Recovery of underwater images based on the attention mechanism and SOS mechanism

Shiwen Li[1,2], Feng Liu[1,3*], and Jian Wei[1,3]
[1] School of Telecommunications and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210003, Jiangsu, China
[e-mail: lishiwen1015@126.com, liuf@njupt.edu.cn]
[2]School of Electronics and Information Engineering, Heyuan Polytechnic, Yuancheng District, Heyuan, 517000, China
[e-mail: lishiwen1015@126.com]
[3]Jiangsu Key Lab of Image Processing and Image Communications, No. 66 Xin Mofan Road, Nanjing(210003), Jiangsu, China
[e-mail: liuf@njupt.edu.cn]
[*]Corresponding author: Feng Liu

Abstract

Underwater images usually have various problems, such as the color cast of underwater images due to the attenuation of different lights in water, the darkness of image caused by the lack of light underwater, and the haze effect of underwater images because of the scattering of light. To address the above problems, the channel attention mechanism, strengthen-operate-subtract (SOS) boosting mechanism and gated fusion module are introduced in our paper, based on which, an underwater image recovery network is proposed. First, for the color cast problem of underwater images, the channel attention mechanism is incorporated in our model, which can well alleviate the color cast of underwater images. Second, as for the darkness of underwater images, the similarity between the target underwater image after dehazing and color correcting, and the image output by our model is used as the loss function, so as to increase the brightness of the underwater image. Finally, we employ the SOS boosting module to eliminate the haze effect of underwater images. Moreover, experiments were carried out to evaluate the performance of our model. The qualitative analysis results show that our method can be applied to effectively recover the underwater images, which outperformed most methods for comparison according to various criteria in the quantitative analysis.

1. Introduction

During continuous explorations of the ocean, the mankind needs to obtain related data of the ocean, while the underwater images are an indispensible source of information. During this process, it is essential to detect the underwater environment by using an underwater optical imaging system. However, the underwater environment is very complicated in general, involving various factors, such as lack of light, diverse types of water bodies and suspended impurities in water. All these factors will significantly affect the results of underwater imaging. As a result, the images captured underwater generally have the problems of low contrast, color distortion and lack of illumination. Therefore, how to effectively process the underwater images has become a popular research subject recently.

At present, researchers try to process the underwater target images mainly using the following three types of methods. The first type is the enhancement method for underwater images, which mainly includes methods such as white balance and histogram equalization. The second type is the recovery method for underwater images, which mainly involves the optical model of underwater images, and the imaging model of underwater image is used for inference, so as to recover the underwater image. The third type is the deep learning method, and with the development of deep learning in recent years, it has played an increasingly more important role in image processing.

These methods have achieved great performances in improving the visual effects of underwater images. However, there are still some problems in the underwater images restored by these methods. For example, color distortion tends to be generated when the underwater images are processed by using the conventional methods. In the meantime, the conventional methods generally have poor robustness, so they may have good performance in recovering some images, while showing poor performance in processing some other images. On the other hand, the deep learning method has poor haze removal ability when processing the underwater images, and it is also difficult to obtain paired datasets. Therefore, the underwater image datasets are mostly synthetic. For example, Fabbri et al. utilized GAN to generate paired underwater image datasets, and in recent years, many researchers have used these datasets to train their networks. We also trained our network using these synthetic datasets.

In this paper, to address various problems of underwater image, such as low brightness, haze effect and color distortion, a deep learning model is designed, which combines the attention mechanism, SOS boosting module and gated fusion module. Moreover, the exponentiated mean local variance, dehazing and color correcting methods are introduced as the loss function of this model, so as to further improve the quality of underwater images. Our goal is not only to maintain the texture and details of the underwater image, but also to improve its brightness and contrast. **Fig. 1** shows the underwater images restored by using our method, from which it can be seen that our method can improve the visual effects of the underwater images.
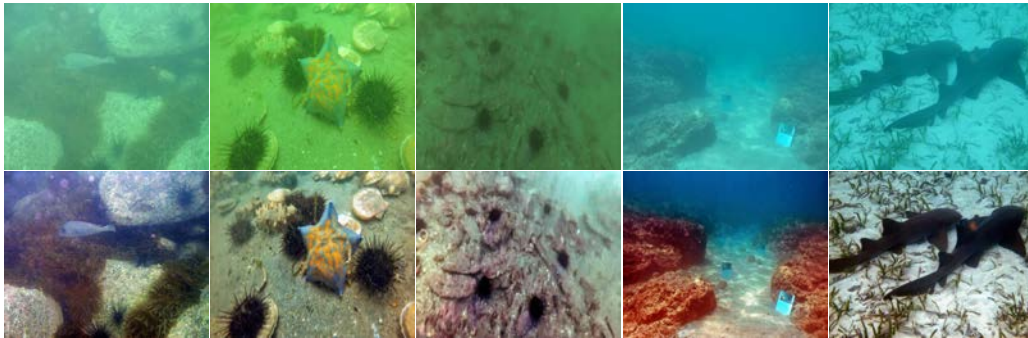
Fig. 1. Underwater images recovered with our method. The original image (top) and the recovered by our method (bottom).

The contributions of this paper are as follows:

(1) The channel attention mechanism is applied to process the underwater images, and we also made improvement to this mechanism for our research. The channel attention mechanism is combined with the SOS boosting module to dehaze the underwater images. In the meantime, the dilated convolution and gated fusion modules are also incorporated to further improve the recovery effectiveness of underwater images.

(2) By using the exponentiated mean local variance as the loss function of model, the denoising of underwater images is carried out while also keeping the edge details of underwater images.

(3) Because the underwater images are dark, the similarity between the underwater target image after dehazing and color correction and the image output by our model is used as the loss function, through which the brightness of the underwater image can be improved.

(4) The experimental results show that the method proposed in this paper can improve the visual effects of the underwater images, and the evaluation indexes adopted in this work include the underwater image quality measurement (UIQM), structural similarity index measurement (SSIM) and peak signal-to-noise ratio (PSNR).

This paper is structured as follows. Section 2 is a brief introduction of related works. The network architecture and loss function are elaborated in Section 3, which also describes the implementation of the algorithm. The qualitative and quantitative analyses are carried out in Section 4, and Section 5 draws conclusions of our work and presents possible future work.

2. Related Works

At current stage, many researchers are investigating the process of underwater images. The methods for processing underwater images include the restoration of underwater images with the optical model, the enhancement of underwater images not using any optical model and the processing of underwater images based on deep learning. Related works of these three methods are summarized as follows.

