

Received November 7, 2020, accepted November 24, 2020, date of publication November 30, 2020,
date of current version December 16, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3041280

Underwater Image Enhancement Based on a Spiral Generative Adversarial Framework

RUYUE HAN¹, YANG GUAN¹, ZHIBIN YU^{1,2}, (Member, IEEE), PENG LIU^{1,3}, (Member, IEEE), AND HAIYONG ZHENG¹, (Member, IEEE)

¹Department of Electronic Engineering, College of Information Science and Engineering, Ocean University of China, Qingdao 266100, China

²Sanya Oceanographic Institution, Ocean University of China, Sanya 572000, China

³Computing Center, Ocean University of China, Sanya 572000, China

Corresponding authors: Zhibin Yu (yuzhibin@ouc.edu.cn) and Peng Liu (liupengouc@ouc.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61701463, and in part by the Key Technology Research and Development Program of Shandong (Public Welfare) under Grant 2019GHY112041.

ABSTRACT Underwater image enhancement has drawn much attention due to the significance of underwater vision. Although considerable progress has been made in this field, a key problem remains unsolved: how can we extract and enhance minutiae while trying to remove the noise caused by scattering and attenuation? To address this limitation, we propose a new underwater image enhancement technique with a novel spiral generative adversarial framework, named Spiral-GAN, which can effectively recover real-world underwater images with more details, vivid colors and better contrast. For steady training and color correction, we include the pixelwise losses that consist of a mean squared error and an angle error in our objective function. In addition, we design our generator with several deconv-conv blocks to preserve the details from the original distorted images. Furthermore, we present a spiral learning strategy for generalizing the enhancing model to effectively recover the real-world underwater images. Finally, we perform a number of qualitative and quantitative evaluations that suggest that our proposed approach can efficiently enhance the quality of underwater images, which can be further used for underwater object detection.

INDEX TERMS Underwater image enhancement, generative adversarial networks, spiral training.

I. INTRODUCTION

Compared with more sophisticated techniques, underwater vision technologies have drawn significant attention for their ability to enhance the quality of underwater images with a low cost of implementation. These technologies play important roles in numerous engineering and research tasks, such as the monitoring of marine species migration and coral reefs [1], [2], inspection of submarine cables and wreckage [3], deep ocean exploration [4], etc. The main challenge in restoring degraded underwater images is caused by strong scattering and absorption phenomena. Various particles, including plankton or dust, may change the direction of light upon collision, which causes underwater images to always suffer from low contrast, color distortion, blur and haze [5]–[7]. Generally, an additional artificial light source may extend the visibility range in the scene. However, such devices may increase the energy overhead. Furthermore, the presence of artificial lighting may lead to another

The associate editor coordinating the review of this manuscript and approving it for publication was Mostafa M. Fouda .

challenge for underwater image processing due to inhomogeneous illumination [8].

In recent years, deep neural networks have been shown to have a powerful ability to address nonlinear problems especially in the field of computer vision [9]. As an important branch of deep neural networks, generative adversarial networks(GANs) have shown great success in various vision tasks [10]–[12] due to the power of adversarial learning. Pix2pix [13], which is an important extension of GAN, showed great potential on image-to-image translation tasks. Zhu *et al.* further extended the idea and proposed CycleGAN, which does not need paired data for training [14]. For underwater image enhancement tasks, CycleGAN is proven to be efficient for paired underwater image synthesis [7].

Enlightened by their work, we consider the underwater image enhancement task as an image-to-image translation problem, and propose an end-to-end GAN-based enhancing model with a noval spiral learning strategy for underwater image enhancement. Fig. 1 displays some results of the proposed method.

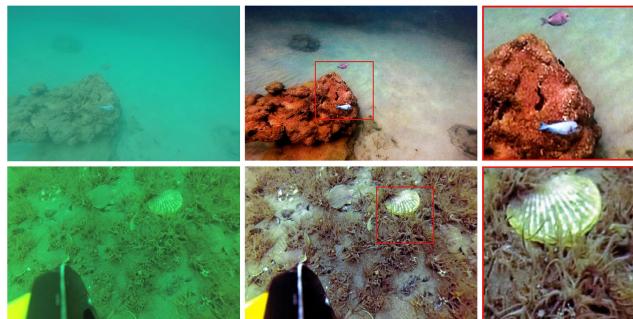


FIGURE 1. Results of our proposed underwater image enhancement technique, Spiral-GAN.

Our contributions are organized as follows:

- We present a novel spiral generative adversarial framework(Spiral-GAN) for the underwater image enhancement task to restoring underwater images. The generator with multiple conv-deconv blocks is designed carefully for preserving more meaningful details in original images.
- We present a novel spiral learning strategy for generalizing the enhancing model to effectively recover the real-world underwater images. In the training process, the proposed spiral training strategy can implicitly increase the training data, allowing the models to learn a more complex mapping, which can alleviate the domain shift problem—the models learned from synthetic underwater paired data often fail to generalize well to real-world underwater data.
- We employ a pixel-level objective function composed by a mean squared error and an angle error to stably train the model for overcoming the over-under exposure problem and avoiding color distortion.

II. RELATED WORK

The aim of underwater restoration or enhancement algorithms is to improve the quality of underwater images. According to the means of the imaging process, we simplify the restoration algorithms into three categories as follows:

A. MODEL-FREE METHODS

Model-free methods typically reallocate the pixel values of a given image to enhance contrast or color correction without modeling the underwater image formation process. There are many typical model-free approaches, including histogram equalization(HE) [15], contrast limited adaptive histogram equalization (CLAHE) [16], the gray-world assumption(GWA) [17], automatic white balance(WB) [18] and color constancy [19], that tried to adjust underwater image pixels in the spatial domain.

Ancuti et al. [20] proposed a fusion-based method by merging two images generated from the input image and the defined weights. Ghani and Isa [21]–[23] improved image contrast and reduced overenhancement and noise introduction by adopting the Rayleigh distribution with the variation

in ICM and UCM. Huang et al. [24] proposed relative global histogram stretching through two-step operations in RGB and CIE-Lab color models. In the frequency domain, according to [25], the high-frequency information of an image generally corresponds to the detail region or edge region, in where the pixel values are vary different with each other. The flat region usually is composed by the low-frequency components. Therefore, by restraining the low-frequency component and amplifying the high-frequency component can improving the underwater images quality in transform domain. Khan et al. [26] introduced a wavelet-based fusion method by using low-pass and high-pass filters to preserve desired frequencies presented in hazy underwater images. Vasamsetti et al. [27] presented a framework of a wavelet-based perspective enhancement technique for underwater images. Most of the model-free approaches are fast but suffer from overenhancement, color distortion and shifting [28].

B. MODEL-BASED METHODS

Model-based methods usually take into account the physics of image processing and light transmission and build physics models. According to some prior assumptions and observations, people can obtain key parameters of the degradation model and then obtain desired results by inverting the model.

He et al. [29] proposed the dark channel prior(DCP) to address the image haze removal issue. Noting the similarity between hazy underwater images and foggy land images on de-scattering, many researchers developed underwater restoration methods based on DCP [5], [30]. Unlike foggy land images, the red channel of underwater images contains mostly pixels whose intensities are close to zero, which will lead to an incorrect transmission map estimation with the standard DCP. Considering this limitation, Drews et al. proposed underwater DCP (UDCP) [31] to apply DCP in the green and blue channels. Galdran et al. proposed a Red Channel Prior(RCP) [32], a variant of the Dark Channel method, for underwater image restoration.

In addition to these methods derived from the DCP algorithm, there are various physical priors for underwater restoration. Peng et al. [33] proposed a method to estimate the depth map of underwater images by combining image blurriness with the image formation model. To enhance the contrast and correct the color distortion of underwater images, Wang et al. proposed the maximum attenuation identification(MAI) [34] method to obtain the depth map from red channel information.

C. DATA-DRIVEN METHODS

Recent years, deep learning has been achieved significant advance especially in the field of vision for its enormous potential in handling non-linear problems [9]. The contemporary learning based methods provide state-of-the-art performance for image enhancement task [35]–[37]. Normally, these data-driven models require large amount of data paired with ground truth to get desired results, but for underwater

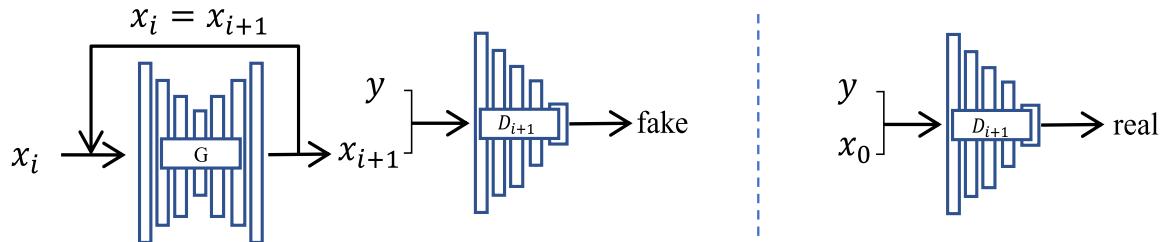


FIGURE 2. Overview of the proposed underwater image enhancement model, consisting of an enhancer G and a group of discriminators Ds whose number is determined by recurrent training cycles. In each cycle, a discriminator $D_{i+1}(i \geq 0)$ learns to distinguish between real and fake images that are generated by G ; meanwhile, the enhancer G tries to generate images indistinguishable from real images to fool the D_{i+1} .

image enhancement, that is a major hindrance towards directly applying a supervised approach to this specific task, because of it is difficult to obtain multiple quality images of the same scenes in real-world underwater environment. To alleviate the problem of lacking of paired training data, many underwater image enhancement studies have investigated the use of synthetic images and reported reasonable success. For example, Li *et al.* proposed WaterGAN [38], a learning-based model, which using a two-stage strategy for finally removing the color casts issues in underwater images. Firstly, WaterGAN takes in-air RGBD images as input and generates corresponding synthetic underwater images as output. And then both synthetic underwater and true color in-air, as well as depth data are fed into correction network for restoring images. Similarly, Fabbri *et al.* [7] proposed UGAN(bsed on Pix2Pix model [13]) and UGAN-P (UGAN with gradient penalty), the authors use CycleGAN as a preprocessing method to reconstruct distorted underwater images based on clean underwater images. After that, the synthetic distorted underwater images are used to prepare paired instances for adversarial training in a supervised pipeline. For enhancing underwater image in real-time, Islam *et al.* [39] proposed a supervised enhancement model, named FUNIE-GAN, whose traing data is prepared using the same procedure suggested by Fabbri *et al.* [7]. Meanwhile, Islam *et al.* also proposed an unsupervised enhancement model with the cycle-consistency loss suggested by Zhu *et al.* [14], in other way to alleviate the problem of obtaining data pairs. In addition to the GAN-based methods mentioned above, there are many CNN-based underwater image enhancement solutions. Li *et al.* [40] take advantage of 12 image enhancement methods to enhance the raw underwater images and then manually select visually appealing image as candidate high-quality reference. After constructing a large-scale paired underwater dataset, the authors proposed a baseline CNN-based enhancing model, named Water-Net, which was supervised to resolve underwater degradation issues. For improving visual perception of underwater imagery, Islam *et al.* present a supervised model, called Deep-SESR [41], the first learning-based approach for tackling the simultaneous enhancement and super-resolution (SESR) problem. As for paired training data,

they used same procedure to produce distorted images as mentioned in [7], [39], followed by Gaussian blurring and bicubic down-sampling.

