

Underwater Image Enhancement by Combining Multi-Attention with Recurrent Residual Convolutional U-Net

Shuqi Wang^{1,2,3}, Zhixiang Chen^{1,2*}, Hui Wang^{1,2}

¹Key Laboratory of Light Field Manipulation and System Integration Applications in Fujian Province, Minnan Normal University, Zhangzhou, 363000, Fujian, China.

²College of Physics and Information Engineering, Minnan Normal University, Zhangzhou, 363000, Fujian, China.

³Key Laboratory of Data Science and Intelligence Application, Minnan Normal University, Zhangzhou, 363000, Fujian, China.

*Corresponding author(s). E-mail(s): zxchenphd@163.com;
Contributing authors: 347423159@qq.com; wangh0802@163.com;

Abstract

The scattering and absorption of light lead to color distortion and blurred details in the captured underwater images. Although underwater image enhancement algorithms have made significant breakthroughs in recent years, enhancing the effectiveness and robustness of underwater degraded images is still a challenging task. To improve the quality of underwater images, we propose a combined multi-attention mechanism and recurrent residual convolutional U-Net (ACU-Net) for underwater image enhancement. First, we add a dual-attention mechanism and convolution module to the U-Net encoder. It can unequally extract features in different channels and spaces and make the extracted image feature information more accurate. Second, we add an attention gate module and recurrent residual convolution module to the U-Net decoder. It helps extract features fully and facilitates the recovery of more detailed information when the image is generated. Finally, we test the subjective results and objective evaluation of our proposed algorithm on synthetic and real datasets. The experimental results show that the robustness of our algorithm outperforms the other five classical algorithms, such as in enhancing underwater images with different color shifts and turbidity. Moreover, it corrects the color bias and improves the contrast and detailed texture of the images.

Keywords: Underwater image, Multi-attention, Recurrent residual convolutional units, Image enhancement, Generative adversarial network

1 Introduction

The exploitation of marine resources has driven the global economy with great promise. However, low visibility underwater imagery significantly impacts the development and utilization of marine resources. Furthermore, light propagates through absorption and scattering in water, which severely

affects the imaging process and leads to blurred images and poor contrast [1]. Therefore, it is important to study clarification techniques such as underwater image enhancement, which will lay the foundation for underwater exploration vehicle research [2], underwater biology [3], archaeology

[4], and the inspection [5] and maintenance of underwater facilities [6].

Existing underwater image enhancement algorithms are broadly classified into two categories: traditional underwater image enhancement algorithms and deep learning of underwater image enhancing algorithms based on deep learning.

Most deep learning methods have made significant breakthroughs in underwater image enhancement that can effectively enhance several images. However, light scattering and absorption occur in water leading to uneven distribution of local features in fogging and color-biased regions in different channels or spaces. To address the above problems of deep learning-based underwater image enhancement algorithms, we propose an innovative underwater image enhancement algorithm (AC-GAN) combining multiple attention and recurrent residual convolution U-Net (ACU-Net). We add a double attention mechanism and a convolution module to the U-Net encoder for generative networks. The attention gates and recursive residual convolution modules are added to the U-Net decoder. Then, the adversarial loss L_{WGAN} , the loss L_1 , the gradient difference loss L_{GDL} , and the perceptual loss L_{CON} are used as loss functions in AC-GAN.

In this paper, we propose a new underwater image enhancement strategy to recover low-quality underwater images by extracting sufficient features without generating excess noise. Our main contributions and work are as follows:

1. We use U-Net with excellent feature extraction capability as the base network for underwater feature extraction. We incorporate our improved multi-attention mechanism and convolution module into the U-Net network. It can maintain the original details of the image while fully extracting the underwater image features.
2. The Convolutional Block Attention Module (CBAM), attention mechanism, and Recurrent Residual convolutional units (RRCU) modules are introduced into the framework of GAN to complement the convolutional structure. It enables the network to pay more attention to pixels and more important channel information in regions where the body of water is more influential.
3. In addition, we experimented with underwater images of different pollution levels and different

color shifts. The subjective and objective evaluation results show that our proposed algorithm can extract the detail features adequately, correct the image color shift more naturally, and recover the details well.

2 Related work

Most traditional underwater image enhancement algorithms are scene-specific and use assumptions and prior knowledge to enhance images. There are problems such as low robustness and poor real-time performance. For example, Singh et al. [7] proposed an exposure-based recursive histogram equalization image enhancement method. This method can obtain a high peak signal-to-noise ratio and low mean square error. However, traditional enhancement methods cannot adaptively improve images' degradation effects due to the variability of the underwater environment. The underwater image enhancement algorithm proposed by Zhang et al. [8] is an extended multiscale Retinex that effectively suppresses the halo phenomenon during image enhancement. This algorithm extends the color recovery of the multiscale Retinex enhancement algorithm to the CIELab color space. However, this method introduces many parameters that lead to poor robustness.

The deep learning-based underwater image enhancement algorithm mainly uses convolutional neural networks. The algorithm automatically extracts underwater image features and obtains the relationship of mapping between original and enhanced underwater images. Thus, the clarity of the underwater image is achieved. Supervised learning-based methods require a large number of paired samples for supervised training. However, it is very difficult to collect enough paired underwater images in practice. Li et al. [9] embedded an underwater image generation model into a generative adversarial network (GAN) structure to enable clear land images to generate turbid underwater images. Chen et al. [10] transformed the turbid underwater image into a clear underwater image by using an underwater image imaging model and then trained a conditional generative adversarial network with the paired dataset to achieve underwater image enhancement. However, due to the difference between the distribution of synthetic images and real underwater images, such methods are not ideal for

real image enhancement. Cycle-consistent adversarial networks (CycleGAN) proposed by Zhu et al. [11] designed a two-way GAN structure and introduced a cycle-consistency loss, which relaxes the effect of underwater image enhancement on paired training sets. It reduces the need for paired training sets. Li et al. [12] proposed a weakly supervised method for underwater image color conversion using CycleGAN. The method designs a multinomial loss function to remove color bias from underwater images while preserving the content information of the original images. However, the method is not effective in improving image contrast.

