



IPMGAN: Integrating physical model and generative adversarial network for underwater image enhancement [☆]

Xiaodong Liu ^{a,*}, Zhi Gao ^b, Ben M. Chen ^{a,c}

^a Department of Electrical and Computer Engineering, National University of Singapore, Singapore 117583, Singapore

^b School of Remote Sensing and Information Engineering, Wuhan University, 430079, China

^c Department of Mechanical and Automation Engineering, Chinese University of Hong Kong, 999077, China

ARTICLE INFO

Article history:

Received 8 February 2020

Revised 10 July 2020

Accepted 13 July 2020

Available online 20 October 2020

Keywords:

Underwater image enhancement

Generative adversarial networks (GANs)

Physical image formation model

ABSTRACT

Autonomous underwater vehicles (AUVs) highly depend on the quality of captured underwater images to perform a variety of tasks. However, compared with everyday images taken in air, underwater images are hazy, with color shift, and in relatively low quality, posing significant challenges to available mature vision algorithms to achieve expected performance. There are, currently, two major lines of approaches to tackle these challenges: the physical image formation model-based and the neural-network-based approaches. In this paper, we propose an integrated approach, where the revised underwater image formation model, i.e., the Akkaynak-Treibitz model, is embedded into the network design for the benefit of combining the advantages of these two approaches. The embedded physical model guides for network learning, and the generative adversarial network (GAN) is adopted for coefficients estimation. We conduct extensive experiments and compare with state-of-the-art approaches quantitatively and qualitatively on nearly all the available underwater datasets, and our method achieves significant improvements.

© 2020 Elsevier B.V. All rights reserved.

1. Introduction

Vision-based autonomous underwater vehicles (AUVs) have become powerful and widely used tools for many tasks such as the ship hull inspection [1], coral reefs monitoring [2], pipeline corrosion inspection [3], seabed mapping [4], and mine hunting [5], etc. To accomplish these tasks, vision algorithms rely heavily on the quality of the underwater images captured by AUVs. However, compared with everyday images taken in air, underwater images are hazy with color shift, and in relatively low quality, making it challenging for available mature vision algorithms to achieve expected performance. Therefore, enhancing the quality of underwater images has become an essential and urgent problem and been attracting more and more attention from many communities.

For the underwater scenario, the marine image quality degradation is mostly caused by wavelength-dependent light absorption and light scattering. In some cases, the red component is with

higher absorption rate, resulting in underwater images appear bluish or greenish, as shown in Fig. 1.

On the other hand, the primary cause of the relatively severe light scattering effect is the larger suspending particles in water, which induce the image to appear blurry with limited visibility [6].

This complex degradation process makes it quite challenging to restore the color and improve the visibility of the underwater images, and the enhancement of underwater images remains an open problem. To improve the quality of underwater images to facilitate the accomplishment of high-level tasks, many pioneering works have been proposed. The majority of these methods could be classified into two lines:

1) **Physical model-based methods.** Some researchers tried to estimate the clear underwater images from the Retinex model [7,8]. However, the Retinex model does not encode the physical description of the image degradation process. Therefore, many researchers utilized the Jaffe-McGlamery image formation model, which models the image degradation process with three components: the direct transmission, forward scattering, and backscatter. Usually, forward scattering is omitted due to its less impact on the image quality. The physical model-based methods, such as those proposed in [9–14], estimate the clear underwater images from the image formation model as Eq. (1), which is considered as the

[☆]The work is supported in part by Peng Cheng Laboratory, and in part by the Key Program of the National Natural Science Foundation of China (Grant No: U1613225).

* Corresponding author.

E-mail addresses: xiaodongliu@u.nus.edu (X. Liu), gaozhi@gmail.com (Z. Gao).



Fig. 1. The original underwater image (left) and the result of IPMGAN (right).

inverse process of image degradation. Due to the ill-posed nature of this problem, assumptions or priors such as the dark channel prior or its variants and haze-line prior have to be introduced for solutions.

However, the aforementioned model is considered as a crude description of the image degradation process and should be further refined. Recently, the authors in [15] proposed the revised underwater image formation procedure as the Akkaynak-Treibitz model. As pointed out in [15], the coefficients in the attenuation and backscatter are distinctly different and cannot be considered the same for simplification. Based on the Akkaynak-Treibitz model, their following up work [16] has shown the improvements and achieved state-of-the-art performance for underwater image enhancement.

