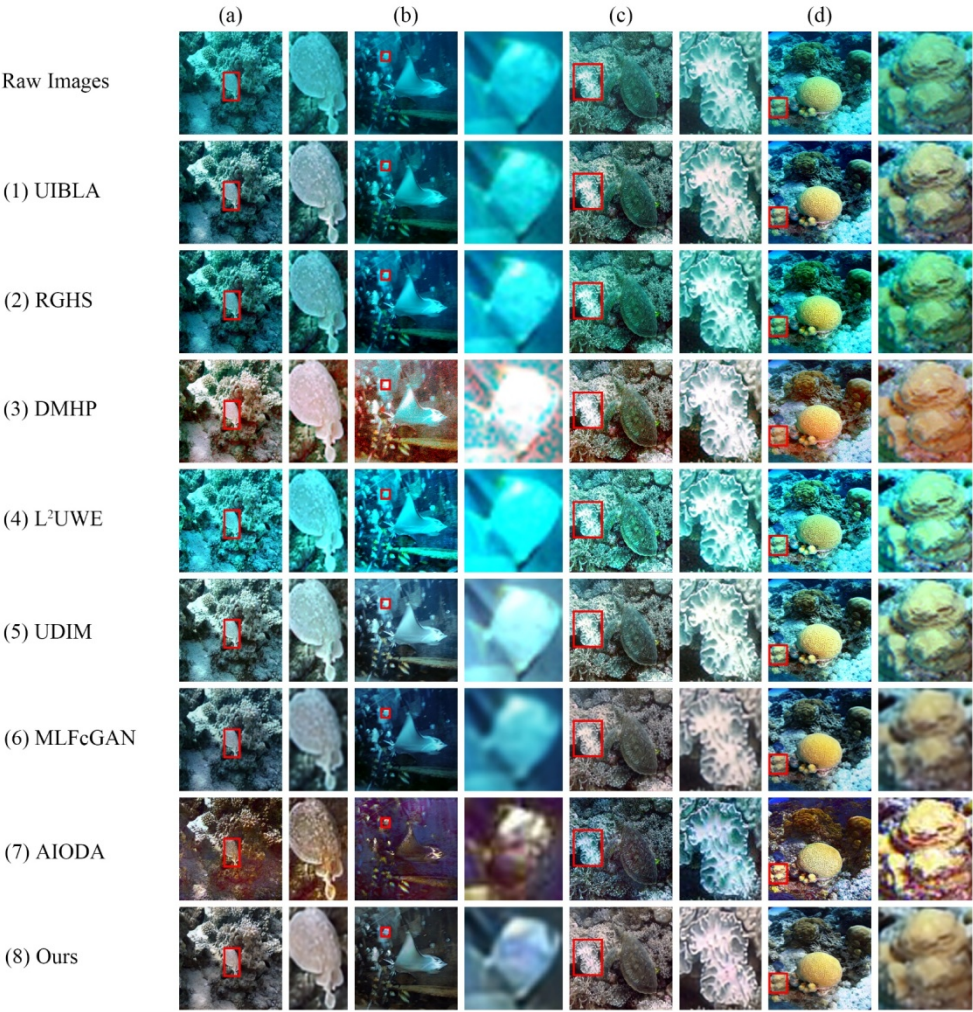


Fig. 11. Comparison of scene details of all methods results

4.5 Ablation Study

This paper also conducted an ablation study to verify the effectiveness of all network components, with an experimental dataset from the non-reference underwater image dataset above, in the following four configurations: (1) Original feature aggregation and image reconstruction network (Ori Net), (2) the method of this paper without channel attention (-w/o CA), (3) the method of this paper without LBP (-w/o LBP), (4) the method of this paper with complete components (Ours). **Fig. 12** shows an enlarged comparison of the results for different components. The method in this paper generates more natural colors with complete components. There are color deviation blocks that are not eliminated in Ori Net and -w/o CA, and -w/o LBP deals with the effect of color deviation blocks comprehensively, but there are texture defects visible to the naked eye. The proposed method not only does not suffer from color deviation, but also has complete texture, correct color, and strong contrast. **Table 3** presents the quantitative comparison of UIQM and UCIQE for all configurations. In all configurations, the method with complete components achieves the best scores on the two indexes, which further proves that each component is beneficial to this network.