Most deep learning-based enhancement networks do the same processing for local features in different channels or spaces. Therefore, the enhancement networks cannot adequately extract underwater image features reflecting different degradation degrees, resulting in a lack of robustness of the model. To address the above, we use the U-Net with excellent feature extraction capability as the generating network for underwater feature extraction. First, we add a dual-attention mechanism and a convolution module to the U-Net encoder of the generative network. It can unequally process features in different channels and spaces, which reduces the amount of computation of the network on redundant information and makes the network focus more on important information. Second, the attention gate and recurrent residual convolution module are added to the U-Net decoder. It can automatically learn the target structure without additional supervision and more effectively enhance the underwater images with different degradation levels. The recurrent residual block sums the features multiple times to help extract global features, which can effectively increase the network depth. Finally, the adversarial loss L_{WGAN} , loss L_1 , gradient difference loss L_{GDL} and perceptual loss L_{CON} are used as loss functions for AC-GAN to make the output image consistent with the real image in terms of color and content.

3 The proposed algorithm

To solve the problems of color bias, fogging, and blurred details in underwater images, we propose an underwater image enhancement algorithm

(AC-GAN). In this algorithm, the generative network part is a combination of multiple attention and ACU-Net.

3.1 Description of AC-GAN

We propose an underwater image enhancement model AC-GAN based on GAN, which is divided into two parts: a generative network and a discriminative network. The structure of the generative network consists of an improved encoder and decoder to perform enhancement operations on degraded underwater images. The structure of the discriminative network is similar to the Markov discriminator to achieve discrimination between the generated image and its corresponding reference image. The proposed algorithm is executed as follows: the AC-GAN algorithm first inputs the distorted underwater image as the input Z of the generator network, and the inference of the network obtains the generated image Z' . Then, the generative image and its corresponding underwater truth image C are taken as the input of the discriminator network. The adversarial loss L_{WGAN} , mean absolute error loss L_1 , gradient difference loss L_{GDL} and perceptual loss L_{CON} are calculated. The loss D-Loss (Fake) of the generative image Z' is obtained by the linear combination of the four-loss functions. The D-Loss and G-Loss are passed backwards to each layer of the generative network to update the parameters of each layer of the network iteratively and finally obtain the underwater image with a sharp, clear, and good defogging effect. The data flow direction of the AC-GAN algorithm is shown in Fig. 1.

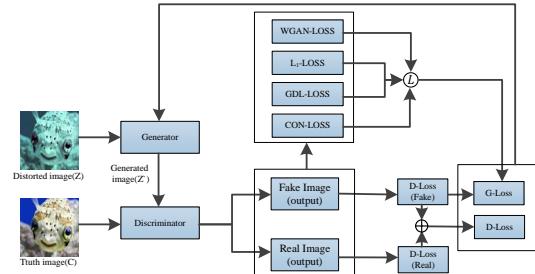


Fig. 1 AC-GAN process.

3.2 Generative and discriminative

Underwater image enhancement is subject to color shift correction, which also prevents details of the

image from being lost during the enhancement process. Therefore, we have improved the U-Net network in the generation network. The CBAM [13] is introduced in the U-Net network encoder, and the Attention Gate Block (AG) [14] and the Recurrent Residual Convolutional Block [15] are introduced in the decoder. The general framework of the model is shown in Fig. 2.

As shown in Fig. 2, the generative network of AC-GAN consists of a modified ACU-Net network, which performs the global feature extraction and learning tasks, respectively. In addition, the discriminative network adopts the structure of PatchGAN [16] based on the Markov model, which is more computationally efficient and has a wider application area compared with the global discriminative algorithm.

Our modified U-Net network model has five layers and the model input pixels are 256×256 color images. The CBAM module is added after each convolutional layer of the encoder for feature extraction. The convolution kernels in the 5-layer compressed path are 64, 128, 256, 512, and 1024, respectively. The size of the 5-layer feature map is 256×256 , 128×128 , 64×64 , 32×32 , and 16×16 . The feature information of the amplification and contraction paths are fused between the same layers using a jump connection. We introduce an attention mechanism module between the jump connections of each layer to improve the ability to learn deep abstract features. The input of the attention gate module is the output of the n^{th} layer of the systolic path and the output of the $n^{th} + 1$ layer after up-sampling. After obtaining the attention probability distribution of the features, the attention gated output is then fused with the up-sampled layer $n^{th} + 1$ output. After each fusion, a normal convolution is performed, followed by a recurrent residual convolution process. The size of the convolution kernels for both the systolic path and the augmented convolution is 3×3 . Each convolutional layer is followed by a nonlinear activation function ReLU. The number of convolutional kernels at the output of this model is three. ACU-Net considers more comprehensive spatial contextual information in image enhancement.

We use Markov model-based PatchGAN architecture for the discriminative network, which can discriminate at the image patch level. The Markov

discriminator outputs an $N \times N$ matrix. Each element represents a relatively large perceptual field in the original image, i.e., corresponds to a region in the original image. It aims to identify whether each region is a real image or an image generated by a generator. The overall decision is made by averaging the realism of all regions. This architecture is important to capture high-frequency features such as local textures and styles efficiently. In addition, this configuration is computationally more efficient because it requires fewer parameters compared to global discriminations. Unlike conventional discriminators that output a scalar value corresponding to real or false, our Patch-GAN discriminator uses five convolutional layers to convert a $256 \times 256 \times 6$ input into a $32 \times 32 \times 1$ output which allows the network model to focus more on image detail information. Furthermore, the output $32 \times 32 \times 1$ feature matrix provides a metric for high-level frequencies, which allows for more efficient acquisition of high-frequency features.