2.1 Studies on the recovery of underwater images based on the optical model

Because the underwater images tend to have haze effect, the land defogging model was directly applied to dehaze the underwater images in the early works on recovery of underwater images. The obtained underwater image recovery model is as shown in Equation (1).

$$(A^c - I^c(x)) = (A^c - J^c(x))t^c(x) \qquad (1)$$

where, A is the background light, c represents the three color channels of red, green and blue, I(x) is the observed image, J(x) is the image that needs to be recovered, and t(x) is the transmission map of underwater image.

In Equation (1), the most critical step is to obtain the transmission map of the underwater image. For estimating the transmission map of underwater image, Chao et al. [1] calculated the transmission map of underwater image using the dark channel prior method proposed by He et al., and then combined the dehazing model in Equation (1) to recover the underwater images. This method fails to consider the attenuation rates of different lights in the underwater images, so the estimated underwater images tend to involve serious color distortions. On the basis of the dark channel theory, Song et al. [2] proposed a new method for estimating the transmission map, which is named the new dark channel prior (NUDCP). Zhou et al.[3] modified the model on the basis of the scene depth map, and their method can be used to estimate the transmission map of underwater image. Because the attenuation of red light is very significant in underwater image, the estimated red-channel image involves serious deviation. In the meantime, during calculation of the transmission map, the background light and other parameters of underwater image need to be estimated, while the estimation of these parameters is very difficult. Therefore, these methods present low robustness during the recovery of underwater images, and have poor performance in processing some underwater images.

2.2 Enhancement of underwater images without the optical model

The enhancement methods for underwater images without the optical model mainly include the histogram prior, Retinex and fusion method. For example, Hummel et al. [4] proposed the histogram equalization (HE) method for process of the underwater images. Zuiderveld et al. improved the HE method, and proposed the contrast limited adaptive histogram equalization (CLAHE) method [5] to process the underwater images. The images restored by directly using the above methods usually have various defects, such as color distortion, high image noise, haze effect and low contrast. Zhang et al. utilized the Retinex theory for processing of underwater images [6], and the brightness and color components are filtered by a bilateral filter and a trilateral filter, which can effectively inhibit halos and artifacts. Ancuti [7] proposed the fusion method to process the underwater images. These methods can significantly improve the contrast and saturation of underwater images, which also have relatively higher scores in qualitative evaluation of images, but the images processed by these methods have big noises. Even though these methods can achieve enhancement of underwater images in a certain degree, most of them can only achieve simple color corrections without effectively removing the haze effects of underwater images due to failure to consider the optical characteristics of underwater images.

2.3 Studies on the processing of underwater images based on deep learning

At present, CNN (convolutional neural network) and GAN (generative adversarial network) are mainly used to process the underwater images. For example, Li et al. [8] and Anwar et al. [9] utilized CNN to process the underwater images. However, these network architectures involve many training parameters. Fu et al. [10] proposed a light weight global-local (G-L) network in 2020, which is a two-branch network to compensate for the global distorted color and local reduced contrast, respectively. At current stage, many researchers have employed

GAN for processing of the underwater images. For example, Chen et al. [11] adopted GAN and combined it with the conventional method to improve the qualities of the recovered underwater images to a certain extent. Liu et al. [12] integrated the optical model and GAN for underwater image processing. Zhou et al. [13] introduced the domain adaptive mechanism and optical model constraint feedback control to process the underwater images. Even though the above methods can effectively process the underwater images, they still have certain problems, such as darkness of the obtained underwater images, and color noise introduced to the generated image.

3. Proposed Method

By combining the strengthen-operate-subtract (SOS) boosting module, the channel attention mechanism and the gated fusion module, this paper presents a Unet architecture. The experimental results prove that this architecture is applicable to dehazing of underwater images. The "Strengthen-Operate-Subtract" enhancement method is integrated to the decoder of our model, which is inspired by the Multi-Scale Boosted Dehazing Network [14]. This is a simple but effective enhanced decoder, which can recover the underwater images step by step. We also improved the channel attention mechanism. Our method is to obtain the mean value and standard deviation of RGB channels of the image first, and then feed the obtained mean value and standard deviation to the fully-connected network. Then, the output of the fully-connected network is added to the average value of the original image to obtain a new average value, which is used as the result of channel attention. Because dilated convolution can cover more adjacent pixels, we adopt dilated convolution smoothing to reduce the meshing artifacts. Finally, we use the gated subnet to determine the importance of different levels of feature maps, and integrate them according to their corresponding importance weights. See **Fig. 2** for the specific network architecture, which mainly consists of the three parts of encoder, the boosted decoder and the gated fusion module.
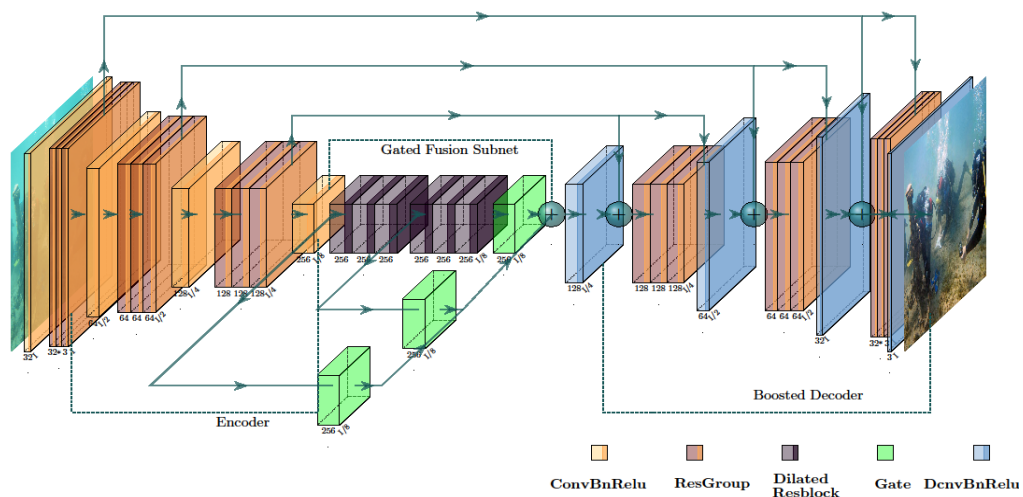

Fig. 2. Underwater images recovered with our method.