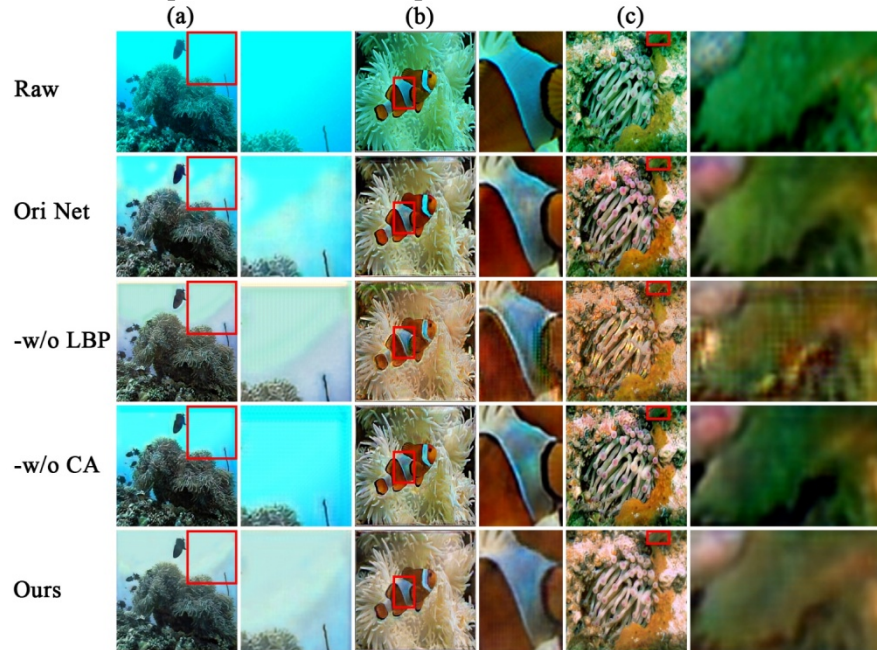


Fig. 12. Detail comparison figure of different components

Table 3. The average results of UIQM and UCIQE for all configurations on the no-reference test dataset. The red values represent the best results, the blue values represent the second-best.

Configuration	Average UIQM	Average UCIQE
Ori Net	2.8959	0.6260
-w/o CA	2.9758	0.6348
-w/o LBP	2.9415	0.6525
Ours	3.0322	0.6748

5. Conclusion

In this paper, an attention-based for multiscale fusion network is proposed for color correction and detail protection of underwater images. The network performs feature fusion on multiple scales of the image, and obtains results that conform to the human visual color after image reconstruction. Different from other methods, this paper introduces the LBP image to protect the detailed information of underwater image in real time, extracts the features of original image and LBP image, and the whole enhancement process has richer feature input. The network introduces channel attention mechanism to make important channels more concerned. Comparative experiments were carried out on the reference dataset and the non-reference dataset. Compared with several different methods, the proposed method has a better comprehensive expression in color, saturation and contrast, and it also achieves outstanding performance in quantitative evaluation. This paper also conducts an ablation study of the relevant modules, which proves that the channel attention mechanism and LBP mechanism used in the network improve the network performance.

In addition, the method in this paper still has some shortcomings. First of all, the resulting image still lacks clarity, which affects the quality of our results. Second, the colors of some scene creatures are not vivid enough in the generated results. Third, the proposed method is not optimal in terms of speed. In future work, the network will be improved to solve existing image quality problems. In addition, considering the running speed, the new network model will have a more lightweight structure and may be applied in other fields.

Acknowledgement

The authors acknowledge the National Natural Science Foundation of China (Grant No. 61772319, 62002200, 62176140), and Shandong Natural Science Foundation of China (Grant No. ZR2020QF012, ZR2021MF068).