2) **Neural-network-based (model-free) methods.** Neural networks have shown the superior capability to learn the latent mapping functions [17,18]. Some researchers have made their attempts to tackle underwater image enhancement with neural networks. Given the synthesized underwater images, the convolutional neural network is trained for small image patches in [19]. Later, Li in [20] introduced the generative adversarial network (GAN) for paired underwater image synthesis based on indoor RGB-D images. Then another CNN is designed and trained for depth estimation and image color correction sequentially based on Euclidean distance. Inspired by the great success of GAN for image translation problems, authors in [21] exploited the conditional generative adversarial network (cGAN) for this task based on the synthesized dataset. To ease the supervised training requirement for the paired dataset, CycleGAN is proposed for image style transformation [22]. Later, it is also introduced to handle the underwater image enhancement in [20,23].

Generally speaking, within the specific datasets, such deep neural-network-based methods outperform the conventional model-based methods due to its powerful learning capability. However, the performance would decrease when the domain gap between the test images and training images is large. What's worse, the network would generate unexpected artifacts, which is detrimental to image quality and the following-up vision tasks. This may be partially due to the end-to-end learning design, which lacks physical model constraints. On the other hand, for model-based methods, to estimate the image attenuation coefficients in the model is non-trivial, where the neural network could be a possible tool to infer parameters and learn the latent related factors. Therefore, it would be quite beneficial to combine these approaches.

In this work, we propose the integrated learning framework for underwater image enhancement. The parameters and coefficients in the image degradation model are learned through the network training to reconstruct the clear underwater images. In addition, the training is under the conditional generative adversarial network framework, where the adversarial loss is helpful in generating genuine restored images with better details. We conduct extensive experiments and compare with state-of-the-art approaches both quantitatively and qualitatively, and our method achieves significant improvements.

The main contributions of our work are summarized as follows:

- 1) We propose a novel physical model integrated network framework for underwater image enhancement based on the Akkaynak-Treibitz physical model [15] and state-of-the-art global and local features fusion net [24]. This design significantly differs from currently existing networks.
- 2) We conduct extensive experiments for performance evaluation quantitatively and qualitatively. Herein, nearly all available competitive methods and datasets have been mentioned and included for comparison, and several major metrics have been utilized for evaluation. Thus, our work can be applied as a milestone and benchmark for the research in this direction. The release of the source code of our method will facilitate future research on this topic.

2. Related works

2.1. Underwater image formation model

The commonly used underwater image formation model is a simplified version, which is analogous to the atmospheric dehazing equation:

$$I_c = J_c e^{-\beta_c z} + B_c^\infty (1 - e^{-\beta_c z}) \quad (1)$$

where the subscript c stands for red, green, blue color channels, I_c is the captured underwater image, J_c is the unattenuated image (i.e. the goal of image restoration). B_c^∞ is the veiling light, β_c stands for the attenuation (including absorption and scattering) coefficients, z is the imaging range. From the view of Jaffe-McGlamery model, the first term $J_c e^{-\beta_c z}$ denotes the direct transmission from the scene radiance. The second term $B_c^\infty (1 - e^{-\beta_c z})$ refers to the backscatter. However, the work [15] showed that using this model to estimate

J_c for recovery may introduce significant errors, where the attenuation coefficients in direct transmission and those in the backscatter are wrongly considered to be the same β_c .

In fact, the coefficients in direct transmission and backscatter are distinct and dependent on different factors. Theoretically, β_c^b and β_c^d are different (the superscript b and d are used to denote for backscatter and direct transmission) and can be expressed as Eqs. (2) and (3) [15]:

$$\beta_c^b = -\ln\left(1 - \frac{B_c(z)}{B_c^\infty}\right)/z \quad (2)$$

$$\beta_c^d = -\ln\left(\frac{I_c(z) - B_c(z)}{J_c}\right)/z \quad (3)$$

$B_c(z)$ is the backscattered signal at distance z . Taking this difference into consideration, the Akkaynak-Treibitz model can be expressed as [15]:

$$I_c = J_c e^{-\beta_c^d(v_d)z} + B_c^\infty \left(1 - e^{-\beta_c^b(v_b)z}\right) \quad (4)$$

where $v_d = \{z, \rho, R, S_c, \beta\}$ and $v_b = \{R, S_c, s, \beta\}$ are the dependent coefficients for β_c^d and β_c^b , respectively. ρ is the scene reflectance, R is the spectrum of ambient light, S_c is the camera sensor spectral response, and s and β are the physical scattering coefficient and beam attenuation coefficient of the water body, respectively.