3.1 SOS Boosting Module

Romano et al. [15] initially proposed the SOS boosting module for image denoising, and its equation is as follows:

$$S^{n+1} = g(I + S^n) - S^n \tag{2}$$

where, S^n is the image prediction result of the n-th iteration, g(.) is the dehazing operation of underwater image, and $I + S^n$ represents the strengthened image by I with haze effect. PoH(.) is defined as:

$$PoH(S) = (1-t)A / S \tag{3}$$

In 2020, Dong et al. [14] used the SOS boosting module to defog images, and proved that this algorithm can improve the defogging performance in theory.

$$PoH(S^{n+1}) < PoH(S^n) \tag{4}$$

In the specific network architecture, we mainly extract the features of underwater images with the encoder, and remove the haze effect of underwater images with the decoder. According to the characteristics of the underwater images, we integrate the channel attention mechanism to the SOS Boosting Module. Its equation is as follows:

$$s^n = G_{\delta n}^n (i^n + (s^{n+1})\uparrow_2) - (s^{n+1})\uparrow_2 \tag{5}$$

where, $\uparrow_2$ represent upsampling, $G_{\delta n}^n$ represents the recovery unit, and in each recovery unit, the residual network architecture with channel attention mechanism is adopted. Its implementation process is as follows. First, the upsampling of feature map s^{n+1} obtained in the previous layer is carried out, and then it is added to the feature map obtained by encoder i^n. Next, the final output is obtained by subtracting the upsampling of feature map s^{n+1} from the $G_{\delta n}^n$ function. The network architecture of $G_{\delta n}^n$ function is a residual group with the channel attention mechanism. At the last layer of the decoder, we adopt a convolutional layer plus the Sigmoid activation function.
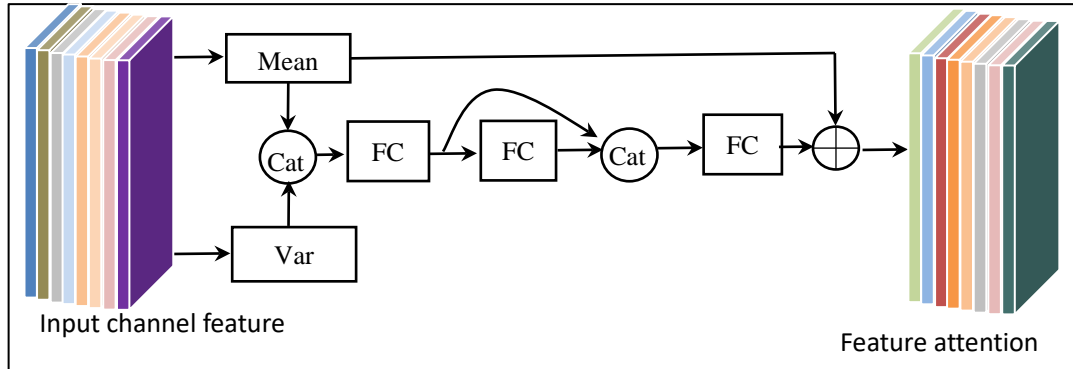
3.2 Channel attention module

Because the underwater image has different attenuations in different channels, and the three channels of RGB (red, green and blue) have certain connections, we mainly focus on the relationship among these channels, and hope the model can automatically learn the importance of different channel features. To this end, we improved the Squeeze and Excitation (SE) module. The channel attention module in this paper is as shown in **Fig. 3**. We conduct squeeze operation of the underwater image first to obtain the channel-level global features, and then get the average (*Avg*) value and standard deviation (Std) of the feature maps of the RGB channels. See Equation (6) for details.

$$F_1 = Relu(MLP(Cat(Avg(F), \ Std(F)))) \tag{6}$$

where, *MLP* represents full connection, and *Cat* represents concatenation operation. According to Equations (6)-(8), the relationship among various channels is learned via three full connections.

Fig. 3. Channel attention module

Finally, a new mean value is obtained by adding the final fully connected output to the original mean value. With such attention mechanism, the model will focus on the underwater image features containing more information, while ignoring less important features. In the meantime, this module also has general features, and is embedded into the SOS boosting module.

$$F_2 = Relu(MLP(F_1))\tag{7}$$

$$\hat{\mu}_J = Sigmoid(MLP(Cat(F_1, F_2) + \mu_I)\tag{8}$$

3.3 Gated fusion module

In this paper, the gated fusion module is used to control the weights of different feature maps. We extract feature maps F_1, F_2 and F_3 first, and then feed these three feature maps to the gated fusion module, as shown in Equation (9).

$$(W_1, W_2, W_3) = Gate(F_1, F_2, F_3)\tag{9}$$

The weights (W_1, W_2, W_3) of these three feature maps are obtained from the gated fusion module. Finally, the weights obtained from the gated fusion module are multiplied by corresponding feature maps to obtain the new feature map, as shown in Equation (10).

$$F_o = W_1 * F_1 + W_2 * F_2 + W_3 * F_3\tag{10}$$

In this paper, our gated fusion module is inspired by the method of Chen et al. [16]-[17]. Feature maps F_1, F_2 and F_3 are input into the gated fusion module, and finally, the output F_o is obtained after fusion of the gated fusion module. The input underwater image is encoded into feature maps via 4 convolutional modules and 3 residual groups with channel attention mechanism. And, feature mapping is downsampled to 1/16 of the size of input image. Symmetrically, the convolutional layer and SOS boosting module are used to upsample the feature maps to the original size of underwater image. In the residual group of network, we introduce the channel attention mechanism, and use the smoothing dilated convolutional layer to replace the regular convolutional layer. The dilation rate for the middle residual block is set as (2, 2, 2, 4, 4, 4, 1). We set the channel number of all middle convolutional layers at 256.

3.4 Loss function

The loss function consists of four parts, as shown in Equation (11):

$$L_{loss} = \lambda_1 L_{L_1} + \lambda_2 L_{ssim} + \lambda_3 L_{dc} + \lambda_4 L_{smooth} \tag{11}$$

where, λ_1, λ_2, λ_3 and λ_4 are the weights of various loss functions.

1). The first term is the L_1 loss function, which can be used to limit the similarity between the target image and the image output by the model. See Equation (12) for specific information:

$$L_{L_1} = \left\| J_c - M(I_c) \right\|_1 \tag{12}$$

where, J represents the target image, I stands for the original underwater image, and $M(I)$ represents the image recovered by the model.

2). L_{ssim} is the structural similarity loss function. We choose the structural similarity loss function to measure the similarity between the image output by model and the target image in structure.

$$L_{ssim} = 1 - SSIM(J_{GT}, M(I)) \tag{13}$$

3). The third term is the exponentiated mean local variance deviation. See Equation (14) for specific information. ε is a very small number, and this number is used to avoid being divided by 0. $\nabla_{x/y}$ represents the operator used to obtain the gradients of recovered underwater image in the horizontal and vertical directions [18][19][20]. The size of ω local block is $r \times r$, and r is set as 3 in this paper. η_s is the exponent that determines the sensitivity to the gradient of J, and we set $\eta_s = 1.2$.

$$E_s(J) = \left\| \nabla_x J \left(\left| \frac{1}{\omega} \sum_\omega \nabla_x J \right|^{\eta_s} + \varepsilon \right)^{-1} \right\|_2^2 + \left\| \nabla_y J \left(\left| \frac{1}{\omega} \sum_\omega \nabla_y J \right|^{\eta_s} + \varepsilon \right)^{-1} \right\|_2^2 \tag{14}$$

The gradient feature is expressed by the local variation, and its deviation indicates the variation correlation in the local patch. η_s is a single exponential used to alter the structure and texture awareness of an underwater image. The smoothness of the underwater image is affected by this value. Because the local variance information is taken into consideration, the exponentiated mean local variance can better present the details and retain the structure. As a result, the capacity of the mean local variance to discriminate texture and structure is quite high.

