

# UW-CycleGAN: Model-Driven CycleGAN for Underwater Image Restoration

Haorui Yan<sup>ID</sup>, Zhenwei Zhang<sup>ID</sup>, Jing Xu, Tingting Wang, Ping An, Aobo Wang, and Yuping Duan<sup>ID</sup>

**Abstract**—The formation of underwater images is a complex physical process that often suffers from various degradation factors, such as blurriness, low contrast, and color casts, which pose challenges for underwater object detection and recognition tasks. Because of the absence of reference images, learning-based methods that rely on unpaired images have been employed to enhance the underwater images. However, these methods may lose their effectiveness in real-world complex underwater environments. In this article, we propose a model-driven cycle-consistent generative adversarial network (CycleGAN) model, which is inspired by the underwater image formation model to estimate the background light, transmission map, scene depth, and attenuation coefficient directly. Comprehensive experiments have demonstrated that our approach surpasses the compared underwater image restoration methods in both qualitative and quantitative aspects, providing restored images with satisfactory color saturation and brightness. We also conduct experiments on underwater object detection to illustrate the effectiveness of our CycleGAN in improving the detection accuracy. All our source codes and data are available at <https://github.com/Duanlab123/UW-CycleGAN>.

**Index Terms**—Cycle-consistent adversarial learning, image decomposition, model-driven learning, object detection, underwater image restoration.

## I. INTRODUCTION

THE rapid expansion of marine exploration and observation has made underwater image restoration and enhancement crucial for the extensive utilization of marine resources. For instance, underwater autonomous navigation tasks require real-time object recognition and detection based on high-quality underwater images. However, underwater imaging undergoes a complex process resulting in image quality degradation issues, such as color distortion, blurry details,

Manuscript received 10 June 2023; revised 27 August 2023; accepted 8 September 2023. Date of publication 15 September 2023; date of current version 4 October 2023. This work was supported in part by the National Natural Science Foundation of China (NSFC) under Grant 12071345. (Haorui Yan and Zhenwei Zhang contributed equally to this work.) (Corresponding author: Yuping Duan.)

Haorui Yan is with the Center for Applied Mathematics, Tianjin University, Tianjin 300072, China.

Zhenwei Zhang is with the School of Mathematics, North University of China, Taiyuan 030051, China.

Jing Xu is with the School of Statistics and Mathematics, Zhejiang Gongshang University, Hangzhou 310018, China.

Tingting Wang, Ping An, and Aobo Wang are with the China Academy of Aerospace Aerodynamics, Beijing 100074, China.

Yuping Duan is with the Laboratory of Mathematics and Complex Systems, Ministry of Education of China, School of Mathematical Sciences, Beijing Normal University, Beijing 100875, China (e-mail: doveduan@gmail.com).

Digital Object Identifier 10.1109/TGRS.2023.3315772

and low contrast [1]. Therefore, the restoration of underwater images remains a challenging and demanding problem.

Various methods have been developed to alleviate underwater image degradation in the field of underwater image restoration and enhancement [2]. Image restoration is an inverse problem that aims to recover degraded images by employing underwater degradation models. Initially, early methods for image restoration focused on utilizing prior knowledge acquired through statistical analysis or assumptions. He et al. [3], [4] discovered the dark channel prior (DCP) by observing that haze-free images have at least one color channel with low-intensity value in the RGB spectrum and combined it with the image formation model. Many underwater image enhancement techniques based on prior knowledge, such as red channel [5] and NUDCP [6], were developed soon after DCP's discovery. Akkaynak and Treibitz [7], [8] proposed Sea-thru to estimate the parameters of the revised underwater model for better scene recovery. Other attempts have been made to restore underwater images, including regularization [9], [10], prior knowledge [3], [11], [12], and color correction [13], [14], [15]. Although these methods can obtain clear images by utilizing observed priors, traditional restoration approaches are difficult to apply to various degradation factors because they require many model parameters that are difficult to estimate. In addition, most prior-based methods are numerical and computationally expensive. Image enhancement employs subjective qualitative criteria to create visually appealing images without relying on a physical model for image formation, such as Retinex models [16], [17] and some fusion methods [13], [18], [19]. With the rapid evolution of deep learning, numerous learning-based methods have been proposed to address the issue of underwater image restoration and enhancement [20], [21], [22]. These techniques can be broadly classified into convolutional neural network (CNN)-based methods and generative adversarial network (GAN)-based methods.

Due to insufficient training data, CNN-based methods have utilized synthesized underwater datasets for training. Li et al. [23] proposed the underwater image enhancement CNN (UWCNN) based on various water types and degradation levels. The same training set as in [23] was used in [24], where an all-in-one network adversarially learned domain-agnostic features to generate enhanced underwater images from degraded images of ten different water types. As an improvement over [23], Duhane et al. [25] introduced object

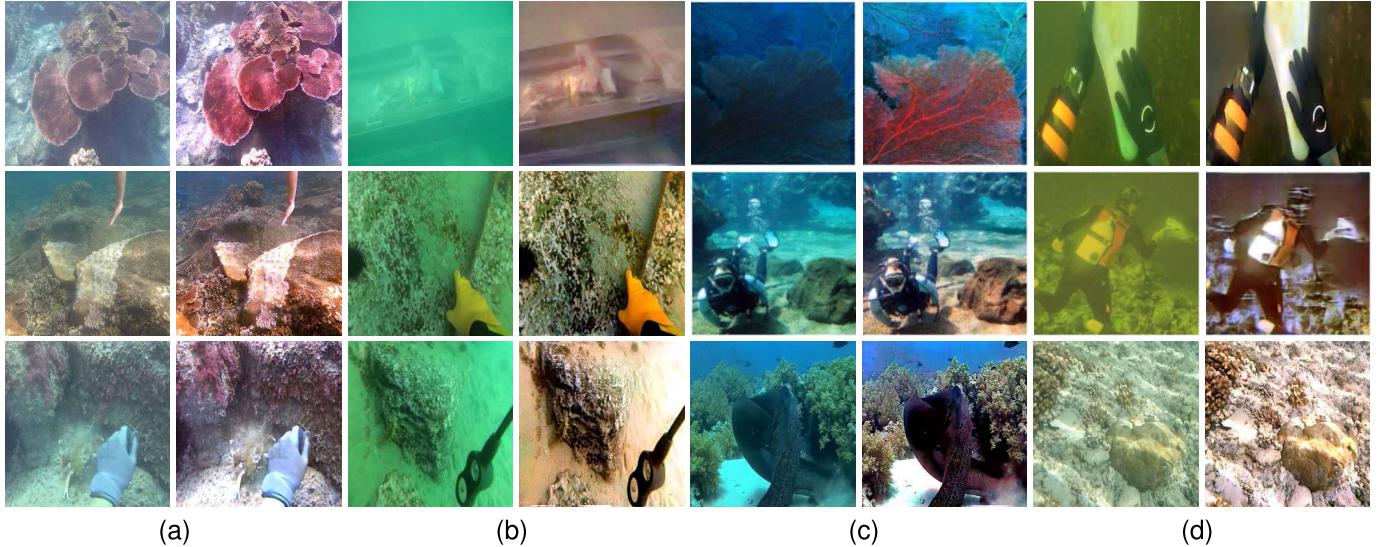


Fig. 1. Restored results of UW-CycleGAN for common underwater degradation. (a) Haze. (b) Green. (c) Blue. (d) Yellow.

blurriness and color shift components to synthesize more precise and underwater-like data. However, solely training on simulation data causes many learning methods to fail when they are applied to real underwater images. To address this issue, Wang et al. [26] proposed a novel two-phase underwater domain adaptation network that simultaneously minimized the gap between synthetic data and real underwater data. Jiang et al. [21] introduced a global-local adversarial mechanism in the reconstruction process to bridge the gap between synthetic and real-world images. To adapt to the real underwater environment, Li et al. [22] proposed the first real-world underwater image benchmark (UIEB) comprising 950 underwater raw images and reference images to meet the demand for sufficient training data. Numerous effective deep learning networks have been developed to enhance underwater images based on the UIEB [22]. Xue et al. [27] introduced the joint luminance and chrominance learning network (JLCL-Net). Wang et al. [28] developed the UIEC<sup>2</sup>-Net, using two color spaces to improve luminance and saturation. Li et al. [29] proposed a network containing multicolor spaces guided by medium transmission to handle color casts and low contrast issues. Qi et al. [30] presented an underwater image co-enhancement network to improve related images in similar environments. Li et al. [31] suggested the comparative learning framework to learn from multiple enhancement reference candidates. Li et al. [32] proposed a human perceptual quality-driven underwater image enhancement framework to achieve better results in human perceptual quality and maintain satisfactory metrics. However, since there still exists a gap between synthetic and real underwater images, the models trained on synthetic data or UIEB [22] often perform unsatisfactorily on other real underwater scenes.

The GAN-based methods [33], [34] improve model generalization abilities by training on unpaired data, thereby proving effective in generating underwater images in an unsupervised manner. The waterGAN [34], which generated synthetic images from RGB-D images and then used them to correct color casts, was a pioneering work. Since then,

many GAN-based methods have been developed to solve the problem of underwater image enhancement. Fabbri et al. [20] utilized cycle-consistent GAN (CycleGAN) directly to generate paired training data, followed by training a fully convolutional encoder-decoder to enhance underwater image quality. Islam et al. [33] designed a fully convolutional conditional GAN-based model for real-time underwater image enhancement known as FUNIE-GAN. Yang et al. [35] introduced a conditional GAN to improve the quality of underwater images. Zhou et al. [36] proposed an unsupervised underwater loop enhancement network (ULENet) to improve turbid underwater images. Based on the underwater image formation model, Chai et al. [37] offered an unsupervised method to estimate the components of the physical image model, such as the scene radiance and the backscatter transmission. Guo et al. [38] used the transformer module to capture global information to train an unsupervised network. Liu et al. [39] developed an object-guided twin adversarial contrastive learning (TACL)-based underwater enhancement method that achieved both visually appealing and task-oriented enhancement. Inspired by transfer learning, Jiang et al. [40] proposed a novel domain adaptation framework for real-world underwater image enhancement that transferred in-air images to real-world underwater image enhancement. To address the problem of joint underwater depth estimation and color correction, Ye et al. [41] suggested an unsupervised adaptation network to handle the joint learning problem, considering it from a novel perspective of style- and feature-level adaptation. Espinosa et al. [42] presented a novel end-to-end deep learning architecture for underwater image enhancement that focuses on solving key image degradation related to blur, haze, and color casts with good inference efficiency. However, most GAN-based methods are data-driven, connecting underwater images with clean images, which lacks the interpretability. Model-driven deep learning methods are well known for their good interpretability and generalization ability.