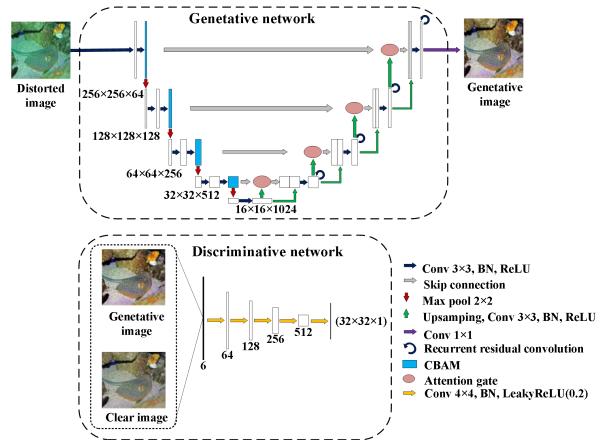


Fig. 2 AC-GAN network structure.

3.2.1 U-Net encoder

Most underwater image enhancement networks treat local features in different channels and different spaces in each image equally. However, due to the scattering of light in water and other reasons, underwater images often have an uneven distribution of local features in different channels and different spaces. To solve the above problem, we add CBAM in the U-Net encoder and introduce AG and RRCU in the decoder.

To extract more accurate image feature information, we add a simple but effective CBAM to the U-Net encoder. CBAM consists of Channel Attention Module (CAM) and Spatial Attention Module (SAM). CAM considers the importance of pixels in different channels, and SAM considers the importance of pixels at different locations in the same channel. Therefore, combining these two attention modules can extract image features more fully.

Channel attention focuses on what is important on the graph. Pooling is done to extract high-level features, and different pooling means that the extracted high-level features are richer. Mean pooling has feedback for every pixel point on the feature graph. Maximum value pooling performs gradient back propagation calculations with feedback from gradients only where the response is greatest in the feature map. First, the feature map F is input into the channel attention module. Then, the input feature map F undergoes maximum global pooling, global average pooling, and share Multi-Layer Perceptron (MLP), and the output features undergo element-wise summation operation. Finally, the output weights $M_c(F)$ of channel attention are generated after the *sigmoid* activation operation. The output feature map F' of the channel attention module is obtained by multiplying $M_c(F)$ and input feature F element by element, as shown in Fig. 3.

Spatial attention is concerned with the places on the graph that have an important role. First, F' is used as the input feature map for the spatial attention model, which is compressed in the channel dimension by maximum global pooling and global average pooling. Second, the extracted two feature maps are maximally subjected to a channel-based merging operation to obtain a two-channel feature map, which is subsequently reduced to a single channel by a 7×7 convolution operation. Then a *sigmoid* function is applied to generate the output weights $M_s(F')$ of the spatial attention module. Finally, $M_s(F')$ is multiplied by the channel-refined feature F' to obtain the final output feature map F'' .

CBAM links channel attention and spatial attention to increase the weight of useful features in channel and space to enhance useful information and suppress useless information. The U-Net encoder incorporates CBAM to reduce the

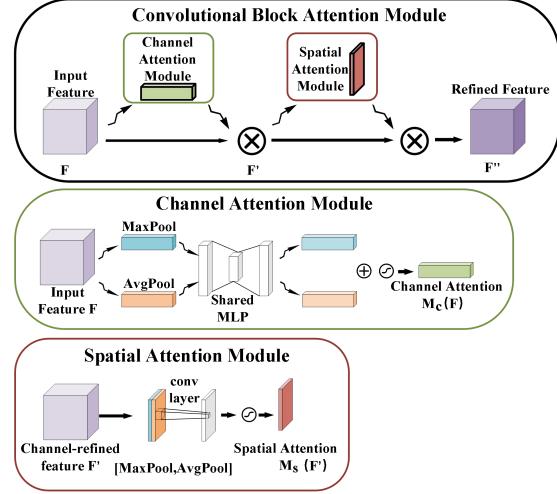


Fig. 3 CBAM module.

amount of computation on redundant information, allowing the network to process features in different channels and spaces unequally, and to enhance underwater images with different degrees of degradation more flexibly.

3.2.2 U-Net decoder

First, we introduce the AG module in the U-Net decoder. This module generates gate information to readjust the weight coefficients of the features at each spatial location. When training with this model, it can inhibit the model from learning task-irrelevant parts while aggravating the learning of task-relevant features, and its framework is shown in Fig. 4. Second, RRCU are used in the encoder and decoder process instead of the traditional *conv+relu* layer. It avoids the gradient disappearance in the back propagation in the deep network structure and effectively increases the network depth, whose framework is shown in Fig. 5.

Introducing AG in the U-Net decoder, the coarse-grained feature maps capture contextual information and highlight the classes and locations of foreground objects. Subsequently, feature maps extracted at multiple scales are merged by skip connections to combine coarse- and fine-grained dense predictions. In addition, AG improves model sensitivity and accuracy by suppressing the activation of features in irrelevant regions. Finally, AG is added at each layer of the U-Net decoder to ensure that attention units at different scales can

influence the response to what is important, allowing key information about the input image to be selectively captured and processed.

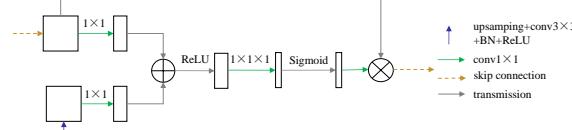


Fig. 4 Attention gate module.

Fig. 5(a) shows the convolution module of the original U-Net model. Fig. 5(b) shows the U-Net convolution module with recurrent cyclic convolution. Fig. 5(c) shows the U-Net convolution module with residual concatenation. To avoid the problem that global features are partially missing during extraction, we use the U-Net convolution module with *residual+recurrent* instead of the traditional *conv+relu* layer in the U-Net decoder. The recurrent residual layer adds up the features to help feature extraction and facilitates its recovery of more detailed information when generating images. Feature accumulation according to different time steps ensures a better and stronger feature representation. One of the recurrent convolution layers usually corresponds to different time-step inputs, and we take two time-steps (timestep=2) to cycle twice, as shown in Fig. 5(d). The way of feature summation with different time steps yields more expressive features, which also helps to extract lower-level features. The improved U-Net decoder helps the training of the deeper network. The improved U-Net decoder helps the training of the deep network and improves the performance of the network in identifying detailed edges without increasing the parameters.

