



OPEN ACCESS

EDITED BY

Ahmad Salman,
National University of Sciences and
Technology (NUST), Pakistan

REVIEWED BY

Peixian Zhuang,
Tsinghua University, China
Khawar Khurshid,
National University of Sciences and
Technology (NUST), Pakistan

*CORRESPONDENCE

Zhibin Yu
✉ yuzhibin@ouc.edu.cn
Shaoda Zhang
✉ Alex@cofoe.com

SPECIALTY SECTION

This article was submitted to
Marine Conservation and
Sustainability,
a section of the journal
Frontiers in Marine Science

RECEIVED 08 June 2022

ACCEPTED 15 December 2022

PUBLISHED 08 February 2023

CITATION

Guan Y, Liu X, Yu Z, Wang Y, Zheng X,
Zhang S and Zheng B (2023) Fast
underwater image enhancement
based on a generative
adversarial framework.
Front. Mar. Sci. 9:964600.
doi: 10.3389/fmars.2022.964600

COPYRIGHT

© 2023 Guan, Liu, Yu, Wang, Zheng,
Zhang and Zheng. This is an open-
access article distributed under the
terms of the [Creative Commons
Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use,
distribution or reproduction in other
forums is permitted, provided the
original author(s) and the copyright
owner(s) are credited and that the
original publication in this journal is
cited, in accordance with accepted
academic practice. No use,
distribution or reproduction is
permitted which does not comply with
these terms.

Fast underwater image enhancement based on a generative adversarial framework

Yang Guan¹, Xiaoyan Liu¹, Zhibin Yu^{1,2*}, Yubo Wang³,
Xingyu Zheng⁴, Shaoda Zhang^{4*} and Bing Zheng^{1,2}

¹Faculty of Information Science and Engineering, Ocean University of China, Qingdao, China, ²Key Laboratory of Ocean Observation and Information of Hainan Province, Sanya Oceanographic Institution, Ocean University of China, Sanya, China, ³School of Life Science and Technology, Xidian University, Xi'an, China, ⁴Cofoe Medical Technology Company Limited, Shenzhen, China

Underwater image enhancement is a fundamental requirement in the field of underwater vision. Along with the development of deep learning, underwater image enhancement has made remarkable progress. However, most deep learning-based enhancement methods are computationally expensive, restricting their application in real-time large-size underwater image processing. Furthermore, GAN-based methods tend to generate spatially inconsistent styles that decrease the enhanced image quality. We propose a novel efficiency model, FSpiral-GAN, based on a generative adversarial framework for large-size underwater image enhancement to solve these problems. We design our model with equal upsampling blocks (EUBs), equal downsampling blocks (EDBs) and lightweight residual channel attention blocks (RCABs), effectively simplifying the network structure and solving the spatial inconsistency problem. Enhancement experiments on many real underwater datasets demonstrate our model's advanced performance and improved efficiency.

KEYWORDS

generative adversarial networks, efficiency, spatial inconsistency, underwater image enhancement, lightweight

1 Introduction

In recent years, underwater robots such as remotely operated vehicles (ROVs) have been widely used in important tasks such as deep-sea exploration (Whitcomb et al., 2000), marine species migration, coral reef monitoring (Shkurti et al., 2012) and the salvage of sunken ships. ROVs use large-size (e.g., 960p (1280 × 960) (Goodman, 2003)

underwater vision data to perform the above engineering tasks (Jenkyns et al., 2015). However, the absorption and scattering of underwater light cause problems such as low contrast, color deviation, and blurred details, seriously affecting the performance of further vision tasks such as exploration, intelligent analysis or subsea operations. Although there has been some progress in the study of underwater image enhancement (Fabbri et al., 2018; Islam et al., 2020; Naik et al., 2021), these models can only process *small-size* (e.g. 256×256) images in real time, so designing a practical underwater enhancement model for real-time *large-size* underwater image processing is still challenging.

Some early underwater image enhancement approaches, such as CLAHE (Pizer et al., 1990), can provide a fast and real-time solution. However, these approaches always suffer from heavy color distortion and noise caused by scattering and absorption (Lu et al., 2017). Due to the success of deep learning, data-driven approaches provide a potential solution for underwater image enhancement (Li et al., 2020; Naik et al., 2021). As one of the most popular areas of deep learning, GAN-based image-to-image (I2I) methods provide a general framework for underwater image enhancement (Islam et al., 2020). However, most I2I-based methods suffer from the contradiction between image quality and processing speed, especially when the input image is large. A larger input image always leads to larger feature maps and greater computational complexity. Some researchers have indicated that the model parameters and processing time could be reduced by using knowledge distillation (Ren et al., 2021). However, for underwater image enhancement tasks, reducing channels and layers is more likely to cause color shifting and distortion (Guo et al., 2019; Han et al., 2020; Zhuang, 2021). Simultaneously improving underwater image quality and translation efficiency is challenging. zhang2022underwater.

To alleviate this problem, FUnIE-GAN is used for upsampling and downsampling to reduce the feature map size and speed up the enhancement process. However, the channel efficiency is not considered in FUnIE-GAN, which makes this approach insufficient for handling the real-time *large-size* underwater enhancement task. Even a low-cost underwater camera can provide a *large-size* video flow (e.g., 720p, 960p or 1080p). These deep learning-based models are unable to be deployed on ROVs due to their high computation costs, which hinder their further applications. In ShuffleNetv2 (Ma et al., 2018), when the input channels are equal to the output channels, the memory access cost (MAC) can be minimized to improve the model inference speed. Inspired by the “equal channel strategy” proposed in ShuffleNetv2, we propose an equal upsampling block (EUB) and equal downsampling block (EDB) to build a more efficient underwater image enhancement model.

It has been shown that many GAN-based models can improve the perceptual quality of underwater images as well as their contrast and color saturation (Han et al., 2020; Islam et al., 2020). However,

GAN-based models can also generate spatially inconsistent styles and introduce artifacts (Li et al., 2019; Wang et al., 2021; Yang et al., 2021b). Based on experiments, we found that a lightweight GAN-based model is more likely to lead to spatial inconsistency due to its limited layer channels. To alleviate this problem without excessively increasing the parameters, we include residual channel attention blocks (RCABs).

To improve the efficiency of the model while maintaining the quality of the enhanced images, we propose an end-to-end, fast and efficient framework for *large-size* underwater enhancement tasks. Our generator is a lightweight encoder and decoder network with equal upsampling blocks (EUBs) and equal downsampling blocks (EDBs). In addition, we introduce a lightweight RCAB in the generator to solve the problem of spatial inconsistency, further improving the visual quality of the enhanced images. Our contributions are summarized as follows:

- We propose an efficient underwater image enhancement model based on a generative adversarial network, namely, FSpiral-GAN, which can handle a 960p enhanced underwater image task at a running speed of 40 FPS with a GPU (Figure 1).
- We propose an equal upsampling block (EUB) and equal downsampling block (EDB) to improve the computational efficiency so that a real-time *large-size* underwater image enhancement task can be handled by minimizing the MAC.
- We propose a novel architecture that includes RCABs and follows the U-Net design principles in the field of underwater image enhancement. Experimental results show that our architecture can effectively solve the problem of spatial inconsistency, and the color transition of images generated by the proposed model is more continuous, without obvious boundaries.
- We provide qualitative and quantitative comparisons with other advanced methods on real underwater datasets. Experimental results prove that the proposed model has fewer floating-point operations (FLOPs) and faster image processing speed than state-of-the-art models while maintaining generated images with good visual quality.

2 Related work

2.1 Traditional underwater image enhancement methods

Traditional underwater image processing methods can be divided into model-free enhancement methods and physical model-based restoration methods.

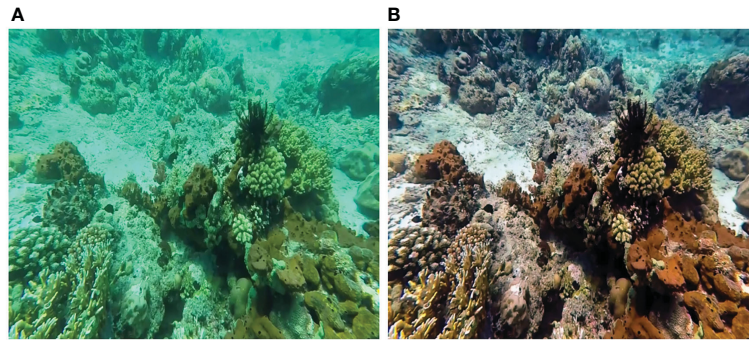


FIGURE 1
(A) Raw underwater image at 960p resolution (1280 × 960), (B) enhanced underwater image generated by our model at a running speed of 40 FPS.

Model-free methods do not consider the degradation principle of underwater imaging and improve the sharpness of the image by adjusting its pixel value. Ancuti et al. (Ancuti et al., 2012) proposed an algorithm based on a fusion strategy that enhances underwater images and videos by using multiscale fusion that combines contrast-enhanced images and color-corrected images. In addition, many researchers have focused on retinex-based methods to improve the quality of underwater images. Fu et al. (Fu et al., 2014) proposed a retinex-based underwater image enhancement algorithm, which includes color correction, layer decomposition, and contrast enhancement. By using bilateral and trilateral filtering in the three channels of an image in the CIELABcolor space, Zhang et al. (Zhang et al., 2017) proposed a retinex-based algorithm to enhance the image. Lu et al. (Lu et al., 2016) enhanced underwater optical images using weighted guided triangular filtering and artificial lighting correction. Zhuang et al. (Zhuang et al., 2021) proposed a Bayesian retinex-based algorithm, which enhances underwater images through multistep priors of reflection and illumination. Most model-free methods are fast but can suffer from overenhancement, color distortion, and color shifting (Lu et al., 2017).

Physical model-based underwater enhancement methods need to estimate the parameters of the underwater imaging model and reconstruct a high-quality image through the inverse degradation process, in which background light and the transmission map need to be estimated. Carlevaris-Bianco et al. (Carlevaris-Bianco et al., 2010) proposed a prior method to estimate the depth of the scene using the attenuation difference between the three color channels of the underwater image to eliminate the influence of underwater haze. Inspired by the image haze removal algorithm using a dark channel prior (He et al., 2010), researchers applied the dark channel prior theory to underwater image processing. By introducing scene depth into the dark channel prior theory and atmospheric scattering model, Cosman (Peng and Cosman, 2017) proposed a scene depth estimation method based on image blur and light absorption for

underwater image restoration. To improve underwater visual quality, Li et al. (Li et al., 2017) proposed an underwater image dehazing algorithm based on the propagation characteristics of light in water and a learning strategy to estimate the transmission map. It is still challenging to restore the color of images containing different underwater scenes with a single model. Akkaynak et al. (Akkaynak and Treibitz, 2019) proved that because the physical model-based underwater image restoration models generally use the atmospheric imaging model and proposed a more accurate physical model Sea-thru. However, the scene depth and optical parameters of underwater scenes are not always available (Islam et al., 2020).