However, there is still a long gap between the synthetic underwater images and realistic ones, owing to the complicated underwater environment and lighting conditions. Because of that, the learned distribution by data-driven methods trained on synthetic underwater images does not always generalize to real-world cases. Additionally, there are other problems, such as lacking details, exhibiting color shifting, in underwater images enhanced by these learning-based models. Addressing these aspects mentioned above is the main focus of this paper.

III. METHODOLOGY

Given a source domain X (distorted images) and a target domain Y (recovered images), our goal is to learn a mapping $F : X \rightarrow Y$ to perform automatic image enhancement. We adopt a conditional GAN-based framework [42] in which the generator(G) tries to learn this mapping by playing an iterative minmax game with the adversarial discriminator. Only by training all the networks together can for the enhancement network G realistically generate clean underwater images.

An overview of the proposed approach is shown in Fig. 2. First, we introduce a novel generator architecture to keep more minute information. Second, we provide a spiral training strategy that can greatly promote the performance of underwater image enhancement tasks. Third, we present a combination of loss functions for our adversarial framework.

A. NETWORK ARCHITECTURES

1) ENHANCEMENT GENERATOR

Our generator network is adapted from UGAN [7] following the principles of U-Net [43]. It is made up of the encoder and decoder network with skip connections between the mirror layers. The skip connection in the generator network has been shown to be effective [7], [13], [39] on many image-to-image translation tasks that require low-level information sharing from inputs and outputs. This skip connection operation is performed by concatenating the outputs produced from a convolutional layer i in the encoder to the outputs of the $n - i$ transpose convolutional layer in the decoder, where n is

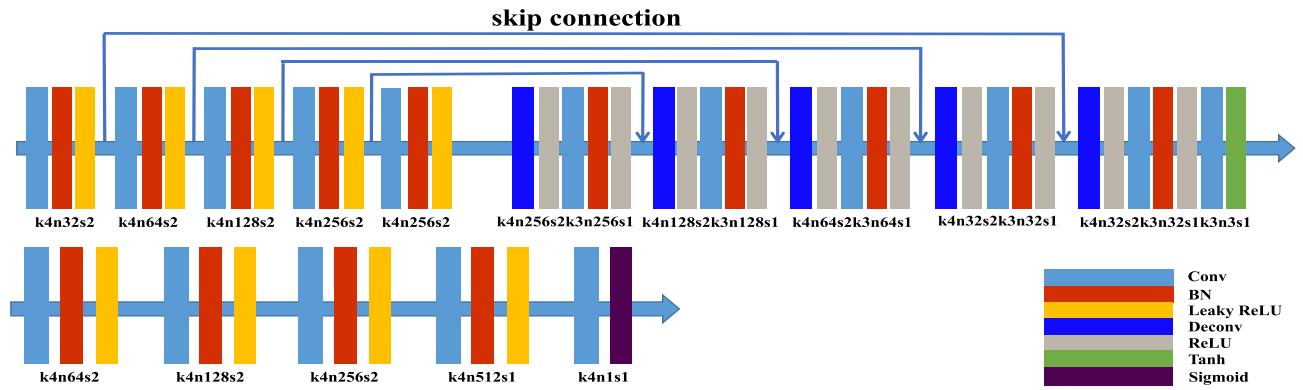


FIGURE 3. The generator and discriminator architecture of the proposed model are shown on the top and bottom row, respectively. “Conv” represents convolution layer. “BN” denotes Batch Normalization. “Deconv” denotes deconvolution layer.

the total number of downsampling convolution layers in the encoder and upsampling convolution layers in the decoder.

However, a main shortage of the encoder-decoder structure is that the feature maps may lose minute information during a sequence of quick downsampling and upsampling operations. For example, the size of the feature map will become 1×1 in the bottleneck of the UGAN, while the input image to the generator is 256×256 . To address this limitation, we decrease the downsampling layers from eight to five in the encoder to reduce the minute information loss. In addition, to minimize potential artifacts generated by the upsampling process, we include conv-deconv blocks, which have proven to be efficient in superresolution tasks [44], [45]. As shown in Fig. 3, each deconv-conv block has an additional convolutional layer with 3×3 filters and a stride of 1 after each transpose convolutional layer with 4×4 filters and a stride of 2 in the decoder. Convolutions with 4×4 filters are applied in all layers in the encoder; then, each convolution layer is followed by a batch normalization [46] and a nonlinear ReLU nonlinearity. The feature map of each layer and other model parameters are shown in Fig. 3, the k4n32s2 denotes a convolution layer has 32 convolution kernels with a kernel size 4 and stride 2. Inspired by the UGAN, our discriminators are adopt similarly with the one in UGAN. The details of a discriminator with five convolutional layers are displayed in Fig. 3, which consists of five downsampling convolutional layers. In each layer of the discriminator, the 4×4 convolutional filters are followed by a Batch Normalization [46] and leaky-ReLU nonlinear activation function [47], the output of last convolution layer as input to the sigmoid activation function.

B. TRAINING WITH THE SPIRAL STRATEGY

The major obstacle to implementing a deep learning model in underwater image enhancement tasks is the lack of paired training data. Although CycleGAN provides a possible solution to provide synthetic paired data, these data still have limited diversity, which may reduce the generalization performance of underwater image enhancement tasks. To alleviate

this problem, we introduce a simple and effective recurrent training strategy. Unlike other approaches that focused on introducing a recurrent submodule in the network structure [48], [49], our method uses this recurrent framework only during the training step. An overview of the general training procedure can be seen in Algorithm 1. During the recurrent training, we use the real distorted images of the training data to train the discriminator and generator at the first iteration of each recurrent operation, but in the remaining iterations of training, we use the fake images generated by the enhancement network in the latest iteration as the input in the next iteration to train the model. We derive inspiration from the idea of spiral learning that suggests that a learner learns more about a subject when the topic is reviewed. We call such a training process “spiral training”, which can gradually teach the generator how to restore the distorted images in each loop. The output of the generator from the previous loop will be used for the input of the next loop while the targets of each loop are the same. Meanwhile, using the generated data as the next step of training data can be considered a variant of data augmentation techniques. This strategy can reduce overfitting problems and improve the model’s generalization ability, which can restore different underwater scenes efficiently.

Please note the proposed spiral training framework is different with recurrent skip connection [50], [51]. The recurrent skip connection is always used in a recurrent neural network (RNN), which is a dynamic model. We always need to feed a RNN with dynamic sequences at both training and test stages. Our spiral GAN is a static model, and we only need to feed the network once at test stage. The spiral feeding metric is only used at training stage to stabilize the adversarial training process and lower the impact of overfitting.

C. OBJECTIVE FUNCTION

The objective of a standard conditional GAN can be expressed as:

$$\min_{G,D} L_{cGAN}(G, D) = \mathbb{E}_{X,Y}[\log D(X, Y)] + \mathbb{E}_X[\log(1 - D(X, G(X)))] \quad (1)$$

Algorithm 1 Training Details of Our Method

Require: the distorted-clean image pairs $\{x, y\}$, the number of total iterations T , the batch size B , the number of the maximum spiral loops N , the discriminator D and the generator G .

Require: initial discriminator parameters θ_0 , initial generator parameters ω_0 .

```

1:  $\theta \leftarrow \theta_0, \omega \leftarrow \omega_0$ 
2: for  $t = 1, \dots, T$  do
3:   Sample image pairs  $\{x_i, y_i\}_{i=1}^B$  from training data.
4:   for  $n = 1, \dots, N$  do
5:      $\{\tilde{y}_i \leftarrow \underbrace{G(\dots G(x_i) \dots)}_n\}_i^B$ 
6:     compute  $\nabla_{\omega}^n$  with total objective(Eq.(7)),  $\omega, \theta \leftarrow Adam(\omega, \theta, \nabla_{\omega}, \nabla_{\theta})$ .
7:      $\{x_i \leftarrow \tilde{y}_i\}_i^B$ 
8:   end for
9: end for
```

where G learns a mapping $\{X\} \rightarrow Y$ by minimizing this objective against an adversarial D that tries to maximize it, and $X(Y)$ means the source(target) domain. Existing approaches have shown that adding an $L2(L1)$ loss can enable G to capture low-level frequencies in the images, generating more realistic results [7], [13], [39]. We also explore this option, adding $L2$ loss in the objective:

$$L_{l2}(G) = \mathbb{E}_{X,Y}[\|Y - G(X)\|_2^2] \quad (2)$$

The $L2$ loss can measure the pixel value difference between real and generated images, for pixels that are far away from the real values while omitting the difference between two close pixels. In other words, $L2$ loss will push the generator to focus on the color distortion, which is a major challenge in underwater image restoration, rather than a bright/dark object.

In the process of propagation, light is scattered by water molecules and impurities to other angles and absorbed. So cameras equipped with lighting are needed to capture pictures in the deep ocean. However, due to the influence of various visibility conditions and underwater environments, the pictures have over-under exposure problems. To further enhance the performance of underwater image enhancement, we include another metric to restrict pixels in the RGB color space to avoid color mismatch called angular loss [52], [53]:

$$L_{angular}(G) = \mathbb{E}_{X,Y}[\angle(Y, G(X))] \quad (3)$$

where \angle denotes the operation that calculates the angular distance between pixels in RGB color space. The angular loss can stably train the model for overcoming the over-under exposure problem and avoiding color distortion.