In this article, we propose a simple yet effective model-driven CycleGAN model for underwater image restoration,

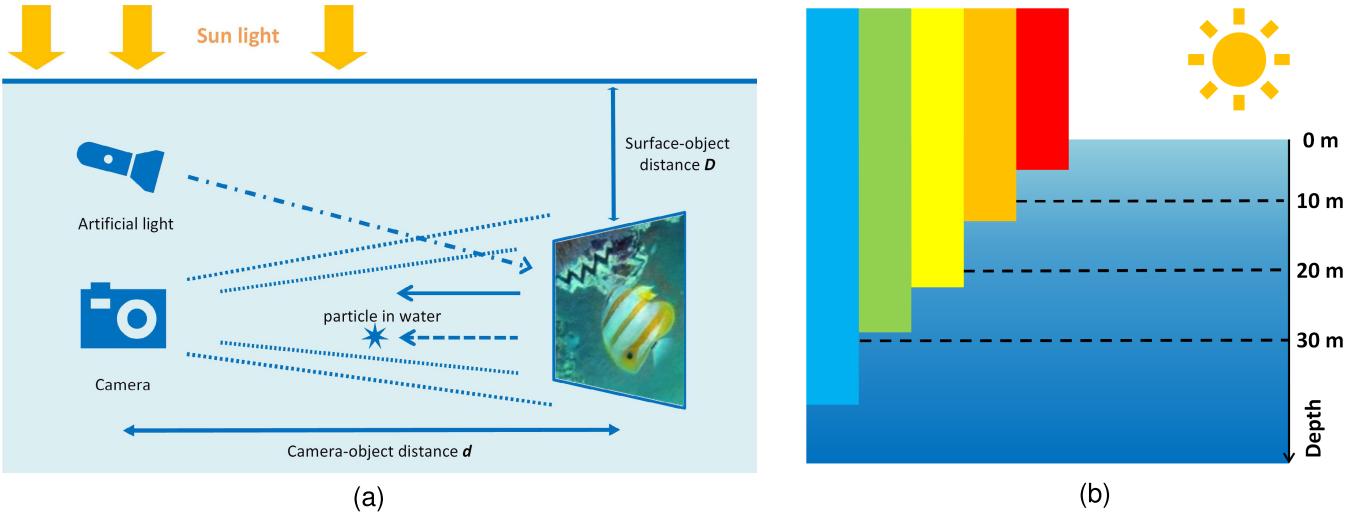


Fig. 2. Underwater image formation. (a) Underwater imaging system. (b) Light attenuation at different water depths.

namely, UW-CycleGAN. Although some GAN models have been developed for underwater image enhancement, most of them ignore the physical properties of real underwater environment leading to the insufficient ability to process real underwater images. To disentangle the complex degradation, we decompose the underwater images as the combination of the clean image, background light, transmission map, scene depth, and attenuation coefficient, which are estimated by the CNN modules in our framework. Then, the complex underwater degradation process is represented as a combination of these subnetworks, which greatly improve the interpretability. More specifically, our UW-CycleGAN model consists of two branches, i.e., degradation–restoration branch and restoration–degradation branch, both of which contain a light estimator, transmission estimator, and depth estimator to evaluate the environmental parameters. The network is jointly trained on unpaired images to generate or capture underwater different light and depth conditions to improve the generalization ability. Fig. 1 shows four typical underwater images and our restoration results, which shows the satisfactory generalization performance in restoring blurred and color distorted underwater images. To sum up, our contributions can be summarized as follows.

- 1) We propose a physics-based underwater image restoration method working on the unpaired data, which estimates the environmental variables rather than the clean images directly to alleviate the ill-posedness existing in the unpaired underwater images.
- 2) Our UW-CycleGAN is self-augmented by generating the underwater images using the estimated environmental variables, including background light, transmission map, attenuation coefficient, and scene depth, which are used to obtain the cycle-consistency loss.
- 3) Extensive experiments are performed on different underwater datasets to demonstrate the effectiveness of our UW-CycleGAN model, which can recover underwater images with saturated color and homogeneous background light.

The remainder of this article is structured as follows. Section II proposes the UW-CycleGAN model based on the

underwater imaging physical model. In Section III, we present numerical experiments by comparing our method with state-of-the-art underwater restoration methods. Section IV provides the application of our model and its limitation. Section V concludes the work and the possible future applications.

## II. MODEL-DRIVEN UNDERWATER CYCLEGAN MODEL

Unlike land photography, underwater images are easily affected by water quality and ambient light, which leads to severe color degradation. Due to the absorption and scattering of light in water, the formation of underwater images becomes a complex physical process. In this article, we propose an accurate physical modeling for underwater images, which is used to guide the development of our UW-CycleGAN model.

### A. Underwater Image Formation Model

The image signal  $I^c$  received by an underwater camera is primarily made up of two components of the direct signal  $D^c$  and the backscatter signal  $B^c$  as follows:

$$I^c = D^c + B^c \quad \text{for } c = r, g, b.$$

The direct signal contains the information about the scene, while the backscatter is an additive signal to degrade the image due to light reflected from particles suspended in the water column. Fig. 2(a) shows an underwater image formation model. The direct signal describes the energy loss of light underwater caused by attenuation. More specifically, homogeneous incident light entering into the water is the major source of illumination in an underwater environment. After penetrating the water depth  $D$ , the incident light with different wavelengths is subjected to varying degrees of attenuation. The attenuation  $\beta^c$  depends on the light wavelength transmitted, where red light is absorbed faster than green and blue, resulting in the blue or green appearance of most underwater images [see Fig. 2(b)]. When penetrating the water of depth  $D$ , the intensity of incident light becomes  $E_S^c e^{-\beta^c D}$  after attenuation. Similarly,  $J^c e^{-\beta^c D}$  is the degraded image at depth  $D$  under the water. Note that the light reflected not only along the water surface to scene propagation path but also along the

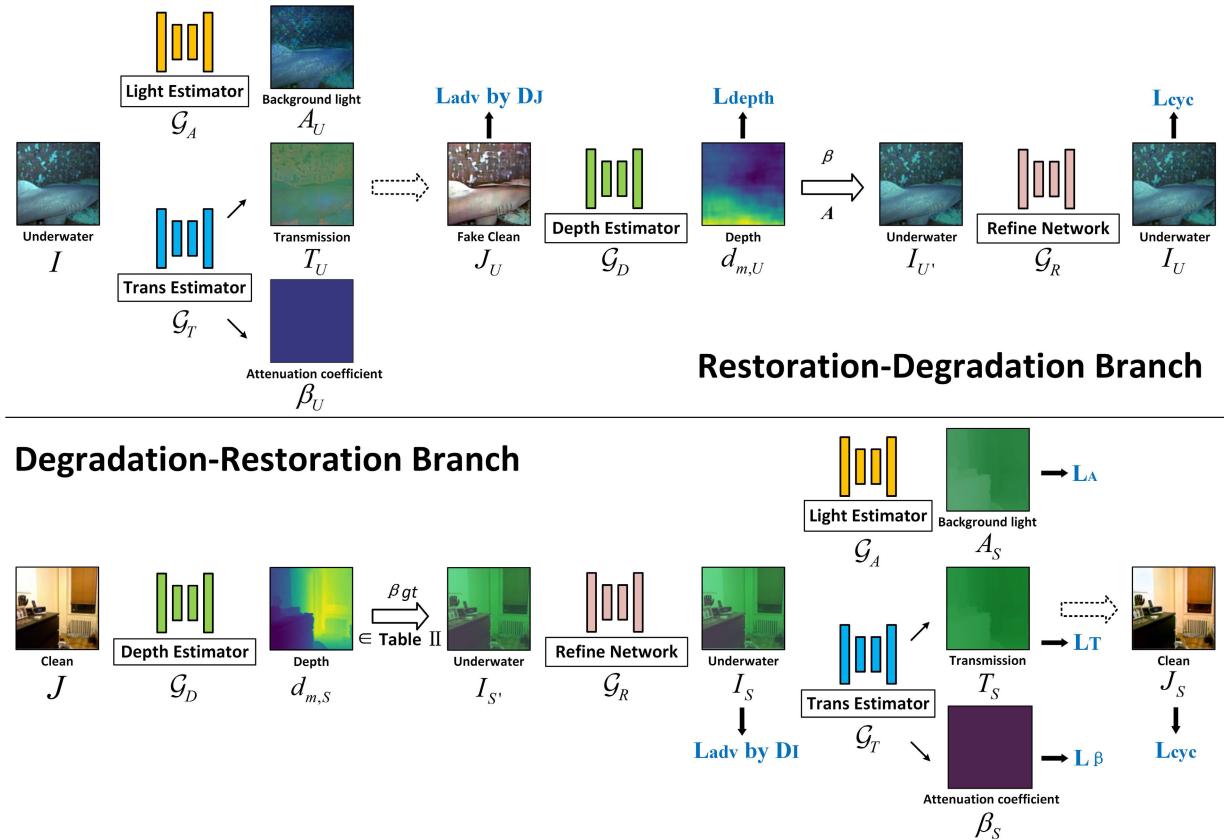


Fig. 3. Illustration of our model-driven CycleGAN structure for unpaired images simulation and restoration.

scene-to-camera route. Therefore, the direct signal captured by the camera can be formulated as follows:

$$D^c = J^c e^{-\beta^c d} e^{-\beta^c d}$$

where  $d$  is the distance from the scene to the camera. Indeed, the transmission map  $t^c$  is defined by  $t^c = e^{-\beta^c d}$ , which represents the percentage of the scene radiance reaching the camera after reflecting in the underwater scene. The backscatter signal refers to the background light  $A^c$  of the surrounding scattered by particles, which can be calculated by

$$B^c = A^c (1 - e^{-\beta^c d}).$$

The background light  $A^c$  is given as  $A^c = \kappa E^c / \beta^c$ , where  $E^c$  is the underwater illumination and  $\kappa$  is a scalar defined by the camera system. The underwater illumination  $E^c$  represents the part of the scene reflected light and ambient light scattered toward the camera by particles in the water. More specifically, the underwater lighting conditions can be considered as follows.

- 1) *Scene Reflected Light*: Based on Beer-Lambert's law, the scene reflected light can be simulated by initial light intensity  $E_P^c$ , light attenuation  $\beta^c$ , and light-traveling distance, which can be written as

$$E_1^c(x) = E_P^c e^{-\beta^c(d(x)+Z)}.$$

The light-traveling distance is approximated by the sum of the scene depth  $d$  and the random distance  $Z$ .

- 2) *Ambient Light*: Due to beam attenuation, different wavelengths of light are attenuated at different water

depths  $D$ , which can be formulated as

$$E_2^c = E_S^c e^{-\beta^c D}.$$

Then, the underwater illumination can be written as follows:

$$E^c(x) = \omega_1 E_1^c(x) + \omega_2 E_2^c$$

where the sum of the weight parameters is subject to 1. Then, the underwater physical formula can be expressed as

$$I^c = J^c e^{-\beta^c D} e^{-\beta^c d} + \kappa E^c (1 - e^{-\beta^c d}) / \beta^c \quad (1)$$

where the background light, wavelength attenuation, scene depth, and water depth are all considered. Given the signal  $I^c$ , our goal is to recover clean image  $J^c$ , which becomes a typical inverse image process.

### B. Network Architecture

For the underwater environments, the desired underwater image and clean image pairs are nearly unreachable. Therefore, the GAN-based methods are used to generate clean images by learning the translation from underwater images to clean images. Among them, the CycleGAN model is a broadly adopted framework to improve the generalization abilities by training on unpaired data. In our framework, the underwater CycleGAN model contains a dewatering network and a rewetting network, which predict the clean images and underwater images from their counterparts, respectively. However, the existing methods often ignore the underwater environmental properties, resulting in a lack of generalization ability in real-world underwater applications.

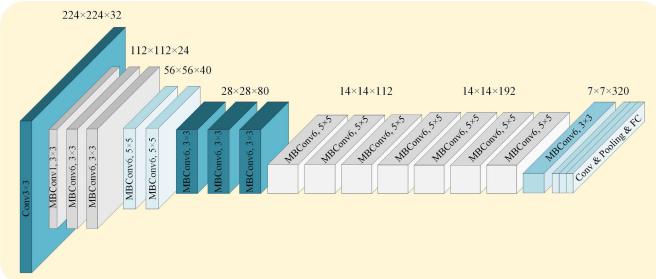


Fig. 4. Backbone of EfficientNet-lite3, which involves 16 MBConv blocks using the linear activation function rather than RELU to retain more feature information.

Thus, we proposed an effective model-driven CycleGAN framework for underwater image restoration, called UW-CycleGAN. Unlike directly establishing the connection between underwater and clean images, our UW-CycleGAN is proposed based on the underwater formation model (1), which can alleviate the ill-posed problems that existed in existing unpaired enhancing methods. As shown in Fig. 3, our UW-CycleGAN method contains two main branches: the restoration–degradation branch and the degradation–restoration branch. The two branches decompose the observed images as a combination of environment variables, which are jointly trained on unpaired images, and the same subnetworks in each branch share the same weights.