4). The fourth term is a dehazing and color correcting loss function.

$$L_{dc} = \left\| J_c^{dc}(x) - M(I) \right\|_1 \tag{15}$$

This loss function measures the similarity between the image output by network and the target image after dehazing and color correction. $J_c^{dc}(x)$ represents the target image after dehazing and color correction[21]. First, we obtain the mean value J_c^{mean} and standard deviation $J_c^{std}(x)$ of three channels, as shown in Equation (16), where, J represents the target image, and μ represents a constant, which is set to 2.5 in this paper.

$$\begin{cases} J_c^{max}(x) = (x)J_c^{mean} + \mu J_c^{std}(x) \\ J_c^{min}(x) = J_c^{mean}(x) - \mu J_c^{std}(x) \end{cases} \tag{16}$$

The above equation can be used to obtain the maximum and minimum values, which are

substituted into Equation (17) to obtain the image $J_c^{cor}(x)$ after color correction.

$$J_c^{cor}(x) = 255 \frac{J_c(x) - J_c^{\min}(x)}{J_c^{\max}(x) - J_c^{\min}(x)} \tag{17}$$

Then, $J_c^{dc}(x)$ is obtained by dehazing $J_c^{cor}(x)$ using the dark channel prior method [22].

4. Experimental results and analysis

In this part, the effectiveness of this network is evaluated via experiments, which mainly includes introduction of the training platform, description of the training datasets, the ablation study, as well as the quantitative and qualitative analyses of the underwater images obtained by different methods. The methods to be compared with ours: Haze Line [23], NUDCP [2], Fusion [7], UWCNN [8], UGAN [24], FUnIE-GAN [25] and G-L [10]. Because the training sets used in Fu's model are basically the same as the training sets in this paper, we did not train their network, and directly used the trained model they provided to test the underwater images.

4.1 Training datasets and experimental setting

Network training set: Dataset has a significant influence on the results obtained by deep learning. In this paper, we mainly use three datasets. One is the Li's dataset [8], which consists of 800 images in total, 720 images were used for training, and the rest 80 were for verifying the results. The second dataset is the Fabbri's dataset [24], which includes 6128 images. This dataset was synthesized on the basis of Cycle GAN, 4902 images from this dataset were used for data training in our experiments, and the rest 800 were used for verification. The third dataset is the enhancing underwater visual perception (EUVP) dataset [25], which was synthesized on the basis of the UGAN. We choose two directories under this dataset. In one directory, the target image data comes from ImageNet, which is named as EUVP_1, 2960 images in this directory were used for training, and 740 were for test. In another directory, the target image data comes from the underwater scenes, which is named as EUVP_2, and there are 2185 images under this directory, 1748 of which were used for training, and the rest 437 were for test.

Setting of training parameters: All images were resized to 256×256×3 for training, and the parameters were $\lambda_1 = \lambda_2 = 1$, $\lambda_3 = 0.5$, $\lambda_4 = 0.05$. The training iterations were 80 (100 iterations were trained in the dataset of Li), and the learning rate was 0.0001. The experiments were carried out on a PC equipped with Intel (R) i9-12900KF processor, and an NVIDIA GTX 3080Ti GPU.

4.2 Qualitative comparison

Next, we compare the visual effects of underwater images obtained by different methods. In the meantime, to verify the robustness of the model proposed in this paper, we also compare and analyze the processing effects of various methods on real-world underwater images. Some of these images are collected by Li et al. [26], and some are from other datasets such as the underwater color cast set (UCCS) [27].

Qualitative comparison on the Li's dataset: In Li's dataset, we randomly choose two images for qualitative comparison, and see **Fig. 4** for specific information. From left to right: the original images, the images obtained with the Haze-line, Fusion, NUDCP, UWCNN, UGAN, FUnIE-GAN, G-L and our method, and the ground-truth images. According to **Fig. 4**,

due to inaccurate estimation of transmission map, the images obtained with the conventional methods tend to have color cast. The recovered images by the Haze-line method are dark, and most details of image are covered. Some images processed with the Fusion methods show reddish color, while the NUDCP methods have weak dehazing ability. The methods based on deep learning can generally provide good performances in processing underwater images, but these methods still have some defects. For instance, the images obtained with the UWCNN and Fu's method have relatively poor visual effects, and the images are dark in general. The methods based on FUnIE-GAN and UGAN can get good results of underwater images, but the images tend to have color cast, and are not very similar to the GT target images. In comparison, the images recovered with our method show great quality in terms of color, brightness and contrast.

Qualitative comparison on the Fabbri's dataset: In the Fabbri's dataset, we also choose two images as the references for qualitative comparison. See **Fig. 5** for the specific comparison results. From left to right: the original images, the images obtained with the Haze-line, Fusion, NUDCP, UWCNN, UGAN, FUnIE-GAN, G-L and our method, and the ground-truth images. According to the images in **Fig. 5**, The color of images restored by the haze-line method shows obvious distortion, while the images obtained by the UGAN method tend to introduce external color noise. In addition, compared with the GT target images, the images obtained by the FUnIE-GAN and Fu's method show low structural similarity.

Fig. 4. Images obtained using different methods on the Li's dataset.