2.2 Data driven-based enhancement methods

Recently, GANs and CNNs have demonstrated powerful capabilities in various image-to-image translation tasks, including image denoising, dehazing, and superresolution (Zhang et al., 2018; Guo et al., 2020; Shao et al., 2020; Wang et al., 2020). Fabbri et al. (Fabbri et al., 2018) first used the image-to-image translation model CycleGAN (Zhu et al., 2017) to generate underwater images and then used the paired data to train an underwater GAN (UGAN). The UGAN architecture is similar to that of U-Net tep2015U and consists of an encoder, decoder, and skip connections. Instead of relying only on the final output feature maps of the encoder to learn all the information of the input image, the spatial information generated by the encoder are retain using this structure. Li et al. (Li et al., 2019) proposed a gated fusion network architecture called WaterNet, which uses images generated by three enhancement methods to help the network learn the most significant features of the input image. The training datasets for WaterNet (Li et al., 2019) contain a variety of different underwater scenes, effectively improving the adaptability of the network. Han et al. (Han et al., 2020) also formulated the

underwater image enhancement task as an application in image translation and proposed Spiral-GAN, which is a novel spiral generative adversarial framework to alleviate the problem of poor generalization performance. To solve the problems of color casts and low contrast caused by wavelength and distance, Li et al. (Li et al., 2021) proposed an underwater image enhancement network that integrates medium transmission-guided and multicolor space. Although these models can improve the visual quality of distorted images, they have high storage requirements and are computationally expensive, which creates a bottleneck problem when real-time underwater exploration equipment is deployed (Naik et al., 2021).

To promote the practical application of deep learning in underwater image enhancement tasks, researchers have focused on designing efficient neural networks. Islam et al. (Islam et al., 2020) considered the nonlinear mapping between distorted underwater images and enhanced underwater images as an example of an application in image-to-image translation tasks and proposed FUnIE-GAN, which uses a simpler generator structure with fewer parameters to achieve fast inference on 256×256 images. Chen et al. (Chen et al., 2019) proposed a single-shot network that can restore the quality of underwater images with higher computational efficiency. Li et al. (Li et al., 2020) proposed a lightweight underwater image and video enhancement model called UWCNN. Since the entire network structure contains only ten convolutional layers, fast and effective training and inference processes can be conducted. Naik et al. (Naik et al., 2021) proposed a shallow neural network architecture to make underwater real-time image enhancement tasks feasible. Yang et al. (Yang et al., 2021b) proposed a lightweight network based on an adaptive feature fusion module and residual module, which can effectively reduce the number of parameters. The aim of the above methods (Chen et al., 2019; Islam et al., 2020; Naik et al., 2021) was to design a lightweight network with acceptable recovery performance. Although these lightweight networks have some effect on improving the speed of the model, they ignore the effect of channel settings on inference speed, and they cannot handle real-time *large-size* underwater image enhancement tasks. In this paper, to alleviate this problem, we consider the underwater image enhancement task as an application in image translation

and use the “equal channel strategy” to process *large-size* underwater images in real time.

3 Proposed model

There is always a trade-off between speed and visual quality in the field of image enhancement. In this paper, we develop a fast and efficient model called FSpiral-GAN that can greatly accelerate the processing speed for *large-size* images and maintain a high quality of the enhanced images. Our model, which is based on a generative adversarial framework, has one generator and N discriminators following the spiral strategy of Spiral-GAN (Han et al., 2020). To improve the efficiency of the model and maintain good quality in the generated images, we design a lightweight generator structure by using the encoder and decoder structure with equal upsampling blocks (EUBs), equal downsampling blocks (EDBs) and RCABs. We also introduce RCABs to alleviate the spatial inconsistency problem that always occurs in the field of underwater image enhancement with GAN-based models (Yang et al., 2021b).

3.1 Network architecture

The overall structure of our model is shown in Figure 2. We follow the training setup of UGAN (Fabbri et al., 2018) and Spiral-GAN (Han et al., 2020) to enhance *large-size* images. To stabilize the adversarial training process and reduce the impact of overfitting, we take advantage of our previous work, Spiral-GAN (Han et al., 2020), and follow the spiral strategy to train the proposed generator and discriminators. Our model has one generator and N discriminators, where N is the number of spiral loops. In addition, these N discriminators are independent of each other without sharing weights. In the i_{th} spiral loop, the output of the generator in the previous loop (the $(i-1)_{th}$ loop) is used as the input to the current loop. The generator G attempt to fool the discriminator D_i ($i = 1, \dots, N$) by making the generated image x_i look more like the true image y , and the discriminator D_i attempts to learn to distinguish between the true image and the generated image. The structure

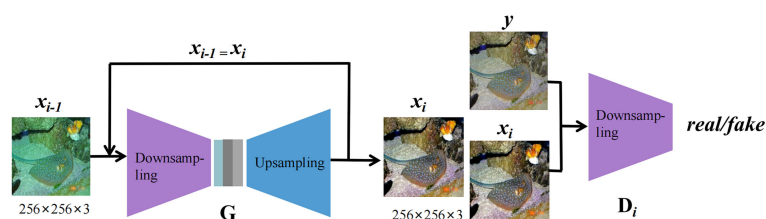


FIGURE 2
The overall structure of our model.

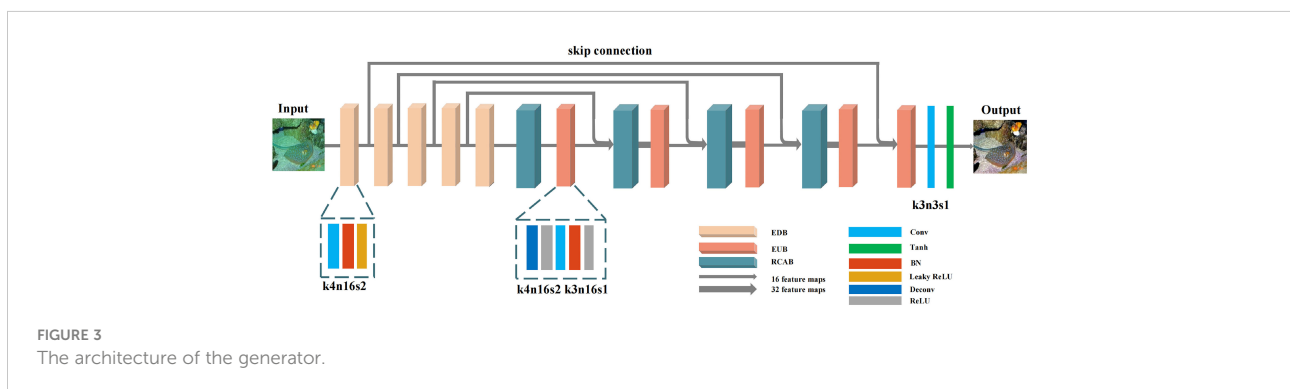
of our discriminator is the same as that described for SpiralGAN (Han et al., 2020). Our discriminator is an extension of PatchGAN (Isola et al., 2017), which can output a true-false matrix instead of a single value. Since there is no further need for discriminators in the test phase, we focus solely on the improvement of the generator while leaving the discriminators unchanged.

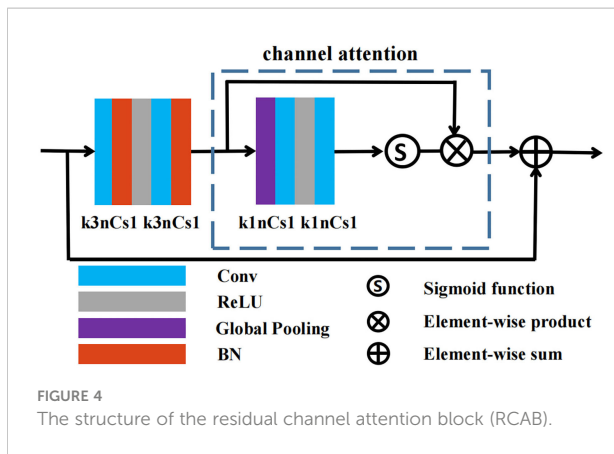
The detailed structure of our generator is shown in Figure 3, which follows the principles of U-Net (Ronneberger et al., 2015). Our generator has an encoder-decoder structure composed of a number of equal upsampling blocks (EUBs), equal downsampling blocks (EDBs) and RCABs with skip connections. When completing experiments, we found that the upsampling and downsampling operations in encoders and decoders can introduce more parameters [49], but these operations have more advantages when dealing with *large-size* images. Please refer to Section 5.4 for more detail.

Enlightened by this point, we design our encoder and decoder structure with equal upsampling blocks (EUBs) and equal downsampling blocks (EDBs) to handle a *large-size* underwater image processing task. Specifically, to avoid quality degradation caused by the change in generator structure, we directly use the upsampling and downsampling blocks of SpiralGAN (Han et al., 2020). In ShuffleNetv2 (Ma et al., 2018), the MAC required to access the intermediate feature maps occupies a large part of the inference speed in the convolutional layers and is a key factor in consumption. It was proposed that minimizing MAC with equal channel width could improve the models inference speed. The influence of MAC on the network inference speed is not considered for the generator in SpiralGAN. The input channels of the convolutional layers are not equal to the output channels in the upsampling and downsampling blocks. Therefore, inspired by this conclusion, we make the input channels and the output channels equal in the downsampling and upsampling blocks to design an efficient generator. We refer to these downsampling and upsampling blocks with equal channels as equal downsampling blocks (EDBs) and equal upsampling blocks (EUBs), respectively. Please note that the skip connection will concatenate the output feature maps of the encoder and decoder in the channel dimension, so the number of input channels of the

deconvolution layer in an EUB will inevitably change. With each skip connection, the input channels will further increase. To keep the total number of channels in the model equal, regardless of the input channels of the deconvolution layer in an EUB, we redesign the generator architecture to force the output channels of the deconvolution layer to be 16. The details of our generator can be found in Figure 3. The width of the line represents the number of input and output feature maps per function block, where $k4n16s2$ denotes a convolutional layer with 4×4 filters, 16 convolution kernels and a stride of 2.

Generated images from GAN-based models for underwater image enhancement often have spatially inconsistent styles (Yang et al., 2021b), which reduces the visual quality of the generated image. To improve the quality of the generated images, we introduce RCAB into a generator that follows the U-Net design principles. We carefully study how the location and the number of RCABs affect the results and propose a configuration with the best underwater enhancement performance. Compared with the RCAB in the encoder, the RCAB in the decoder can learn the input of the previous layer and the input of the skip connections to reconstruct the enhanced underwater image. The feature map information that can be learned is richer, which is beneficial for RCAB when extracting important features. The size of the output feature maps of the deconvolution layer is twice its original size. To make the model have a faster inference speed, we place RCABs before each deconvolutional layer in the four EUBs in the decoder. This combination can effectively solve the spatial inconsistency phenomenon by allowing the color transitions of the enhanced images to be more continuous. An RCAB is a functional block that integrates channel attention (CA) into a residual block (Zhang et al., 2018). Through the experiments, we found that the lack of CA can lead to undersaturation. A full RCAB can alleviate the issue of spatial inconsistency. The structure of the RCAB is shown in Figure 4, and we can see the parameter settings of the convolutional layers in the RCAB. If the size of the input of the RCAB is $H \times W \times C$, H denotes the height, W denotes the width, and C denotes the number of feature maps, where $k3nC_s1$ denotes a convolution layer with 3×3 filters, C convolution kernels and a stride of 1. In the





RCAB, the number of input feature maps is equal to the number of output feature maps. We show that some images exhibit spatial inconsistency with undesirable artifacts in Figure 5. From Figure 5E, we can see that when RCABs are not introduced into the proposed model (w/o RCAB), there are obvious edges and artifacts in the water portion of the enhanced images. This phenomenon also exists in the enhanced images of other GAN-based models, such as UGAN-P, FUnIE-GAN, and Spiral-GAN (Figures 5B–D). When we introduced RCABs into the proposed model, the color of the image after RCAB processing, as shown in Figure 5F, is more continuous and realistic, without obvious boundaries. This demonstrates the superiority of RCAB in dealing with the spatial inconsistency problem. We also consider the success of UResNet [26] by adding a batch norm (BN) to the ResBlock modules in the field of superresolution to improve the visual quality of the images. We add a BN layer after the first and second convolutional

layers of each RCAB to increase the contrast of the generated image. Please note that although each RCAB incurs additional computational cost, the experimental results show that 4 RCABs only slightly affect the processing speed. Please refer to Section 5.3 for more details.