Following the spiral strategy mentioned in section III-B, Equations 1, 2 and 3 can be updated as follows:

$$\begin{aligned} & \min_{G,D} \max L_{spiral_cGAN}(G, D) \\ & = \mathbb{E}_{X,Y}[\log D(X, Y)] \end{aligned}$$

$$+ \sum_N \mathbb{E}_X[\log(1 - D(X, \underbrace{G(\dots G(X) \dots)}_n))]) \quad (4)$$

$$L_{spiral_l2}(G) = \sum_N \mathbb{E}_{X,Y}[\|Y - \underbrace{G(\dots G(X) \dots)}_n\|_2^2] \quad (5)$$

$$L_{spiral_angular}(G) = \sum_N \mathbb{E}_{X,Y}[\angle(Y, \underbrace{G(\dots G(X) \dots)}_n)] \quad (6)$$

where N denotes the maximum spiral loops. Combining all the above mentioned loss functions, we obtain our final objective function as follows:

$$\begin{aligned} \mathcal{L} = \min_G \max_D & L_{spiral_cGAN}(G, D) \\ & + \lambda_1 L_{spiral_l2}(G) + \lambda_2 L_{spiral_angular}(G) \end{aligned} \quad (7)$$

During the experiment, we found that keeping the values of spiral-L2 and spiral-Angular at 10:1 is better. For keeping this ratio, we set $\lambda_1 = 100, \lambda_2 = 0.1$.

Although we not propose $L2$ loss and angular loss, we combine these losses with the spiral training strategy to stably train the model for overcoming the over-under exposure problem and avoiding color distortion. So the $L_{spiral_angular}$ and L_{spiral_l2} are difference from $L2$ loss and angular loss.

IV. EXPERIMENTS

A. DATASET INTRODUCTION

To evaluate the performance of our algorithm, we selected four large-scale real-world underwater datasets as test sets, details are as follows:

1) URPC DATASET

The URPC dataset¹ was collected by Dalian University of Technology and Zhangzidao Joint Lab and includes 4757 real-world underwater images with various distorted degree. All images in URPC have a greenish or blue-greenish hue. The entire dataset is used as test set, 1522 images are resized to a size of 1280×768 , 3235 images are resized to 768×512 .

2) EUVP DATASET

EUVP is an underwater image collection [39] that contains a large amount of real underwater images captured in different locations under various visibility conditions. This set suffers from heavy noise, and most of the images show a blue appearance. We randomly selected 4129 blue hue images and resized these images to 768×512 for testing.

3) UNDERWATER-MOT DATASET

The Underwater-MOT dataset is a self-collected dataset, collected from an outdoor pool over different time periods, which contains 1172 real-word underwater images suffering from the different levels of degradation. Due to changes

¹<http://en.cnurpc.org/>

in illumination changing and the concentrations of plankton, these images in the Underwater-MOT have greenish and bright-yellow tones. A few sample images from the Underwater-MOT dataset are shown in Fig. 6. All the images are resized to 1280×768 as the test set.

4) RUIE DATASET

The RUIE dataset [55] is a large-scale Real-world Underwater Image Enhancement dataset, it aims to evaluate the performance of the enhancement models in image visibility quality, color casts and object detection. We select the Underwater Image Quality Set (UIQS) and Underwater Color Cast Set (UCCS) to evaluate the performance of models in image visibility quality and color casts. We select entire UCCS subset(300 images)and 726 distortion seriously images of UIQS. All the images are resized to 512×512 as the test set.

B. MEASUREMENTS

Because there is no veritable scene that is available as the reference for real-world underwater images, we employ a non-reference metric UIQM [56], which is usually used for underwater image quality evaluation [39], [41]. This metric is composed of three attribute measures: the underwater image colorfulness measure (UICM), the underwater image sharpness measure(UISM), and the underwater image contrast measure (UIConM). Each attribute measure can be used separately for evaluating one aspect of the underwater image degradation. A higher score indicates the result is more consistent with human visual perception. The UIQM is expressed as follows:

$$UIQM = c_1 \times UICM + c_2 \times UISM + c_3 \times UIConM \quad (8)$$

where the colorfulness, sharpness, and contrast measures are linearly combined. We set the three parameters $c_1 = 0.0282$, $c_2 = 0.2953$, and $c_3 = 3.5753$ as the original work.

C. TRAINING DETAILS

Our training data are taken from the paired dataset [7], including 6128 paired images generated by CycleGAN. All training images are 256×256 resolutions and then randomly flipped in the horizontal and vertical directions throughout the entire training process. In all experiments, we set $\lambda_1 = 100$, $\lambda_2 = 0.1$, a batch size of 32, spiral loops as 10 and the Adam Optimizer [57] with learning rate $1e^{-4}$. Our model is trained on a GeForce GTX 1080 Ti for 10 recurrent times and 100 epochs.

D. COMPARISON AND EVALUATION

In this section, we conduct four groups of qualitative and quantitative experiments on four underwater datasets and one group of color accuracy test to demonstrate the generalization and effectiveness of the proposed method. To verify the superiority of our method in a fair manner, we select both conventional enhancement/restoration models

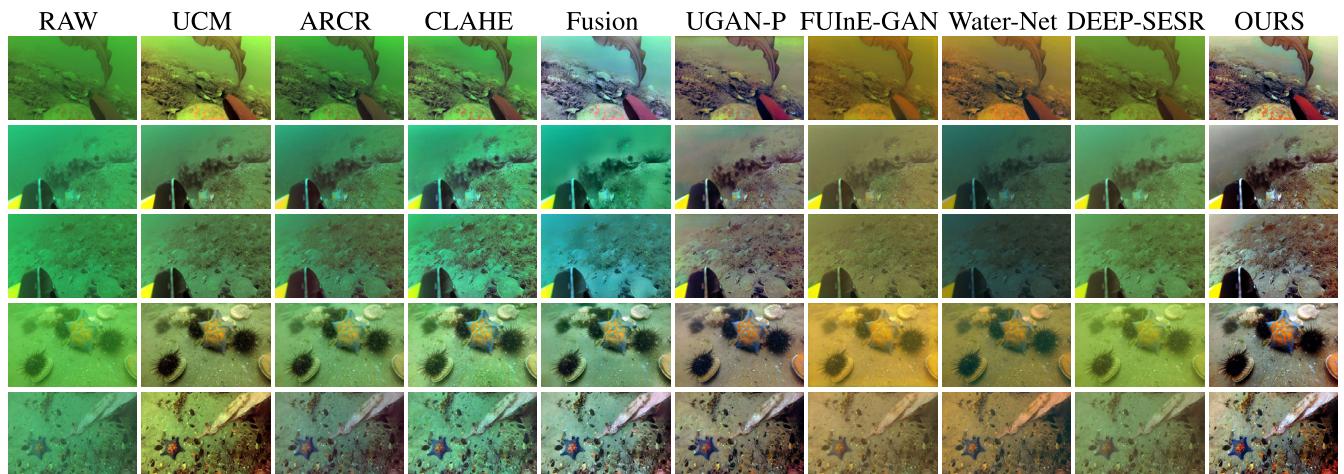
and recently developed learning-based methods as our comparisons. We consider four representative traditional underwater image enhancement/restoration algorithms, including unsupervised color correction method(UCM) [54], red channel prior(ARCR) [32], contrast limited adaptive histogram equalization(CLAHE) [16], and a fusion-based method(Fusion) [20]. For data-driven methods, we select four supervised learning-based methods as well, covering UGAN-P [7], FUNIE-GAN [39], Water-Net [40] and Deep-SESR [41]. All data-driven models with initial setups are trained on the dataset provided by UGAN-P, except for Deep-SESR, only the inference model is available, and only the enhancement performance is considered during the experiment. Meanwhile, we select UGAN-P as the baseline for evaluation. Next, we will discuss the performance results of all comparison algorithms on the four real-world underwater datasets in detail.

1) DISCUSSION ON URPC DATASET

The quantitative results of different underwater enhancement algorithms mentioned above on the URPC dataset are shown in Table 1. As shown in Table 1, our method achieves the best performance in terms of UIQM and UIConM, which indicates the enhanced underwater images by the proposed method have a vivid color, better sharpness and contrast, and the most visually appealing appearance. The unanimous conclusion can be deducted from Fig. 4. For traditional algorithms, ARCR hardly enhance the underwater images, particularly challenging images with serious degradation. Although UCM and CLAHE can remove some effects of scattering, they have insufficient saturation and contrast. Meanwhile, they cannot deal with color casting well, especially when the water is seriously greenish. Fusion can effectively remove the haze effect caused by light scattering and notably improve the image brightness, but this approach tends to generate oversaturation and color shifting. For data-driven methods, FUNIE-GAN has resolved the greenish color cast to a limited extent, but FUNIE-GAN aggravates the scattering effect. For Water-Net, it can increase the contrast and retain more details, but for some images (such as the images in the 2nd and 3rd rows), it cannot deal with the color casting well, and can even turn the water type into dusty blue. As for Deep-SESR method, it has slight improvement on resolving green color casting and removing the effect caused by scatter, but the visibility, color and details are still insufficient. Both UGAN-P and the proposed method can effectively remove the greenish background casting and the scattering effect. Compared with the baseline UGAN-P, our method has better visibility, cleaner details and fewer artifacts. For example, in our recovered underwater images, the texture details of sea plants and urchins are visually cleaner, and the color of starfish is more vivid. Thus, the qualitative comparison demonstrates that the results of our method are the best visually appealing among the compared results on the URPC dataset.

TABLE 1. Quantitative comparison of different methods on URPC dataset. The first and second best scores are colored red and blue, respectively.

Methods	None Data-driven Methods				Data-driven Methods				
	UCM	ARCR	CLAHE	Fusion	UGAN-P	FUNIE-GAN	Water-Net	Deep-SESR	Ours
UISM	7.0733	7.0944	7.0634	7.1215	6.7198	7.0876	6.9830	7.0521	6.8354
UICM	3.1178	1.6050	3.7077	1.4864	1.0901	-0.2439	1.6105	0.6512	2.7179
UIConM	0.5910	0.5337	0.5681	0.5615	0.6194	0.4359	0.4885	0.3728	1.0920
UIQM	4.2512	4.0910	4.1621	4.2152	4.2406	3.6445	3.8540	3.4339	5.9995

**FIGURE 4.** Subjective comparisons of different methods on URPC dataset. From left to right are raw underwater images, and the results of UCM [54], ARCR [32], CLAHE [16], Fusion [20], UGAN-P [7], FUNIE-GAN [39], Water-Net [40], Deep-SESR [41], the proposed Spiral-GAN.**TABLE 2.** Quantitative comparison of different methods on EUVP dataset. The first and second best scores are colored red and blue, respectively.