*1) Restoration–Degradation Branch:* The restoration–degradation branch performs the process of restoring underwater images and degrading recovered clean images, which follows the cycle of “underwater → clean → underwater.” It learns the process of underwater restoration and degradation by employing four subnetworks, i.e., the transmission estimation network  $\mathcal{G}_T$ , the scene depth estimation network  $\mathcal{G}_d$ , the background light estimation network  $\mathcal{G}_A$ , and the refine network  $\mathcal{G}_R$ , all of which are built up using the structure of EfficientNet-lite3 [43], as shown in Fig. 4. Although the subnetworks are based on the same structure, aiming at different subtasks, different loss functions are used to extract the corresponding features.

a) *Restoration: from underwater image to clean image:*  
 Based on the physical model, the clean image can be calculated by the transmission map and the background light. Given an underwater image  $I^c$ , the transmission estimation network is trained to estimate the transmission map  $t_U^c$  and the attenuation coefficient  $\beta_{II}^c$ , as follows:

$$(\beta_{IJ}^c, t_{IJ}^c) = \mathcal{G}_T(I^c).$$

Similarly, we apply the light estimation network  $\mathcal{G}_A$  to predict the background light of the underwater image  $I^c$  as

$$A_U^c = \mathcal{G}_A(I^c).$$

Then, the restored image can be obtained by calculating the following formula:

$$J_{II}^c = \left( I^c - A_{II}^c (1 - t_{II}^c) \right) / t_{II}^c.$$

The restored  $J_U^c$  is then fed into the discriminator of clean images  $\mathcal{D}_J$  to judge whether it belongs to the clean image domain or not. The discriminators used in our framework are

based on the LSGAN [44] due to its promising stability and visual quality. According to the definition of transmission map, the scene depth  $d_{p,U}$  of the underwater image can be calculated by the physical formula as follows:

$$d_{p,U} = -\ln t_U^c / \beta_U^c. \quad (2)$$

b) *Degradation: from clean image to underwater image:* To obtain the cycle loss of underwater images  $I^c$ , the clean image  $J_U^c$  can be degraded into an underwater image according to the underwater imaging physical model. Before this, the scene depth can be predicted from the depth estimation network  $\mathcal{G}_d$  as

$$d_{m,U} = \mathcal{G}_d(J_U^c).$$

Then, we can obtain the estimated underwater image  $\hat{I}_U^c$  from the image formation model as follows:

$$\hat{I}_U^c(x) = J_U^c(x)e^{-\beta_U^c d_{m,U}(x)} + A_U^c(1 - e^{-\beta_U^c d_{m,U}(x)}). \quad (3)$$

Afterward, the underwater image  $I_U^c$  can be derived from the refine network  $\mathcal{G}_R$  as follows:

$$I_U^c(x) = \mathcal{G}_R(\hat{I}_U^c(x)) \quad (4)$$

where  $I_U^c(x)$  should be consistent with the input underwater image  $I^c$ . Different from previous CycleGAN-based methods [20], [34] that directly synthesize an underwater image from the input clean image, the refine network  $\mathcal{G}_R$  works as an image-to-image translation network that maps the estimated underwater images to underwater images followed the distribution of real underwater images. Our refine networks  $\mathcal{G}_R$  in both restoration-degradation branch and degradation-restoration branch employ the same network architecture and are optimized by the discriminator loss.

2) *Degradation–Restoration Branch*: The degradation–restoration branch performs the process of degrading clean images and restoring simulated underwater images, which follows the cycle of “clean → underwater → clean.”

a) *Degradation: from clean image to underwater image:*  
 We feed a group of the clean images  $J^c$  and the corresponding depth images  $d_{\text{ref}}$  into the depth estimation network. Given the parameters including the attenuation coefficients  $\beta_{\text{ref}}^c$ , the transmission  $t_{\text{ref}}^c$ , and the background light  $A_{\text{ref}}^c$ , we can estimate the degradation images  $\hat{I}_S^c$  of the clean images  $J^c$  based on the image model (1) to cover different underwater degradation situations. Then, the simulated underwater images  $\hat{I}_S^c$  were input to the refined network  $\mathcal{G}_R$  to obtain the simulated images  $I_S^c$ . After that, the discriminator of underwater images  $\mathcal{D}_I$  to identify whether the simulated underwater images  $I_S^c$  belong to the underwater environment.

b) *Restoration: from underwater image to clean image:* To restore the clean image from the simulated image, the proposed framework estimates three physical variables ( $\beta_S^c, t_S^c$ ),  $A_S^c$  from the networks  $\mathcal{G}_T(I_S^c)$  and  $\mathcal{G}_A(I_S^c)$ . Based on the physical model, the clean image can be calculated through the two physical variables of the transmission map and the background light as follows:

$$J_s^c = (I_s^c - A_s^c(1 - t_s^c)) / t_s^c.$$

The scene depth  $d_{p,s}$  can be calculated by the physical formula as  $d_{p,s} = -\ln t_s^c / \beta_s^c$ .

### C. Loss Functions

In our framework, we train the proposed networks together to perform the restoration and degradation cycles. The cycle-consistency loss and the adversarial training loss are employed to penalize the content consistency and data distribution, respectively.

The cycle-consistency loss imposes that the intermediate image transferred from one domain to another should be able to transfer back. In our framework, the degraded image  $I_U^c$  and the restored image  $J_S^c$  should be consistent with their images  $I^c$  and  $J^c$ , respectively, which can be defined as

$$L_{\text{cyc}} = \|I_U^c - I^c\|_1 + \|J_S^c - J^c\|_1.$$

The adversarial learning loss evaluates whether a generated image belongs to a specific domain. In other words, it penalizes the latent recovered images, which should be visually realistic and follow the same distribution as images in training sets. For the networks  $\{\mathcal{G}_T, \mathcal{G}_A\}$  and the discriminator  $\mathcal{D}_J$ , the adversarial loss is given as follows:

$$\begin{aligned} L_{\text{adv}}(\mathcal{D}_J) &= (\mathcal{D}_J(J^c) - 1)^2 + (\mathcal{D}_J(J_U^c))^2 \\ L_{\text{adv}}(\mathcal{G}_T, \mathcal{G}_A) &= (\mathcal{D}_J(J_U^c) - 1)^2. \end{aligned} \quad (5)$$

Here,  $J^c$ ,  $J_U^c$ , and  $\mathcal{D}_J$  are the real clean image, recovered result from  $\{\mathcal{G}_T, \mathcal{G}_A\}$ , and the discriminator of clean image, respectively. The adversarial loss for the underwater refine network  $\mathcal{G}_R$  and the corresponding discriminator  $\mathcal{D}_I$  can be defined as follows:

$$\begin{aligned} L_{\text{adv}}(\mathcal{D}_I) &= (\mathcal{D}_I(I^c) - 1)^2 + (\mathcal{D}_I(I_S^c))^2 \\ L_{\text{adv}}(\mathcal{G}_R) &= (\mathcal{D}_I(I_S^c) - 1)^2 \end{aligned} \quad (6)$$

where  $I^c$  and  $I_S^c$  are the real underwater image and the simulated underwater image, respectively.

The underwater environmental loss contains the loss functions for the physical variables in the degradation–restoration branch, where the reference values of these variables are known in advance.

- 1) The attenuation supervision loss penalizes the difference between  $\beta_{\text{ref}}^c$  and  $\beta_S^c$

$$L_\beta = (\beta_S^c - \beta_{\text{ref}}^c)^2.$$

- 2) The transmission supervision loss minimizes the difference between the predicted result  $t_S^c$  and  $t_{\text{ref}}^c$  by

$$L_T = \|t_S^c - t_{\text{ref}}^c\|_1.$$

- 3) The background light supervision loss optimizes the background light  $A_S^c$  by

$$L_A = \|A_S^c - A_{\text{ref}}^c\|_1.$$

- 4) The depth loss penalizes the difference between the estimated depth and the given depth as follows:

$$\begin{aligned} L_{\text{depth}} &= \frac{1}{2} \|d_{\text{ref}} - d_{p,S}\|_1 + \frac{1}{2} \|d_{\text{ref}} - d_{m,S}\|_1 \\ &\quad + \frac{1}{2} \|d_{p,U} - d_{m,U}\|_1. \end{aligned} \quad (7)$$

Thus, the underwater environmental parameter loss is defined as the combination of the attenuation supervision loss,

TABLE I  
PARAMETERS USED TO GENERATE THE SYNTHETIC  
UNDERWATER IMAGE DATASET

Notation	Description	Range
$D$	Water depth	[5m, 20m]
$d$	Transmission distance	NYU-V1 dataset [48]
$\beta^c$	Attenuation coefficients	Table II
$E_S^c$	Air light intensity	[0.7, 1]
$E_p^c$	Light intensity	[0.7, 1]
$\omega_a, \omega_b$	Weights of lighting	$\omega_a \in [0, 1]$ and $\omega_a + \omega_b = 1$
$\kappa$	Camera system parameter	[0.7, 1.1]
$t^c$	Transmission map	$t^c = e^{-\beta^c d}$
$E^c$	Underwater illumination	$E^c(x) = \omega_1 E_1^c + \omega_2 E_2^c(x)$
$A^c$	Background light	$A^c = \kappa E^c / \beta^c$

the transmission supervision loss, the background light supervision loss, and the depth loss as follows:

$$L_{\text{environ}} = L_\beta + L_T + L_A + L_{\text{depth}}. \quad (8)$$

Finally, we obtain the following weighted loss for our UW-CycleGAN model as follows:

$$L = \lambda_{\text{cyc}} L_{\text{cyc}} + \lambda_{\text{adv}} L_{\text{adv}} + \lambda_{\text{environ}} L_{\text{environ}} \quad (9)$$

where  $\lambda_{\text{cyc}}$ ,  $\lambda_{\text{adv}}$ , and  $\lambda_{\text{environ}}$  are the weights used to balance different terms.

### III. IMPLEMENTATION AND EVALUATION

This section describes the implementation details and experimental settings of our UW-CycleGAN model. We validate the advantages of the cycle structure and the model-driven approach by comparing them with the state-of-the-art traditional and learning-based underwater image restoration methods.

#### A. Synthetic and Real Underwater Image Datasets

1) *Synthetic Underwater Image Datasets*: Due to the insufficient availability of real-world paired underwater images, the CNN-based methods are frequently trained on synthetic underwater images [23], [25]. Li et al. [23] constructed a synthetic underwater image, which consists of nine types of water with each type containing 1000 images randomly selected from the NYU-V2 dataset [45]. Based on the synthetic dataset, Li et al. [23] developed a CNN model to directly recover the underwater images. Unlike previous studies [23], [24], [26], which assumed  $A^c$  to be a globally constant function, we simulate underwater images covering a diverse set of background light and water depth using the underwater imaging physical model (1), which can cover the real underwater environment. We use the NYU-V1 dataset as a benchmark dataset to generate various underwater images, which contains a total of 2284 RGB images and their corresponding depth maps. Therefore, we use the parameters presented in Table I to construct a synthetic underwater image dataset, where ten types of attenuation coefficients are used to cover the diversified types of seawater (see Table II).

TABLE II  
COEFFICIENTS  $e^{-\beta^c}$  VALUES FOR SYNTHESIZING TEN UNDERWATER IMAGE TYPES

Channels \ Types	I	IA	IB	II	III	1	3	5	7	9
R	0.805	0.804	0.830	0.800	0.750	0.750	0.710	0.670	0.620	0.550
G	0.961	0.955	0.950	0.925	0.885	0.885	0.820	0.730	0.610	0.460
B	0.982	0.975	0.968	0.940	0.890	0.875	0.800	0.670	0.500	0.290

2) *Real Underwater Image Dataset*: Underwater image datasets play a crucial role in the development of underwater image processing. However, collecting sufficient underwater images is challenging, resulting in incomplete underwater image datasets. For training, we use the UIEB dataset [22] as input underwater images and the NYU-V1 RGB-D dataset [45] as input clean images. The selected datasets enable better study of the underwater and clean domains. For testing, we conduct experiments on seven publicly available real-world underwater datasets: UIEB [22], EUVP [33], UIEB-60 [22], EUVPUN [33], SQUID [46], U45 [47], and UIQS [48]. Specifically, we use 90 paired images from UIEB and 100 paired images from EUVP for comparison. We randomly select 100 images from EUVPUN, 20 images from UIQS, and all real underwater images of U45 and UIEB-60, which are datasets with no reference.