3.2 Objective

Generative adversarial networks (GANs) contain at least one generator G and discriminator D . A typical conditional generator (Isola et al., 2017) attempts to fool the discriminator by learning the mapping of the source domain to the target domain to make the generated image look more realistic. Let x denote the underwater image and y denote the ground truth image without distortion. G attempts to learn the mapping from x to y . The loss function of a standard conditional GAN can be represented by

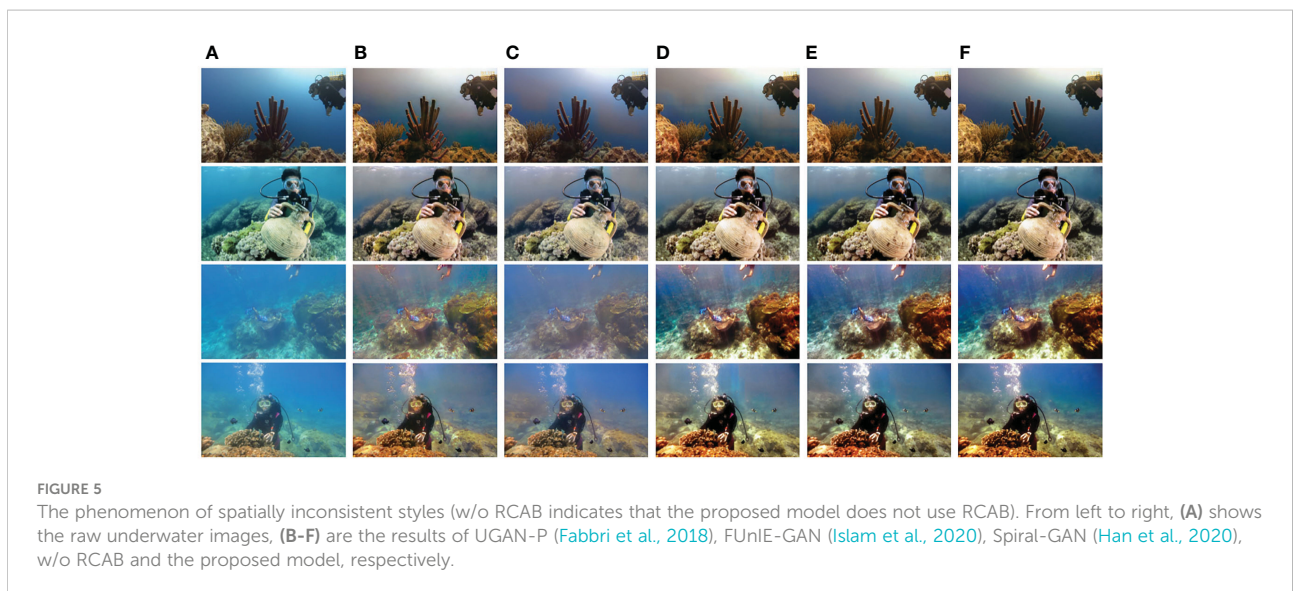
$$\min_{G,D} \max_{x,y \sim p_{data}(x,y)} [\log D(x,y)] + \mathbb{E}_{x \sim p_{data}(x,y)} [\log (1 - D(x, G(x)))] \tag{1}$$

where $p_{data}(x,y)$ represents the joint distribution of x and y .

The L_2 loss computes the L_2 distance between the generated image and ground truth. The L_2 loss amplifies the gap between the larger error and the smaller error and will make the network focus on distorted colors. To restore the distortion color, we adopt the L_2 loss function:

$$L_{L_2}(G) = \mathbb{E}_{x,y \sim p_{data}(x,y)} [\|y - G(x)\|_2] \tag{2}$$

Due to the absorption and scattering of light under water when collecting images from the deep ocean, the camera is equipped with lighting equipment, and the collected datasets are



affected by light, leading to overexposure problems. To remove this effect, we use the angular loss (Sidorov, 2020). The angular loss calculates the angular distance of the generated image and the ground truth image in the RGB color space, stabilizes the training process, and prevents the generated image from undergoing color distortion. The function can be written as follows:

$$L_{angular}(G) = \mathbb{E}_{x,y \sim p_{data}(x,y)} [\mathcal{L}(y, G(x))] \quad (3)$$

To alleviate the lack of diversity in the underwater paired data generated by CycleGAN [54], Spiral-GAN introduces a spiral strategy. Following the spiral strategy in Spiral-GAN, the above three loss functions can be expressed as:

$$\min_{G,D} \max_{G,D} L_{spiral_cGAN}(G, D) = \mathbb{E}_{x,y \sim p_{data}(x,y)} [\log D(x, y)] + \sum_N \mathbb{E}_{x \sim p_{data}(x,y)} [\log (1 - D(x, \underbrace{G(\dots G(x) \dots)}_n))] \quad (4)$$

$$L_{spiral_L2}(G) = \sum_N \mathbb{E}_{x,y \sim p_{data}(x,y)} [\|y - \underbrace{G(\dots G(x) \dots)}_n\|_2] \quad (5)$$

$$L_{spiral_angular}(G) = \sum_N \mathbb{E}_{x,y \sim p_{data}(x,y)} [\mathcal{L}(y, \underbrace{G(\dots G(x) \dots)}_n)] \quad (6)$$

where N is the number of spiral loops. x and y represent the input image and the corresponding ground truth image, respectively. L_{spiral_cGAN} , L_{spiral_L2} and $L_{spiral_angular}$ represent the conditional GAN loss, L2 loss and angular loss after applying the spiral strategy, respectively. $\underbrace{G(\dots G(x) \dots)}_n$ indicates the generated images of generator G in the n spiral loop. The final objective function of our proposed FSpiral-GAN is as follows:

$$\mathcal{L} = \min_G \max_D L_{spiral_cGAN}(G, D) + \lambda_1 L_{spiral_L2}(G) + \lambda_2 L_{spiral_angular}(G) \quad (7)$$

where λ_1 and λ_2 are the weights of L_{spiral_L2} and $L_{spiral_angular}$ respectively.

4 Experiments

4.1 Underwater datasets

4.1.1 EUVP Dataset

The EUVP dataset is a large dataset collected by FUNIE-GAN researchers (Islam et al., 2020). It includes more than 12K paired images and 8K unpaired images of good and poor visual quality. The underwater images were taken during experiments and explorations under different underwater scenarios and water quality conditions. These images contain a variety of water quality types and scenes. We randomly selected 4129 images for our model testing.

4.1.2 UIEB dataset

The UIEB Dataset (Li et al., 2019) includes images of underwater scenes taken by Li et al. and images collected from the internet. It contains images with varying degrees of distortion and complex scenes, including coral and marine life. We selected 890 images with reference images for testing, and all images were resized to 768×512. Please note that we highlight the best scores in bold type in Tables 1–10.

4.1.3 RUIE dataset

The RUIE Dataset (Liu et al., 2020) contains a large number of underwater images divided into an underwater image quality set (UIQS), underwater color case set (UCCS) and underwater higher-level task-driven set (UHTS) due to different evaluation objectives. The UIQS is used to evaluate the ability of underwater image enhancement algorithms to improve image quality. The UCCS is intended to test the ability of the algorithms to correct the color cast. The UHTS is used to test the ability of algorithms to improve visual perception. We tested our model on the UCCS of the RUIE dataset, and all images were resized to 512×512.

4.1.4 D3 dataset

The D3 dataset (Akkaynak and Treibitz, 2019) contains 68 images taken from various angles of a single reef scene along with corresponding depth maps. We selected all images for our model testing, and all images were resized to 7968×5312.

4.2 Evaluation metrics

We use the peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM) (Hore and Ziou, 2010) to assess the quality of the generated images. The PSNR and SSIM are two full-reference metrics used to measure the reconstruction quality and similarity of the reference images. We also use the full reference metric patch-based contrast quality index (PCQI) (Wang et al., 2015) to assess the quality of images with contrast changes. The closer the contrast of the generated image is to that of the reference image, the higher the PCQI score.

We use UIQM (Panetta et al., 2015) to evaluate underwater image quality, which is made up of the underwater colorfulness measure (UICM), underwater image sharpness measure (UISM) and underwater image contrast measure (UIConM). The UIQM metrics can be written as follows:

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

According to the original work, we set the three parameters as $c_1 = 0.0282$, $c_2 = 0.2953$, and $c_3 = 3.5753$. We use underwater color image quality evaluation (UCIQE) (Yang and Sowmya, 2015), which consists of a linear combination of color intensity, saturation, and contrast, to evaluate the degree of low contrast, blur, and color casts in underwater images. We also use a metric

TABLE 1 Quantitative evaluation on the UIEB dataset.

Method	CLAHE	OMGD	FUnIE-GAN	WaterNet	UWCNN	Shallow-UWnet	Spiral-GAN	Ours
SSIM	0.8459	0.8002	0.7481	0.6902	0.6977	0.8198	0.8435	0.8450
PSNR	18.58	18.51	16.91	15.06	14.61	18.18	19.02	19.16
PCQI	0.9783	0.6366	0.7263	0.6027	0.6017	0.6693	0.8440	0.8226

The bold indicates the best scores.

TABLE 2 Quantitative evaluation on the UCCS subset of the RUIE dataset.

Method	CLAHE	OMGD	UGAN-P	FUnIE-GAN	WaterNet	UWCNN	Shallow-UWnet	Spiral-GAN	Ours
UISM	6.521	6.749	6.611	6.739	6.623	6.606	6.703	6.544	6.473
UICM	1.791	2.917	1.905	1.131	2.955	0.864	0.742	2.892	3.336
UIComM	0.743	0.681	0.699	0.603	0.626	0.580	0.512	0.724	0.807
UIQM	4.631	4.511	4.498	4.179	4.277	4.050	3.830	4.602	4.892
UCIQE	0.493	0.5181	0.547	0.4957	0.4999	0.4579	0.4397	0.6137	0.6148
CCF	20.14	14.01	18.85	14.51	12.68	12.24	11.83	27.64	26.43

The bold indicates the best scores.