Methods	None Data-driven Methods				Data-driven Methods				
	UCM	ARCR	CLAHE	Fusion	UGAN-P	FUNIE	Water-Net	Deep-SESR	Ours
UISM	7.1974	7.2408	7.1910	7.1581	6.8265	7.1282	7.0058	7.2899	6.8531
UICM	2.9687	3.3103	1.4047	0.2259	2.3179	0.6767	3.6741	1.3091	3.6651
UIConM	0.6727	0.5668	0.5420	0.5954	0.6080	0.4387	0.5089	0.3596	0.7728
UIQM	4.6142	4.2580	4.1010	4.3595	4.2549	3.6925	3.9920	3.4754	4.8899

2) DISCUSSION ON EUVP DATASET

Table 2 summarizes the average UIQM results on the EUVP dataset and a few representative comparisons are presented in Fig. 5. Although the haze in raw underwater images can be removed by ARCR method, the color, details and contrast are not as good as our enhanced results. CLAHE hardly removes the effects of scattering and rarely increases contrast in raw images. The UCM and fusion methods can effectively resolve the haze effect in raw images and increase contrast. However, in some cases, these two methods easily turn the water into undesired type(such as the images in the 4th column, UCM turns water type into black hue, and Fusion turns it into azure blue). The results of traditional methods show that these methods have limitations in handling different degraded underwater images and hardly obtain all desired results at the same time. For deep learning algorithms, FUNIE-GAN and Deep-SESR methods have slight positive effectiveness on resolving the effect of scattering and only change the color of the water. Water-Net introduces an

unexpected gray hue; in some images (such as the images in the 4th column), it also turns the water type into azure blue. The results of UGAN-P have improved the visibility and visually made these enhanced images resemble images captured from clean water, as our method does. However, compared with our method, UGAN-P still lacks details. As demonstrated in Table 2, our method stands out among the compared methods in terms of UIQM and UIConM values.

3) DISCUSSION ON UNDERWATER-MOT DATASET

The comparison examples on Underwater-MOT dataset are displayed in Fig. 6. According to the degradation degree of input images, we simply divide the raw images into five groups, as demonstrated in the first column in Fig. 6; from top to bottom, the distorted level worsens. For traditional enhancement or restoration algorithms, UCM and ARCR have a slight influence on improving all degraded inputs. The results of CLAHE and Fusion show that these methods are effective for removing the effects caused by light scattering,

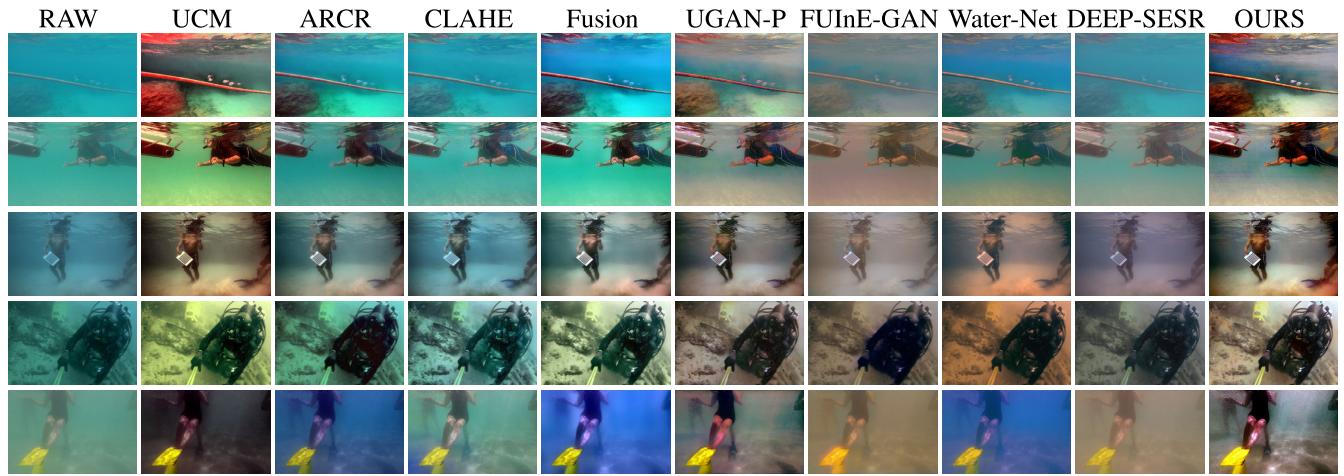


FIGURE 5. Subjective comparisons of different methods on EUVP dataset. From left to right are raw underwater images, and the results of UCM [54], ARCR [32], CLAHE [16], Fusion [20], UGAN-P [7], FUInE-GAN [39], Water-Net [40], Deep-SESR [41], the proposed Spiral-GAN.

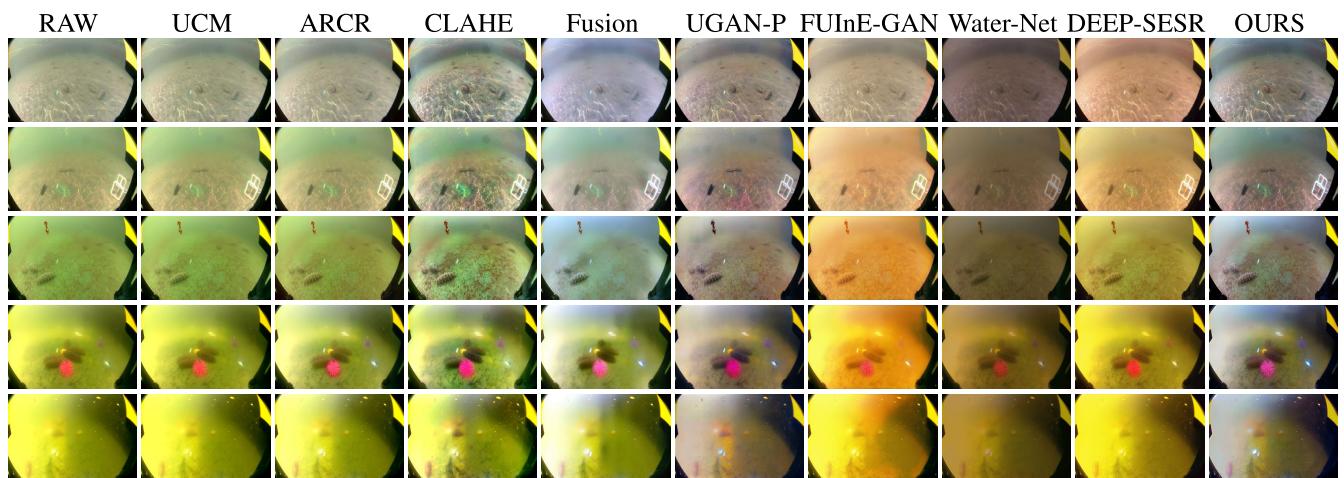


FIGURE 6. Subjective comparisons of different methods on Underwater-MOT dataset. From left to right are raw underwater images, and the results of UCM [54], ARCR [32], CLAHE [16], Fusion [20], UGAN-P [7], FUInE-GAN [39], Water-Net [40], Deep-SESR [41], the proposed Spiral-GAN.

TABLE 3. Quantitative comparison of different methods on Underwater-MOT dataset. The first and second best scores are colored red and blue, respectively.

Methods	None Data-driven Methods				Data-driven Methods				
	UCM	ARCR	CLAHE	Fusion	UGAN-P	FUInE-GAN	Water-Net	Deep-SESR	OURS
UISM	7.0738	7.1205	7.0531	7.5175	6.8926	7.1840	7.3724	7.1718	7.3663
UICM	0.8791	1.3068	1.8884	2.3161	1.9047	0.9198	1.1463	1.1545	2.8532
UIConM	0.3063	0.2656	0.3642	0.3603	0.3887	0.2962	0.3020	0.2650	0.4000
UIQM	3.2088	3.0890	3.4381	3.5734	3.4790	3.2065	3.2890	3.0977	3.6857

but CLAHE fails to rectify the greenish and yellowish hue in images, while Fusion often oversaturates bright regions in the scene. For data-driven methods, FUInE-GAN and Deep-SESR have slight positive effect on enhancing the quality of underwater images and even aggravate the scattering effect. In addition, these two methods also introduce considerable artifacts and color casts (e.g., the reddish or orange tint). As for Water-Net, undesirable gray tones are always introduced. UGAN-P is able to remove haze-like effects well

and rectify the water color in these challenge images (e.g., the last two inputs). However, compared with our method, the images enhanced by UGAN-P still suffer from blurred vision (e.g. the pink sea urchin). Table 3 reports the quantitative results of different methods on Underwater-MOT set. Our SpiralGAN receives the highest average score in terms of UICM, UIConM and UIQM, which indicates our method produces the best results from a subjective perspective.

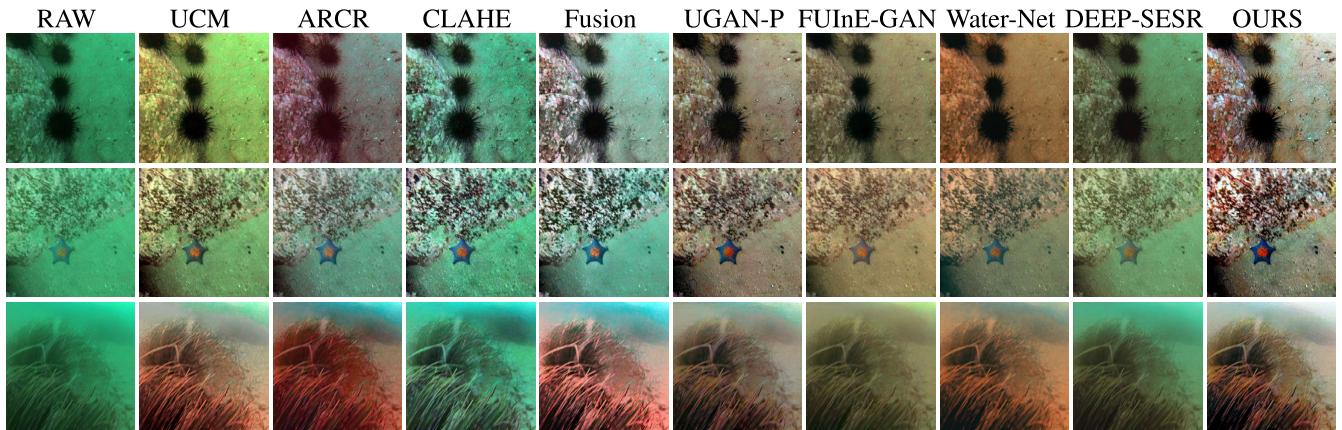


FIGURE 7. Subjective comparisons of different methods on UIQS subset of RUIE. From left to right are raw underwater images, and the results of UCM [54], ARCR [32], CLAHE [16], Fusion [20], UGAN-P [7], FUInE-GAN [39], Water-Net [40], Deep-SESR [41], the proposed Spiral-GAN.