Though the physical model in Eq. (4) has shown to be the most accurate physical model, the parameter estimation still requires modeling and data fitting work [16]. The authors in [16] suggested employing the supreme learning ability of neural networks for parameter estimation. Therefore, in this paper, we will design the network based on the Akkaynak-Treibitz physical model, and the components and coefficients are adaptively estimated by neural network modules.

2.2. Generative Adversarial Network (GAN)

In [25], GAN is designed to produce realistic-looking images given the random noise vector. GAN trains a generator network G and a discriminator network D via an adversarial process. The training process alternately optimizes G and D , which compete against each other. The generator G is trained to produce samples to fool the discriminator D , and D is trained to distinguish real from fake images provided by G . The adversarial loss function for this mini-max game is derived from the standard cross-entropy cost:

$$\mathcal{L}_{\text{GAN}} = \min_G \max_D \frac{1}{2} (E_{J_c} \log D(J_c) + E_{I_c} \log(1 - D(G(I_c)))) \quad (5)$$

Eq. (5) is straightforward for mathematical analysis, but it can not guarantee stable results, causing the network to suffer from the mode collapse problem. Therefore, many modifications and improvements on the adversarial loss have been developed, such as the least square GAN (LSGAN) [26], Wassertein GAN (WGAN) [27], and Wassertein GAN with gradient penalty (WGAN-GP) [28] which further enforced the gradient penalty of the discriminator to satisfy the Lipschitz constraint for better performance.

Different from the basic GAN, the conditional GAN takes conditional variables as input and has been demonstrated beneficial for image-to-image translation problems, such as night-to-day, edge-to-object translation tasks [29]. Similarly, underwater image enhancement could also be formulated as an image-to-image translation problem. Therefore, for the underwater image enhancement task, the attenuated underwater image I_c is fed into the network as conditional information, as shown in Fig. 2.

2.3. Neural-network-based underwater image enhancement

Due to the great success of the deep convolution neural networks (CNNs) in many computer vision tasks, some researchers have attempted to apply the neural network to handle the underwater image enhancement task. Wang et al. [19] trained the CNN for underwater image color correction for small image patches 32×32 . Later, based on the encoder-decoder network structure, Li et al. in [20] proposed the two-stage CNN for depth estimation and color correction sequentially.

To generate more realistic images, the adversarial loss and GAN framework is introduced recently. In [21], the adversarial loss and L_1 distance loss were adopted to train the GAN, given the paired underwater images. In [30], the cycle consistency loss was introduced into the GAN framework, and the network was trained on the unpaired dataset. Lu et al. in [23] added the structural similarity metric (SSIM) loss for end-to-end training on the unpaired dataset. The con-current work, which is similar to ours, is presented in [31]. The authors in [31] incorporated the physical model into the network design as well. However, there still exist significant differences. First of all, the authors in [31] still utilized the simplified model instead of the revised one. Secondly, in our work, the components in the model are estimated via state-of-the-art global and local feature fusion network structure [24]. The comparison is made and presented in Section 4.2, demonstrating the advantages of our design.

3. Our method: IPMGAN

The overview of the proposed method is shown in Fig. 3. The Akkaynak-Treibitz physical image degradation model is considered as the guidance to design the generator. The discriminator is trained to distinguish the ground-truth images and the restored ones.

3.1. Network structure

3.1.1. The revised physical model-based generator

Based on Eq. (4), the clear underwater image J_c can be referred as:

$$J_c = \frac{I_c - B_c^\infty \left(1 - e^{-\beta_c^b(v_b)z}\right)}{e^{-\beta_c^d(v_d)z}} \quad (6)$$

To simplify the notation, we denote $T = e^{-\beta_c^d(v_d)z}$ and $S = e^{-\beta_c^b(v_b)z}$. Therefore, J_c can be reconstructed as:

$$\hat{J}_c = \frac{I_c - \hat{B}_c^\infty \left(1 - \hat{S}\right)}{\hat{T}} \quad (7)$$

where we denote \hat{T} and \hat{S} as the estimations of the transmission map ($e^{-\beta_c^d(v_d)z}$) and the scattering map ($e^{-\beta_c^b(v_b)z}$). \hat{B}_c^∞ denotes the estimation of the veiling light. Therefore, to obtain the estimation of the clear underwater image \hat{J}_c , \hat{B}_c^∞ , \hat{T} , \hat{S} are required. To this end, the generator is designed with three branches to estimate the unknowns, respectively.