### B. Quantitative Evaluation

Image quality assessment is a crucial aspect of underwater image enhancement and restoration. Evaluation metrics for image quality can be categorized based on whether a reference image exists or not. For test sets with reference images, we utilize common metrics such as peak signal-to-noise ratio (PSNR), structural similarity (SSIM), and mean square error (mse) to evaluate the quality of our generated clean images for paired underwater images. The higher the PSNR and SSIM, or the lower the mse, the better the quality of the recovered result and the closer it is to the reference image.

Due to the limited availability of reference images for real-world underwater scenarios, underwater images usually do not have reference clean images. Thus, we use two commonly used nonreference metrics to evaluate the performance of underwater results, i.e., underwater image quality measure (UIQM) [49] and underwater color image quality evaluation (UCIQE) [50]. Besides, two recent reference-free underwater image quality assessment metrics CCF [51] and frequency domain underwater image quality assessment metric (FDUM) [52] are used for evaluation. For these metrics, a higher score indicates better human visual perception. More specifically, the UIQM combines underwater image colorfulness measurement (UICM), underwater image sharpness measurement (UISM), and underwater image contrast measurement (UIConM) as follows:

$$\text{UIQM} = c_1 \times \text{UICM} + c_2 \times \text{UISM} + c_3 \times \text{UIConM} \quad (10)$$

where  $c_1-c_3$  are parameters used to balance the three preferences. The UCIQE is designed to quantify nonuniform color cast, blurring, and noise, which is a weighted sum of

chroma ( $\sigma_c$ ), saturation ( $\text{con}_l$ ), and contrast ( $\mu_s$ ) based on CIELab color space as

$$\text{UCIQE} = w_1 \times \sigma_c + w_2 \times \text{con}_l + w_3 \times \mu_s \quad (11)$$

with  $w_1-w_3$  being the constant weights. The CCF combines the colorfulness metric, contrast metric, and haze density metric together to predict the color loss caused by absorption, the blurring caused by forwarding scattering, and the haze caused by backward scattering

$$\text{CCF} = u_1 \times \text{Col} + u_2 \times \text{Con} + u_3 \times \text{Fog} \quad (12)$$

where  $u_1-u_3$  are the constant weights. Similarly, the FDUM is defined as a linear combination of colorfulness, contrast, and sharpness by weighted summation as follows:

$$\text{FDUM} = v_1 \times \text{Col} + v_2 \times \text{Con} + v_3 \times \text{Sharp} \quad (13)$$

where Col, Con, and Sharp indicate the colorfulness, contrast, and sharpness, respectively, and  $v_1-v_3$  are the constant weights.

### C. Implementation Details

The framework is implemented using PyTorch and trained on a Linux workstation equipped with GeForce RTX 3090 GPU. Inspired by [53], we construct the discriminator in the same way, which outputs a binary map instead of a binary value. We employ the Adam optimizer [54] to optimize the proposed networks, with  $\beta_1 = 0.9$  and  $\beta_2 = 0.999$ . The training images are cropped to a size of  $256 \times 256$  to feed into the proposed model. The total epoch is set to 800, and the initial learning rate is set to  $\text{lr} = 1 \times 10^{-4}$ . The learning rate decreases by a factor of 0.8 after every 100 epochs. The training batch size is set to 2.

### D. Comparison Methods

We compare the performance of our proposed UW-CycleGAN with eight state-of-the-art underwater image enhancement or restoration methods on diverse and challenging test datasets. The comparison algorithms include three traditional algorithms: Rank1 [12], ACDC [14], and MMLE [55]; three CNN-based algorithms: Water-Net [22], UWCNN [23], and Ucolor [29]; and two GAN-based algorithms: FUNIE-GAN [33] and TACL [39]. It is worth noting that Water-Net [22], Ucolor [29], and FUNIE-GAN [33] are all retrained on the UIEB dataset. Both the TACL [39] and the proposed UW-CycleGAN are based on CycleGAN, and TACL uses UIEB and BSD500 for training, while UW-CycleGAN uses UIEB and NYU-V1 as training sets. The UWCNN [23] was trained using simulations of multiple water types, due to the lack of training code, we download the publicly trained model.

TABLE III

EVALUATIONS OF COMPARED METHODS ON THE PAIRED TESTING SAMPLES, WHERE ALL COMPARISON METHODS ARE RETRAINED ON THE UIEB DATASET. THE BEST, SECOND BEST, AND THIRD BEST RESULTS ARE DENOTED IN BOLD, UNDERLINED, AND ITALIC, RESPECTIVELY

Datasets Methods	UIEB						EUVP					
	PSNR	SSIM	MSE	UIQM	CCF	FDUM	PSNR	SSIM	MSE	UIQM	CCF	FDUM
ACDC [15]	17.38	0.79	1.44	1.26	26.86	0.43	14.14	0.62	3.09	1.22	32.09	0.42
Rank1 [13]	17.58	0.81	1.31	1.63	22.62	0.61	14.46	0.70	2.92	1.45	27.72	0.56
MMLE [58]	16.71	0.79	1.96	1.48	<u>38.09</u>	<u>0.69</u>	12.76	0.62	3.97	1.54	44.67	0.66
UWCNN [24]	13.05	0.63	4.02	0.89	11.70	0.31	16.06	0.71	2.53	1.28	18.62	0.38
Water-Net [23]	18.78	0.82	1.13	1.73	18.03	0.44	<i>17.46</i>	0.72	1.96	1.69	27.51	0.49
Ucolor [30]	19.44	<u>0.83</u>	0.91	<i>1.74</i>	19.37	0.44	15.54	0.72	1.55	<i>1.70</i>	21.32	0.57
F-GAN [34]	<b>20.50</b>	0.81	0.71	1.69	26.28	0.68	16.58	0.72	2.05	1.56	30.63	0.60
TACL [36]	20.88	0.82	0.71	<u>1.77</u>	21.44	0.56	<u>18.31</u>	<b>0.73</b>	1.48	<b>2.24</b>	27.29	0.67
UW-CycleGAN	<b>21.14</b>	<b>0.83</b>	<b>0.59</b>	<b>2.24</b>	<b>50.10</b>	<b>0.86</b>	<b>18.94</b>	0.73	<b>0.89</b>	2.13	<b>60.77</b>	<b>0.72</b>

TABLE IV

EVALUATIONS OF COMPARED METHODS ON UNPAIRED TESTING SAMPLES, WHERE ALL COMPARISON METHODS ARE RETRAINED ON THE UIEB DATASET. THE BEST, SECOND BEST, AND THIRD BEST RESULTS ARE DENOTED IN BOLD, UNDERLINED, AND ITALIC, RESPECTIVELY

Datasets Methods	U45				UIQS				UIEB-60				EUVPUN			
	UIQM	UCIQE	CCF	FDUM												
ACDC [15]	1.03	0.35	<u>30.65</u>	0.47	0.73	0.34	<u>26.16</u>	0.34	1.03	0.38	19.42	0.33	1.14	0.41	26.98	0.45
Rank1 [13]	1.57	<u>0.48</u>	25.37	<u>0.70</u>	1.51	<u>0.48</u>	22.74	0.58	1.36	0.40	17.05	0.51	1.48	<u>0.42</u>	23.18	<u>0.66</u>
MMLE [58]	1.38	<b>0.49</b>	<b>44.80</b>	<b>0.88</b>	1.58	<b>0.51</b>	<u>36.10</u>	<b>0.66</b>	1.53	0.38	<u>31.64</u>	<u>0.54</u>	1.35	0.31	<u>38.98</u>	<b>0.75</b>
UWCNN [24]	0.68	0.31	14.56	0.33	0.88	0.23	13.99	0.25	0.63	0.29	9.02	0.22	0.66	0.31	12.16	0.31
Water-Net [23]	1.57	0.37	23.26	0.53	1.28	0.42	17.97	0.40	<u>1.63</u>	<u>0.41</u>	12.89	0.35	<u>1.71</u>	0.39	20.34	0.51
Ucolor [30]	1.72	0.39	15.37	0.48	<u>1.69</u>	0.41	14.70	0.36	<u>1.63</u>	0.34	12.88	0.32	1.59	0.39	14.86	0.47
F-GAN [34]	1.45	0.39	25.98	0.61	1.46	0.39	22.91	0.49	<u>1.61</u>	0.38	<u>21.62</u>	<u>0.57</u>	1.44	0.38	24.58	0.57
TACL [36]	1.85	0.42	25.31	0.57	<u>1.59</u>	0.39	21.36	0.46	1.58	<u>0.41</u>	14.14	0.14	<u>1.91</u>	0.43	21.05	0.52
UW-CycleGAN	<b>2.38</b>	<b>0.49</b>	<b>46.86</b>	0.73	<b>2.48</b>	0.48	<b>40.85</b>	0.62	<b>2.10</b>	<b>0.49</b>	<b>47.22</b>	<b>0.73</b>	<b>2.48</b>	<b>0.51</b>	<b>39.15</b>	0.63

### E. Comparison With State-of-the-Art Methods

In this section, we evaluate the proposed UW-CycleGAN model by comparing it with several state-of-the-art underwater image restoration and enhancement methods. We evaluate the effectiveness of our model on both paired and unpaired datasets with varying degrees of quality degradation, such as hazy images and color distortion. For the fairness of comparison, we retrained all the competing methods on the same UIEB dataset.

First, we provide the quantitative comparison results among our method and other compared methods on reference metrics and nonreference metrics in Tables III and IV. As presented, UW-CycleGAN not only exceeds in terms of metrics with references but also obtains high scores on nonreferenced metrics, demonstrating the strengths of our model-driven strategy. Based on the results in Tables III and IV, we can conclude the following observations.

- 1) On the paired datasets UIEB and EUVP, the Ucolor and Water-Net provide satisfactory underwater enhancement effects compared to the traditional methods, i.e., ACDC, Rank1, and MMLE. The two GAN models (FUnIEGAN and TACL) outperform the supervised learning methods. Also, our UW-CycleGAN attains the best PSNR and SSIM, showing the highest similarity with the reference images.
- 2) For the rest unpaired datasets, it can be observed that our UW-CycleGAN has absolute advantages and far exceeds other methods in terms of nonreferenced metrics

(UIQM, UCIQE, CCF, and FDUM), illustrating that our restored images possess the best image quality.

Although the datasets UIEB and EUVP provide reference images for underwater images, these reference images are also generated by image restoration methods. In particular, the reference images of UIEB are the best selected from existing enhancement methods, while the reference images of EUVP are estimated by the GAN model. Thus, both PSNR and SSIM metrics are unable to accurately evaluate the quality of image restoration. The nonreference metrics are more suitable for evaluating underwater images, which emphasize the color saturation, image sharpness, and other information of the recovered images. Obviously, the nonreference metrics in Tables III and IV confirm the advantages of our UW-CycleGAN in processing real-world underwater images such that the combination of the physical model and CycleGAN model can work well in complex underwater environments.

The visual comparisons on paired datasets and unpaired datasets (i.e., UIEB and EUVP) are shown in Figs. 5 and 6, respectively. By observing the visual results, we have the following conclusions.