TABLE 4. Quantitative comparison of different methods on UIQS of RUIE dataset. The first and second best scores are colored red and blue, respectively.

Methods	None Data-driven Methods				Data-driven Methods				
	UCM	ARCR	CLAHE	Fusion	UGAN-P	FUInE-GAN	Water-Net	Deep-SESR	OURS
UISM	6.6748	6.5485	6.5971	6.5823	6.6364	6.6809	6.5988	6.8835	6.5684
UICM	5.2897	6.4299	2.2634	7.3441	1.9189	1.0265	3.7179	2.5614	3.0599
UIConM	0.6278	0.6387	0.6423	0.6059	0.6136	0.5032	0.6733	0.4436	0.6745
UIQM	4.3649	4.3987	4.3085	4.3171	4.2076	3.8008	4.1033	3.6908	4.4373

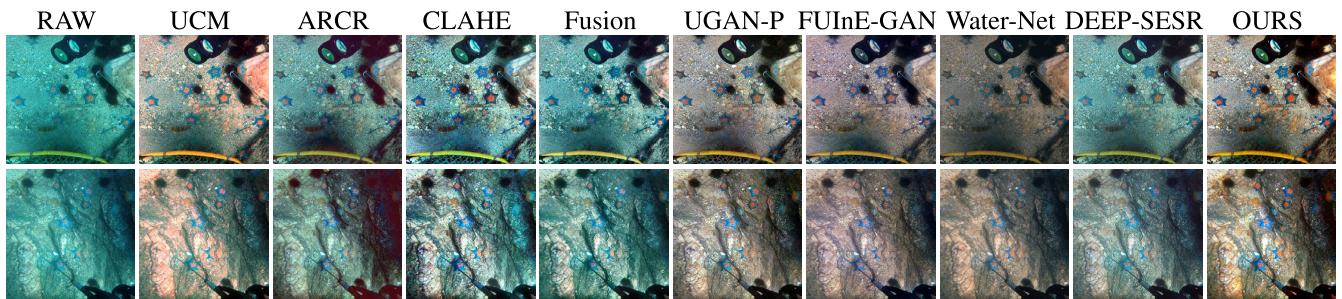


FIGURE 8. Subjective comparisons of different methods on UCCS subset of RUIE. From left to right are raw underwater images, and the results of UCM [54], ARCR [32], CLAHE [16], Fusion [20], UGAN-P [7], FUInE-GAN [39], Water-Net [40], Deep-SESR [41], the proposed Spiral-GAN.

4) DISCUSSION ON RUIE DATASET

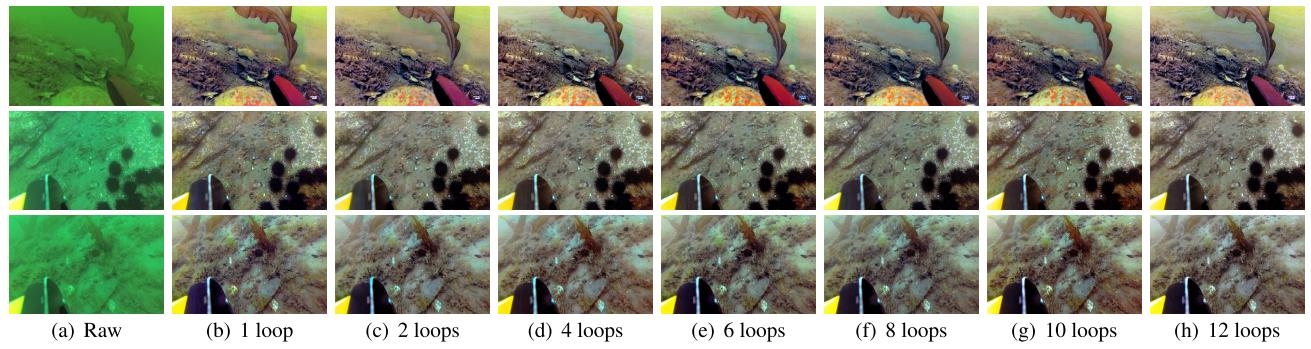
The comparison examples of the UIQS of RUIE dataset are displayed in Fig. 7, the original images suffered the serious distortion and have over-under exposure problem. The images generated by conventional methods remove the green hue in a limit extent. As the baseline, the enhancement images of UGAN-P have lower color saturation. Deep-SESR hardly has positive effectiveness on removing scattering effect and green hue casts. Although FUInE-GAN and Water-Net can remove the green hue, they also decreases the contrast and brightness. Our model achieve the best performance in visibility quality, we can improve the contrast and brightness at the same time, it means our model can improved over-under exposure problem. The quantitative comparisons of the results of UIQS are shown in Table 4, our

method achieves the best performance in terms of UIQM and UIConM, which means that the underwater images enhanced by our method have the most visually appealing appearance.

For the comparative experiment on the UCCS of RUIE dataset, we aim to evaluate the ability of correcting color cast for enhancement models, as shown in Table 4, we have the highest score in terms of UICM, UIConM and UIQM, which means the generated images of our model can correct the color cast and improve the contrast. From the Fig. 8, we can see that the UGAN-P, Water-Net and FUInE-GAN can correct the color cast, but these two methods reduces contrast and brightness compared to the UGAN-P. ARCR, CLAHE and Fusion hardly remove the blue tones in images, but they effectively remove the haze-like effect. UCM easily

TABLE 5. Quantitative comparison of different methods on UCCS of RUIE dataset. The first and second best scores are colored red and blue, respectively.

Methods	None Data-driven Methods				Data-driven Methods				
	UCM	ARCR	CLAHE	Fusion	UGAN-P	FUNIE-GAN	Water-Net	Deep-SESR	OURS
UISM	6.6118	6.5182	6.5340	6.5574	6.6110	6.6978	6.5672	6.7497	6.5218
UICM	0.8791	1.3068	1.8884	2.3161	1.9047	0.9198	1.1463	1.1545	2.8532
UIConM	0.6745	0.6875	0.7288	0.7006	0.6988	0.6031	0.6243	0.5489	0.7326
UIQM	4.4771	4.5014	4.5972	4.5742	4.4977	4.1660	4.2561	4.0027	4.6244

**FIGURE 9.** The color correction results. From left to right are raw underwater image, and the results of UCM [54], ARCR [32], CLAHE [16], Fusion [20], UGAN-P [7], FUNIE-GAN [39], Water-Net [40], Deep-SESR [41], the proposed Spiral-GAN, and reference image.**FIGURE 10.** Ablation study of the effect of spiral times. Excepting for loop times, all parameter settings are fixed(L2 loss+Angular loss+GAN loss+proposed generator). From left to right are raw underwater images, and the results of (a) 1 loop, (b) 2 loops, (c) 4 loops, (d) 6 loops, (e) 8 loops, (f) 10 loops and (g) 12 loops.

introduces color distortion and over exposure. Similar to the performance on the UIQS of our model, the images enhanced by our proposed method have higher contrast, high color saturation and higher brightness.

Qualitative and quantitative experiments on four real-world datasets demonstrate the effectiveness and generalization of the proposed Spiral-GAN. Furthermore, for verifying the color accuracy of the proposed method, we conduct a group of color accuracy test, the results can be found in Fig 9. Comparing raw colour card(shown in first column of Fig 9) and target card(the last column), some color patches(especially red patch) change appear in the underwater color checker chart image due to the effects of light scattering and absorbtion. Fig 9 shows that the color patches restored by our method are visually appealing, even if they are far away from the lens, which demonstrates that our method can restore underwater images to a relatively genuine color among the compared methods.

E. THE ABLATION STUDY

To demonstrate the effectiveness of each component of the proposed model, we implement an ablation study. The power

of the spiral training can be verified by comparing Fig. 10. Insufficient spiral training loops may lead to artifacts and mosaics. The image enhanced with L_2 loss (Fig. 11(b)) shows clearer details and brighter colors than that obtained with L_1 loss (Fig. 11(d)). There is a certain loss in color brightness and over-under exposure (e.g. the scallops and echini) when we remove the angular loss by comparing (Fig. 11(b)) and (Fig. 11(c)). The enhanced image exhibits an oversaturation of color (e.g., the color of scallops and seaweed) when we replace our generator with the one used in UGAN-P by comparing (Fig. 11(b)) and (Fig. 11(e)). Note that with the help of the proposed spiral training strategy and a combination of our loss functions, the finding shown in (Fig. 11(e)) indicate a better performance than those in (Fig. 11(f)) obtained by a pure UGAN-P [7]. The quantitative comparison of different loop numbers and other parameters can be found in Table 6.

Fig. 12 shows the outputs of each spiral loop during the training and test stages. We find that the halos and color distortion phenomenon increase as the loop number increases. These images with different scales of halos and the color distortion, which can be considered a variant of data augmentation, can practically enlarge the adverse effects that may be

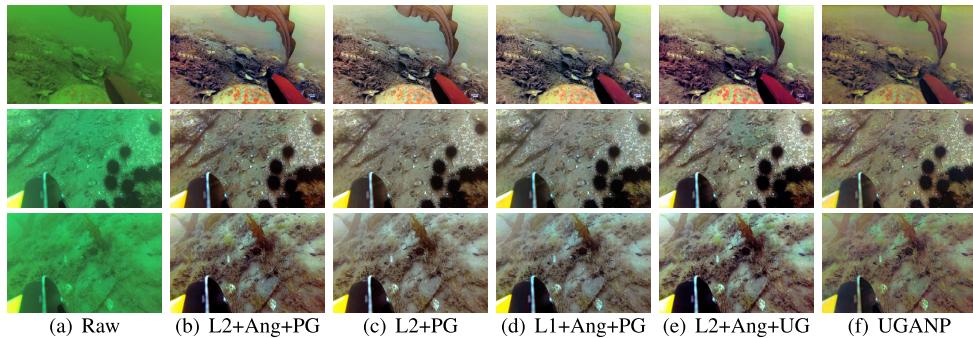


FIGURE 11. Ablation study of the effect of parameter settings excepting for spiral times. The number of spiral loop is fixed to 10 times. From left to right are (a) raw underwater images, and the results of (b) the best model of ours, L2+Angular+proposed generator (c) L2+proposed generator, (d) L1+Angular+proposed generator, (e) L2+Angular+generator of UGAN-P, (f) original UGAN-P.