To obtain \hat{T} and \hat{S} , the multi-scale local feature and global feature extraction and fusion network [24] is adopted. This network structure is based on the encoder-decoder architecture, where the global features are dynamically fused to multi-scale local features with fusion units, and the hybrid features are fed to the corresponding decoder layer with skip connections. The hybrid features encoding both the low-level knowledge (such as edges or corners) from the local features and the semantic knowledge (such as the layout of the scene) from the global features, are

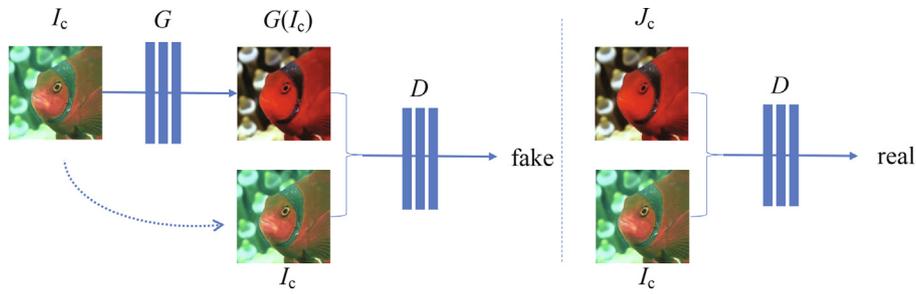


Fig. 2. cGAN for underwater image enhancement. The generator G is trained to enhance the underwater images I_c . The discriminator D is trained to classify between the enhanced images $G(I_c)$ and the ground-truth image J_c . Unlike the basic GANs, both the generator and discriminator will observe the input underwater image I_c .

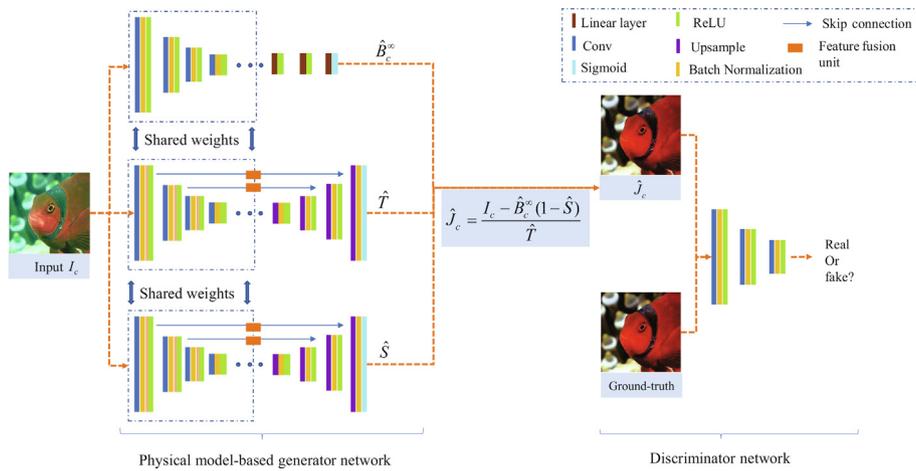


Fig. 3. The overview of the IPMGAN network structure which integrates the Akkaynak-Treibitz model under the generative adversarial network framework.

beneficial for the estimation of the transmission map and scattering map. Since both T and S are in the range of $(0, 1)$, the last activation layer in the network is replaced by Sigmoid layer.

For the estimation of veiling light \hat{B}_c^∞ , the up-sample layers in the decoder are modified as the fully-connected layers. Similarly, the Sigmoid is utilized as the range of B_c^∞ is within $(0, 1)$.

As all branches involve the process of feature extraction and encoding, therefore, the weights for the encoder of all three branches are shared for memory efficiency. The overview of the network is shown in Fig. 3. Details of the data flow and generation in the three branches of the generator is depicted in Fig. 4.