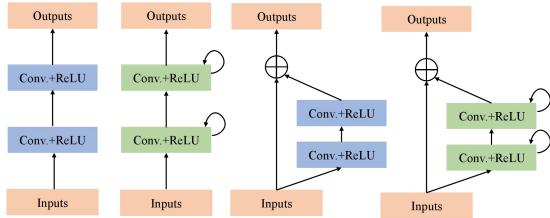


Fig. 5 ^(a) Different variant of convolutional and recurrent convolutional units. (a) Forward convolutional units. (b) Recurrent convolutional block. (c) Residual convolutional unit. (d) RRCU.

3.3 Loss function

(1) Adversarial loss

The traditional GAN loss function takes the logarithm of the loss function to calculate the distance between the probability distribution of the original data and the generated data. However, there is the problem of gradient disappearance. In this case, we use L_{WGAN} loss. When the datasets of the two distributions overlap less, the Wasserstein distance still shows the distance between the two distributions.

The CAR-GAN algorithm is based on a generative adversarial network model. Its adversarial loss function L_{WGAN} as shown in equation (1):

$$L_{WGAN}(G, D) = \mathbb{E}[D(I^C)] - \mathbb{E}[D(G(I^D))] + \lambda_{GP} \mathbb{E}_{\hat{x} \sim \mathbb{P}_{\hat{x}}}[(\|\nabla_{\hat{x}} D(\hat{x})\|_2 - 1)^2], \quad (1)$$

where $\mathbb{P}_{\hat{x}}$ is defined as the samples along a straight line between point pairs from the original data and generator distribution, λ_{GP} is a weighing factor, I^C is the underwater image without distortion, I^D is the same image with distortion.

(2) L_1 loss

To give G some ground truth, the low-frequency features in the image need to be fully extracted. We use the L_1 regularization term loss. Let G extract samples and features from the global similarity space of L_1 sense [16, 17]. The L_1 loss term is shown in equation (2):

$$L_1 = \mathbb{E}[\|I^C - (G(I^D))\|_1]. \quad (2)$$

(3) Gradient difference loss

Generative models often produce relatively blurred images. To solve this problem, we use gradient difference loss L_{GDL} to improve the gradient prediction of the image by directly penalizing the generator to achieve a sharpened image. The predicted image of I^C $I^P = G(I^D)$. Parameter α is an integer greater than or equal to 1. The gradient difference loss L_{GDL} as follows:

$$L_{GDL}(I^C, I^P) = \sum_{i,j} |I_{i,j}^C - I_{i-1,j}^C| - |I_{i,j}^P - I_{i-1,j}^P|^{\alpha} |I_{i,j}^C - I_{i,j-1}^C| - |I_{i,j}^P - I_{i,j-1}^P|^{\alpha}. \quad (3)$$

(4) Image content loss

Inspired by [18, 19], we add a content loss term to the target and let G generate an enhanced

image with similar content to the real image. Then, we input the generated and original high-resolution images into the VGG19 network for feature extraction. Here, only a portion of these features is extracted using the VGG19 network. The image content function $\Phi(\cdot)$ is a high-frequency feature extracted from the 5th block of the conv2 layer of the VGG19 network that is trained in advance. The image content loss is as follows:

$$L_{CON}(G) = \mathbb{E}[||\Phi(I^C) - \Phi((G(I^D))||_2]. \quad (4)$$

Finally, we use the following objective function for pairwise training to determine the overall loss of AC-GAN L_{AC} as a linear combination of the above four losses to speed up the convergence of the network parameters. It guides G -learning to improve the quality of the generated images so that they are close to real clear images in terms of structure, color, and detail.

$$\begin{aligned} L_{AC}(G) &= \min_G \max_D L_{WGAN}(G, D) + \lambda_1 L_1 \\ &\quad + \lambda_2 L_{GDL}(I^C, I^P) + \lambda_3 L_{CON}(G), \end{aligned} \quad (5)$$

where, $\lambda_1 = 100$, $\lambda_2 = 10$, and $\lambda_3 = 30$ are the scale factors that we empirically adjusted to hyper parameters.

4 Experimental analysis and discussion of results

To demonstrate the effectiveness of the AC-GAN, we conducted a series of comparison experiments. The comparison methods include (1) A physically-based model: the dark underwater channel prior (UDCP) [20]. (2) Based on the non-physical model: Underwater Image Enhancement by Attenuated Color Channel Correction and Detail Preserved Contrast Enhancement (ACDC) [21]. (3) Data-driven models based on: underwater GAN with gradient penalty [22] and Simultaneous Enhancement and Super-Resolution (SESR) [23] and Fast Underwater Image Enhancement for Visual Perception (FUNIE-GAN) [24].

Due to the complexity of the underwater environment, real underwater image datasets are limited. Therefore, we select paired underwater images from the synthetic underwater image

dataset EUVP [24] to train AC-GAN. First, we use the synthetic underwater image dataset EUVP and real-world underwater image dataset RUIE [25] to test the feasibility and robustness of our proposed algorithm. Then, we performed a subjective and objective comparative analysis of the experimental results for each method. The edge detection with the Canny operator is used to verify the enhancement effect of the methods on fogging and uneven illumination in different water environments and the image contrast, texture details, and contour details enhancement. Finally, we perform ablation experiments on AC-GAN.

4.1 Subjective evaluation

We randomly selected four underwater degraded images to compare with the enhancement results of the six algorithms, and the result plots are shown in Fig. 6.