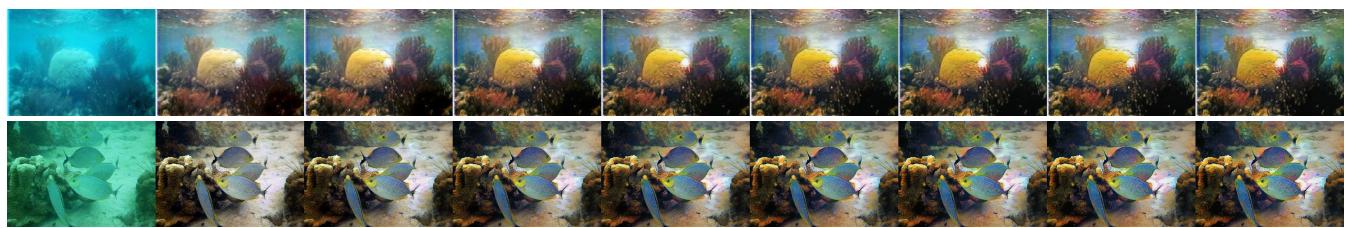


FIGURE 12. The intermediate results presentation of an 8-spiral experiment. The first row is training sample and the second row is testing sample. From left to right are inputs, the results of first loop, second loop, ..., eighth loop.

TABLE 6. Quantitative comparison of ablation tests on the URPC dataset.

Recurrent time + Loss	UIQM
1 loop+L2+Angular+GAN	4.4563
2 loops+L2+Angular+GAN	4.5066
4 loops+L2+Angular+GAN	4.4983
6 loops+L2+Angular+GAN	4.5366
8 loops+L2+Angular+GAN	4.8179
10 loops+L2+Angular+GAN	5.9995
12 loops+L2+Angular+GAN	4.6367
10 loops+L2+GAN	4.5385
10 loops+L1+Angular+GAN	4.5228
10 loops+L2+Angular+GAN+UGAN	4.5418

caused by the generator. Thus, the spiral strategy can further increase the sensitivity and enhance the robustness of our model in the training stage. However, additional loops in the test stage will cost more time and reduce the performance for underwater image enhancement. For this reason why we only feed the generator multiple loops in the training stage. In the test stage, we just need to feed the underwater image to the generator once to obtain our result.

F. APPLICATION TESTS

In this section, we utilize three applications, underwater object detection [58], SIFT [59] and Canny [60], to further evaluate the performance of the proposed method. We select Yolov3 [58] with the pretrained weights on COCO dataset [61] as our detection model. Although most of the underwater targets are not labeled in COCO, we can still use

the detection model to locate the position of the “person” in underwater images. A few detection results are shown in Fig. 13. For some severely degraded images, the proposed enhancement algorithm can effectively improve the detection accuracy. We also evaluate the effect of our proposed model by using Canny [60] and SIFT [59] to extract key points and contours from original and enhanced underwater images. The results can be found in Fig. 14, Fig. 15, respectively. Table 7 shows the number of matching key points calculated from images shown in Fig. 14. The application tests demonstrate that the enhanced images have more key matching points and edge detection features than the original input image. Although the intention of designing the generator is to imporve the quality of the generated images instead of deliberately pursuing lightweight networks and faster inference-time, our model still achieved comparable result in speed compared with the state-of-art data-driven methods. Table 8 shows the runtime² of different methods averaged on 1521 images of size 1280×768. For conventional methods, only the codes of CPU version are available.

V. DISCUSSION

Inspired by CycleGAN [14] and FUnIE-GAN-UP [39], we also conduct one group of unsupervised experiments with the help of the cycle-consistency loss. The architecture of generator and discriminator are same as the our proposed supervised model. Our objective function only

²Runtime is measured on a PC with an Nvidia GTX 1080Ti GPU and Intel(R) Xeon(R) CPU E5-2609 v4 @ 1.70GHz.

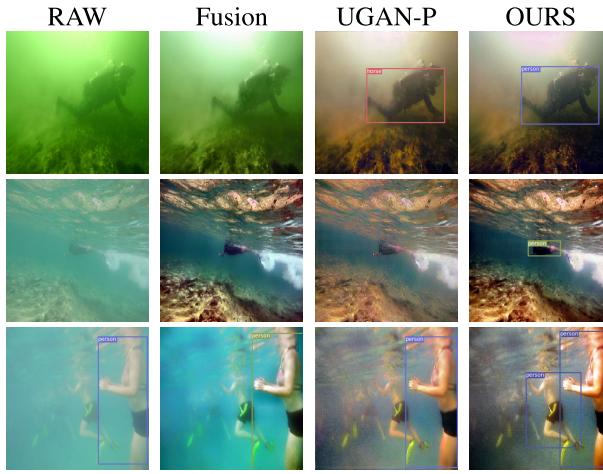


FIGURE 13. Object detection results. From left to right are raw underwater images, and the results of Fusion [20], UGAN-P [7] and the proposed.

TABLE 7. Matching key points results. The best result is in red whereas the second best one is in blue.

Method	picture1	picture2	picture3
RAW	36	15	19
Fusion [20]	248	172	331
UGAN-P [7]	256	157	269
OURS	357	422	529

consists of adversarial loss of standard conditional GAN and cycle-consistent loss. As suggested by CycleGAN, the formula for the cycle-consistency loss function is as follows:

$$L_{cyc}(G, F) = \mathbb{E}_{X, Y}[\|X - F(G(X))\|_1] + \mathbb{E}_{X, Y}[\|Y - G(F(Y))\|_1] \quad (9)$$

where G learns a mapping $\{X\} \rightarrow Y$ by minimizing this objective, F learns a mapping $\{Y\} \rightarrow X$ by minimizing this objective. Following the spiral strategy mentioned in section III-B, Equations 9 can be updated as follows:

$$L_{spiral_cyc}(G, F) = \mathbb{E}_{X, Y}[\|X - \underbrace{F(G(\dots F(\underbrace{G(X)}_n \dots)))}_n\|_1] + \mathbb{E}_{X, Y}[\|Y - \underbrace{G(F(\dots G(\underbrace{F(Y)}_n \dots)))}_n\|_1] \quad (10)$$

Therefore, our full object is:

$$\mathcal{L} = \min_{G, F} \max_{D_Y, D_X} L_{spiral_cGAN}(G, D_Y) + L_{spiral_cGAN}(F, D_X) + \lambda_{spiral_cyc}(G, F) \quad (11)$$

where D_Y aims to distinguish between the real clean samples Y and the generated samples by $G(X)$, and the D_X aims to distinguish between the real distortion samples X and the generated samples by $F(Y)$. Here we set $\lambda_{cyc} = 10$ by empirical tuning. Our training data are taken from the unpaired instances of EUVP dataset [39],

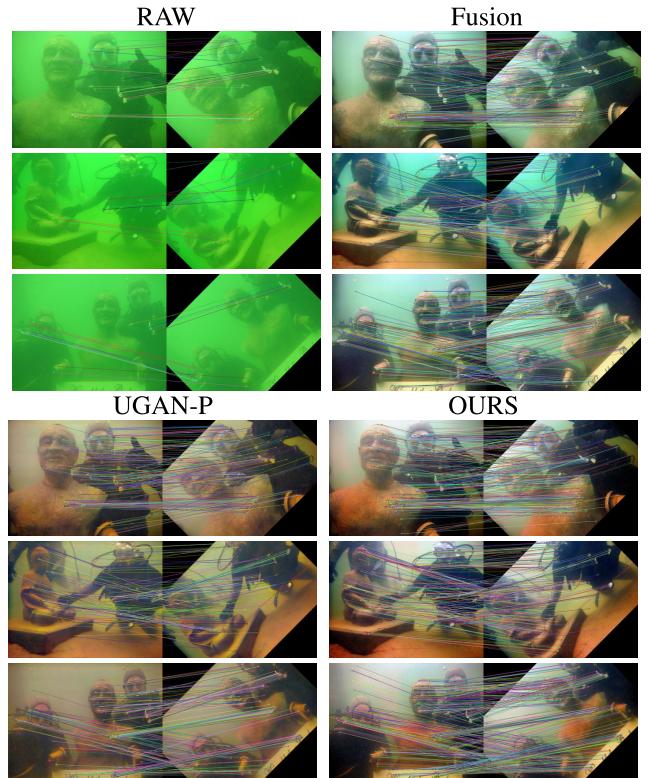


FIGURE 14. Keypoint matching results. From left to right are raw underwater images, and the results of Fusion [20], UGAN-P [7] and the proposed.

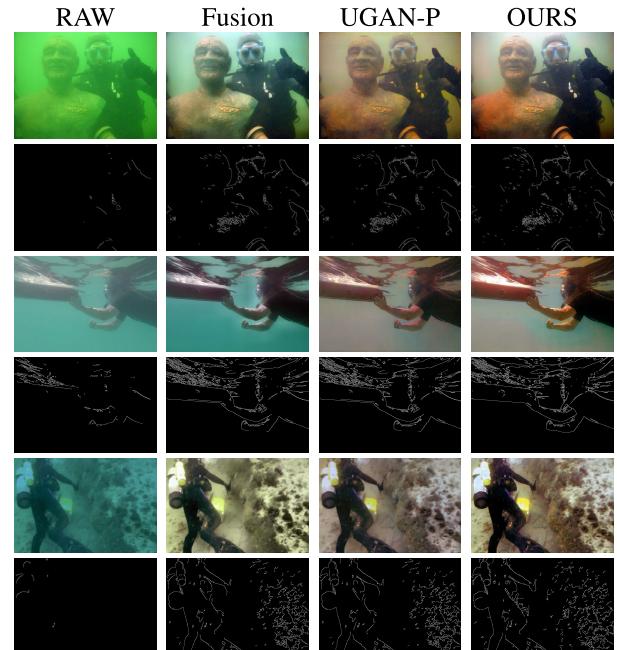


FIGURE 15. Canny Edge Detection results. From left to right are raw underwater images, and the results of Fusion [20], UGAN-P [7] and the proposed.

including 3195 poor quality images and 3140 good quality images. The spiral loop is set to 10.