TABLE 3 Quantitative evaluation on the EUVP dataset.

Method	CLAHE	OMGD	UGAN-P	FUnIE-GAN	WaterNet	UWCNN	Shallow-UWnet	Spiral-GAN	Ours
UISM	7.191	6.968	6.827	7.128	6.984	7.139	7.298	6.873	6.864
UICM	1.405	2.354	2.318	0.677	4.335	1.219	0.889	3.438	4.370
UIComM	0.542	0.560	0.608	0.439	0.531	0.473	0.354	0.767	0.852
UIQM	4.101	4.125	4.255	3.693	4.083	3.834	3.446	4.868	5.146
UCIQE	0.440	0.4857	0.530	0.4313	0.4812	0.4356	0.3975	0.6166	0.6213
CCF	13.68	11.46	14.33	10.06	9.75	8.90	9.30	22.87	21.08

The bold indicates the best scores.

TABLE 4 Frames per second (FPS) of different methods on the underwater video.

Method	CLAHE	OMGD	FUnIE-GAN	WaterNet	UWCNN	Shallow-UWnet	Spiral-GAN	Ours
1280×960	137.7	20.4	13.1	1.5	10.6	7.1	22.4	40.0
256×256	820.7	185.0	36.1	15.2	267.7	68.4	222.8	246.3

The bold indicates the best scores.

TABLE 5 Parameters and FLOPs of different methods on the 1280×960 underwater video.

Method	CLAHE	OMGD	FUnIE-GAN	WaterNet	UWCNN	Shallow-UWnet	Spiral-GAN	Ours
Parameters	–	137347	4215843	1090668	39972	219456	4993539	133987
FLOPs(G)	–	53	233	2679	98	406	138	20

The bold indicates the best scores.

TABLE 6 Quantitative evaluation on 7968×5312 images.

Methods	FPS	UCIQE	UIQM
Fusion	0.0061	0.653	4.630
Sea-thru	0.0003	0.674	4.783
Ours	0.9118	0.667	4.690

The bold indicates the best scores.

TABLE 7 The result of mutual information.

Method	CLAHE	OMGD	FUnIE-GAN	WaterNet	UWCNN	SUWnet	Spiral-GAN	Ours
score	0.547	0.6416	0.623	0.691	0.701	0.695	0.644	0.722

The bold indicates the best scores.

TABLE 8 PCQI, SSIM, PSNR, and FPS comparison on the UIEB dataset using different RCAB numbers.

Method	PCQI	SSIM	PSNR	FPS
N=3	0.8107	0.844	18.961	41
N=4(Ours)	0.8226	0.845	19.164	40
N=5	0.7942	0.843	18.972	32

The bold indicates the best scores.

TABLE 9 Comparison of RCAB and BN using the UIEB dataset. Method w/o RCAB w/o CA w/o BN.

Method	w/o RCAB	w/o CA	w/o BN	ours
PCQI	0.8203	0.8062	0.8203	0.8226
SSIM	0.837	0.839	0.841	0.845
PSNR	18.99	18.13	19.12	19.16
FPS	41.7	40.4	40.7	40.0

The bold indicates the best scores.

TABLE 10 Comparison of upsampling and downsampling operations during underwater video processing.

Input size	equal channel	upsampling and downsampling	Parameters	FLOPs(G)	FPS(img/s)
256 × 256	×	✓	3.303M	2.806	242
1280 × 960			3.303M	52.61	32.24
256 × 256	✓	×	0.110M	6.794	95.12
1280 × 960			0.110M	127.4	4.96
256 × 256	✓	✓	0.134M	1.086	246.3
1280 × 960			0.134M	20.37	40.32

based on imaging analysis of underwater absorption and scattering properties, called CCF (Wang et al., 2018), which effectively quantifies absorption and scattering-induced color shifts and blurring.

We consider FLOPs (He et al., 2016; Tan and Le, 2019) and frames per second (FPS) to evaluate the efficiency of our model.

4.3 Implementation details

In our experiment, we use a training dataset from UGAN (Fabbri et al., 2018), with a total of 6128 pairs of images. The size of the input image is 256×256 . We use the ADAM optimizer with a learning rate of $1e^{-4}$. The batch size is set to 32, and spiral loops $N=10$, $\Lambda_1 = 100$ and $\Lambda_2 = 0.1$ according to the settings of Spiral-GAN. The reduction ratio r in the channel attention is set to 1. We trained our model on the PyTorch framework for 100 epochs on a GeForce GTX 1080 Ti.

4.4 Comparative experiments with other methods

To evaluate the performance of our model, we select 10 state-of-the-art methods: CLAHE (Pizer et al., 1990), Fusion

(Ancuti et al., 2012) and Sea-thru (Akkaynak and Treibitz, 2019), three traditional methods; OMGD (Ren et al., 2021), a fast image translation method; and UGAN-P (Fabbri et al., 2018), FunIE-GAN (Islam et al., 2020), WaterNet (Li et al., 2019), UWCNN (Li et al., 2020), Shallow-UWnet (Naik et al., 2021), and Spiral-GAN (Han et al., 2020), deep learning-based underwater image enhancement methods.

4.4.1 Experiments with ground truth

The UIEB dataset is a manually collected dataset, in which the reference image is the best image selected by the volunteers from all underwater enhancement experiments. Figure 6 shows a qualitative comparison of the results between the different methods on the UIEB dataset. As shown in Figures 6B–D, image contrast is increased when using CLAHE, but the color cast is not corrected. For OMGD, a limited amount of scatter is removed, the contrast is only slightly improved, and an additional blue cast is introduced. When using FUNIE-GAN, the color distortion is corrected well and the contrast of the image is slightly improved. However, when enhancing an image that contains green water, additional colors are introduced, such as earthen yellow (1st and 4th and 6th rows); in addition, minimal elimination of the blurring caused by the light scattering of tiny particles is achieved (4th and 6th rows). As shown in Figure 6E, a new grayish-blue hue (1st row) is introduced by WaterNet, and

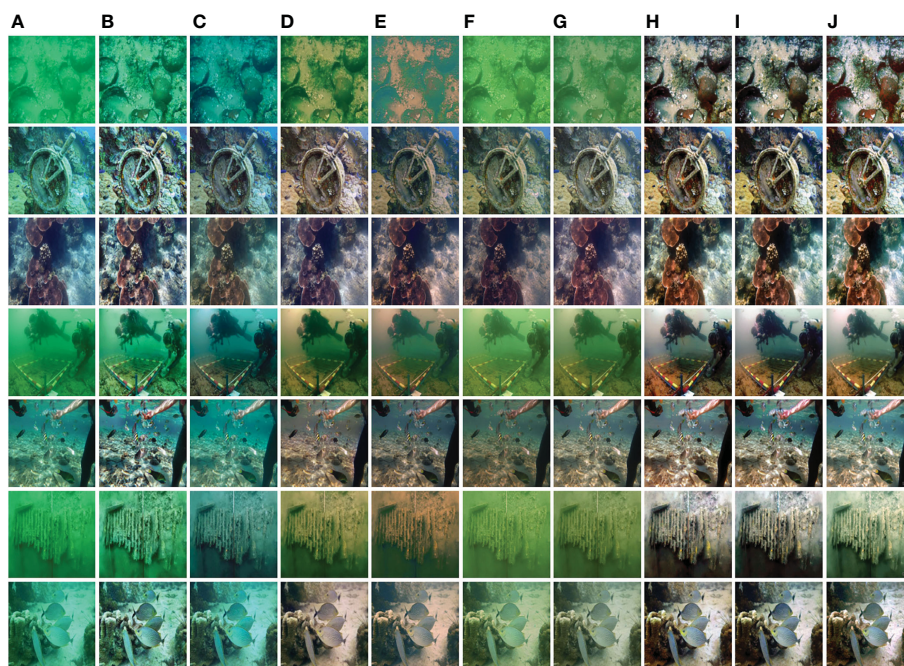


FIGURE 6

Visual comparisons on the UIEB dataset. From left to right, (A) shows the raw underwater images, (B–I) are the results of CLAHE (Pizer et al., 1990), OMGD (Ren et al., 2021), FUNIE-GAN (Islam et al., 2020), WaterNet (Li et al., 2019), UWCNN (Li et al., 2020), Shallow-UWnet (Naik et al., 2021), Spiral-GAN (Han et al., 2020) and the proposed model, respectively, and (J) shows the reference images.

minimal changes are observed for the removal of scattering and improvement of contrast. As shown in Figure 6F, images enhanced using UWCNN show minimal improvements in contrast, are not deblurred and greenish tones are not ideally enhanced. For Shallow-UWnet, the color cast is simply corrected, the contrast is not improved, and the scattering is not resolved, as shown in Figure 6G. As shown in Figures 6H, I, the proposed model and Spiral-GAN are effective in enhancing contrast, removing blur, and correcting color cast, but when using Spiral-GAN, a slight red deviation tends to be generated in the image (3_{rd} and 4_{th} rows). We also used PSNR and SSIM for quantitative analysis, and the results are shown in Table 1. PSNR represents the degree of global pixel difference, so the color of the enhanced image has a greater effect on PSNR. We achieved the highest PSNR and PCQI value using our model. This result verifies that the enhanced image from our model is closer to the reference image in color and contrast.

4.4.2 Experiments without ground truth

The UCCS subset of the RUIE dataset is a dataset that evaluates the ability to correct the distorted color of enhancement models. This subset contains 100 images for blue, green, and blue-green water types. The images in the EUVP dataset have complex water types and underwater scenes and have various camera angles. We use these images to measure the performance in restoring the color and generalization performance of the models.

The comparisons on the UCCS subset are shown in Figure 7. For each type of water in Figure 7, we present two representative images. The pictures from top to bottom are blue-green, bluish, and

greenish tones. As shown in Figures 7B–G, the use of CLAHE and OMGD have little effect on color restoration, and both methods have difficulty removing scattering effects. UGAN-P, FUnIE-GAN, and WaterNet can be used to restore color but only slightly improve contrast and resolve the blurring caused by scattering. The results from WaterNet have a gray tone. The visual results from UWCNN are similar to those of WaterNet on the UIEB dataset: the color cast is only slightly corrected, and the visibility is only slightly improved. Shallow-UWnet has a certain effect on restoring color, as shown in Figure 7H. As shown in Figures 7I, J, the proposed model and Spiral-GAN can be used to remove blue and green tones and improve contrast. Due to the lack of reference images, we have to use the UIQM, UCIQE, and CCF, which have no reference matrix, to measure the quality of the generated images. In Table 2, we can see that our model outperforms other methods on most quality metrics. We achieve the best UIQM results on the UCCS subset using our model. A high UIQM value means that the underwater images have high color saturation and contrast.