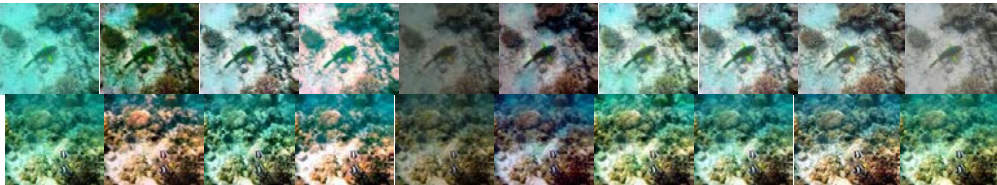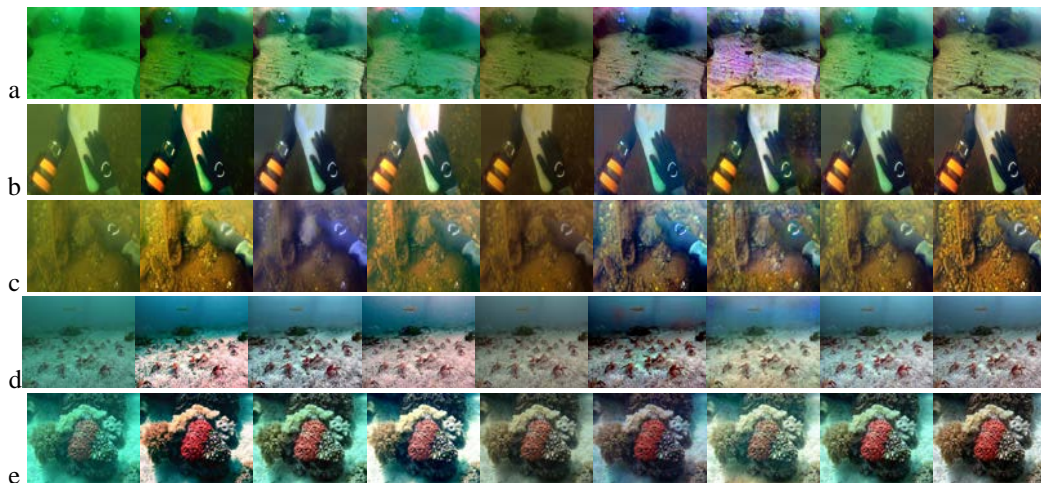**Fig. 5.** Images obtained using different methods on the Fabbri's dataset.

Fig. 6. Images obtained using different methods on EUVP_1 dataset.

Fig. 7. Images obtained using different methods on EUVP_2 dataset.

Because we have introduced the SOS boosting module to dehaze the underwater images, the images recovered with our method show high structural similarity to the GT target images, and they also have better visual effects than the images obtained with other methods.

Qualitative comparison on the EUVP dataset: For images in the dataset directories of EUVP_1 and EUVP_2, the visual results are presented in **Fig. 6** and **Fig. 7**, respectively. From left to right: the original images, the images obtained with the Haze-line, Fusion, NUDCP, UWCNN, UGAN, FUnIE-GAN, G-L and our method, and the ground-truth images. The images under the two directories have very similar characteristics, so we analyze these two directories together. According to **Fig. 6** and **Fig. 7**, the images recovered with the Haze-line method also have low brightness and reddish color, and some images recovered with NDCP also show color cast. The deep learning methods do not present significant difference in visual effects in processing images in these two directories. In comparison, we have introduced the channel attention mechanism and the loss functions, which can help improve the brightness of underwater images and address the color cast problem of underwater images.

Fig. 8. Images obtained using different methods on the real-world underwater images.

Qualitative comparison on the real-world underwater scenes data: In addition to comparison and analysis on the training sets, we also conducted tests on some public images in order to verify the robustness and adaptation of the proposed model, and these images were collected from the Internet and sorted out by Li et al.

See **Fig. 8** for the specific comparative results. From left to right: the original images, the images obtained with the Haze-line, Fusion, NUDCP, UWCNN, UGAN, FUnIE-GAN, G-L and our method. We can see that the images obtained with the conventional methods have poor performances, which still have the color cast problem and poor overall dehazing effects. For example, among the images recovered by the Haze-line method, the image in **Fig. 8 (a)** is greenish, while those in **Fig. 8 (b)** and **Fig. 8 (e)** are darker. The images recovered by the method of Song et al. are reddish, and according to the images recovered with this method in **Fig. 8 (a)**, **Fig. 8 (c)** and **Fig. 8 (d)**, the dehazing effects of this method are also poor. The deep learning methods can provide better dehazing performances, but tend to introduce color cast. For instance, after being processed with the UGAN and FUGAN methods, the images in **Fig. 8(a)** and **Fig. 8(e)** present color cast. Furthermore, the deep learning methods, including Fu's method and Li's method, have relatively poor dehazing performances. Overall, the images

recovered with our method not only have better visual effects, but also better dehazing performances than other methods.

Qualitative comparison on the UCCS dataset: This dataset consists of 300 images in total, including the underwater images with different degradation types. We compare the results of processing the images from this dataset using different methods, and the results are shown in **Fig. 9**. From left to right: the original images, the images obtained with the Haze-line, Fusion, NUDCP, UWCNN, UGAN, FUnIE-GAN, G-L and our method.
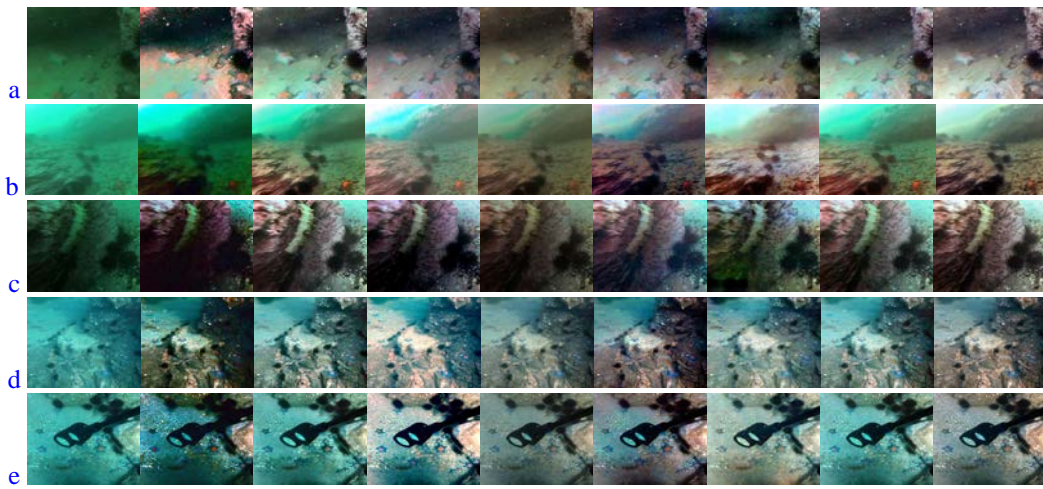

Fig. 9. Images obtained using different methods on the UCCS dataset.

Through comparison and analysis, we can see that the improvement by the conventional methods is relatively poor, and the obtained images have serious color cast. In comparison, the deep learning methods have much better performances on the improvement of colors, so they have stronger robustness and adaptation than the conventional methods. However, some of the deep learning methods still have some problems in processing these underwater images. For example, the methods of Li et al. and Fu et al. have relatively weak performances, and some details of the images cannot be well detected. The images processed with the UGAN and FUGAN methods are very clear, but these images tend to introduce external noises and generate color cast. Yet, the proposed method in this paper has achieved great performances in terms of contrast and brightness. Compared with other methods, our method has stronger robustness and adaptation ability.