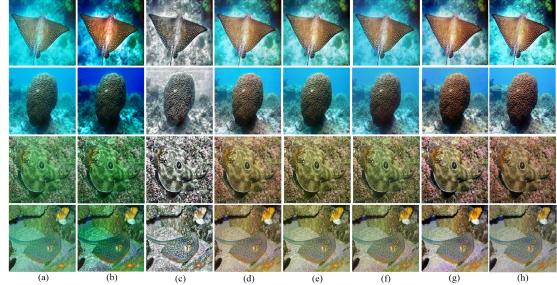


Fig. 6 Subjective comparisons on synthetic underwater images. (a) Distorted. (b) UDCP. (c) ACDC. (d) UGAN. (e) SESR. (f) FunIE-GAN. (g) Our result. (h) G.Truth.

Physical model-based methods recover underwater images by constructing degradation models and estimating model parameters. Many physical parameters and underwater optical properties are required, making these methods difficult to implement. Due to the lack of rich training data, these methods based on the dark channel prior perform poorly in marine scenes. As shown in Fig. 6(b), UDCP appears to overcompensate for color and fails to correct green tones in the image. Non-physical models do not have complex underwater physical parameters and produce better objective details by processing image pixels, such as the ACDC algorithm. However, the ACDC algorithm is less effective in enhancing the color shift correction of underwater images, as shown in

Fig. 6(c). The data-driven-based models UGAN, SESR, and FUnIE-GAN perform better. However, UGAN, SESR, and FUnIE-GAN all suffer from different degrees of color under-correction, as shown in Fig. 6(d)-(f). Overall, Fig. 6(h) shows our method’s clearest and closest results to the real image.

Image processing operations such as target detection and feature point matching pairs require a high level of image detail. However, existing underwater image enhancement methods often correct only for color shifts, and image details are still not recovered. We rely on canny edge detection [26] to evaluate visibility detail edge recovery for images with slight deterioration in visibility processed by AC-GAN and different underwater image enhancement algorithms. Since the non-physical model-based and physical model-based correction of color bias is less effective, we chose three learning-based methods similar to ours for visibility edge recovery evaluation. Fig. 7 shows that the top row is the input image, the middle row is the corresponding edge detection map, and the bottom row is the local detail map. From Fig. 7(a), we observe that the original image edges are difficult to detect due to the strong scattering. For these lightly contaminated scenes, the edge detection maps of the underwater images enhanced by UGAN and SESR methods do not increase much in the retrieved edges compared with the original image edge detection maps. It indicates that the UGAN and SESR methods are less effective in detail recovery, although they correct the color bias. Fig. 7(d) shows that the FUnIE-GAN edge map appears with local edges that do not exist in the edge map of the real clear image, and local noise is found in the FUnIE-GAN enhanced image after zooming in on the local map. As observed in Fig. 7(e), the edge detection map shows that our enhancement method has more visible edges and is closer to the edge texture of the clear image in the image space. This indicates that the proposed AC-GAN can effectively correct the color bias and reveal more details of the image structure without generating additional noise.

To verify the generalization ability of the model in the complex water environment and effectively illustrate the enhancement effect of the model on the images in real scenes. We select ten underwater images with different turbidity and

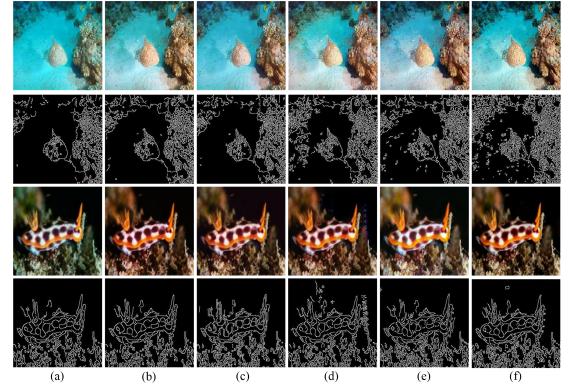


Fig. 7 Edge detection experiment. (a) Distorted. (b) UGAN. (c) SESR. (d) FUnIE-GAN. (e) Our result. (f) G.Truth.

different degrees of color shift from the RUIE-UCCS publicly available underwater real image dataset for the experiment. These images are classified into five categories: bluish underwater images, greenish underwater images, greenish-bluish underwater images, and deep-blue and haze underwater images.

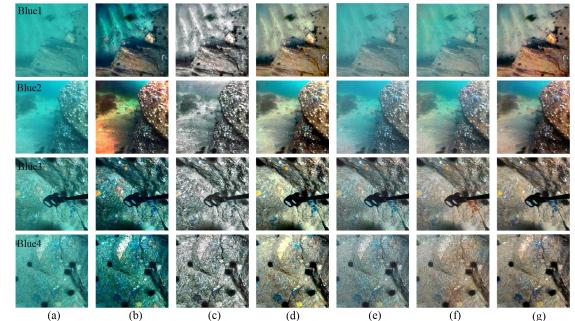


Fig. 8 Subjective comparisons on the bluish underwater images. From left to right are raw underwater images and the result of UDCP, ACDC, UGAN, SESR, FUnIE-GAN, and the proposed method.

The overall contrast of the enhanced images by the UDCP algorithm is improved, but its color compensation is too strong, resulting in blurred image details. The overall contrast of the enhanced images by the ACDC algorithm is significantly improved, but its recovered image color is distorted. The enhancement of SESR and FUnIE-GAN algorithms are ineffective in enhancing the deep-blue and haze underwater images, as shown in Fig. 11. The improvement for bluish and greenish underwater images with deeper pollution of Blue1 and Blue2 is not obvious, as shown in

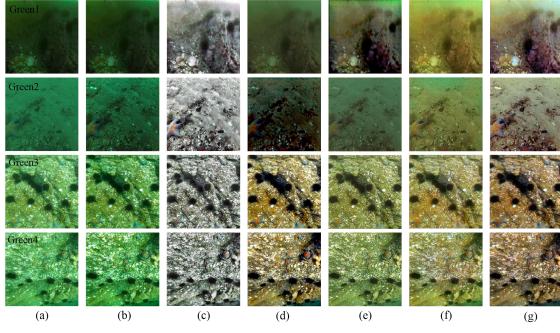


Fig. 9 Subjective comparisons on the greenish underwater images. From left to right are raw underwater images and the result of UDCP, ACDC, UGAN, SESR, FUnIE-GAN, and the proposed method.