Figure 8 shows the comparison examples on the EUVP dataset. As shown in Figures 8B–H, the use of CLAHE and OMGD have little influence on the removal of the blurs caused by scattering, and they have difficulty restoring color. UGAN-P, FUnIE-GAN and UWCNN improve the image contrast, but these methods are not good at correcting color casts while handling images with green water types. WaterNet performs the worst when handling images with blue water types (1_{st} and 2_{nd} rows). Shallow-UWnet does little to improve contrast and remove scatter and only fades the color of the water in the generated image. As shown in Figures 8I, J, both Spiral-GAN and the proposed model can correct the color cast as well as the

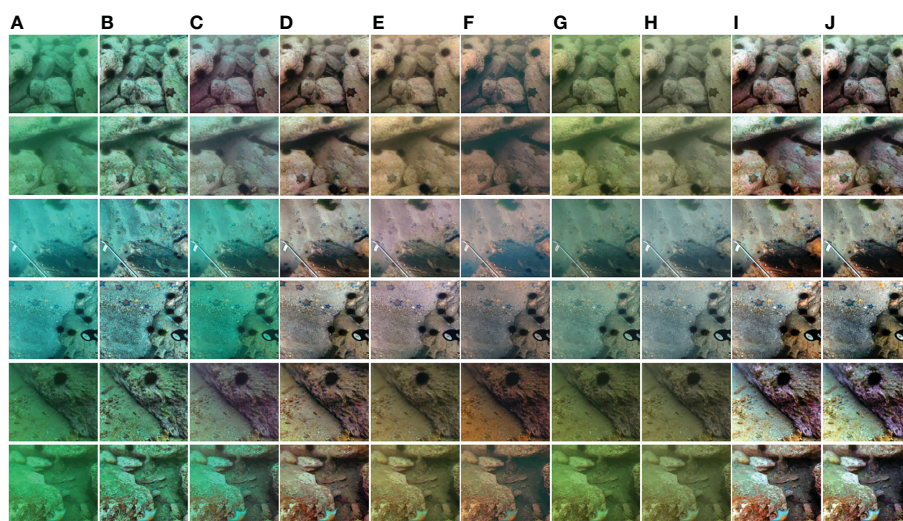


FIGURE 7

Visual comparisons on the UCCS subset of the RUIE dataset. From left to right, (A) shows the raw underwater images, (B–J) are the results of CLAHE (Pizer et al., 1990), OMGD (Ren et al., 2021), UGAN-P (Fabbri et al., 2018), FUnIE-GAN (Islam et al., 2020), WaterNet (Li et al., 2019), UWCNN (Li et al., 2020), Shallow-UWnet (Naik et al., 2021), Spiral-GAN (Han et al., 2020) and the proposed model, respectively.

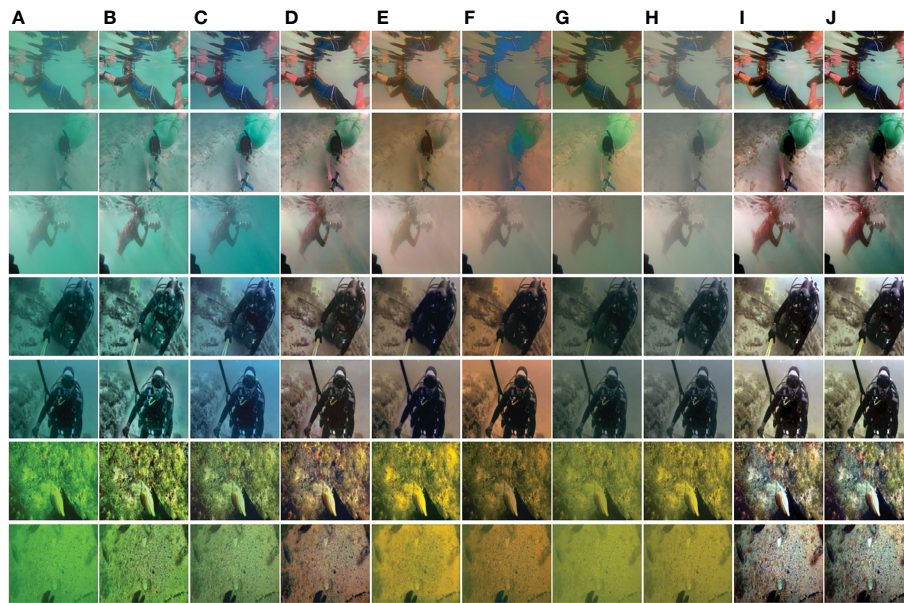


FIGURE 8

Visual comparisons on the EUVP dataset. From left to right, (A) shows the raw underwater images, (B–J) are the results of CLAHE (Pizer et al., 1990), OMGD (Ren et al., 2021), UGAN-P (Fabbri et al., 2018), FUnIE-GAN (Islam et al., 2020), WaterNet (Li et al., 2019), UWCNN (Li et al., 2020), Shallow-UWnet (Naik et al., 2021), Spiral-GAN (Han et al., 2020) and the proposed model, respectively.

contrast of the images. We find that many GAN-based models, such as the UGAN-P, FUnIE-GAN, and Spiral-GAN models, as shown in Figures 8D, E and Figure 8I, generate images with spatially inconsistent styles (1_{st} row). In contrast, the color transitions of images generated by the proposed model are more continuous, without obvious boundaries, which proves the effectiveness of the RCAB. In Table 3, we can see that the images generated by the proposed model have the highest UIQM and UCIQE, indicating that our method is most consistent with human visual perception.

4.4.3 Experiments on underwater video

In this section, we validate our model on a real underwater video to verify the running speed on the *large-size* and *small-size* images. We download a diving video from YouTube. The size of each frame in the video is 1280×960. We resize the video to 256×256 size and test the inference speed of different models on the two video sizes. A comparison of the FPS results of different methods on underwater video is shown in Table 4. Table 5 shows the number of parameters and FLOPs when different models process the 1280×960 videos. Because CLAHE is a traditional method, the GPU version of CLAHE is not available. Three inputs are generated for WaterNet by applying white balance (WB), histogram equalization (HE), and gamma correction (GC) to the input image. Since these three methods only have code in MATLAB, we only consider the time to enter the neural network when calculating the FPS.

In Table 4, we can see that all fast underwater image enhancement models are capable of processing 256×256 videos in real time, but only the proposed model can process 1280×960 videos in real time because our model takes into account the effect of MAC on the network inference speed. The subjective results of the model processing 1280×960 video are shown in Figure 9. The performance of all methods in different frames (1_{st} to the 4_{th} rows) in the same scene are consistent, and there are no flicker artifacts. As shown in Figure 9B, although the contrast can be significantly improved when using CLAHE, color bias cannot be corrected. As shown in Figure 9C, the degree of color cast intensified when using OMGD. When using FUnIE-GAN and WaterNet, there is little effect on removing scattering effects, but these methods can restore color well, as shown in Figures 9D, E. The degree of color cast is weakened using UWCNN and Shallow-UWnet, as shown in Figures 9F, G. As shown in Figures 9H, I, both Spiral-GAN and the proposed model increase contrast and improve visibility, but Spiral-GAN makes the enhancement results spatially inconsistent, which represents a discontinuous color transition. The proposed model introduces RCAB to solve spatial inconsistency problems. In Table 5, we can see that among the deep learning models, our model has the fewest FLOPs for handling *large-size* images. Compared with the baseline model Spiral-GAN, the proposed model is compressed to 37.3× parameters, 6.9× FLOPs and increased the FPS 1.8× when dealing with *large-size* images. These results show that the enhancement effect of our model is superior to that of Spiral-GAN while running nearly as fast as Spiral-GAN.

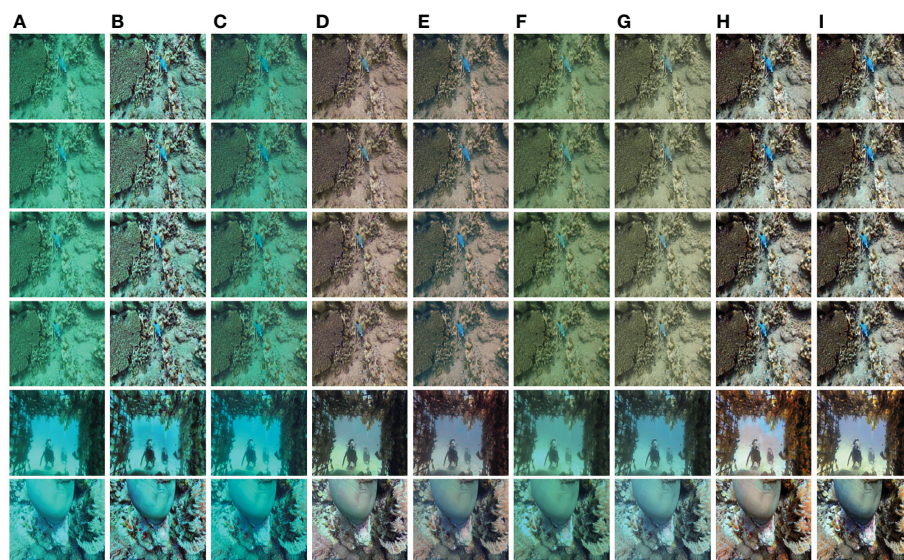


FIGURE 9

Visual comparisons of the 1280×960 frames in the underwater video. From left to right, (A) shows the raw underwater images, (B–I) are the results of CLAHE (Pizer et al., 1990), OMGD (Ren et al., 2021), FUnE-GAN (Islam et al., 2020), WaterNet (Li et al., 2019), UWCNN (Li et al., 2020), Shallow-UWnet (Naik et al., 2021), Spiral-GAN (Han et al., 2020) and the proposed model, respectively.

4.4.4 Experiments on D3 dataset

The Figure 10 shows the results of the comparison experiment on the D3 dataset, the size of each image in the dataset is 7968×5312. It can be seen from Figures 10B–D that all three methods can effectively remove the color cast and improve the contrast and color saturation of the images. As can be seen from Table 6, The proposed model approximates the sea-thru method on most of the evaluation metrics and has obvious speed advantages in processing high-resolution images.

4.5 Color accuracy test

Most underwater image enhancement methods can enhance underwater image contrast while neglecting color correction. We design an experiment to determine how close the enhanced color is to the original color. Figure 11A shows the raw underwater image, and Figure 11J represents the reference image. The results are shown in Figures 11B–I. Due to the absorption and scattering of light, the original color card has a color cast and low contrast

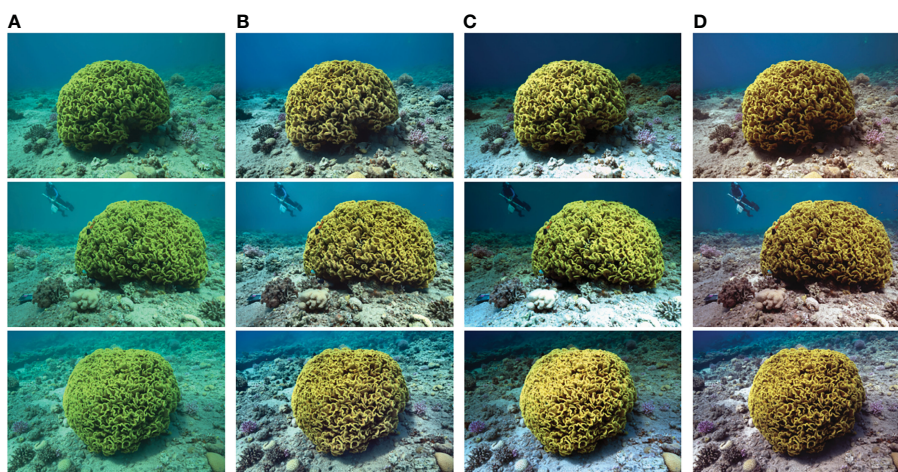


FIGURE 10

Visual comparisons on the 7968×5312 images in the D3 dataset. From left to right, (A) shows the raw underwater images, (B–D) are the results of Fusion (Ancuti et al., 2012), Sea-thru (Akkaynak and Treibitz, 2019) and the proposed model, respectively.