References

- [1] Sousa. J. P, Ferreira. B. M, Cruz. N. A, "Guidance of an autonomous surface vehicle for underwater navigation Aid," in *Proc. of 2018 IEEE/OES Autonomous Underwater Vehicle Workshop*, pp. 1–6, 2018. [Article \(CrossRef Link\)](#)
- [2] Petillot. Y. R, Antonelli. G, Casalino. G, et al, "Underwater robots: From remotely operated vehicles to intervention-autonomous underwater vehicles," *IEEE Robotics & Automation Magazine*, vol. 26, no. 2, pp. 94–101, 2019. [Article \(CrossRef Link\)](#)
- [3] Zhang. Z, Wu. Y, "Adaptive fuzzy tracking control of autonomous underwater vehicles with output constraints," *IEEE Transactions on Fuzzy Systems*, vol. 29, no. 5, pp. 1311–1319, 2021. [Article \(CrossRef Link\)](#)
- [4] Li. C. Y, Guo. J. C, Cong. R. M, et al, "Underwater image enhancement by dehazing with minimum information loss and histogram distribution prior," *IEEE Transactions on Image Processing*, vol. 25, no. 12, pp. 5664–5677, 2016. [Article \(CrossRef Link\)](#)
- [5] Hu. H, Zhao. L, Li. X, et al, "Underwater image recovery under the nonuniform optical field based on polarimetric imaging," *IEEE Photonics Journal*, vol. 10, no. 1, pp. 1–9, 2018. [Article \(CrossRef Link\)](#)
- [6] Akkaynak. D, Treibitz. T, "A revised underwater image formation model," in *Proc. of the IEEE conference on computer vision and pattern recognition*, pp. 6723–6732, 2018. [Article \(CrossRef Link\)](#)