Input and output of the generator network.

In Fig. 4, the input is the original underwater image, outputs are the estimated scattering map, transmission map, and veiling light by the three branches, separately. In this paper, all the training and test images are resized to 256×256 due to memory limitations. Therefore, the input dimension is $256 \times 256 \times 3$. The scattering map S and transmission map T are of the same size as the input image: $256 \times 256 \times 3$. The veiling light B_c^∞ is of 1×3 , which represents different values for three color channels.

Data flow and connections.

The shared encoder consists of 8 convolution layers with stride = 2, and the final global feature vector with dimension $1 \times 1 \times 512$ is passed to scattering map branch, transmission map branch, and the veiling light branch simultaneously. Scattering map branch and transmission map branch consist of 8 upsample-convolution layers, and veiling light branch consists of three linear layers. All the parameters in the convolution and linear layers are provided in Fig. 4. Besides, the shared encoder is also connected with the scattering map branch and transmission map branch with feature fusion units via skip connection at each

resolution, as the hybrid features encoding both global and local features are beneficial to the pixel-to-pixel estimation tasks [24]. More details about the feature fusion unit can be found in [24].

3.1.2. The discriminator

The discriminator we used is based on the 70×70 PatchGAN [29], which classifies the 70×70 image patches as real or fake. Then, the judgment for the whole image is obtained based on the average of all patches. Compared to the conventional full-image discriminator, the patch-based setting gives more attention to the high-frequencies content in the image with fewer parameters, which is favored by many researchers.

3.2. Objective functions

The training objective is to minimize the distances between the ground-truth and the reconstructed images. In this work, we measure the distance between the ground-truth and reconstructed images in three levels: i.e., pixel level, structure level, and image distribution level.

To measure the per-pixel distance, the L_1 distance loss is adopted:

$$L_{L_1}(J_c, G(I_c)) = E_{I_c, J_c} [\|J_c - G(I_c)\|_1] \tag{8}$$

where I_c is the input image, $G(I_c)$ is the reconstructed image, J_c stands for the ground-truth image.

To measure the structural difference between $G(I_c)$ and J_c , the structural similarity for each patch is calculated as:

$$SSIM(p) = \frac{2\mu_a\mu_b + C_1}{\mu_a^2 + \mu_b^2 + C_1} \frac{2\sigma_{ab} + C_2}{\sigma_a^2 + \sigma_b^2 + C_2} \tag{9}$$