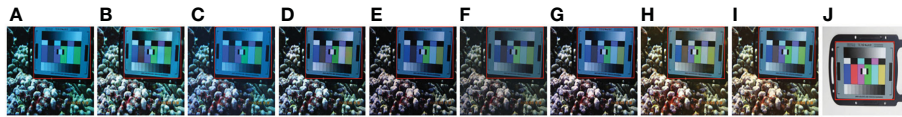


FIGURE 11

The visual results of the color accuracy test. From left to right, (A) shows the raw underwater images, (B–I) are the results of CLAHE (Pizer et al., 1990), OMGD (Ren et al., 2021), FUnIE-GAN (Islam et al., 2020), WaterNet (Li et al., 2019), UWCNN (Li et al., 2020), Shallow-UWnet (Naik et al., 2021), Spiral-GAN (Han et al., 2020) and the proposed model, respectively, and (J) shows the reference images.

problems compared with the ground truth image. The images enhanced by CLAHE, OMGD, and FUnIE-GAN still have serious color casts. The WaterNet, Shallow-UWnet and UWCNN methods can be used to remove blue color cast better, but these methods have little effect on improving contrast. Both our model and Spiral-GAN can be used to recover colors and improve contrast and brightness. Compared to the Spiral-GAN model, our model can be used to recover the blue sky better. This test verifies that our model can be used to correct the distorted color and enhance the image with a better perception effect.

To further demonstrate that the image generated by our method is close to the reference image, we design an experiment to calculate the MI (Kraskov et al., 2004) between the generated image and the reference image. Mutual information is a fundamental magnitude to measure the relationship between random variables (Belghazi et al., 2018). MI can also be used to measure the true correlations between variables (Kinney and Atwal, 2014). Please note that there is no pixelwise matching between the raw image and the reference image. To minimize the effects of the backgrounds, we crop only the color card part of the color bar in the generated image and the reference image to calculate the MI, and the red box part is the color card part of the color bar. The experimental results are shown in Table 7. Our model has the highest score, indicating that the generated image of our model has the strongest correlation with the reference image.

5 Ablation study

In this section, we discuss the impact of different configurations on performance, including the network width, RCAB and BN and EUB and EDB ablation experiments.

5.1 Network width

Generally, a network will perform better with more channels. Here, we evaluate the influence of different feature numbers C in EUB and EDB on network performance by choosing $C=8, 16$ and 32 . The ablation results on the UIEB dataset are shown in Table 11 and Figure 12. As shown in Figure 12B, when the number of channels is 8, the network cannot learn features well, and some artifacts will appear, such as artifacts above the hat and water ripples (shown in the red box). At the same time, some details are missing in the enhanced image, and the SSIM values decrease. When the channel number changes from 16 to 32, we observe a small gain in the PCQI, PSNR, and SSIM indices with dramatic decreases in running speed, as shown in Table 11. Considering the balance of image quality and model speed, we use 16 as a default value for the number of channels.

5.2 RCAB numbers

In this section, we verify that the model can achieve optimal performance by introducing 4 RCABs. We test $N \in \{3, 4, 5\}$ RCABs in the generator. The experimental results are shown in Table 8 and Figure 13. From Figure 13B, we can see that when the decoder has only 3 RCABs, some generated images are reddish, such as the face of the stone statue in the first row and the diver's body in the second row. When the decoder has 5 RCABs, the generated images have obvious artifacts, such as under the sea in the second row and above the box in the third row in Figure 13D. When the model has 4 RCABs, there are no artifacts or reddish color in the generated images. As seen from Table 8, when the number of RCABs is 4, the quality of the generated images of the model is significantly

TABLE 11 PCQI, SSIM, PSNR, and FPS comparison on the UIEB dataset using different network widths.

Method	PCQI	SSIM	PSNR	FPS
C=8	0.7909	0.839	19.28	55
C=16	0.8226	0.845	19.164	40
C=32	0.8402	0.854	19.462	25

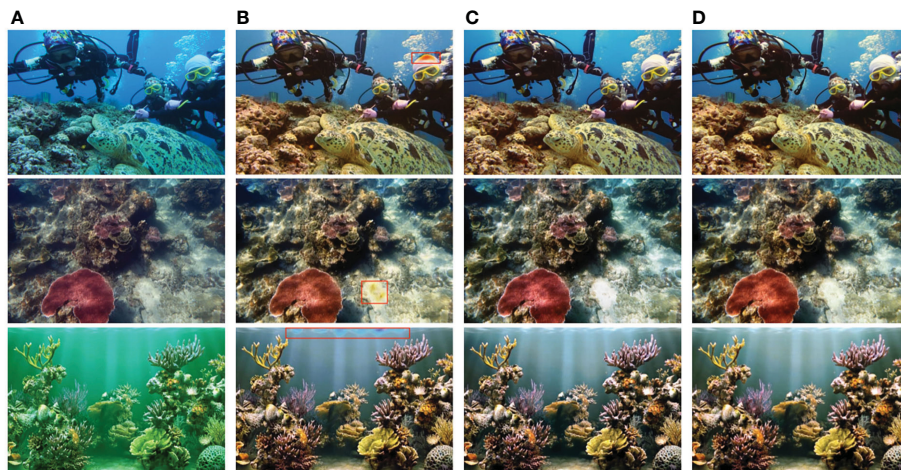


FIGURE 12
Visual comparisons with different network widths C in EDB and EUB on the UIEB dataset. From left to right, (A) shows the raw underwater images, and (B–D) are the results of $C = 8$, $C = 16$ and $C = 32$, respectively.

improved. Therefore, when the decoder includes 4 RCABs, the model can obtain the best enhancement performance.

5.3 RCAB and BN layers

In this section, we verify the effect of RCABs and BN layers in the proposed model. As a lightweight network, our model can handle a large batch size of up to 32. Thus, a BN is a more suitable solution for providing normalization for training data. Practically, the BN layers work well when restoring underwater image details and improving contrast (Liu et al., 2019). Our

ablation experiments of RCAB, CA and BN on FSpiral-GAN are shown in Figure 14 and Table 9. As shown in Figures 14B, E, an RCAB can alleviate spatial inconsistency by allowing the color transitions of enhanced images to be more consistent (1st and 2nd rows), and the generated images tend to be more continuous without obvious boundaries. After the image passes through the BN layer, its color distribution is normalized, so our model with an RCAB and BN tends to correct the reddish color deviation and improve contrast. The enhancement results of the model introducing BN and CA comprehensively improved PCQI, SSIM, and PSNR scores. As shown in Table 9, an RCAB with a BN or CA only slightly decreases the FPS.

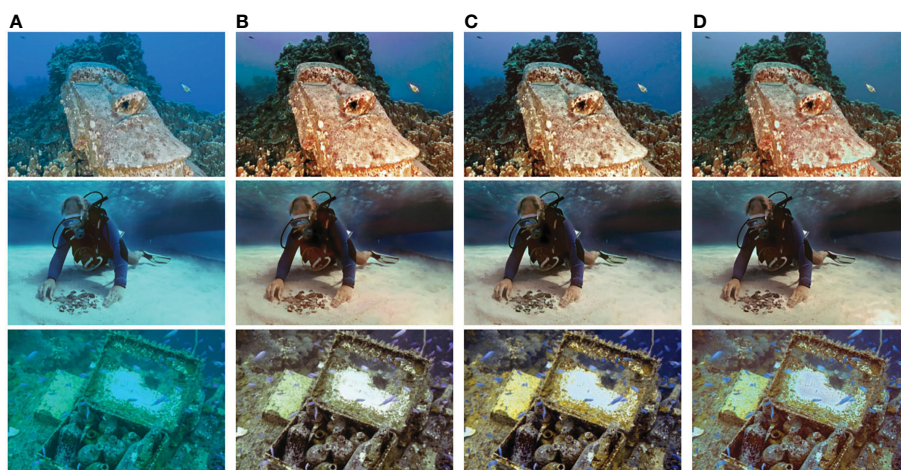


FIGURE 13
Visual comparisons with different RCAB numbers N in the generator on the UIEB dataset. From left to right, (A) shows the raw underwater images, and (B–D) are the results of $N = 3$, $N = 4$ (Ours) and $N = 5$, respectively.

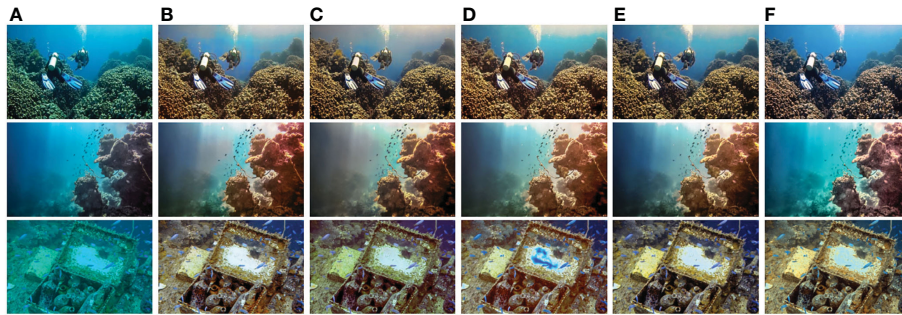


FIGURE 14

Visual comparisons of ablation experiments on RCAB and BN layers on UIEB datasets. Note: w/o RCAB indicates that the proposed model does not use RCAB, w/o CA indicates that the proposed model uses RCAB and does not use CA, w/o BN indicates that the proposed model uses RCAB and does not use BN. From left to right, (A) shows the raw underwater images, (B–E) are the results of w/o RCAB, w/o CA, w/o BN and the proposed model, respectively, and (J) shows the reference images.

5.4 EUB and EDB

We design an experiment to explore how much the EUB and EDB can change parameters, FLOPs and FPS on *small-size* and *large-size* images, respectively. The experimental results are shown in Table 10. When upsampling and downsampling operations and equal channels are used simultaneously in Table 10, we use EUB and EDB. In Table 10, when the equal channel operation is used in EUB and EDB, the model has 0.134M parameters. The equal channel operations reduce approximately 3M parameters compared to the model that does not introduce equal channel operation, and it also significantly reduces FLOPs, and increases FPS from 32 to 40 for large-size images. Meanwhile, we find that the network with upsampling and downsampling operations can also speedup the network inference. Although upsampling and downsampling operations increased approximately 0.24M parameters, they can significantly decrease FLOPs and increase the processing speed for both *small-size* and *large-size* images. These results demonstrate that upsampling and downsampling operations have more advantages in dealing with *large-size* images and that the operations of equal channels can make the generator efficient and enable networks to deal with images larger than 960 p (1280×960). Therefore, a network with EUB and EDB is suitable for processing *large-size* images.