The Quantitative comparisons of the results of URPC dataset are shown in Table 9, where the values in red

TABLE 8. Inference-time(in second) and memory requirement comparisons. The best result is in red whereas the second best one is in blue.

Method	Inference(s)	Memory(MB)	Platform
UCM [54]	51.3662	∅	Python(CPU)
CLAHE [16]	0.0076	∅	Python(CPU)
ARCR [32]	0.3406	∅	MATLAB(CPU)
Fusion [20]	2.7584	∅	MATLAB(CPU)
UGAN-P [7]	0.0484	218.08	TensorFlow(GPU)
FUnIE-GAN [39]	0.0655	16.08	Keras(GPU)
Water-Net [40]	0.2867	4.16	Tensorflow(GPU)
OURS	0.0380	19.05	PyTorch(GPU)

TABLE 9. Quantitative comparison of Unpaired training on URPC dataset.

Methods	CycleGAN	FUnIE-GAN-UP	OURS
UISM	6.9563	7.0261	7.0139
UICM	2.2523	0.7275	5.0026
UIConM	0.5436	0.4231	0.7041
UIQM	4.0612	3.6081	4.7297

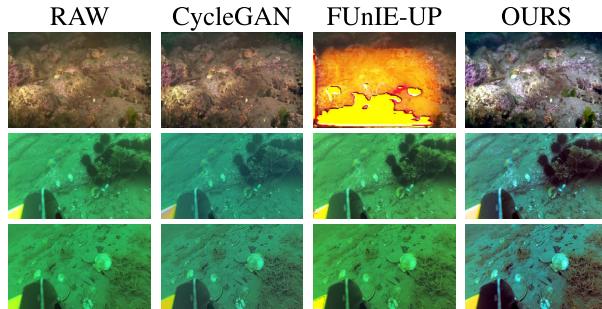


FIGURE 16. The qualitative performance comparison on the dataset URPC of unpaired training. From left to right are raw underwater images, and the results of CycleGAN [14], FUnIE-GAN-UP [39] and Ours.

represent the best results. As shown in Table 9, without the help of L2 loss and angular loss, our framework still achieves the best performance among three methods in terms of UICM, UIConM and UIQM, which means that the underwater images enhanced by our method have a vivid color, better contrast and sharpness, and the most visually appealing appearance. The experimental results show that our model also has great potential on unpaired dataset. The comparison examples of the URPC dataset are displayed in Fig. 16. As the baseline, CycleGAN just turns the green water type into cyan water type, meanwhile, it improves the contrast of the generated images and the color saturation to a limited extent. For FUnIE-UP, the enhancement effect is limited and some enhanced images are seriously distorted (*e.g.*, the first row). Compared with the CycleGAN and FUnIE-UP, the enhancement images of our proposed method have the best contrast and color saturation, the color of starfishes are more vivid, and the contrast of sea cucumbers are more obviously.

We also conduct experiments in the EUVP dataset. we select the validation folder of the EUVP unpaired dataset, which contains 330 images, all images are resized to 512×512 for testing. Table 10 and Fig. 17 show the

TABLE 10. Quantitative comparison of unpaired training on EUVP dataset.

Methods	CycleGAN	FUnIE-GAN-UP	OURS
UISM	6.8795	6.5656	6.6662
UICM	2.8475	3.2167	6.0158
UIConM	0.6577	0.5865	0.7094
UIQM	4.4977	4.1263	4.6745

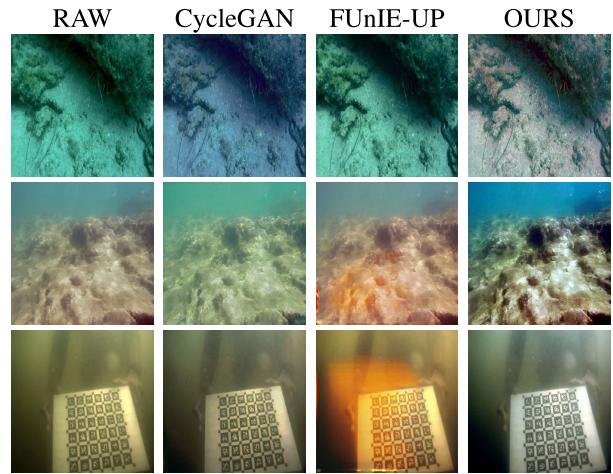


FIGURE 17. The qualitative performance comparison on the EUVP dataset of unpaired training. From left to right are raw underwater images, and the results of CycleGAN [14], FUnIE-GAN-UP [39] and Ours.

quantitative and qualitative performance among three methods. Our framework still achieves the best performance in terms of UICM, UIConM and UIQM. As shown in Fig. 17, the generated images by CycleGAN accentuates the color of the water, such as the first and second input images. In some images, FUnIE-UP increased distortion degree. Our model improve the over-under exposure problem and increase contrast and brightness. Through the performance of our model in the above experimental, we believe that our model still has great potential in unpaired training.

At present, our work mainly focuses on enhancing the details and colors of underwater images, without considering the depth and inherent optical parameters of the underwater scene during the training process. Despite effectiveness to improve the contrast, colors and details of underwater images, our method is not always very good for some challenging scenes. For example, the generated images from these severely degraded are easily boosting noise and over-saturated. And for these images with a large area of water background, it is difficult to completely eliminate the long-distance blur effect caused by backscatter in the image. Future work aims at addressing these short-comings.

VI. CONCLUSION

In this paper, we present a simple yet efficient GAN-based model for underwater image enhancement. With the help of the proposed spiral training strategy and employed pixel-level loss function, the underwater enhancing model can generally and effectively transform underwater images with various

quality into clean images with richer details and colors. Our ablation study shows the effectiveness of each module of the proposed model. The qualitative and quantitative evaluation results have shown that the proposed model performs as well and often better than the state-of-the-art models. We believe our framework may have the potential to improve translation quality for other low-level image translation tasks, such as image de-noising and de-hazing. We leave this for our further work.