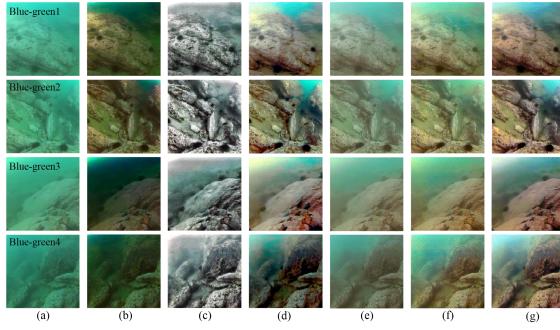


Fig. 10 Subjective comparisons on the greenish bluish underwater images. From left to right are raw underwater images and the result of UDCP, ACDC, UGAN, SESR, FUnIE-GAN, and the proposed method.

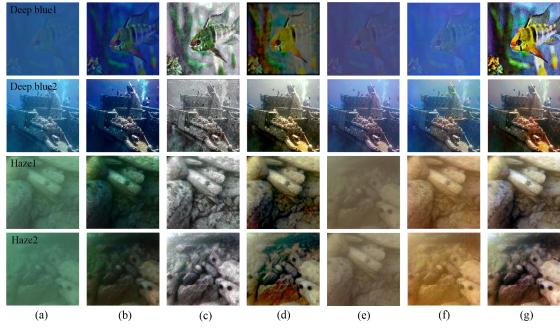


Fig. 11 Subjective comparisons on the deep-blue and haze underwater images. From left to right are raw underwater images and the result of UDCP, ACDC, UGAN, SESR, FUnIE-GAN, and the proposed method.

Fig. 8 and Fig. 9. Its improvement in greenish bluish underwater images is certain, but there is the problem of unnatural subjective vision and poor contrast, as shown in Fig. 10. The UGAN algorithm and our proposed algorithm enhance images significantly better than several other algorithms. However, the enhancement effect of the

UGAN algorithm on the heavily polluted greenish underwater images is average, such as Green1 and Green2 in Fig. 8. The UGAN algorithm significantly improves the haze and greenish-bluish underwater images' color shift. However, there exist unnatural color-corrected backgrounds, such as unnatural local exposures and abnormally dark areas in Blue-Green4 in Fig. 10. In contrast, our algorithm-enhanced images eliminate the effect of color bias, have higher visibility and more natural and realistic colors, and enhance the contrast and details of the image.

4.2 Objective evaluation

The subjective vision shows that the AC-GAN algorithm is more effective in enhancing underwater images with different degrees of degradation. To further validate the effectiveness of AC-GAN, we analyzed it from an objective perspective. We selected two Full-Reference metrics, including Peak-Signal to Noise Ratio (PSNR) [27] and Structural Similarity Index Measurement (SSIM) [27] to evaluate the quality of the underwater images in Fig. 6. In addition, we further selected four non-reference metrics, including Information Entropy (IE) [28], Patch-based Contrast Quality Index (PCQI) [29], Underwater Image Quality Metric (UIQM) [30], and Underwater Color Image Quality Evaluation (UCIQE) [31] to evaluate the quality of the real underwater images of Fig. 8–Fig. 11.

First, IE is the average number of information describing the color richness of underwater images. It reflects the richness of image information, and the larger value of information entropy represents the richer image information, the clearer image, and the better quality. Second, PCQI mainly evaluates the contrast perception of underwater images by human eyes from an objective perspective. Finally, UIQM and UCIQE are more comprehensive metrics designed specifically to evaluate the quality of underwater images. Higher values of PCQI, UCIQE, and UIQM indicate better image quality.

	UDCP	ACDC	UGAN	SESR	FUnIE	Ours method
image1	14.2165	17.8667	26.5786	25.9292	26.0606	27.2468
image2	16.4124	19.7087	25.3870	21.2684	24.2953	25.7835
image3	19.8496	19.4080	26.1746	28.2917	27.4836	25.8003
image4	25.5438	17.4711	23.2075	24.4714	23.3143	22.5326
average	19.0056	18.6137	25.3369	24.9902	25.2884	25.3408

Table 1 Comparison of PSNR[27] evaluation indicators.

4.3 Application in high-level computer vision tasks

4.3.1 Evaluation of object detection results

To demonstrate that our algorithm reduces the effect of color bias on the underwater visual system, we verified the effectiveness of our algorithm in enhancing underwater recovery image recognition using the Google Vision API (<https://cloud.google.com/vision/>). As shown in Fig. 12, fish in the original image were recognized as broad biological categories, whereas the enhanced image was recognized more specifically as clownfish and anemonefish. In addition, most of the metrics values were also improved. It can be seen that underwater images enhanced by our algorithm improve the accuracy of target detection. In addition, underwater images have more interfering factors that affect tag recognition. Therefore, we evaluated the label detection of real underwater images and algorithm-enhanced underwater recovery images, including label name and correct rate. As shown in Fig. 12(a), the Google Vision API recognizes the two sharks in the figure as a whole as one animal and incorrectly recognizes the rock and human legs and flippers as a single animal. After our algorithm enhances the image, the misrecognition of Google Vision API is effectively corrected. It is worth mentioning that these labels are indeed present in the image.

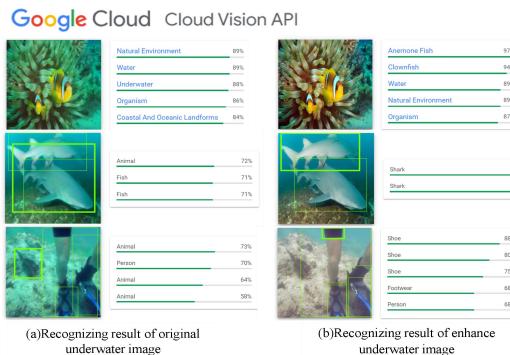


Fig. 12 A sample of improving target detection results of Google Vision API.