5.5 Spiral loops

Figure 15 shows the test performance of the proposed model with different spiral loops. The spiral strategy proposed in Spiral-GAN (Han et al., 2020) is to learn more features of underwater images by repeatedly learning the output of the previous loop, which improves the visual quality of the generated images and the generalization performance of the underwater image enhancement model. From Figures 15B–F, we can see that with only two loops, the model can remove the blue color cast well, but the resulting image has lower contrast and color saturation. As the number of loops increases, although the resulting image has higher contrast, color saturation, and sharper details, the resulting image has a red color cast, as shown in Figure 15E. As shown in Figure 15F, when the number of model loops is set to 10, the red color cast is the resulting image has been corrected and has visually pleasing contrast details.

6 Applications

To further explore the potential of our model for underwater image feature extraction and matching, we first perform a key point matching experiment with SIFT (Lowe, 2004) to verify the ability of our model to recover local features. Figure 16 and Table 12 show that our model can greatly increase performance when key point matching between two underwater pictures compared with CLAHE.

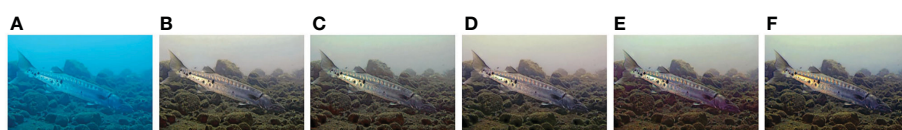


FIGURE 15

Visual results with different spiral loops. From left to right, (A) shows the raw underwater images, and (B–F) are the results of second loops, fourth loops, sixth loops, eighth loops and tenth loops (Ours), respectively.

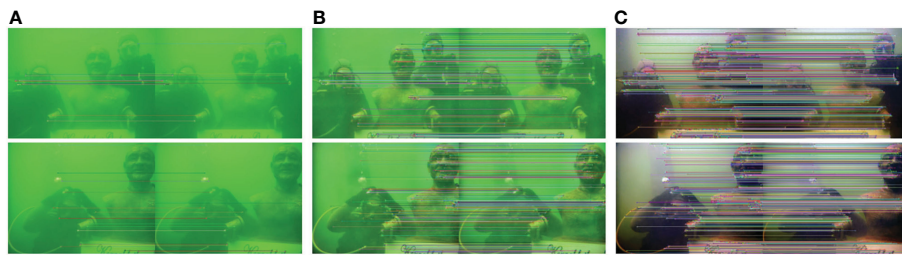


FIGURE 16
The visual results of keypoint matching. From left to right, (A) shows raw underwater images, and (B, C) are the results of CLAHE and the proposed model, respectively.

TABLE 12 The results of keypoint matching.

Method	Original	CLAHE	Ours
picture1	29	269	1193
picture2	25	308	85

To further verify the application of our model in object detection (Bhogale, 2020; Zhu et al., 2021), we use a pretrained YOLOv3 (Redmon and Farhadi, 2018) model on the COCO dataset as our detection model. Although many objects in the underwater scene are not included in the COCO dataset, YOLOv3 can be used to correctly detect humans with the help of our model (Figure 17). We also conduct a Canny (Canny, 1986)-based underwater object edge detection test. Figure 18 shows that more edge features can be generated using the image enhanced by our model than with the Canny operator, indicating that our model is effective in restoring edge features.

7 Conclusion

In this paper, we propose an efficient lightweight underwater image enhancement model for *large-size* underwater images. Our generator is a lightweight encoder and decoder network with RCABs, EUBs and EDBs. We analyze the advantages of upsampling and downsampling operations when dealing with *large-size* images and propose EUB and EDB to handle a real-time *large-size* underwater image enhancement task. The EUB and EDB follow the practical guidelines of ShuffleNetv2 for efficient network design, and the input channels are equal to the output channels of all convolutional layers in the EUB and EDB.

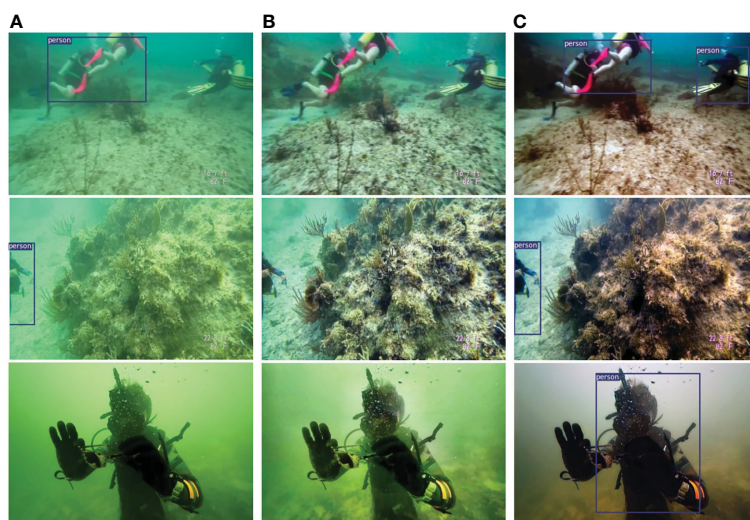


FIGURE 17
Visual results of object detection. From left to right, (A) shows the raw underwater images, and (B, C) are the results of CLAHE and the proposed model, respectively.

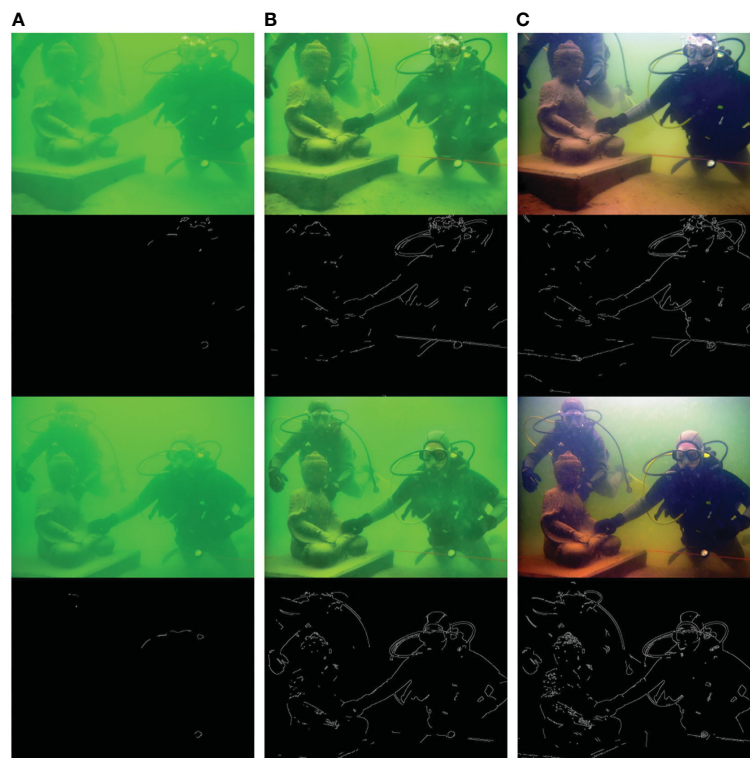


FIGURE 18

The visual results of Canny edge detection. From left to right, (A) shows the raw underwater images, and (B, C) are the results of CLAHE and the proposed model, respectively.

In addition, we introduce a lightweight RCAB to solve the problem of spatial inconsistency. The color transitions of the images generated by the proposed model are more continuous, without obvious boundaries. Qualitative and quantitative tests show that our model is highly efficient and can handle a real-time *large-size* underwater image processing task while maintaining good quality in the enhanced images.

The proposed method can take into account the speed and the quality of the generated image at the same time and can be used as preprocessing for some vision-based tasks that have a demand for high speed, such as SLAM (Yang et al., 2021a) or object detection. We will leave these tasks for future work.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

YG performed the experiments and wrote the manuscript. XL and YW revised the manuscript. ZY provided the ideas and

revised the article. XZ, SZ, and BZ provided advices and GPU devices for parallel computing. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by Hainan Province of China under Grant Number ZDKJ202017, the National Natural Science Foundation of China (Grant No. 62171419) and the Finance Science and Technology Project of and the Natural Science Foundation of Shandong Province of China (Grant No. ZR2021LZH005).

Conflict of interest

Authors XZ and SZ are/were employed by Cofee Medical Technology Company Limited.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer MY declared a past co-authorship with one of the authors BZ to the handling editor.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or

claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmars.2022.964600/full#supplementary-material>