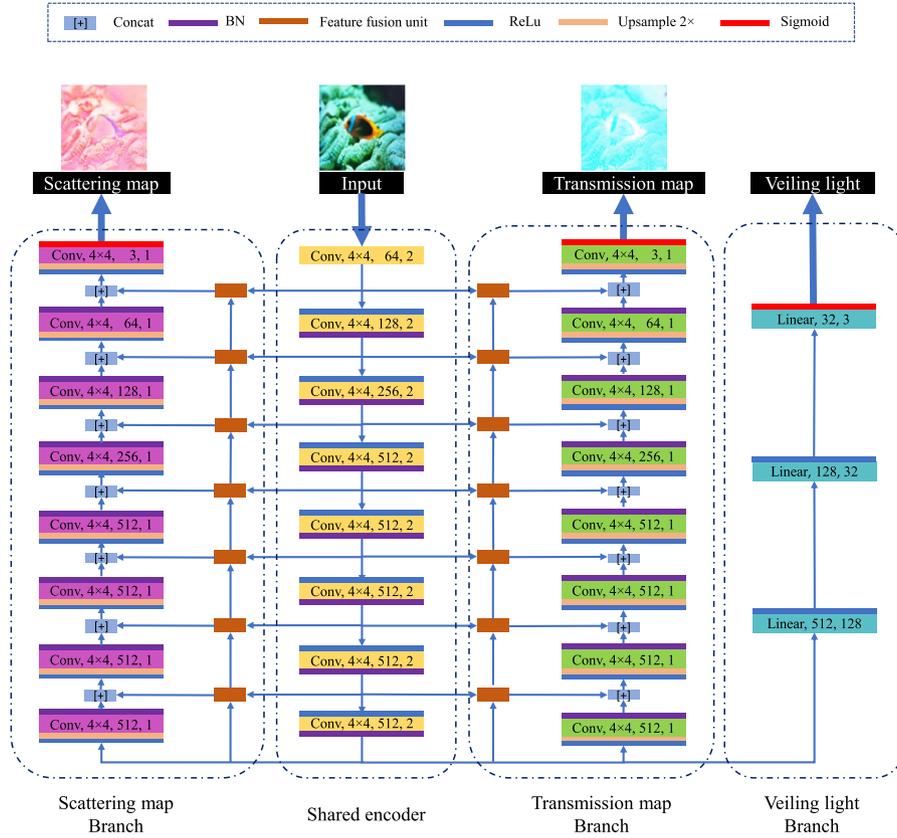


Fig. 4. Details of the three branches in the generator to illustrate the data generation. Tensor concatenation, batch normalization layer, feature fusion unit [24], ReLU layer, upsample layer, Sigmoid layer, convolution layer and linear layer are depicted in different colors. Parameters in linear layer are number of input and output channels, respectively. Parameters in convolution layer are: kernel size, number of output channels, stride.

where a, b are the image patch with the same location centered at pixel p in $G(I_c)$ and J_c , respectively. The default patch size is 13. μ_a, μ_b are the mean values, σ_a, σ_b are the standard deviations, σ_{ab} is the covariance of patch a and patch b . C_1 and C_2 are the parameters for numerical stability, whose default values are 0.02 and 0.03, respectively. Therefore, the SSIM loss between $G(I_c)$ and J_c can be expressed as:

$$\mathcal{L}_{SSIM}(J_c, G(I_c)) = 1 - \frac{1}{N} \sum (\text{SSIM}(p)) \quad (10)$$

where N is the total number of pixels.

To minimize the distribution gap between the reconstructed images and the ground-truth images, the adversarial loss is adopted. In this paper, the WGAN-GP adversarial loss [28] is adopted in the conditional setting. The adversarial loss conditioned on the input I_c is denoted as:

$$\begin{aligned} \mathcal{L}_{\text{cGAN-GP}} = & E_{I_c, J_c} [D(I_c, J_c)] - E_{I_c} [D(I_c, G(I_c))] \\ & + \lambda_{\text{GP}} E_j \left[\left(\|\nabla_{\hat{I}} D(\hat{I})\|_2 - 1 \right)^2 \right] \end{aligned} \quad (11)$$

where \hat{I} is the sample along the lines between the generated image $G(I_c)$ and the ground-truth images J_c . λ_{GP} is the weight factor for gradient penalty.

The overall objective function \mathcal{L}^* is:

$$\mathcal{L}^* = \min_G \max_D \lambda_G \mathcal{L}_{\text{cGAN-GP}}(G, D) + \lambda_1 \mathcal{L}_{L_1}(G) + \lambda_2 \mathcal{L}_{\text{SSIM}}(G) \quad (12)$$

where λ_G, λ_1 and λ_2 are the weight factors for the GAN loss, L_1 distance loss and SSIM loss, respectively.

4. Experiments

In this section, the datasets and the training parameters are presented first. Then the comparisons against state-of-the-art methods are conducted on extensive underwater datasets. Lastly, the ablation study is discussed.

Datasets. To make a fair comparison, we have done a detailed search for the publicly available datasets with paired images (i.e., the ground-truth images are available). To our knowledge, three datasets are made publicly available, and are suitable for performance evaluation. The summary of these datasets are shown in Table 1.

- 1) UGAN dataset. The dataset proposed in [21] is synthesized with CycleGAN. CycleGAN is first trained to learn the image degradation function. The clear underwater images with good white balance and visibility were collected from the ImageNet, which function as the ground-truth images. The trained CycleGAN downgrades the clear underwater images to generate the corresponding degraded counterparts. There are 6128 image pairs in total. In our experiment, we randomly choose 5500 images as the training set and the remaining 628 images for test.
- 2) EUVP dataset. The dataset proposed in [32] is synthesized with a similar procedure as the UGAN dataset. Differently, it contains three distinct categories: a) synthesized underwater dark scenes (3081 pairs for training, 483 for test); b) images collected from ImageNet are degraded (3801 pairs for training, 459 for test); c) other images from real underwater scenes (4002 pairs for training, 325 for test).