4.3 Quantitative comparison

Since subjective evaluation is based on observation, it tends to be affected by personal factors. To further evaluate the proposed algorithm, we also conduct quantitative evaluation.

Comparison in terms of PSNR and SSIM: These are full-reference image quality criteria. **Table 1** reports the Peak Signal to Noise Ratio (PSNR) and Structural Similarity (SSIM) values of images obtained with different methods on Fabbri's dataset, Li's dataset, EUVP_1 and EUVP_2, in which, the score in bold font represents the optimal value achieved on this criterion.

Table 1. PSNR and SSIM evaluation results of underwater images on the Fabbri's dataset, Li's dataset, EUVP_1 and EUVP_2. The best performances values are in bold.

Methods	Fabbri's	Li's	EUVP_1	EUVP_2
Haze-line [23]	21.2465 / 0.5958	21.3901 / 0.6392	20.1828 / 0.5893	21.0490 / 0.5985
NUDCP [2]	19.5184 / 0.6703	19.8471 / 0.6884	19.7038 / 0.6852	19.7554 / 0.6852
Fusion [7]	21.2465 / 0.7196	21.3901 / 0.7761	20.1828 / 0.7221	21.0490 / 0.7149
UWCNN [8]	19.5184 / 0.7512	19.8471 / 0.7551	19.7038 / 0.7129	19.7554 / 0.7523
UGAN [24]	18.6496 / 0.6860	19.2363 / 0.7490	19.7770 / 0.6558	19.4072 / 0.6947
FUnIE-GAN[25]	19.3247 / 0.7085	21.0145 / 0.6817	19.8879 / 0.6469	19.4227 / 0.7106
G-L[10]	18.2640 / 0.7381	**21.2474 / 0.8224**	20.0884 / 0.7105	22.4449 / 0.7743
ours	**24.6645 / 0.8410**	21.1002 / 0.8195	**21.3009 / 0.7704**	**24.7011 / 0.8333**

As shown in **Table 1**, due to the introduced noise during processing, the images obtained by UGAN and UWCNN method have low PSNR. The images obtained by the G-L and our method show high scores in structural similarity, while those recovered with the NUDCP and Haze-line methods show low structural similarity. Our method achieves the best performances on both PSNR and SSIM. It has the best performance on the Fabbri's dataset, and it achieves the mean PSNR of 24.6645, and the mean SSIM of 0.8410. According to the quantitative comparison on the other three datasets, it still outperforms the other methods. Because our method can effectively remove the haze effects of underwater images, which can help highlight the texture and details of image, it shows significant advantages in the quality evaluation based on SSIM and PSNR.

Comparison in terms of UIQM and UCIQE: The underwater image quality measurement (UIQM) [28] and underwater color image quality evaluation (UCIQE) [29] are two no-reference image quality criteria. Generally speaking, higher UIQM and UCIQE values represent better quality of underwater images. **Table 2** shows the UIQM values of underwater images on the Fabbri's dataset, Li's dataset, EUVP_1 and EUVP_2. According to **Table 2**, the conventional methods, including Haze-line, Fusion and NUDCP, have the lowest UIQM scores. According to the visual effect analysis in Section 4.2, we can see that the images obtained with the conventional methods have serious color cast and are darker, which results in low contrast. In the meantime, the texture details of these images are not rich, so the UIQM values are relatively low. The images obtained with our method and UGAN have higher UIQM values. For example, on the Fabbri's dataset, Li's dataset, EUVP_1 and EUVP_2, the UIQM values of underwater images obtained by our method are 3.0554, 3.1436, 3.0482 and 3.1349, respectively. In the meantime, based on the qualitative analysis in **Fig. 6**, we can see that the G-L and UWCNN methods have weaker dehazing ability, and the images recovered by the two methods are also darker. Moreover, their texture details cannot be well presented, so the UIQM values are lower. Among the images obtained with the UGAN method, even though some color noises are introduced in some parts, they still have higher contrast according to the qualitative analysis, so their UIQM values are also higher. Introducing the loss function can help improve the contrast of underwater images. Therefore, our method has achieved high scores in UIQM.

Table 2 shows the UCIQE values of the underwater images obtained by different methods. The scores in bold font represent the best results. According to **Table 2**, we can see that the conventional methods have good performances in UCIQE. The images recovered with

the Haze-line method have the highest UCIQE scores, while those obtained with the UWCNN method have the lowest UCIQE scores. The images obtained with other methods do not present much difference in the UCIQE value. However, according to the qualitative comparison results in Section 4.2, the images recovered by the Haze-line method have great performances on chroma and saturation, so they have the highest UCIQE scores. Yet, these images have serious color cast problem, and the texture details are not clear due to darkness of images. Therefore, we can infer that the UCIQE criterion does not consider color cast and other problems, so this criterion has certain defects [30]. At the same time, the underwater images restored by the deep learning methods are also affected by the training dataset. The datasets used for training in this paper are synthetic, which has a certain impact on the UICQE values of the restored images.

Table 2. UIQM and UCIQE evaluation results of underwater images on the Fabbri's dataset, Li's dataset, EUVP_1 and EUVP_2. The best and second best performance values are in bold

Methods	Fabbri's	Li's	EUVP_1	EUVP_2
Haze-line [23]	2.1622 / **0.6779**	2.2990 / **0.6538**	2.0979 / **0.6965**	2.1635 / **0.6842**
NUDCP [2]	2.4082 / **0.6254**	2.4414 / 0.6123	2.3888 / **0.6505**	2.4888 / **0.6359**
Fusion[7]	2.8630 / 0.5915	2.9330 / 0.6017	2.8676 / 0.6129	2.8904 / 0.5998
UWCNN [8]	2.9825 / 0.5440	**3.0521** / 0.5478	2.9553 / 0.5644	3.0296 / 0.5516
UGAN [24]	**3.0900** / 0.6013	2.8990 / 0.6294	**3.0841** / 0.6221	**3.1547** / 0.6057
FUnIE-GAN[25]	3.0197 / 0.6091	2.9316 / 0.6153	2.9316 / 0.6128	3.0278 / 0.6156
G-L[10]	2.9404 / :0.6022	2.8843 / **0.6278**	2.8961 / 0.6197	2.9690 / 0.6097
ours	**3.0554** / 0.5954	**3.1436** / 0.6014	**3.0482** / 0.6183	**3.1349** / 0.6043

Finally, we conduct quantitative analysis on the UCCS dataset [27] to verify the adaptation ability and robustness of our model. Because the dataset does not contain paired images, we only conduct analysis and test based on the two criteria of UIQM and UICQE. The UIQM and UICQE values of the images obtained with different methods are listed in **Table 3**. According to **Table 3**, the underwater images obtained with our method and Fu's method have the higher UIQM scores, with the mean values of UIQM reaching 3.1313 and 3.1681, respectively. In comparison, the images obtained with the Haze-line and FUnIE-GAN method have lower UIQM scores. The underwater images obtained with the Haze-line and NUDCP methods have higher UCIQE scores, and their mean values are 0.6294 and 0.6203, respectively. However, with the defects of UCIQE criteria considered, and the qualitative comparison in Section 4.2 combined, our method has some advantages in improving the brightness of underwater images, keeping the texture details and achieving color balance. Therefore, our method outperforms the other algorithms on the basis of the comprehensive evaluation.