- 1) The restored images of the traditional methods exhibit obvious overestimated (e.g., MMLE) or inadequate-estimated (e.g., ACDC) color saturation, making them difficult to adapt to different water environments.
- 2) The enhancement effects of Water-Net and Ucolor are limited. Although compared to the aforementioned methods, TACL and FUnIE GAN provide visually

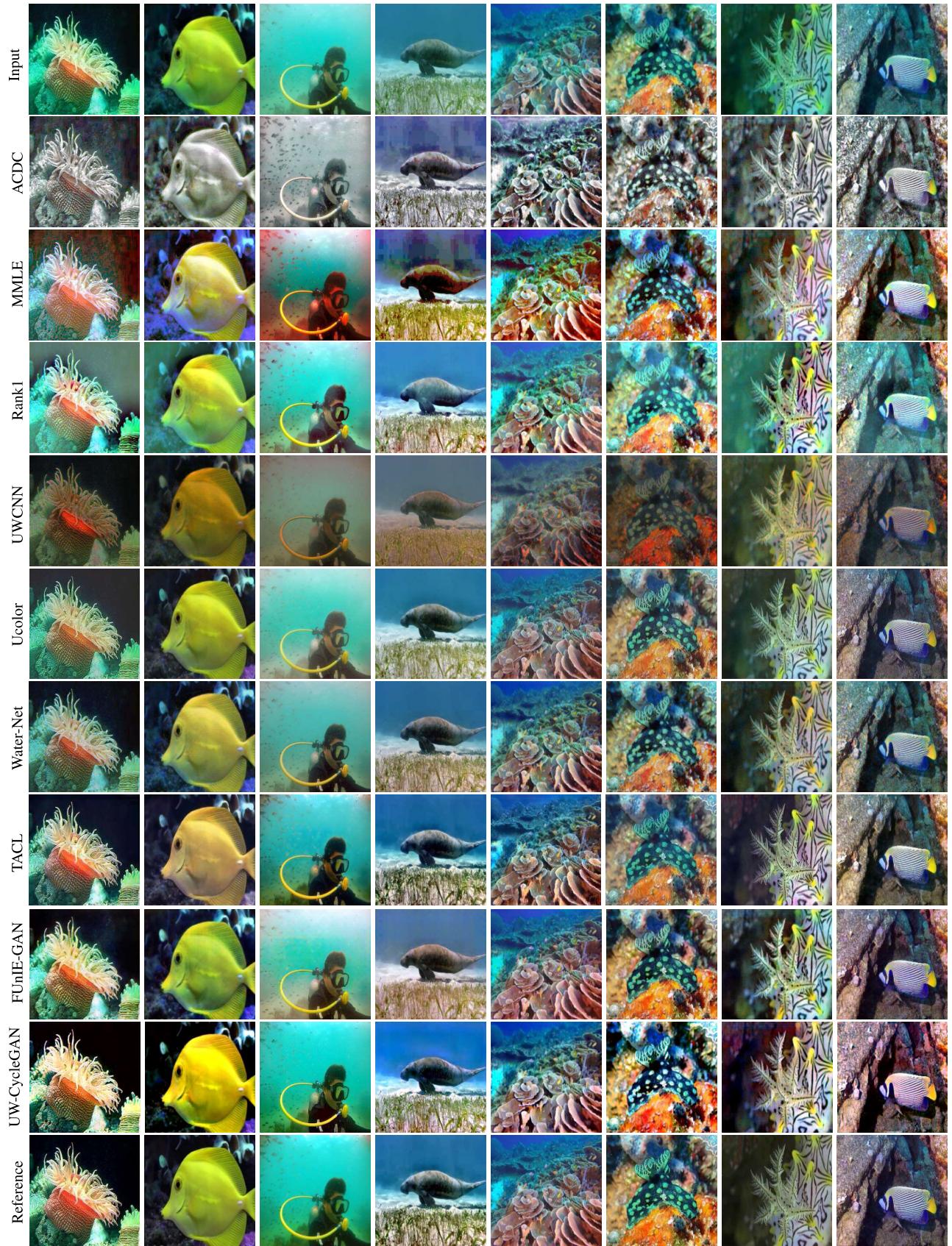


Fig. 5. Visual comparison among the competing methods on paired datasets. The underwater images are listed in the first row, and the reference images are in the last row. Rows 2–9 are the results of comparison methods, and ours are in the second row from the bottom.

satisfying results, they fail to recover the structural details of underwater scenes and introduce artifacts.

3) Our UW-CycleGAN model can successfully recover the clean images from the observed underwater images

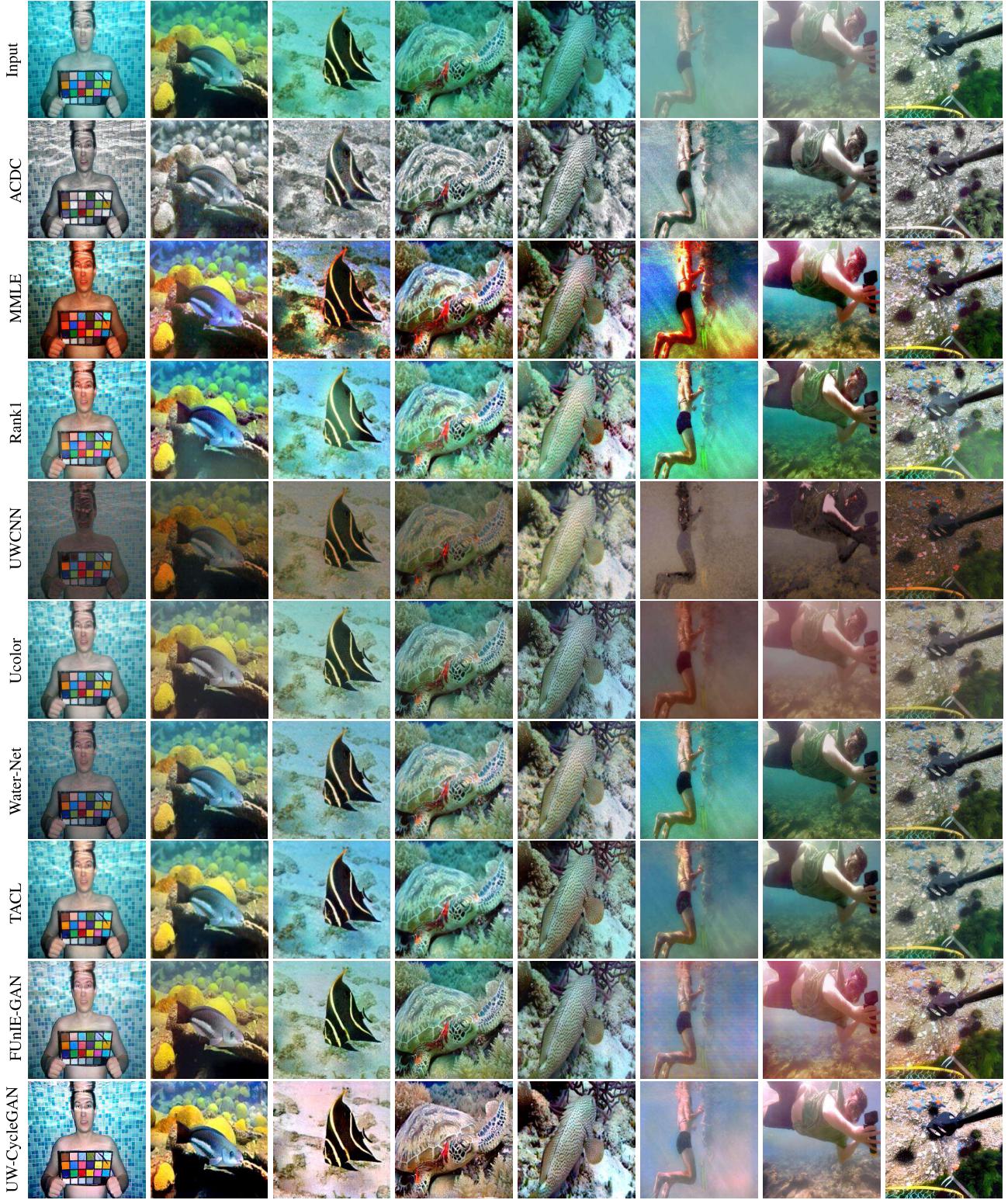


Fig. 6. Visual comparison among the competing methods on unpaired datasets. The underwater images are listed in the first row, rows 2–9 are results of comparison methods, and ours are in the last row.

under various degradation. It provides restored images without loss of color information and with good texture details, demonstrating the effectiveness of the model-driven UW-CycleGAN approach.

Obviously, compared to the generative models, FUnIE-GAN and TACL, our UW-CycleGAN presents better generalization

abilities and achieves better visual performance. For images with color distortion, our model can remove the background light of the given distorted underwater images and obtain images with saturated color, demonstrating the good adaptation ability of our model in dealing with real-world underwater images, which are in accordance with the nonreference

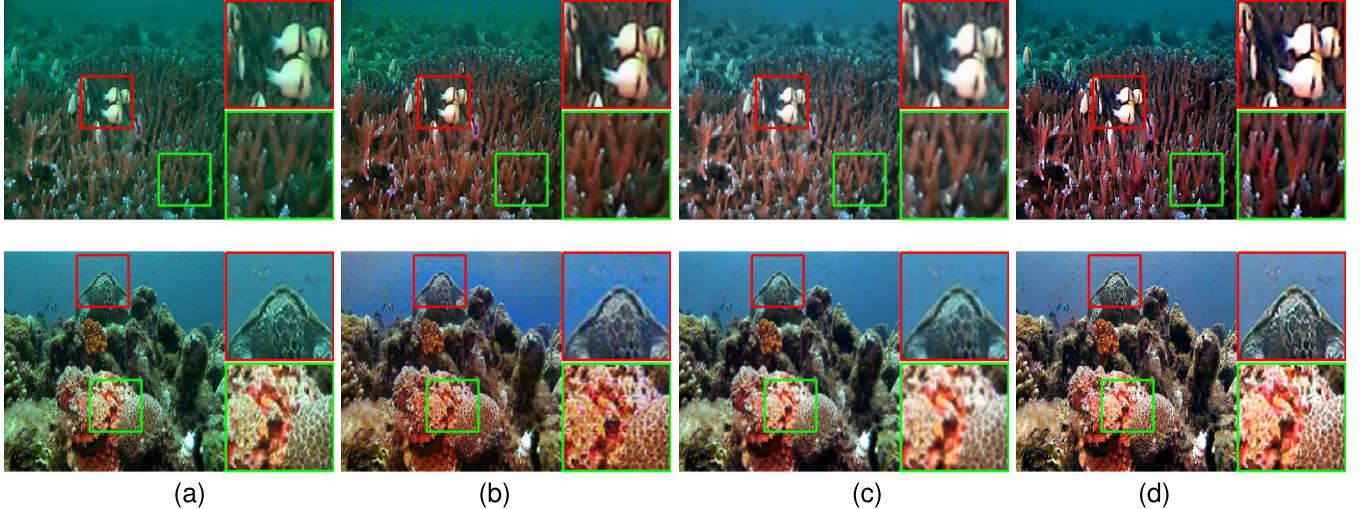


Fig. 7. Visual comparison with FUNIE-GAN [33] and TACL [39]. By enlarging the local details of the above instances, the advantage of UW-CycleGAN in detail processing is demonstrated incisively and vividly. (a) Input. (b) FUNIE-GAN. (c) TACL. (d) UW-CycleGAN.

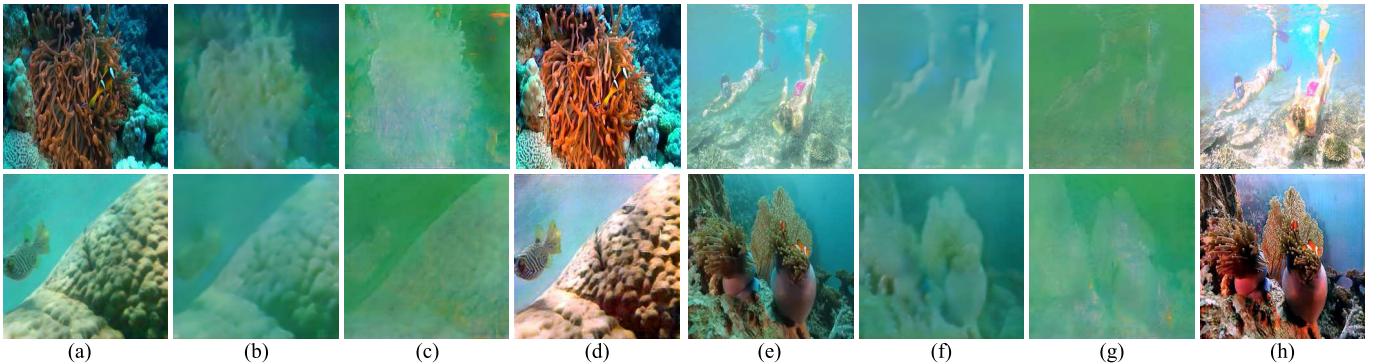


Fig. 8. Results of background light and transmission map. (a) and (e) Input images. (b) and (f) Estimated background light (A). (c) and (g) Transmission map (T). (d) and (h) Restored underwater images (results).

metrics. In summary, UW-CycleGAN demonstrates outstanding generalization capabilities and produces appealing restored images with vibrant-colored reflections on both paired and unpaired datasets, outperforming other methods.