4.3.2 Evaluation of text detection results

To better apply underwater image enhancement techniques to underwater relic exploration and underwater trash brand statistics, etc. We test the recognition accuracy of text information in underwater enhanced images. As shown in Fig. 13, we conducted OCR experiments on the Google Vision API. We chose a real and artificially rendered underwater pollution image for validation. For the real underwater pollution image, we chose a free-diving flipper image with text as a typical experiment. In the experimental example, OCR incorrectly recognized the letters as ON, while the underwater enhanced image corrected this misrecognition. Artificially rendered underwater pollution image. We chose an image with text markers in the lower right corner as the experimental example. OCR fails to recognize any text for the underwater pollution image, while the underwater enhanced image accurately recognizes the corresponding text. We can see that the underwater image enhanced by our algorithm can recognize more textual information with higher accuracy by OCR detection. The experimental results show that our algorithm can improve the accuracy and clarity of text recognition in underwater-enhanced images.

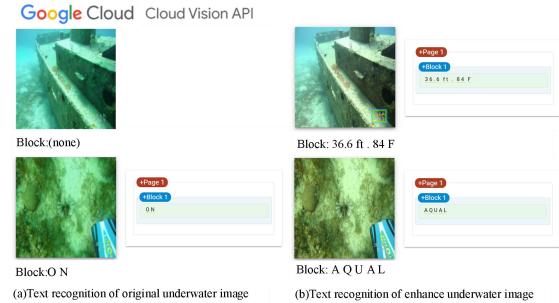


Fig. 13 A sample of improving the text recognition result of Google Vision API.

4.4 Ablation Research

To verify the effectiveness of ACU-Net, we conducted comparison experiments. U-Net was trained as a generative network. The enhanced underwater image details and local exposures were recovered better as the number of epoch rounds increased, as shown in Fig. 14(b)-(c). In contrast,

ACU-Net trained as a generative network, the underwater images enhanced by the first epoch round and the underwater images enhanced by the 50th round of U-Net have similar results, as shown in Fig. 14(d)-(e). The local exposure and details of the degraded underwater images enhanced by the 50th epoch round of ACU-Net are recovered. It indicates that using multiple attention mechanism and recurrent residual convolution in the encoder and decoder process can extract the image features more fully and retain the image details better, which helps to train a deeper network structure and avoid gradient unlearning.

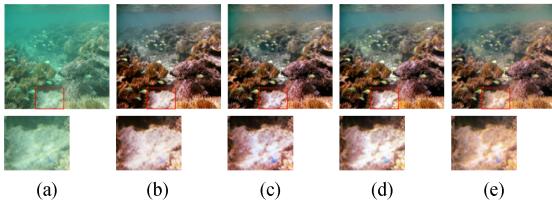


Fig. 14 U-Net and ACU-Net comparison experiment. (a) Original image. (b) U-Net epoch 0. (c) U-Net epoch 50. (d) ACU-Net epoch 0. (e) ACU-Net epoch 50.

We further demonstrate the contribution of each loss term of AC-GAN to the enhancement model, and we perform ablation experiments. For example, in Fig. 15(b), we observe that the L_1 helps to generate sharper images. Similarly, the L_{GDL} helps in color correction, and the local green color is corrected as shown in Fig. 15(c). And the L_{CON} helps to provide finer texture details, such as sharper fisheye details, as shown in Fig. 15(d).

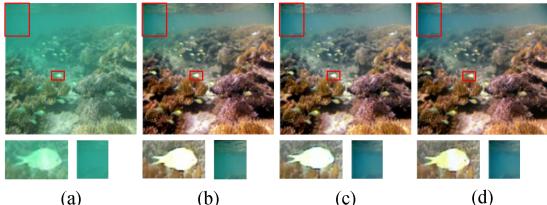


Fig. 15 Ablation experiment. (a) Input. (b) L_1 loss. (c) L_1 loss, L_{GDL} loss. (d) L_1 loss, L_{GDL} loss, L_{CON} loss. (e) ACU-Net epoch 50.

5 Conclusions

To eliminate the influence of light absorption and scattering on imaging, the enhanced underwater images represent the ocean scene information completely. We propose ACU-Net for underwater image enhancement. Combining the multi-attention and recurrent residual convolution of U-Net can fully extract the image features and better recover the image details. We train our model with synthetic datasets, and the trained model can effectively enhance underwater images with different color shifts and turbidity. Experiments show that our algorithm has good feature extraction ability and robustness for underwater images. Our proposed algorithm can correct color distortion, improve the contrast of underwater images, and recovers image details.

Contributions

Shuqi Wang: Conceptualization, Methodology, Data curation, Software, Writing - original draft, Writing - review & editing. **Zhixiang Chen:** Conceptualization, Methodology, Supervision, Writing - review & editing. **Hui Wang:** Supervision, Writing - review & editing.

Ethical statement

I testify on behalf of all the co-authors:

- * This material has not been published in whole or in part elsewhere.
- * The manuscript is not currently being considered for publication in another journal.
- * All authors have been personally and actively involved in substantive work leading to the manuscript, and will hold themselves jointly and individually responsible for its content.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the Natural Science Foundation of Zhangzhou under Grant

ZZ2020J33, the Natural Science Foundation of Fujian under Grant 2023J01924, and the National Natural Science Foundation of China under Grant 62001199.

Data availability statement

The data that support the findings of this study are available from the corresponding author Z. Chen and the first author S. Wang, upon reasonable request.