Table 3. UIQM and UCIQE scores of images in the UCCS dataset by different methods. The best and second best performance values are in bold

Methods	H-Line	NUDCP	Fusion	UWCNN	UGAN	FGAN	G-L	Ours
UIQM	2.4331	3.1167	3.1214	2.8864	3.0310	3.0302	**3.1681**	**3.1313**
UICQE	**0.6294**	**0.6203**	0.6041	0.5037	0.6061	0.5722	0.5944	0.5862

To sum it up, compared with the other seven methods, the underwater images recovered by our method have the highest scores of SSIM and PSNR. According to the qualitative analysis, these images also have higher brightness, contrast and saturation. Therefore, the model proposed in this paper can well improve the visual effects of underwater images.

Fig. 10. Images recovered by different network architectures. From left to right: the original images, the images recovered without the attention mechanism, the SOS module, and the images recovered with the attention mechanism + SOS module.

4.4 Ablation study

4.4.1 Analysis on the influence of different network architectures

According to **Fig. 10**, the images obtained by the network architecture without introducing the dehazing mechanism have strong haze effects. As shown in the third column of **Fig. 10**, the colors of images obtained by the architecture without introducing the attention mechanism are not balanced, and there is slight color cast. In comparison, after introducing the attention mechanism and dehazing mechanism, the haze effects of images can be more effectively removed, and the color cast is also alleviated.

Table 4. SSIM and PSNR scores of images in Li's dataset using different network architectures.

Methods	Without attention mechanism	Without SOS	With attention mechanism + SOS
SSIM	0.7337	0.7545	**0.8195**
PSNR	19.2162	20.4159	**21.1002**

Table 4 shows the SSIM and PSNR scores of images in the Li's dataset obtained with different network architectures. Because the images in Li's dataset were captured from the real-world underwater scenes, we conducted ablation study on this dataset. According to **Table 4**, we can see that the images recovered by introducing the dehazing mechanism and attention mechanism have good performances in terms of SSIM and PSNR. Therefore, from both qualitative and quantitative perspectives, introduction of the attention mechanism and dehazing mechanism can help improve the visual effects and scores of the quantitative analysis indexes of the underwater images to a certain extent.
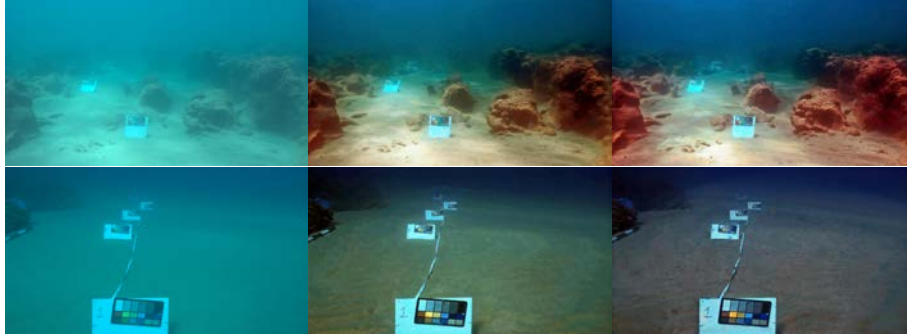
Fig. 11. Images recovered using different loss function. From left to right: the original images, images recovered With $L_{L1}+L_{ssim}$ and with $L_{L1}+L_{ssim}+L_{smooth}+L_{ce}$, respectively.

Table 5. SSIM and PSNR scores of images in Li's dataset using different loss function.

Methods	$L_{L1}+L_{ssim}$	$L_{L1}+L_{ssim}+L_{smooth}+L_{dc}$
SSIM	0.8140	**0.8195**
PSNR	20.7920	**21.1002**

4.4.2. Analysis of the influence of loss function

The loss function in Equation (15) is used to weaken the haze effect and improve the brightness of underwater images. The loss function in Equation (14) is used for denoising of underwater images. According to **Fig. 11**, by introducing the two loss functions, the brightness of image is improved, and the haze effect is weakened. From the quantitative perspective, see **Table 5** for specific information, introduction of these two loss functions can improve the SSIM and PSRN scores of underwater images. The main reason is that by introducing the smooth loss function, the edge details of the underwater image can be preserved to the great extent while also denoising the image.

5. Conclusion

This paper introduces a method using CNN to recover underwater images in detail. First, according to the characteristics of underwater images, we design a deep learning model by combining the SOS mechanism, channel attention mechanism and gated subset. In the meantime, considering the noise of underwater images, we use the traditional edge preservation filtering method as the loss function of model. Due to the low brightness of underwater images, we introduce a loss function to improve the brightness of the image. Finally, we trained our model, compare and analyze the images obtained by our model with those by the other seven methods. We conduct both quantitative and qualitative analyses. From the perspective of qualitative analysis, the images recovered with our method have good brightness, and the texture details can also be well preserved. According to the quantitative analysis, compared with most methods, our method has great performance in accordance with various evaluation criteria. Even though the deep learning methods have great advantages compared with the conventional methods, the images obtained by the deep learning methods still have some problems due to restriction of datasets. In the future, we need to address these

problems in processing underwater images, such as the blurring effect of underwater images caused by shaking of underwater photography and the influence of underwater turbulences. Therefore, for better image recovery, the blurring effect needs to be removed. However, because paired datasets were difficult to obtain during our work, we will focus on the synthesis of datasets and the elimination of blur effects in underwater images.

Acknowledgement

References

[1] L. Chao and M. Wang, "Removal of water scattering," in *Proc. of 2010 2nd International Conference on Computer Engineering and Technology*, vol. 2, pp. V2-35-V2-39, 2010. Article (CrossRef Link)

[2] W. Song, Y. Wang, D. Huang, A. Liotta, and C. Perra, "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, March 2020. Article (CrossRef Link)

[3] J. Zhou, T. Yang, W. Ren, D. Zhang, and W. Zhang, "Underwater image restoration via depth map and illumination estimation based on a single image," *Opt. Express*, vol. 29, no. 18, pp. 29864–29886, 2021. Article (CrossRef Link)

[4] R. Hummel, "Image enhancement by histogram transformation," *Comput. Graph. Image Process.*, pp. 184–195, 1975. Article (CrossRef Link)

[5] K. Zuiderveld, "Contrast limited adaptive histogram equalization," *Graph. Gems*, vol. 38, pp. 474–485, 1994.