#### F. Comparison With the GAN-Based Methods

To further highlight the benefits of our CycleGAN model, we compare its restoration performance with two GAN-based methods: FUNIE-GAN and TACL. The FUNIE-GAN is a GAN model composed of a generator and a discriminator, while the TACL is a CycleGAN-based model that directly estimates the restored images from underwater images. The visual comparisons in Fig. 7 indicate that although FUNIE-GAN and TACL have some restoration effect, their results still exhibit a veil-like appearance with light green tints and obscuring details. In contrast, our UW-CycleGAN effectively removes the impact of water, even processing local details very well. The FUNIE-GAN result on the second row is affected by artifacts, leading to inhomogeneous background, which is not presented in our results. Due to the physical model used in our framework, our UW-CycleGAN can produce superior results compared to those GAN-based methods that directly estimate clean images.

#### G. Estimation of Background Light and Transmission Map

As our model is a model-driven approach, it has a great dependency on the physical process of underwater imaging. We obtain the improved results by integrating the intermediate components estimated by the subnetworks in accordance with the physical model. The background light plays a crucial role in determining the success of image restoration. Many model-based methods [5], [11], [14], [56] assume that background light is uniform, which contradicts its physical properties and results in inaccurate estimations. Indeed, the background light behaves like a smooth function that varies smoothly with the location. The transmission map represents the proportion of scene radiance that reaches the camera, serving as another essential component of the underwater imaging model. Since the transmission map is a function associated with attenuation coefficients and scene depth, it can also convey scene information.

In Fig. 8, we showcase several examples where the background lights and transmission maps of the input underwater images are estimated by two separate subnetworks. As demonstrated, our model effectively estimates smooth background lights, retaining color information and texture features. In addition, the transmission maps provide insight into the underwater scene information. The accurate estimation

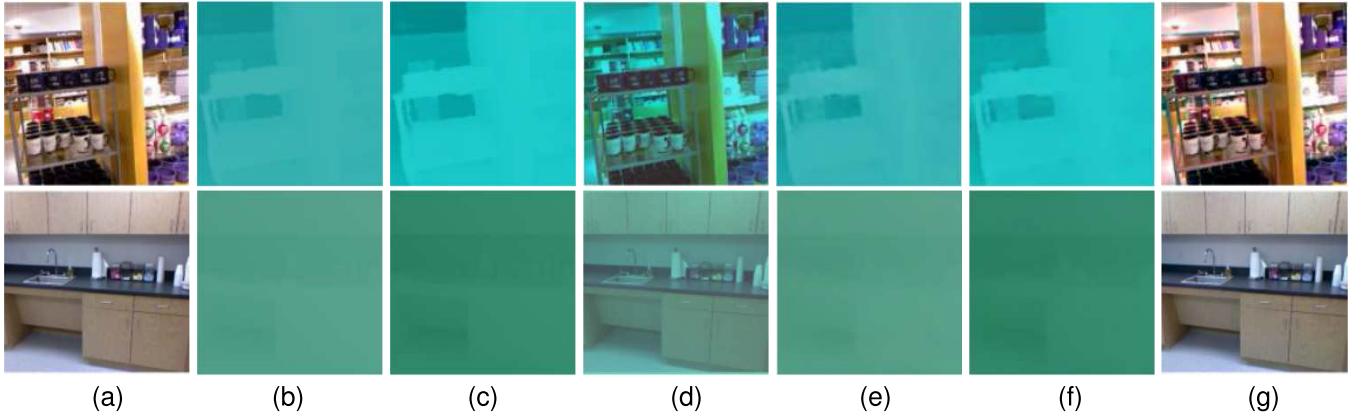


Fig. 9. Underwater image simulation from clean images, where  $I_S$  is the simulated underwater image,  $A_S$  and  $T_S$  are the estimated background light and transmission map, respectively, and  $A_{\text{ref}}$  and  $T_{\text{ref}}$  are the reference of  $A_S$  and  $T_S$ , respectively. (a)  $J$ . (b)  $A_{\text{ref}}$ . (c)  $T_{\text{ref}}$ . (d)  $I_S$ . (e)  $A_S$ . (f)  $T_S$ . (g)  $J_S$ .

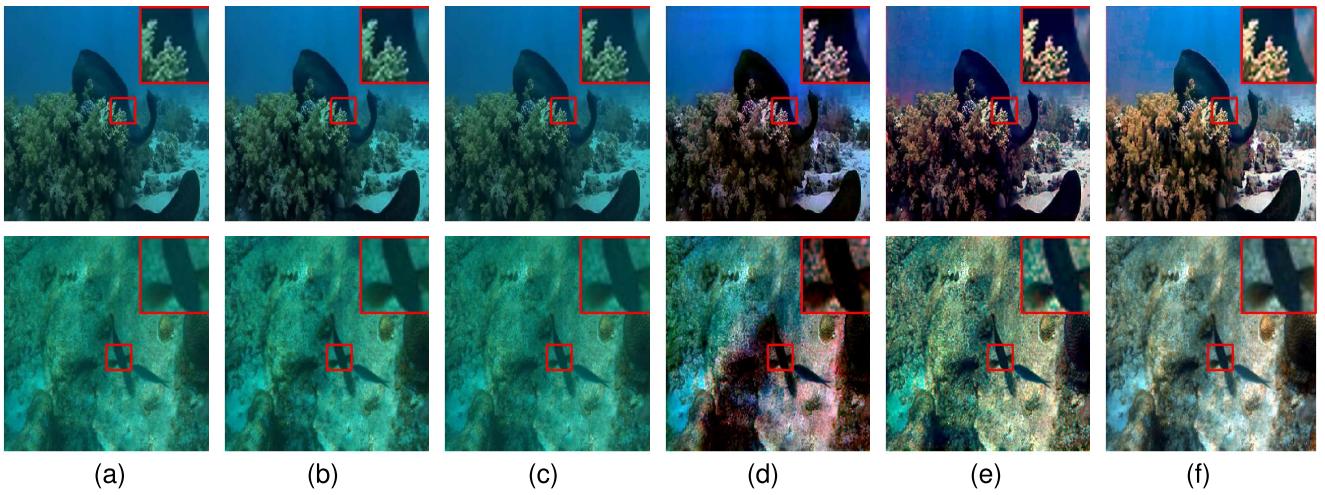


Fig. 10. Visual results of the ablation study. (a) Input. (b)–(f) Restoration results of different models (Model A, Model B, Model C, Model D, and UW-CycleGAN, respectively).

TABLE V  
ABLATION STUDY OF LOSS FUNCTION ON THE UIEB DATASET. THE BEST RESULTS ARE DENOTED IN BOLD

Items	Environmental Loss				Evaluation Metrics					
	$L_{\text{depth}}$	$L_T$	$L_\beta$	$L_A$	PSNR	MSE	SSIM	UIQM	CCF	FDUM
Model A	✗	✗	✗	✗	15.52	2.01	0.59	1.96	25.18	0.42
Model B	✓	✗	✗	✗	17.61	1.86	0.67	2.02	27.30	0.63
Model C	✓	✓	✗	✗	18.24	1.19	0.71	2.06	37.71	0.71
Model D	✓	✓	✓	✗	18.71	0.90	0.75	2.21	39.04	0.71
UW-CycleGAN	✓	✓	✓	✓	<b>21.14</b>	<b>0.59</b>	<b>0.83</b>	<b>2.24</b>	<b>50.10</b>	<b>0.86</b>

of background light and transmission map underscores the efficacy of our model. To further verify the accuracy of the estimated physical variables, we compared the estimated background light and transmission map with the reference images on the simulated underwater images. As shown in Fig. 9, we can observe that the estimations are visually similar to the references, which confirms the effectiveness of the subnetworks  $\mathcal{G}_A$  and  $\mathcal{G}_T$ .

#### H. Ablation Study

In this section, we conduct ablation studies to demonstrate the effectiveness of the environmental supervision loss. Our environmental loss function  $L_{\text{environ}}$  is the linear combination

of the supervision loss function  $L_{\text{depth}}$ ,  $L_T$ ,  $L_\beta$ , and  $L_A$ , which are used to guide the subnetworks to quickly learn and restore simulated images, providing a good foundation for restoring real underwater images. In our ablation experiment, we sequentially introduced the loss function  $L_{\text{depth}}$ ,  $L_T$ ,  $L_\beta$ , and  $L_A$  and train the corresponding network models on the UIEB dataset, named after Models A–D. As can be seen in Table V, the performance of our model is significantly improved by introducing the supervision losses. By gradually introducing the supervised loss function  $L_{\text{depth}}$ ,  $L_T$ ,  $L_\beta$ , and  $L_A$ , all the evaluation metrics are improved step by step. Two typical visual results are shown in Fig. 10, which verified the effectiveness of the transmission estimation network and the

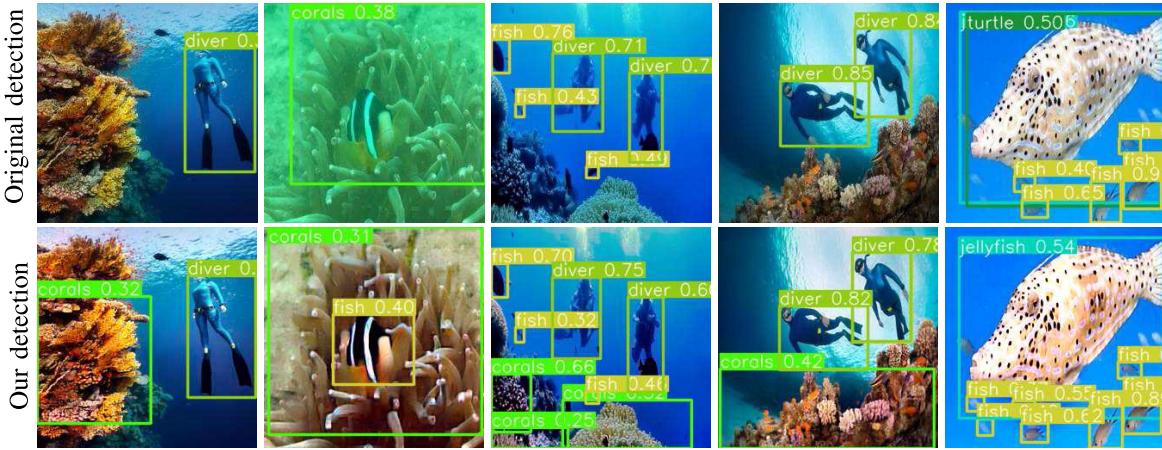


Fig. 11. Visual comparison of underwater object detection on the original images and recovered images of our UW-CycleGAN, where the missing objects are identified using the recovered images.



Fig. 12. Visual comparison of underwater object detection on the original images and recovered images of our UW-CycleGAN, where the reliability of the detection is improved on the recovered images.

background light estimation network in recovering the high-quality images.

### I. Inference Time Analysis

Last but not least, we compare the inference time of our UW-CycleGAN with other learning-based methods on different datasets in Table VI. As presented, the results clearly highlight the efficiency of our UW-CycleGAN, underscoring its practical advantages. The inference time of UW-CycleGAN consumes the least inference time across different datasets, consistently completing the task in less than a second. The remarkable speed positions it as the fastest method among the comparative methods, demonstrating its advantages in real applications.