References

- [1] R. Schettini, S. Corchs, Underwater image processing: state of the art of restoration and image enhancement methods. EURASIP journal on advances in signal processing **2010**, 1–14 (2010)
- [2] F. Bonin, A. Burguera, G. Oliver, Imaging systems for advanced underwater vehicles. Journal of Maritime Research **8**(1), 65–86 (2011)
- [3] M. Ludvigsen, B. Sortland, G. Johnsen, H. Singh, Applications of geo-referenced underwater photo mosaics in marine biology and archaeology. Oceanography **20**(4), 140–149 (2007)
- [4] A. Dudhane, P. Hambarde, P. Patil, S. Murala, Deep underwater image restoration and beyond. IEEE Signal Processing Letters **27**, 675–679 (2020)
- [5] P. Hambarde, S. Murala, A. Dhall, Uw-gan: Single-image depth estimation and image enhancement for underwater images. IEEE Transactions on Instrumentation and Measurement **70**, 1–12 (2021)
- [6] M.R. Khan, A. Kulkarni, S.S. Phutke, S. Murala, in *2023 International Joint Conference on Neural Networks (IJCNN)* (IEEE, 2023), pp. 1–8
- [7] K. Singh, R. Kapoor, S.K. Sinha, Enhancement of low exposure images via recursive histogram equalization algorithms. Optik **126**(20), 2619–2625 (2015)
- [8] S. Zhang, T. Wang, J. Dong, H. Yu, Underwater image enhancement via extended multi-scale retinex. Neurocomputing **245**, 1–9 (2017)
- [9] J. Li, K.A. Skinner, R.M. Eustice, M. Johnson-Roberson, Watergan: Unsupervised generative network to enable real-time color correction of monocular underwater images. IEEE Robotics and Automation Letters **3**(1), 387 – 394 (2018)
- [10] Chen, Xingyu, Junzhi, Kong, Shihan, Zhengxing, Fang, Towards real-time advancement of underwater visual quality with gan. IEEE Transactions on Industrial Electronics **66**(12), 9350–9359 (2019)
- [11] J.Y. Zhu, T. Park, P. Isola, A.A. Efros, Unpaired image-to-image translation using cycle-consistent adversarial networks. Proceedings of the IEEE international conference on computer vision. pp. 2223–2232 (2017)
- [12] C.G. Chongyi Li, Jichang Guo, Emerging from water: Underwater image color correction based on weakly supervised color transfer. IEEE Signal processing letters **25**(3), 323–327 (2018)
- [13] S. Woo, J. Park, J.Y. Lee, I.S. Kweon, in *Proceedings of the European conference on computer vision (ECCV)* (2018), pp. 3–19
- [14] O. Oktay, J. Schlemper, L.L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N.Y. Hammerla, B. Kainz, et al., Attention u-net: Learning where to look for the pancreas. arXiv preprint arXiv:1804.03999 (2018)
- [15] M.Z. Alom, M. Hasan, C. Yakopcic, T.M. Taha, V.K. Asari, Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation. arXiv preprint arXiv:1802.06955 (2018)
- [16] P. Isola, J.Y. Zhu, T. Zhou, A.A. Efros, in *Proceedings of the IEEE conference on computer vision and pattern recognition* (2017), pp. 1125–1134

- [17] X. Yu, Y. Qu, M. Hong, in *International Conference on Pattern Recognition* (Springer, 2018), pp. 66–75
- [18] J. Johnson, A. Alahi, L. Fei-Fei, in *European conference on computer vision* (Springer, 2016), pp. 694–711
- [19] A. Ignatov, N. Kobyshev, R. Timofte, K. Vanhoey, L. Van Gool, in *Proceedings of the IEEE International Conference on Computer Vision* (2017), pp. 3277–3285
- [20] P. Drews, E. Nascimento, F. Moraes, S. Botelho, M. Campos, in *Proceedings of the IEEE international conference on computer vision workshops* (2013), pp. 825–830
- [21] W. Zhang, Y. Wang, C. Li, Underwater image enhancement by attenuated color channel correction and detail preserved contrast enhancement. *IEEE Journal of Oceanic Engineering* (2022)
- [22] C. Fabbri, M.J. Islam, J. Sattar, in *2018 IEEE International Conference on Robotics and Automation (ICRA)* (IEEE, 2018), pp. 7159–7165
- [23] M.J. Islam, P. Luo, J. Sattar, Simultaneous enhancement and super-resolution of underwater imagery for improved visual perception. arXiv preprint arXiv:2002.01155 (2020)
- [24] M.J. Islam, Y. Xia, J. Sattar, Fast underwater image enhancement for improved visual perception. *IEEE Robotics and Automation Letters* **5**(2), 3227–3234 (2020)
- [25] R. Liu, X. Fan, M. Zhu, M. Hou, Z. Luo, Real-world underwater enhancement: Challenges, benchmarks, and solutions under natural light. *IEEE transactions on circuits and systems for video technology* **30**(12), 4861–4875 (2020)
- [26] J. Canny, A computational approach to edge detection. *IEEE Transactions on pattern analysis and machine intelligence* **8**(6), 679–698 (1986)
- [27] C. Li, S. Anwar, F. Porikli, Underwater scene prior inspired deep underwater image and video enhancement. *Pattern Recognition* **98**, 107038 (2020)
- [28] W. Zhang, L. Dong, X. Pan, J. Zhou, L. Qin, W. Xu, Single image defogging based on multi-channel convolutional msrccr. *IEEE Access* **7**, 72492–72504 (2019)
- [29] S. Wang, K. Ma, H. Yeganeh, Z. Wang, W. Lin, A patch-structure representation method for quality assessment of contrast changed images. *IEEE Signal Processing Letters* **22**(12), 2387–2390 (2015)
- [30] K. Panetta, C. Gao, S. Agaian, Human-visual-system-inspired underwater image quality measures. *IEEE Journal of Oceanic Engineering* **41**(3), 541–551 (2015)
- [31] M. Yang, A. Sowmya, An underwater color image quality evaluation metric. *IEEE Transactions on Image Processing* **24**(12), 6062–6071 (2015)