[6] S. Zhang and T. Wang et al., "Underwater image enhancement via extended multi-scale retinex," *Neurocomputing*, vol. 245, pp. 1–9, 2017. Article (CrossRef Link)

[7] C. Ancuti, C. O. Ancuti, T. Haber, P. Bekaert, "Enhancing underwater images and videos by fusion," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, pp. 81–88, 2012. Article (CrossRef Link)

[8] C. Li, S. Anwar, and F. Porikli, "Underwater scene prior inspired deep underwater image and video enhancement," *Pattern Recogn*, vol. 98, 2020. Article (CrossRef Link)

[9] S. Anwar, C. Li, and F. Porikli, "Deep Underwater Image Enhancement," *ArXiv180703528*, Jul. 2018. Article (CrossRef Link)

[10] X. Fu and X. Cao, "Underwater image enhancement with global-local networks and compressed-histogram equalization," *Signal Processing: Image Communication*, vol. 86, 2020. Article (CrossRef Link)

[11] X. Chen and J. Yu et al., "Towards real-time advancement of underwater visual quality with gan," *IEEE Transactions on Ind. Electron*, vol. 66, no. 12, pp. 9350 - 9359, 2019. Article (CrossRef Link)

[12] X. Liu, Z. Gao, and B. M. Chen, "IPMGAN: integrating physical model and gener- ative adversarial network forunderwater image enhancement," *Neurocomputing*, vol. 453, pp. 538-551, 2021. Article (CrossRef Link)

[13] Y. Zhou, K. Yan and X. Li, "Underwater Image Enhancement via Physical-Feedback Adversarial Transfer Learning," *IEEE Journal of Oceanic Engineering*, vol. 47, no. 1, pp. 76-87, Jan. 2022. Article (CrossRef Link)

[14] H. Dong and J. Pan et al., "Multi-scale boosted dehazing network with dense feature fusion," in *Proc. of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 2157-2167, 2020. Article (CrossRef Link)

[15] Y. Romano and M. Elad, "Boosting of image denoising algorithms," *SIAM Journal on Imaging Sciences*, vol. 8, no. 2, pp. 1187–1219, Jan. 2015. Article (CrossRef Link)

[16] D. Chen and M. He et al., "Gated context aggregation network for image dehazing and deraining," in *Proc. of 2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 1375-1383, 2019. Article (CrossRef Link)

[17] L. Chen, Q. S. Sun, and F. Wang, "Attention-adaptive and deformable convolutional modules for dynamic scene deblurring," *Information Sciences*, vol. 546, pp. 368–377, 2021. Article (CrossRef Link)

[18] J. Xu, Y. Hou, and D. Ren et al., "Star: A structure and texture aware retinex model," *IEEE Transactions on Image Process*, vol. 29, pp. 5022–5037, 2020. Article (CrossRef Link)

[19] M. Li, J. Liu, and W. Yang et al., "Structure-revealing low-light image enhancement via robust retinex model," *IEEE Transactions on Image Processing*, vol. 27, no. 6, pp. 2828-2841, June 2018. Article (CrossRef Link)

[20] B. Cai, X. Xu, and K. Guo, "A joint intrinsic extrinsic prior model for retinex," in *Proc. of 2017 IEEE Int. Conf. on Comput. Vis. (ICCV)*, pp. 4000-4009, 2017. Article (CrossRef Link)

[21] X. Fu and P. Zhuang et al., "A retinex-based enhancing approach for single underwater image," in *Proc. of 2014 IEEE International Conference on Image Processing (ICIP)*, pp. 4572-4576, 2014. Article (CrossRef Link)

[22] K. He, S. Jian, and X. Tang, "Single image haze removal using dark channel prior," *Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 12, pp. 2341-2353, Dec. 2011. Article (CrossRef Link)

[23] D. Berman, D. Levy, and S. Avidan 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, 1 Aug. 2021. Article (CrossRef Link)

[24] C. Fabbri, M. Islam, and J. Sattar, "Enhancing underwater imagery using generative adversarial networks," in *Proc. of 2018 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 7159-7165, 2018. Article (CrossRef Link)

[25] M. J. Islam and Y. Xia et al., "Fast underwater image enhancement for improved visual perception," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 3227-3234, April 2020. Article (CrossRef Link)

[26] C., Li, C. Guo, W. Ren, R. Cong, et al., "An underwater image enhancement benchmark dataset and beyond," *IEEE Transactions on Image Processing*, vol. 29, pp. 4376-4389, 2020. Article (CrossRef Link)

[27] R. Liu, X. Fan, M. Zhu, M. Hou and Z. Luo, "Real-World Underwater Enhancement: Challenges, Benchmarks, and Solutions Under Natural Light," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 12, pp. 4861-4875, Dec. 2020. Article (CrossRef Link)

[28] K. Panetta, C. Gao and S. Agaian, "Human-Visual-System-Inspired Underwater Image Quality Measures," *IEEE Journal of Oceanic Engineering*, vol. 41, no. 3, pp. 541-551, July 2016. Article (CrossRef Link)

[29] M. Yang and A. Sowmya, "An Underwater Color Image Quality Evaluation Metric," *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 6062-6071, Dec. 2015. Article (CrossRef Link)

[30] Z. Liang, Y. Wang, X. Ding, Z. Mi, and X. Fu, "Single underwater image enhancement by attenuation map guided color correction and detail preserved dehazing," *Neurocomputing*, vol. 425, no. 15, pp. 160–172, 2021. Article (CrossRef Link)

SHIWEN LI received the master degree in electrical circuit and system from the South China Normal University, China, in 2009, and is currently working toward the Ph.D. degree in signal and information processing at Nanjing University of Posts and Telecommunications, Nanjing, China. His Current research interests include image/video processing, image restoration, underwater image enhancement and deep learning.

FENG LIU received the M.S. and Ph.D. degrees from the School of Electronic and Optical Engineering, Nanjing University of Science and Technology, Nanjing, China, in 1993 and 1997, respectively. He is currently a Full Professor with the College of Telecommunications and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing. His research interests include machine learning, visual tracking, and intelligent video analysis. He is a member of the IEEE Signal Processing Society.

JIAN WEI is is an senior engineer at the Jiangsu Province Key Laboratory on Image Processing and Image Communications, Nanjing University of Posts and Telecommunications, Nanjing, China. He received his BS degree in signal and information processing from Nanjing University of Posts and Telecommunications in 2019. His research interests include machine learning and visual tracking.