## IV. APPLICATION AND DISCUSSION

### A. Evaluation on Object Detection Task

Due to various degradation situations in underwater images, object detection methods may fail to accurately identify targets. Due to the ability of our UW-CycleGAN to enhance the quality of underwater images, the accuracy of target recognition methods has been obviously improved. The well-known YOLO [57] has been broadly used for object detection because of its fast speed and high accuracy. Building upon

TABLE VI  
INFERENCE TIME PER IMAGE (S) OF DIFFERENT LEARNING-BASED METHODS ON TEST SETS

Methods	Datasets	UIEB	EUVP	U45	UIEB-60	EUVPUN
UWCNN [24]		1.789	1.705	1.789	1.789	1.691
Water-Net [23]		2.196	2.603	1.634	4.981	4.456
Ucolor [30]		5.741	2.873	2.923	14.951	5.792
F-GAN [34]		2.728	2.138	2.345	7.693	5.792
TACL [36]		1.574	1.893	1.610	2.135	2.087
UW-CycleGAN		<b>0.789</b>	<b>0.783</b>	<b>0.810</b>	<b>0.907</b>	<b>0.858</b>

YOLOv5 [58], the newly introduced YOLOv8 incorporates the refined structures such as a decoupled head, anchor-free, and task alignment learning strategy. We employ the YOLOv8 on the RUOD [59] dataset to validate the effectiveness of our UW-CycleGAN. The RUOD dataset contains a total of 14 000 underwater images of general underwater scenarios, which are classified into ten target categories, i.e., holothurian, echinus, scallop, starfish, fish, corals, diver, cuttlefish, turtles, and jellyfish. We select 100 images from the RUOD dataset with different underwater scenarios to be the test dataset, for which both original underwater images and the recovered images by our UW-CycleGAN are used as the input of YOLOv8 for object detection. The visual performance is

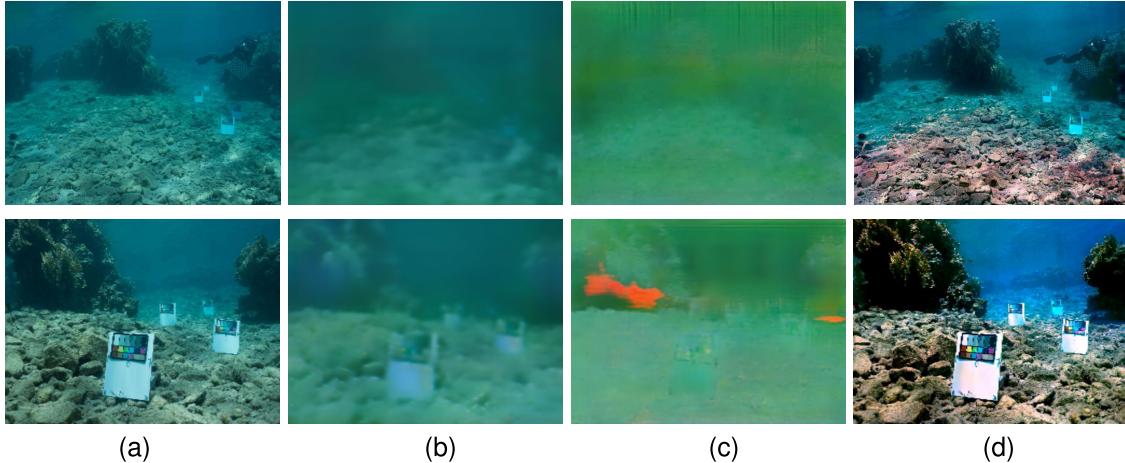


Fig. 13. Illustration of failure examples. (a) Input. (b) Estimated background light (A). (c) Estimated transmission map (T). (d) Restored image (UW-CycleGAN).

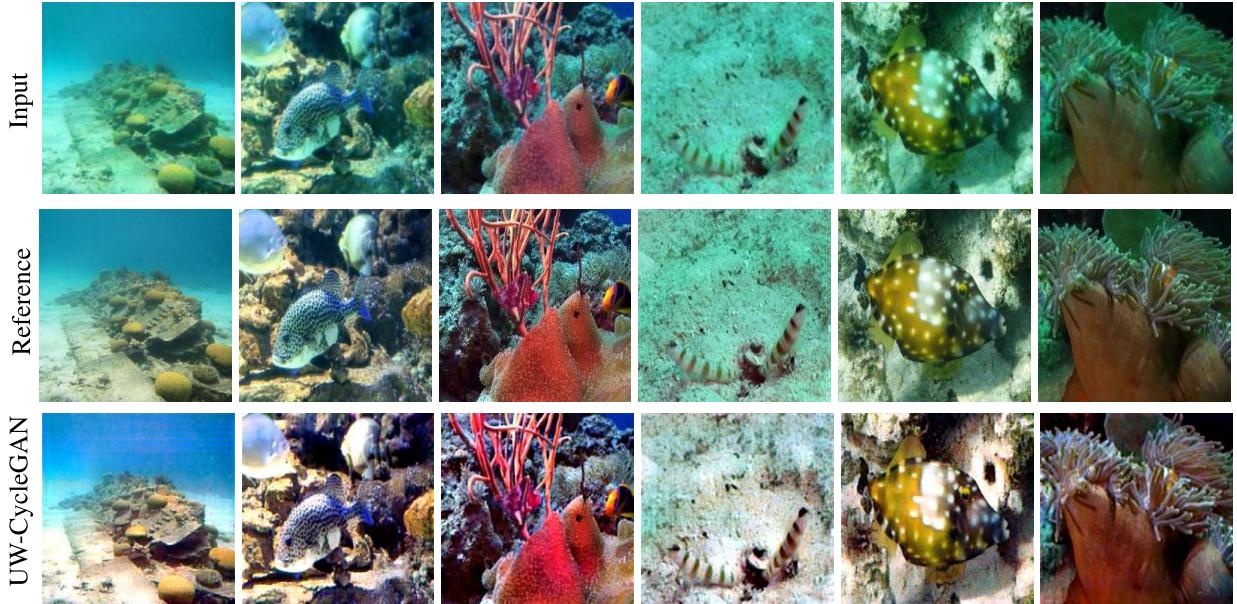


Fig. 14. Comparison of the reference images and recovered results of UW-CycleGAN.

shown in Figs. 11 and 12. Fig. 11 shows the examples that our UW-CycleGAN can help YOLOv8 to identify more objects from the background. On the other hand, Fig. 12 presents the examples that our UW-CycleGAN can improve the confidence of the recognition. Besides, we also list both mean average precision (mAP) and Box accuracy in Table VII, where the mAP represents the mean average precision and Box represents the ratio of the number of predicted bounding boxes to the number of the labeled bounding boxes. It can be seen that the recovered images by our UW-CycleGAN are more conducive to object detection with an average 10% improvement on mAP and 16% improvement on Box, especially echinus, diver, and holothurian gaining the best improvement in mAP, and starfish, fish, and cuttlefish obtaining the best improvement in Box.

#### B. Limitation and Discussion

While UW-CycleGAN can accurately estimate transmission maps and background lights, producing visually appealing clean images with good generalization capabilities across

TABLE VII  
OBJECT DETECTION PERFORMANCE OF THE IMAGES RESTORED BY OUR UW-CYCLEGAN

Class	Images	Original Image		Restored Image	
		mAP(%)	Box(%)	mAP(%)	Box(%)
Holothurian		0.155	0.503	0.408	0.563
Echinus		0.207	0.761	0.703	0.720
Scallop		0.295	0.628	0.310	0.699
Starfish		0.366	0.418	0.484	0.780
Fish		0.233	0.332	0.408	0.663
Corals		0.351	0.480	0.428	0.629
Diver		0.481	0.610	0.772	0.742
Cuttlefish		0.533	0.496	0.780	0.824
Turtle		0.612	0.825	0.859	0.843
Jellyfish		0.345	0.309	0.523	0.518
Overall		0.462	0.529	<b>0.567</b>	<b>0.698</b>

various datasets, its performance is still affected by the underwater environments and may be failed for certain specific images. Fig. 13 presents two examples where our method falls short. In these instances, our model is unable to restore scenes in a long distance, resulting in partially restored outcomes. In fact, it is a common limitation for

TABLE VIII

QUANTITATIVE COMPARISON METRICS OF THE REFERENCE OF PAIRED SETS AND RESULTS GENERATED BY OUR UW-CYCLEGAN

Metrics Datasets	UCIQE	$\sigma_c$	$con_l$	$\mu_s$
UIEB Reference	0.453	0.216	0.910	0.395
UIEB UW-CycleGAN	<b>0.502</b>	<b>0.293</b>	<b>0.956</b>	<b>0.397</b>
EUVP Reference	0.438	0.170	0.837	0.500
EUVP UW-CycleGAN	<b>0.522</b>	<b>0.269</b>	<b>0.972</b>	<b>0.503</b>

underwater image restoration methods, as shown in the first row of Fig. 7. A possible explanation is that our UW-CycleGAN employs adversarial training based on the in-air training set consisting of indoor images with relatively limited scene depth. When our trained model is applied to underwater images, it performs effectively in closer ranges, leaving the farther areas inadequately addressed. Indeed, underwater single-image depth prediction has consistently been a challenging task due to the unstructured and hazardous nature of undersea environments [60].

Another challenge of underwater image restoration is the absence of accurate reference images. The reference images in existing paired datasets, such as UIEB and EUVP, are chosen from the outcomes of the existing underwater image restoration methods. As a result, they may not fully represent the actual underwater scene. In fact, the clean images produced by our method often exhibit better visual quality than the reference images, as shown in Fig. 14 and Table VIII. Concretely, the second row of Fig. 14 highlights the limited accuracy of the reference images. In comparison, our method generates much satisfactory results, restoring true colors and textures. Moreover, our UW-CycleGAN achieves higher UCIQE values compared to the reference images, indicating better restored images with less color deviation and good image contrast.

## V. CONCLUSION

In this article, we introduced a novel model-driven CycleGAN approach for underwater image restoration, which employs two jointly trained branches to decouple the complex underwater degradation process. By integrating the physical formulation with deep networks, our framework disentangles the intricate degradation and trains each subnetwork to focus on a single physical variable. We carried out a series of qualitative and quantitative evaluation experiments, showcasing the satisfactory generalization performance of our proposed model in restoring blurred and color-distorted underwater images. Moreover, we conducted experiments on underwater object detection to validate the positive significance of our CycleGAN in improving the qualities of underwater images. However, our method does have limitations. For instance, the depth estimator may lose effectiveness for long-distance underwater scenes, which subsequently impacts the estimation of background light and transmission maps, rendering our method unable to remove color degradation at long-distance scenarios. In the future, we aim to enhance our model's performance in the aforementioned failure scenarios and explore its applicability

in more practical downstream tasks, including underwater object detection, autonomous underwater vehicle (AUV) navigation applications, and marine trash identification.