- [7] Akkaynak. D, Treibitz, T, “Sea-thru: A method for removing water from underwater images,” in *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1682–1691, 2019. [Article \(CrossRef Link\)](#)
- [8] Sánchez-Ferreira. C, Coelho. L. S, Ayala. H. V. H, et al, “Bio-inspired optimization algorithms for real underwater image restoration,” *Signal Processing: Image Communication*, vol. 77, no. 1, pp. 49–65, 2019. [Article \(CrossRef Link\)](#)
- [9] Dai. C, Lin. M, Wu, X, et al, “Single underwater image restoration by decomposing curves of attenuating color,” *Optics & Laser Technology*, vol. 123, no. 1, 2020, Article No. 105947. [Article \(CrossRef Link\)](#)
- [10] Berman. D, Levy. D, Avidan. S, et al, “Underwater Single Image Color Restoration Using Haze-Lines and a New Quantitative Dataset,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 8, pp. 2822–2837, 2021. [Article \(CrossRef Link\)](#)
- [11] Kar. A, Dhara. S. K, Sen. D, et al, “Zero-Shot Single Image Restoration Through Controlled Perturbation of Koschmieder's Model,” in *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 16205–16215, 2021. [Article \(CrossRef Link\)](#)
- [12] Ancuti. C. O, Ancuti. C, Vleeschouwer. C. D, et al, “Color balance and fusion for underwater image enhancement,” *IEEE Transactions on image processing*, vol. 27, no. 1, pp. 379–393, 2018. [Article \(CrossRef Link\)](#)
- [13] Gao. S. B, Zhang. M, Zhao. Q, et al, “Underwater image enhancement using adaptive retinal mechanisms,” *IEEE Transactions on Image Processing*, vol. 28, no. 11, pp. 5580–5595, 2019. [Article \(CrossRef Link\)](#)
- [14] Ancuti. C. O, Ancuti. C, De. Vleeschouwer. C, et al, “Color channel compensation (3C): A fundamental pre-processing step for image enhancement,” *IEEE Transactions on Image Processing*, vol. 29, no. 1, pp. 2653–2665, 2019. [Article \(CrossRef Link\)](#)
- [15] Hegde. D, Desai. C, Tabib. R, et al, “Adaptive cubic spline interpolation in CIELAB color space for underwater image enhancement,” *Procedia Computer Science*, vol. 171, no. 1, pp. 52–61, 2020. [Article \(CrossRef Link\)](#)
- [16] W. Song, Y. Wang, D. Huang, et al, “Enhancement of underwater images with statistical model of background light and optimization of transmission map,” *IEEE Transactions on Broadcasting*, vol. 66, no. 1, pp. 153–169, 2020. [Article \(CrossRef Link\)](#)
- [17] Liang. Z, Wang. Y, Ding. X, et al, “Single underwater image enhancement by attenuation map guided color correction and detail preserved dehazing,” *Neurocomputing*, vol. 425, no. 1, pp. 160–172, 2021. [Article \(CrossRef Link\)](#)
- [18] Zhang. W, Dong. L, Zhang. T, et al, “Enhancing underwater image via color correction and bi-interval contrast enhancement,” *Signal Processing: Image Communication*, vol. 90, no. 1, 2021, Article No. 116030. [Article \(CrossRef Link\)](#)
- [19] Chen. C, Chen. Q, Xu. J, et al, “Learning to see in the dark,” in *Proc. of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3291–3300, 2018. [Article \(CrossRef Link\)](#)
- [20] Li. C, Cong. R, Hou. J, et al, “Nested network with two-stream pyramid for salient object detection in optical remote sensing images,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, pp. 11, pp. 9156–9166, 2019. [Article \(CrossRef Link\)](#)
- [21] Ren. W, Liu. S, Ma. L, et al, “Low-light image enhancement via a deep hybrid network,” *IEEE Transactions on Image Processing*, vol. 28, no. 9, pp. 4364–4375, 2019. [Article \(CrossRef Link\)](#)
- [22] Guo. C, Li. C, Guo. J, et al, “Zero-reference deep curve estimation for low-light image enhancement,” in *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1780–1789, 2020. [Article \(CrossRef Link\)](#)
- [23] Li. J, Wang. N, Zhang. L, et al, “Recurrent Feature Reasoning for Image Inpainting,” in *Proc. of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7760–7768, 2020. [Article \(CrossRef Link\)](#)
- [24] Li. J, Skinner. K. A, Eustice. R. M, et al, “WaterGAN: Unsupervised generative network to enable real-time color correction of monocular underwater images,” *IEEE Robotics and Automation letters*, vol. 3, no. 1, pp. 387–394, 2018. [Article \(CrossRef Link\)](#)