REFERENCES

- [1] O. Hoegh-Guldberg, P. J. Mumby, A. J. Hooten, R. S. Steneck, P. Greenfield, E. Gomez, C. D. Harvell, P. F. Sale, A. J. Edwards, K. Caldeira, N. Knowlton, C. M. Eakin, R. Iglesias-Prieto, N. Muthiga, R. H. Bradbury, A. Dubi, and M. E. Hatziolos, "Coral reefs under rapid climate change and ocean acidification," *Science*, vol. 318, no. 5857, pp. 1737–1742, 2007. [Online]. Available: <https://science.sciencemag.org/content/318/5857/1737>
- [2] F. Shkurti, A. Xu, M. Meghjani, J. C. Gamboa Higuera, Y. Girdhar, P. Giguere, B. B. Dey, J. Li, A. Kalmbach, C. Prahacs, K. Turgeon, I. Rekleitis, and G. Dudek, "Multi-domain monitoring of marine environments using a heterogeneous robot team," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Oct. 2012, pp. 1747–1753.
- [3] B. Bingham, B. Foley, H. Singh, R. Camilli, K. Delaporta, R. Eustice, A. Mallios, D. Mindell, C. Roman, and D. Sakellariou, "Robotic tools for deep water archaeology: Surveying an ancient shipwreck with an autonomous underwater vehicle," *J. Field Robot.*, vol. 27, no. 6, pp. 702–717, Nov. 2010.
- [4] L. Whitcomb, D. R. Yoerger, H. Singh, and J. Howland, "Advances in underwater robot vehicles for deep ocean exploration: Navigation, control, and survey operations," in *Robotics Research*, J. M. Hollerbach and D. E. Koditschek, Eds. London, U.K.: Springer, 2000, pp. 439–448.
- [5] H. Lu, Y. Li, L. Zhang, and S. Serikawa, "Contrast enhancement for images in turbid water," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 32, no. 5, pp. 886–893, 2015.
- [6] C.-Y. Li, J.-C. Guo, R.-M. Cong, Y.-W. Pang, and B. Wang, "Underwater image enhancement by dehazing with minimum information loss and histogram distribution prior," *IEEE Trans. Image Process.*, vol. 25, no. 12, pp. 5664–5677, Dec. 2016.
- [7] C. Fabbri, M. J. Islam, and J. Sattar, "Enhancing underwater imagery using generative adversarial networks," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2018, pp. 7159–7165.
- [8] Y.-T. Peng and P. C. Cosman, "Underwater image restoration based on image blurriness and light absorption," *IEEE Trans. Image Process.*, vol. 26, no. 4, pp. 1579–1594, Apr. 2017.
- [9] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, vol. 141, no. 5, pp. 1097–1105.
- [10] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 2672–2680.
- [11] X. Chen, Y. Duan, R. Houthooft, J. Schulman, I. Sutskever, and P. Abbeel, "Infogan: Interpretable representation learning by information maximizing generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, 2016, pp. 2172–2180.
- [12] A. Odena, C. Olah, and J. Shlens, "Conditional image synthesis with auxiliary classifier gans," in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 2642–2651.
- [13] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-Image translation with conditional adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 5967–5976.
- [14] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired Image-to-Image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2242–2251.
- [15] R. Hummel, "Image enhancement by histogram transformation," *Comput. Graph. Image Process.*, vol. 6, no. 2, pp. 184–195, 1977.
- [16] S. M. Pizer, R. E. Johnston, J. P. Erickson, B. C. Yankaskas, and K. E. Muller, "Contrast-limited adaptive histogram equalization: Speed and effectiveness," in *Proc. 1st Conf. Visualizat. Biomed. Comput.*, 1990, p. 337, doi: [10.1109/vbc.1990.109340](https://doi.org/10.1109/vbc.1990.109340).
- [17] G. Buchsbaum, "A spatial processor model for object colour perception," *J. Franklin Inst.*, vol. 310, no. 1, pp. 1–26, Jul. 1980.
- [18] Y.-C. Liu, W.-H. Chan, and Y.-Q. Chen, "Automatic white balance for digital still camera," *IEEE Trans. Consum. Electron.*, vol. 41, no. 3, pp. 460–466, 1995.
- [19] J. van de Weijer, T. Gevers, and A. Gijsenij, "Edge-based color constancy," *IEEE Trans. Image Process.*, vol. 16, no. 9, pp. 2207–2214, Sep. 2007.
- [20] C. Anutti, C. O. Anutti, T. Haber, and P. Bekaert, "Enhancing underwater images and videos by fusion," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 81–88.
- [21] A. S. A. Ghani and N. A. M. Isa, "Underwater image quality enhancement through Rayleigh-stretching and averaging image planes," *Int. J. Nav. Archit. Ocean Eng.*, vol. 6, no. 4, pp. 840–866, Dec. 2014.
- [22] A. S. Abdul Ghani and N. A. Mat Isa, "Underwater image quality enhancement through integrated color model with Rayleigh distribution," *Appl. Soft Comput.*, vol. 27, pp. 219–230, Feb. 2015.
- [23] A. S. Abdul Ghani and N. A. Mat Isa, "Enhancement of low quality underwater image through integrated global and local contrast correction," *Appl. Soft Comput.*, vol. 37, pp. 332–344, Dec. 2015.
- [24] D. Huang, Y. Wang, W. Song, J. Sequeira, and S. Mavromatis, "Shallow-water image enhancement using relative global histogram stretching based on adaptive parameter acquisition," in *Proc. Int. Conf. Multimedia Model.* Bangkok, Thailand: Springer, Feb. 2018, pp. 453–465.
- [25] M. H. Asmare, V. S. Asirvadam, and A. F. M. Hani, "Image enhancement based on contourlet transform," *Signal, Image Video Process.*, vol. 9, no. 7, pp. 1679–1690, Oct. 2015.
- [26] A. Khan, S. S. A. Ali, A. S. Malik, A. Anwer, and F. Meriaudeau, "Underwater image enhancement by wavelet based fusion," in *Proc. IEEE Int. Conf. Underwater Syst. Technology: Theory Appl. (USYS)*, 2016, pp. 83–88.
- [27] S. Vasamsetti, N. Mittal, B. C. Neelapu, and H. K. Sardana, "Wavelet based perspective on variational enhancement technique for underwater imagery," *Ocean Eng.*, vol. 141, pp. 88–100, Sep. 2017.
- [28] H. Lu, Y. Li, Y. Zhang, M. Chen, S. Serikawa, and H. Kim, "Underwater optical image processing: A comprehensive review," *Mobile Netw. Appl.*, vol. 22, no. 6, pp. 1204–1211, Dec. 2017.
- [29] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 12, pp. 2341–2353, Dec. 2011.
- [30] J. Y. Chiang and Y.-C. Chen, "Underwater image enhancement by wavelength compensation and dehazing," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1756–1769, Apr. 2012.
- [31] P. Drews Jr, E. do Nascimento, F. Moraes, S. Botelho, and M. Campos, "Transmission estimation in underwater single images," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops*, Dec. 2013, pp. 825–830.
- [32] A. Galdran, D. Pardo, A. Picón, and A. Alvarez-Gila, "Automatic red-channel underwater image restoration," *J. Vis. Commun. Image Represent.*, vol. 26, pp. 132–145, Jan. 2015.
- [33] Y.-T. Peng, X. Zhao, and P. C. Cosman, "Single underwater image enhancement using depth estimation based on blurriness," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2015, pp. 4952–4956.
- [34] N. Wang, H. Zheng, and B. Zheng, "Underwater image restoration via maximum attenuation identification," *IEEE Access*, vol. 5, pp. 18941–18952, 2017.
- [35] Y.-S. Chen, Y.-C. Wang, M.-H. Kao, and Y.-Y. Chuang, "Deep photo enhancer: Unpaired learning for image enhancement from photographs with GANs," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 6306–6314.
- [36] C. Guo, C. Li, J. Guo, C. C. Loy, J. Hou, S. Kwong, and R. Cong, "Zero-reference deep curve estimation for low-light image enhancement," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 1780–1789.
- [37] H.-U. Kim, Y. J. Koh, and C.-S. Kim, "Pienet: Personalized image enhancement," in *Proc. Eur. Conf. Comput. Vis.*, 2020, pp. 374–390.
- [38] J. Li, K. A. Skinner, R. M. Eustice, and M. Johnson-Roberson, "WaterGAN: Unsupervised generative network to enable real-time color correction of monocular underwater images," *IEEE Robot. Autom. Lett.*, vol. 3, no. 1, pp. 387–394, Jan. 2018.
- [39] M. J. Islam, Y. Xia, and J. Sattar, "Fast underwater image enhancement for improved visual perception," *IEEE Robot. Autom. Lett.*, vol. 5, no. 2, pp. 3227–3234, Apr. 2020.
- [40] C. Li, C. Guo, W. Ren, R. Cong, J. Hou, S. Kwong, and D. Tao, "An underwater image enhancement benchmark dataset and beyond," *IEEE Trans. Image Process.*, vol. 29, pp. 4376–4389, Feb. 2020.

- [41] M. Jahidul Islam, P. Luo, and J. Sattar, "Simultaneous enhancement and super-resolution of underwater imagery for improved visual perception," 2020, *arXiv:2002.01155*. [Online]. Available: <http://arxiv.org/abs/2002.01155>
- [42] M. Mirza and S. Osindero, "Conditional generative adversarial nets," 2014, *arXiv:1411.1784*. [Online]. Available: <http://arxiv.org/abs/1411.1784>
- [43] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2015, pp. 234–241.
- [44] L. Galteri, L. Seidenari, M. Bertini, and A. D. Bimbo, "Deep generative adversarial compression artifact removal," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 4836–4845.
- [45] S. Iizuka, E. Simo-Serra, and H. Ishikawa, "Globally and locally consistent image completion," *ACM Trans. Graph.*, vol. 36, no. 4, pp. 1–14, Jul. 2017.
- [46] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," 2015, *arXiv:1502.03167*. [Online]. Available: <http://arxiv.org/abs/1502.03167>
- [47] A. L. Maas, A. Y. Hannun, and A. Y. Ng, "Rectifier nonlinearities improve neural network acoustic models," in *Proc. ICML*, 2013, vol. 30, no. 1, p. 3.
- [48] X. Yang, H. Mei, J. Zhang, K. Xu, B. Yin, Q. Zhang, and X. Wei, "DRFN: Deep recurrent fusion network for single-image super-resolution with large factors," *IEEE Trans. Multimedia*, vol. 21, no. 2, pp. 328–337, Feb. 2019.
- [49] M. Chu, Y. Xie, J. Mayer, L. Leal-Taixé, and N. Thuerey, "Learning temporal coherence via self-supervision for GAN-based video generation," 2018, *arXiv:1811.09393*. [Online]. Available: <http://arxiv.org/abs/1811.09393>
- [50] A. Graves, "Generating sequences with recurrent neural networks," 2013, *arXiv:1308.0850*. [Online]. Available: <http://arxiv.org/abs/1308.0850>
- [51] J.-Y. Liu and Y.-H. Yang, "Denoising auto-encoder with recurrent skip connections and residual regression for music source separation," in *Proc. 17th IEEE Int. Conf. Mach. Learn. Appl. (ICMLA)*, Dec. 2018, pp. 773–778.
- [52] O. Sidorov, "Artificial color constancy via Googlenet with angular loss function," *Appl. Artif. Intell.*, vol. 5, pp. 1–13, Dec. 2020.
- [53] O. Sidorov, "Conditional GANs for multi-illuminant color constancy: Revolution or yet another approach?" in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2019.
- [54] K. Iqbal, M. Odetayo, A. James, R. Abdul Salam, and A. Zawawi Hj Talib, "Enhancing the low quality images using unsupervised colour correction method," in *Proc. IEEE Int. Conf. Syst., Man Cybern.*, Oct. 2010, pp. 1703–1709.
- [55] R. Liu, X. Fan, M. Zhu, M. Hou, and Z. Luo, "Real-world underwater enhancement: Challenges, benchmarks, and solutions under natural light," *IEEE Trans. Circuits Syst. Video Technol.*, early access, Jan. 3, 2020, doi: [10.1109/TCSVT.2019.2963772](https://doi.org/10.1109/TCSVT.2019.2963772).
- [56] K. Panetta, C. Gao, and S. Agaian, "Human-Visual-System-Inspired underwater image quality measures," *IEEE J. Ocean. Eng.*, vol. 41, no. 3, pp. 541–551, Jul. 2016.
- [57] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*. [Online]. Available: <http://arxiv.org/abs/1412.6980>
- [58] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," 2018, *arXiv:1804.02767*. [Online]. Available: <http://arxiv.org/abs/1804.02767>
- [59] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, Nov. 2004.
- [60] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vols. PAMI-8, no. 6, pp. 679–698, Nov. 1986.
- [61] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft COCO: Common objects in context," in *Computer Vision*, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds. Cham, Switzerland: Springer, 2014, pp. 740–755.



RUYUE HAN received the B.Tech. degree in communication engineering from Xiangtan University, China, in 2015. She is currently pursuing the master's degree with the Ocean University of China. Her research interests include underwater image processing, image restoration, and artificial neural network.



YANG GUAN received the B.S. degree in electronic information engineering from the Weifang University, Weifang, China, in 2019. She is currently pursuing the master's degree with the Ocean University of China. Her researches include underwater image processing, object detection, and the applications with generative adversarial neural network.



ZHIBIN YU (Member, IEEE) received the B.S. degree in thermal energy and power engineering from the Harbin Institute of Technology, Harbin, China, in 2005, and the M.S. degree in computer engineering and the Ph.D. degree in electrical engineering from Kyungpook National University, Daegu, South Korea, in 2011 and 2016, respectively. In 2016, he joined the Department of Electronic Engineering, Ocean University of China, where he is currently an Associate Professor. His research interests include underwater image processing, artificial neural networks, and underwater 3-D reconstruction.



PENG LIU (Member, IEEE) received the B.S. degree in electronic information engineering and the master's degree in telecommunication and information system from the Ocean University of China, Qingdao, China, in 2004 and 2007, respectively. In 2007, he joined the Computing Center, Ocean University of China, where he is currently a Lecturer. He is currently pursuing the Ph.D. degree with the School of Information Science and Engineering, Ocean University of China. His current research interests include image processing, computer vision, and machine learning.



HAIYONG ZHENG (Member, IEEE) received the B.S. degree in electronic information engineering and the Ph.D. degree in ocean information sensing and processing from the Ocean University of China, Qingdao, China, in 2004 and 2009, respectively. In 2009, he joined the Department of Electronic Engineering, Ocean University of China, where he is currently a Professor. His research interests include image processing, computer vision, and machine learning.