## REFERENCES

- [1] M. Jian, X. Liu, H. Luo, X. Lu, H. Yu, and J. Dong, "Underwater image processing and analysis: A review," *Signal Process., Image Commun.*, vol. 91, pp. 116–128, Feb. 2021.
- [2] R. Schettini and S. Corchs, "Underwater image processing: State of the art of restoration and image enhancement methods," *EURASIP J. Adv. Signal Process.*, vol. 2010, no. 1, pp. 1–14, Dec. 2010.
- [3] 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.
- [4] K. He, J. Sun, and X. Tang, "Guided image filtering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 6, pp. 1397–1409, Jun. 2013.
- [5] 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.
- [6] W. Song, Y. Wang, D. Huang, A. Liotta, and C. Perra, "Enhancement of underwater images with statistical model of background light and optimization of transmission map," *IEEE Trans. Broadcast.*, vol. 66, no. 1, pp. 153–169, Mar. 2020.
- [7] D. Akkaynak and T. Treibitz, "A revised underwater image formation model," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 6723–6732.
- [8] D. Akkaynak and T. Treibitz, "Sea-thru: A method for removing water from underwater images," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 1682–1691.
- [9] X. Ding, Z. Liang, Y. Wang, and X. Fu, "Depth-aware total variation regularization for underwater image dehazing," *Signal Process., Image Commun.*, vol. 98, pp. 116–128, Oct. 2021.
- [10] G. Hou, J. Li, G. Wang, Z. Pan, and X. Zhao, "Underwater image dehazing and denoising via curvature variation regularization," *Multimedia Tools Appl.*, vol. 79, nos. 27–28, pp. 20199–20219, Jul. 2020.
- [11] J. Liu, R. W. Liu, J. Sun, and T. Zeng, "Rank-one prior: Real-time scene recovery," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 7, pp. 8845–8860, Jul. 2023.
- [12] J. Liu, R. W. Liu, J. Sun, and T. Zeng, "Rank-one prior: Toward real-time scene recovery," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 14802–14810.
- [13] C. O. Ancuti, C. Ancuti, C. De Vleeschouwer, and P. Bekaert, "Color balance and fusion for underwater image enhancement," *IEEE Trans. Image Process.*, vol. 27, no. 1, pp. 379–393, Jan. 2018.
- [14] W. Zhang, Y. Wang, and C. Li, "Underwater image enhancement by attenuated color channel correction and detail preserved contrast enhancement," *IEEE J. Ocean. Eng.*, vol. 47, no. 3, pp. 718–735, Jul. 2022.
- [15] J. Yuan, Z. Cai, and W. Cao, "TEBCF: Real-world underwater image texture enhancement model based on blurriness and color fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4204315.
- [16] S. Zhang, T. Wang, J. Dong, and H. Yu, "Underwater image enhancement via extended multi-scale retinex," *Neurocomputing*, vol. 245, pp. 1–9, Jul. 2017.
- [17] M. Li, J. Liu, W. Yang, X. Sun, and Z. Guo, "Structure-revealing low-light image enhancement via robust retinex model," *IEEE Trans. Image Process.*, vol. 27, no. 6, pp. 2828–2841, Jun. 2018.
- [18] P. Zhuang, J. Wu, F. Porikli, and C. Li, "Underwater image enhancement with hyper-Laplacian reflectance priors," *IEEE Trans. Image Process.*, vol. 31, pp. 5442–5455, 2022.
- [19] R. Chen, Z. Cai, and W. Cao, "MFFN: An underwater sensing scene image enhancement method based on multiscale feature fusion network," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4205612.
- [20] 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.
- [21] Z. Jiang, Z. Li, S. Yang, X. Fan, and R. Liu, "Target oriented perceptual adversarial fusion network for underwater image enhancement," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 10, pp. 6584–6598, Oct. 2022.
- [22] C. Li et al., "An underwater image enhancement benchmark dataset and beyond," *IEEE Trans. Image Process.*, vol. 29, pp. 4376–4389, 2020.

- [23] C. Li, S. Anwar, and F. Porikli, "Underwater scene prior inspired deep underwater image and video enhancement," *Pattern Recognit.*, vol. 98, pp. 107–138, Feb. 2020.
- [24] P. M. Uplavikar, Z. Wu, and Z. Wang, "All-in-one underwater image enhancement using domain-adversarial learning," in *Proc. CVPR Workshops*, 2019, pp. 1–8.
- [25] A. Dudhane, P. Hambarde, P. Patil, and S. Murala, "Deep underwater image restoration and beyond," *IEEE Signal Process. Lett.*, vol. 27, pp. 675–679, 2020.
- [26] Z. Wang, L. Shen, M. Xu, M. Yu, K. Wang, and Y. Lin, "Domain adaptation for underwater image enhancement," *IEEE Trans. Image Process.*, vol. 32, pp. 1442–1457, 2023.
- [27] X. Xue, Z. Hao, L. Ma, Y. Wang, and R. Liu, "Joint luminance and chrominance learning for underwater image enhancement," *IEEE Signal Process. Lett.*, vol. 28, pp. 818–822, 2021.
- [28] Y. Wang, J. Guo, H. Gao, and H. Yue, "UIEC<sup>2</sup>-Net: CNN-based underwater image enhancement using two color space," *Signal Process., Image Commun.*, vol. 96, pp. 116–126, Aug. 2021.
- [29] C. Li, S. Anwar, J. Hou, R. Cong, C. Guo, and W. Ren, "Underwater image enhancement via medium transmission-guided multi-color space embedding," *IEEE Trans. Image Process.*, vol. 30, pp. 4985–5000, 2021.
- [30] Q. Qi et al., "Underwater image co-enhancement with correlation feature matching and joint learning," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 3, pp. 1133–1147, Mar. 2022.
- [31] K. Li et al., "Beyond single reference for training: Underwater image enhancement via comparative learning," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 33, no. 6, pp. 2561–2576, Jun. 2023.
- [32] M. Li, Y. Lin, L. Shen, Z. Wang, K. Wang, and Z. Wang, "Human perceptual quality driven underwater image enhancement framework," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5634415.
- [33] 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.
- [34] 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.
- [35] M. Yang, K. Hu, Y. Du, Z. Wei, Z. Sheng, and J. Hu, "Underwater image enhancement based on conditional generative adversarial network," *Signal Process., Image Commun.*, vol. 81, pp. 115–123, Feb. 2020.
- [36] W.-H. Zhou et al., "Deep images enhancement for turbid underwater images based on unsupervised learning," *Comput. Electron. Agricult.*, vol. 202, pp. 107–122, Nov. 2022.
- [37] S. Chai, Z. Fu, Y. Huang, X. Tu, and X. Ding, "Unsupervised and untrained underwater image restoration based on physical image formation model," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2022, pp. 2774–2778.
- [38] Z. Guo, D. Guo, Z. Gu, H. Zheng, B. Zheng, and G. Wang, "Unsupervised underwater image clearness via transformer," in *Proc. OCEANS, Chennai*, Feb. 2022, pp. 1–4.
- [39] R. Liu, Z. Jiang, S. Yang, and X. Fan, "Twin adversarial contrastive learning for underwater image enhancement and beyond," *IEEE Trans. Image Process.*, vol. 31, pp. 4922–4936, 2022.
- [40] Q. Jiang, Y. Zhang, F. Bao, X. Zhao, C. Zhang, and P. Liu, "Two-step domain adaptation for underwater image enhancement," *Pattern Recognit.*, vol. 122, pp. 108–114, Feb. 2022.
- [41] X. Ye et al., "Deep joint depth estimation and color correction from monocular underwater images based on unsupervised adaptation networks," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 11, pp. 3995–4008, Nov. 2020.
- [42] A. R. Espinosa, D. McIntosh, and A. B. Albu, "An efficient approach for underwater image improvement: Deblurring, dehazing, and color correction," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. Workshops (WACVW)*, Jan. 2023, pp. 206–215.
- [43] M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proc. Int. Conf. Mach. Learn.*, 2019, vol. 4, no. 8, pp. 6105–6114.
- [44] X. Mao, Q. Li, H. Xie, R. Y. K. Lau, Z. Wang, and S. P. Smolley, "Least squares generative adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2813–2821.
- [45] N. Silberman, D. Hoiem, P. Kohli, and R. Fergus, "Indoor segmentation and support inference from RGBD images," in *Proc. ECCV*, 2012, vol. 7576, no. 5, pp. 746–760.
- [46] D. Berman, D. Levy, S. Avidan, and T. Treibitz, "Underwater single image color restoration using haze-lines and a new quantitative dataset," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 8, pp. 2822–2837, Aug. 2021.
- [47] H. Li, J. Li, and W. Wang, "A fusion adversarial underwater image enhancement network with a public test dataset," 2019, *arXiv:1906.06819*.
- [48] A. Duarte, F. Codevilla, J. D. O. Gaya, and S. S. C. Botelho, "A dataset to evaluate underwater image restoration methods," in *Proc. OCEANS, Shanghai*, Apr. 2016, pp. 1–6.
- [49] 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.
- [50] M. Yang and A. Sowmya, "An underwater color image quality evaluation metric," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 6062–6071, Dec. 2015.
- [51] Y. Wang et al., "An imaging-inspired no-reference underwater color image quality assessment metric," *Comput. Electr. Eng.*, vol. 70, pp. 904–913, Aug. 2018.
- [52] N. Yang, Q. Zhong, K. Li, R. Cong, Y. Zhao, and S. Kwong, "A reference-free underwater image quality assessment metric in frequency domain," *Signal Process., Image Commun.*, vol. 94, May 2021, Art. no. 116218. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0923596521000503>
- [53] 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. 1125–1134.
- [54] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.
- [55] W. Zhang, P. Zhuang, H.-H. Sun, G. Li, S. Kwong, and C. Li, "Underwater image enhancement via minimal color loss and locally adaptive contrast enhancement," *IEEE Trans. Image Process.*, vol. 31, pp. 3997–4010, 2022.
- [56] D. Berman, T. Treibitz, and S. Avidan, "Diving into haze-lines: Color restoration of underwater images," in *Proc. Brit. Mach. Vis. Conf. (BMVC)*, 2017, vol. 1, no. 2, p. 2.
- [57] J. Redmon, S. K. Divvala, R. B. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," *CoRR*, vol. abs/1506.02640, pp. 1–10, Jun. 2015.
- [58] X. Zhu, S. Lyu, X. Wang, and Q. Zhao, "TPH-YOLOv5: Improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshops (ICCVW)*, Oct. 2021, pp. 2778–2788.
- [59] X. Chen, Y. Lu, Z. Wu, J. Yu, and L. Wen, "Reveal of domain effect: How visual restoration contributes to object detection in aquatic scenes," 2020, *arXiv:2003.01913*.
- [60] P. Hambarde, S. Murala, and A. Dhall, "UW-GAN: Single-image depth estimation and image enhancement for underwater images," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–12, 2021.



**Haorui Yan** received the B.S. degree in information and computational science from the Hebei University of Technology, Tianjin, China, in 2021. She is currently pursuing the M.S. degree in mathematics with Tianjin University, Tianjin.

Her research interests include computer vision and image restoration.



**Zhenwei Zhang** received the Ph.D. degree in mathematics from Tianjin University, Tianjin, China, in 2023.

He is currently working with the School of Mathematics, North University of China, Taiyuan, China. His research interests include computer vision and image processing.



**Ping An** received the B.S. degree in communication engineering from Jilin University, Changchun, China, in 2018, and the M.S. degree in electrical and information engineering from Tianjin University, Tianjin, China, in 2021.

She is currently with the China Academy of Aerospace Aerodynamics, Beijing, China. Her research interests include few-shot learning, object detection, and computer vision.



**Jing Xu** received the Ph.D. degree in applied mathematics from the Chinese Academy of Sciences, Beijing, China, in 2007.

She is currently a Professor with the School of Statistics and Mathematics, Zhejiang Gongshang University, Hangzhou, China. Her research interests include interdisciplinary research and medical image processing using artificial intelligence algorithms to assist doctors in diagnosing diseases.



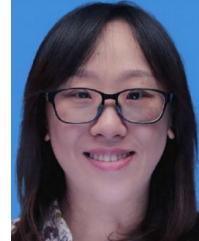
**Aobo Wang** received the M.S. degree from the College of Shipbuilding Engineering, Harbin Engineering University, Harbin, China, in 2015.

He is currently a Senior Engineer of unmanned underwater vehicles (UUVs) with the China Academy of Aerospace Aerodynamics (CAAA), Beijing, China. His research interests include the overall design of UUVs.



**Tingting Wang** received the M.S. degree from the School of Automation Science and Electrical Engineering, Beihang University, Beijing, China, in 2017.

She is currently an Engineer of unmanned underwater vehicles (UUVs) at the China Academy of Aerospace Aerodynamics (CAAA), Beijing. Her research interests include intelligent guidance and navigation, target detection, and recognition-based UUVs.



**Yuping Duan** received the Ph.D. degree from Nanyang Technological University, Singapore, in 2012.

She is currently a Professor with the School of Mathematical Sciences, Beijing Normal University, Beijing, China. Her research interests include numerical optimization, computer vision, and image processing.