- [25] Li. C, Guo. J, Guo. C, "Emerging from water: Underwater image color correction based on weakly supervised color transfer," *IEEE Signal processing letters*, vol. 25, no. 3, pp. 323–327, 2018. [Article \(CrossRef Link\)](#)
- [26] Hou. M, Liu. R, Fan. X, et al, "Joint residual learning for underwater image enhancement," in *Proc. of 2018 25th IEEE International Conference on Image Processing*, pp. 4043–4047, 2018. [Article \(CrossRef Link\)](#)
- [27] Silberman. N, Hoiem. D, Kohli. P, et al, "Indoor segmentation and support inference from rgb-d images," in *Proc. of European Conference on Computer Vision*, Springer, Berlin, Heidelberg, pp. 746–760, 2012. [Article \(CrossRef Link\)](#)
- [28] Jaffe. J. S, "Computer modeling and the design of optimal underwater imaging systems," *IEEE Journal of Oceanic Engineering*, vol. 15, no. 2, pp. 101–111. 1990. [Article \(CrossRef Link\)](#)
- [29] Jamadandi. A, Mudanagudi. U, "Exemplar-based underwater image enhancement augmented by wavelet corrected transforms," in *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp. 11–17, 2019. [Article \(CrossRef Link\)](#)
- [30] Desai. C, Tabib. R. A, Reddy. S. S, et al, "RUIG: Realistic Underwater Image Generation Towards Restoration," in *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp. 2181–2189, 2021. [Article \(CrossRef Link\)](#)
- [31] Fabbri. C, Islam. M. J, Sattar. J, "Enhancing underwater imagery using generative adversarial networks," in *Proc. of 2018 IEEE International Conference on Robotics and Automation*, pp. 7159–7165, 2018. [Article \(CrossRef Link\)](#)
- [32] Deng. J, Dong. W, Socher. R, et al, "Imagenet: A large-scale hierarchical image database," in *Proc. of 2009 IEEE conference on computer vision and pattern recognition*, pp. 248–255, 2009. [Article \(CrossRef Link\)](#)
- [33] Panetta. K, Gao. C, Aguin. S, "Human-visual-system-inspired underwater image quality measures," *IEEE Journal of Oceanic Engineering*, vol. 41, no. 3, pp. 541–551, 2016. [Article \(CrossRef Link\)](#)
- [34] Yang. M, Sowmya. A, "An underwater color image quality evaluation metric," *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 6062–6071, 2015. [Article \(CrossRef Link\)](#)
- [35] Peng. Y. T, Cosman. P. C, "Underwater image restoration based on image blurriness and light absorption," *IEEE transactions on image processing*, vol. 26, no. 4, pp. 1579–1594, 2017. [Article \(CrossRef Link\)](#)
- [36] Huang. D, Wang. Y, Song. W, et al, "Shallow-water image enhancement using relative global histogram stretching based on adaptive parameter acquisition," in *Proc. of International conference on multimedia modeling*, Springer, Cham, pp. 453–465, 2018. [Article \(CrossRef Link\)](#)
- [37] Marques. T. P, Albu. A. B, "L²uwe: A framework for the efficient enhancement of low-light underwater images using local contrast and multi-scale fusion," in *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp. 538–539, 2020. [Article \(CrossRef Link\)](#)
- [38] Chen. X, Zhang. P, Quan. L, et al, "Underwater Image Enhancement based on Deep Learning and Image Formation Model," *arXiv:2101.00991v2*, 2021. [Article \(CrossRef Link\)](#)
- [39] Liu. X, Gao. Z, Chen. B. M, "MLFcGAN: Multilevel feature fusion-based conditional GAN for underwater image color correction," *IEEE Geoscience and Remote Sensing Letters*, vol. 17, no. 9, pp. 1488–1492, 2020. [Article \(CrossRef Link\)](#)
- [40] Uplavikar. P. M, Wu. Z, Wang. Z, "All-in-One Underwater Image Enhancement Using Domain-Adversarial Learning," in *Proc. of CVPR Workshops*, pp. 1–8, 2019. [Article \(CrossRef Link\)](#)



ZhiXiong Huang received his bachelor's degree in digital media technology from Guilin University of Electronic Technology, Guilin, China in 2020. Currently studying for a master's degree in the School of Information and Electronic Engineering, Shandong Technology and Business University, Yantai, Shandong. His research interests include computer vision, image processing, and machine learning.



Jinjiang Li received the B.S. and M.S. degrees in computer science from Taiyuan University of Technology, Taiyuan, China, in 2001 and 2004, respectively, the Ph.D. degree in computer science from Shandong University, Jinan, China, in 2010. From 2004 to 2006, he was an assistant research fellow at the institute of computer science and technology of Peking University, Beijing, China. From 2012 to 2014, he was a Post-Doctoral Fellow at Tsinghua University, Beijing, China. He is currently a Professor at the school of computer science and technology, Shandong Technology and Business University. His research interests include image processing, computer graphics, computer vision and machine learning.



Zhen Hua received the B.S. and M.S. degrees in electrical automation from Taiyuan University of Technology, Taiyuan, China, in 1989 and 1992, respectively, the Ph.D. degree in electronic information engineering from China University of Mining and Technology, Beijing, China, in 2008. She is currently a professor at Shandong Technology and Business University. Her research interests include computer aided geometric design, information visualization, virtual reality and image processing.