References

- Akkaynak, D., and Treibitz, T. (2019). "Sea-Thru: A method for removing water from underwater images," in 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). (Long Beach, CA, USA: IEEE) 1682–1691.
- Ancuti, C., Ancuti, C. O., Haber, T., and Bekaert, P. (2012). "Enhancing underwater images and videos by fusion," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. New York City, USA: IEEE, 81–88.
- Belghazi, M. I., Baratin, A., Rajeshwar, S., Ozair, S., Bengio, Y., Courville, A., et al. (2018). "Mutual information neural estimation," in *International conference on machine learning* (Birmingham, United Kingdom: PMLR), 531–540.
- Bhogale, K. (2020). Data-free knowledge distillation for segmentation using data-enriching gan. *arXiv preprint arXiv* 2011, 809. doi: 10.48550/arXiv.2011.00809
- Canny, J. (1986). A computational approach to edge detection. *IEEE Trans. Pattern Anal. Mach. Intell.* 6, 679–698. doi: 10.1109/TPAMI.1986.47678514767851
- Carlevaris-Bianco, N., Mohan, A., and Eustice, R. M. (2010). "Initial results in underwater single image dehazing," in *Oceans Mts/IEEE Seattle* (Seattle, USA: IEEE).
- Chen, X., Yu, J., Kong, S., Wu, Z., Fang, X., and Wen, L. (2019). Towards real-time advancement of underwater visual quality with gan. *IEEE Trans. Ind. Electron.* 66, 9350–9359. doi: 10.1109/TIE.2019.2893840
- Fabbri, C., Islam, M. J., and Sattar, J. (2018). "Enhancing underwater imagery using generative adversarial networks," in *Proceedings of the IEEE International Conference on Robotics and Automation*. 7159–7165 8460552 (New York City, USA: IEEE).
- Fu, X., Zhuang, P., Huang, Y., Liao, Y., Zhang, X.-P., and Ding, X. (2014). "A retinex-based enhancing approach for single underwater image," in *International Conference on Image Processing*. 4572–4576 2015A (New York City, USA: IEEE).
- Goodman, J. (2003). "Development of the 960p stereoscopic video format," in *Stereoscopic displays and virtual reality systems X (International society for optics and photonics)* (Washington, USA: SPIE) 5006, 187–194.
- Guo, Y., Chen, J., Wang, J., Chen, Q., Cao, J., Deng, Z., et al. (2020). "Closed-loop matters: Dual regression networks for single image super-resolution," in 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). (Seattle, WA, USA: IEEE) 5407–5416. doi: 10.1109/CVPR42600.2020.00545
- Guo, Y., Li, H., and Zhuang, P. (2019). Underwater image enhancement using a multiscale dense generative adversarial network. *IEEE J. Oceanic Eng.* 45, 862–870. doi: 10.1109/JOE.2019.2911447
- Han, R., Guan, Y., Yu, Z., Liu, P., and Zheng, H. (2020). "Underwater Image Enhancement Based on a Spiral Generative Adversarial Framework," in *IEEE Access*, vol. 8, 218838–218852. doi: 10.1109/ACCESS.2020.3041280
- He, K., Sun, J., and Tang, X. (2010). Single image haze removal using dark channel prior. *IEEE Trans. Pattern Anal. Mach. Intell.* 33, 2341–2353. doi: 10.1109/TPAMI.2010.168
- He, K., Zhang, X., Ren, S., and Sun, J. (2016). "Deep residual learning for image recognition," in 2016 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). (Las Vegas, Nevada, USA: IEEE) 770–778.
- Hore, A., and Ziou, D. (2010). "Image quality metrics: Psnr vs. ssim," in *International Conference on Pattern Recognition*. 2366–2369 (New York City, USA: IEEE).
- Islam, M. J., Xia, Y., and Sattar, J. (2020). Fast underwater image enhancement for improved visual perception. *IEEE Robotics Automation Lett.* 5, 3227–3234. doi: 10.1109/LRA.2020.29747109001231
- Isola, P., Zhu, J.-Y., Zhou, T., and Efros, A. A. (2017). "Image-to-image translation with conditional adversarial networks," in 2017 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). (Honolulu, Hawaii, USA: IEEE) Vol. 1125–1134, 8100115.
- Jenkyns, R., Cheng, J., Kershtien, S., Lavalley, T., Qi, D., Schouten, R., et al. (2015). "Ship to shore: Rov dive logging and data acquisition," in *OCEANS 2015* (Washington: IEEE) 2016, 1–4. doi: 10.23919/OCEANS.2015.7404608
- Kinney, J. B., and Atwal, G. S. (2014). Equitability, mutual information, and the maximal information coefficient. *Proc. Natl. Acad. Sci.* 111, 3354–3359. doi: 10.1073/pnas.1309933111
- Kraskov, A., Stögbauer, H., and Grassberger, P. (2004). Estimating mutual information. *Phys. Rev.* 69 (6), 066138. doi: 10.1103/PhysRevE.69.066138
- Li, C., Anwar, S., Hou, J., Cong, R., Guo, C., and Ren, W. (2021). Underwater image enhancement via medium transmission-guided multi-color space embedding. *IEEE Trans. Image Process.* 30, 4985–5000. doi: 10.1109/TIP.2021.3076367
- Li, C., Anwar, S., and Porikli, F. (2020). Underwater scene prior inspired deep underwater image and video enhancement. *Pattern Recognition* 98, 107038. doi: 10.1016/j.patcog.2019.107038
- Li, C., Guo, J., Guo, C., Cong, R., and Gong, J. (2017). A hybrid method for underwater image correction. *Pattern Recognition Lett.* 94, 62–67. doi: 10.1016/j.patrec.2017.05.023L201762
- Li, C., Guo, C., Ren, W., Cong, R., Hou, J., Kwong, S., et al. (2019). An underwater image enhancement benchmark dataset and beyond. *IEEE Trans. Image Process.* 29, 4376–4389. doi: 10.1109/TIP.2019.29552418917818
- Liu, R., Fan, X., Zhu, M., Hou, M., and Luo, Z. (2020). Real-world underwater enhancement: Challenges, benchmarks, and solutions under natural light. *IEEE Trans. Circuits Syst. Video Technol.* 30, 4861–4875. doi: 10.1109/TCSVT.2019.29637728949763
- Liu, P., Wang, G., Qi, H., Zhang, C., Zheng, H., and Yu, Z. (2019). "Underwater image enhancement with a deep residual framework," (New York City, USA : IEEE), 94614–94629. doi: 10.1109/ACCESS.2019.2928976
- Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision* 60, 91–110. doi: 10.1023/B:VISI.0000029664.99615.94
- Lu, H., Li, Y., Xu, X., Li, J., Liu, Z., Li, X., et al. (2016). Underwater image enhancement method using weighted guided trigonometric filtering and artificial light correction. *J. Visual Communication Image Representation* 38, 504–516. doi: 10.1016/j.jvcir.2016.03.0292016
- Lu, H., Li, Y., Zhang, Y., Chen, M., Serikawa, S., and Kim, H. (2017). Underwater optical image processing: a comprehensive review. *Mobile Networks Appl.* 22, 1204–1211. doi: 10.1007/s11036-017-0863-4
- Ma, N., Zhang, X., Zheng, H.-T., and Sun, J. (2018). "Shufflenet v2: Practical guidelines for efficient cnn architecture design," in *Proceedings of the European Conference on Computer Vision 2018*. (Munich, Germany: IEEE), 116–131. doi: 10.1007/978-3-030-01264-9_8
- Naik, A., Swarnakar, A., and Mittal, K. (2021). "Shallow-uwnet: Compressed model for underwater image enhancement (student abstract)," in *Proceedings of the AAAI Conference on Artificial Intelligence*. (AAAI Press/Palo Alto, California USA) 35, 15853–15854.
- Panetta, K., Gao, C., and Agaian, S. (2015). Human-visual-system-inspired underwater image quality measures. *IEEE J. Oceanic Eng.* 41, 541–551. doi: 10.1109/JOE.2015.24699157305804
- Peng, Y.-T., and Cosman, P. C. (2017). Underwater image restoration based on image blurriness and light absorption. *IEEE Trans. Image Process.* 26, 1579–1594. doi: 10.1109/TIP.2017.26638467840002
- Pizer, S. M., Johnston, R. E., Erickson, J. P., Yankaskas, B. C., and Muller, K. E. (1990). Contrast-limited adaptive histogram equalization: speed and effectiveness. *Proc. First Conf. Visualization Biomed. Computing*. 337. doi: 10.1109/VBC.1990.109340

- Redmon, J., and Farhadi, A. (2018). Yolov3: An incremental improvement. *arXiv preprint arXiv*. doi: 10.48550/arXiv.1804.027672018YOLOv3
- Ren, Y., Wu, J., Xiao, X., and Yang, J. (2021). "Online multi-granularity distillation for gan compression," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*. Held virtually (New York City, USA: IEEE) 6793–6803 ren2021online.
- Ronneberger, O., Fischer, P., and Brox, T. (2015). "U-Net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-assisted Intervention*. 234–241 (Cham, Switzerland: Springer). doi: 10.1007/978-3-319-24574-4282015U
- Shao, Y., Li, L., Ren, W., Gao, C., and Sang, N. (2020). "Domain adaptation for image dehazing," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE/New York City, USA, 2808–2817. doi: 10.1109/CVPR42600.2020.00288
- Shkurti, F., Xu, A., Meghiani, M., Higuera, J. C. G., Girdhar, Y., Giguere, P., et al. (2012). Multi-domain monitoring of marine environments using a heterogeneous robot team. *Proc. IEEE/RISJ Int. Conf. Intelligent Robots Syst.*, 1747–1753 6385685. doi: 10.1109/IROS.2012.6385685
- Sidorov, O. (2020). Artificial color constancy via googlenet with angular loss function. *Appl. Artif. Intell.* 34, 643–655. doi: 10.1080/08839514.2020.17306302019Artificial
- Tan, M., and Le, Q. (2019). "Efficientnet: Rethinking model scaling for convolutional neural networks," in *International Conference on Machine Learning*. 6105–6114 2019EfficientNet (Birmingham, United Kingdom: PMLR).
- Wang, Y., Guo, J., Gao, H., and Yue, H. (2021). Uiec 2-net: Cnn-based underwater image enhancement using two color space. *Signal Processing: Image Communication* 96, 116250. doi: 10.1016/j.image.2021.116250wang2021uiec
- Wang, Y., Huang, H., Xu, Q., Liu, J., Liu, Y., and Wang, J. (2020). "Practical deep raw image denoising on mobile devices," in *Proceedings of the European Conference on Computer Vision*. 1–16 2020Practical (Cham, Switzerland: Springer).
- Wang, Y., Li, N., Li, Z., Gu, Z., Zheng, H., Zheng, B., et al. (2018). An imaging-inspired no-reference underwater color image quality assessment metric. *Comput. Electrical Eng.* 70, 904–913. doi: 10.1016/j.compeleceng.2017.12.006wang2018imaging
- Wang, S., Ma, K., Yeganeh, H., Wang, Z., and Lin, W. (2015). A patch-structure representation method for quality assessment of contrast changed images. *IEEE Signal Process. Lett.* 22, 2387–2390. doi: 10.1109/LSP.2015.2487369wang2015patch
- Whitcomb, L., Yoerger, D. R., Singh, H., and Howland, J. (2000). "Advances in underwater robot vehicles for deep ocean exploration: Navigation, control, and survey operations," in *Robotics research* (Berlin/Heidelberg, Germany: Springer) 439–448. doi: 10.1007/978-1-4471-0765-1_53
- Yang, A. J., Cui, C., Bârsan, I. A., Urtasun, R., and Wang, S. (2021a). "Asynchronous multi-view slam," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. 5669–5676 (IEEE).
- Yang, H.-H., Huang, K.-C., and Chen, W.-T. (2021b). "Laffnet: A lightweight adaptive feature fusion network for underwater image enhancement," in *Proceedings of the IEEE International Conference on Robotics and Automation*. (New York City, USA: IEEE). 685–692. doi: 10.1049/ipr2.120612021LAFFNet
- Yang, M., and Sowmya, A. (2015). An underwater color image quality evaluation metric. *IEEE Trans. Image Process.* 24, 6062–6071. doi: 10.1109/TIP.2015.2491020
- Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B., and Fu, Y. (2018). "Image super-resolution using very deep residual channel attention networks," in *Proceedings of the European Conference on Computer Vision*. (New York City, USA: IEEE), 245, 286–301.
- Zhang, S., Wang, T., Dong, J., and Yu, H. (2017). Underwater image enhancement via extended multi-scale Retinex. *Neurocomputing* 245, 1–9. doi: 10.1016/j.neucom.2017.03.029
- Zhuang, P. (2021). "Retinex underwater image enhancement with multiorder gradient priors," in *2021 IEEE International Conference on Image Processing (ICIP)*. 1709–1713 (IEEE).
- Zhuang, P., Li, C., and Wu, J. (2021). Bayesian Retinex underwater image enhancement. *Eng. Appl. Artif. Intell.* 101, 104171. doi: 10.1016/j.engappai.2021.104171
- Zhu, J.-Y., Park, T., Isola, P., and Efros, A. A. (2017). "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*. (New York City, USA: IEEE), 2223–2232. doi: 10.1109/ICCV.2017.2448237506
- Zhu, B., Song, Q., Yang, L., Wang, Z., Liu, C., and Hu, M. (2021). "Cpm r-cnn: Calibrating point-guided misalignment in object detection," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. (New York City, USA: IEEE) 3248–3257. doi: 10.48550/arXiv.2003.03570