Table 1

Summary of the number of images for current datasets. EUVP-1, EUVP-2, EUVP-3 denotes the EUVP dark, scenes and ImageNet dataset receptively.

	UGAN	EUVP-1	EUVP-2	EUVP-3	Li
# Training images	5500	3081	3801	4002	780
# Test images	628	483	459	325	110

Table 2

PSNR for all methods evaluated on all the datasets. EUVP-1, EUVP-2, EUVP-3 denotes the EUVP dark, scenes and ImageNet dataset receptively.

	UGAN Dataset	Li Dataset	EUVP-1	EUVP-2	EUVP-3
EUFB	16.3331	22.3296	16.3624	16.2701	17.2915
UVE	15.5490	14.7532	14.6330	15.5743	16.5140
MBIE	14.5422	13.3378	13.1518	14.5320	13.0404
WCID	11.8857	11.6471	11.1719	11.8797	13.2368
UWCNN	15.4190	13.8526	18.3899	17.7377	15.2322
UGAN	18.6562	18.3731	20.2567	19.2276	18.5281
CycleGAN	22.3160	18.1203	19.4939	23.7342	22.1755
WaterGAN	20.7450	16.7586	19.2024	20.8330	20.3246
IPMGAN	23.5439	22.1555	21.4981	28.1323	23.3435

Table 3

SSIM for all methods evaluated on all the datasets.

	UGAN Dataset	Li Dataset	EUVP-1	EUVP-2	EUVP-3
EUFB	0.5379	0.8339	0.5328	0.5300	0.6910
UVE	0.4505	0.6703	0.4245	0.4880	0.6246
MBIE	0.4119	0.3457	0.3048	0.4105	0.2610
WCID	0.1955	0.3688	0.0850	0.1916	0.4056
UWCNN	0.6127	0.5258	0.5828	0.6756	0.5964
UGAN	0.5702	0.6804	0.6415	0.6095	0.5597
CycleGAN	0.7464	0.7103	0.6088	0.8157	0.7415
WaterGAN	0.7285	0.5488	0.5917	0.6795	0.7231
IPMGAN	0.8142	0.8146	0.6944	0.9197	0.8029

Table 4

UIQM* for all methods evaluated on all the datasets.

	UGAN Dataset	Li Dataset	EUVP-1	EUVP-2	EUVP-3
EUFB	0.4420	0.5519	0.4693	0.3543	0.4695
UVE	0.3600	0.6108	0.5823	0.2442	0.3649
MBIE	1.0000	1.0000	1.0000	1.0000	1.0000
WCID	0.5837	0.5528	0.6484	0.4222	0.5498
UWCNN	0.5157	0.6494	0.5553	0.3632	0.5590
UGAN	0.4683	0.5994	0.5806	0.3552	0.5060
CycleGAN	0.5355	0.5881	0.5631	0.3548	0.5848
WaterGAN	0.5242	0.6734	0.4863	0.4182	0.5748
IPMGAN	0.5857	0.7535	0.6675	0.4520	0.6223

Table 5

UCIQE* for all methods evaluated on all the datasets.

	UGAN Dataset	Li Dataset	EUVP-1	EUVP-2	EUVP-3
EUFB	0.9370	0.9588	0.9419	0.9789	0.9363
UVE	0.9104	0.8603	0.9057	0.8929	0.9118
MBIE	1.0000	1.0000	1.0000	1.0000	1.0000
WCID	0.9522	0.8968	0.9682	0.9343	0.9510
UWCNN	0.6180	0.6811	0.7662	0.7239	0.5596
UGAN	0.9309	0.9238	0.8736	0.9366	0.9297
CycleGAN	0.9024	0.8682	0.9315	0.8968	0.9030
WaterGAN	0.8857	0.8357	0.8787	0.8464	0.8764
IPMGAN	0.9169	0.9394	0.8714	0.9253	0.9140

3) Li dataset. The following procedures construct the dataset proposed in [33]. The authors firstly evaluated the current underwater image enhancement methods on many underwater images. Therein, all the enhanced results were assessed by humans, whom were required to mark the

enhanced results. Then, the enhanced image with the highest score was selected as the ground-truth. In other words, based on all the enhanced results of state-of-the-art methods, the ones, which are favored by human and consistent with human vision, are chosen as the ground-truth images.

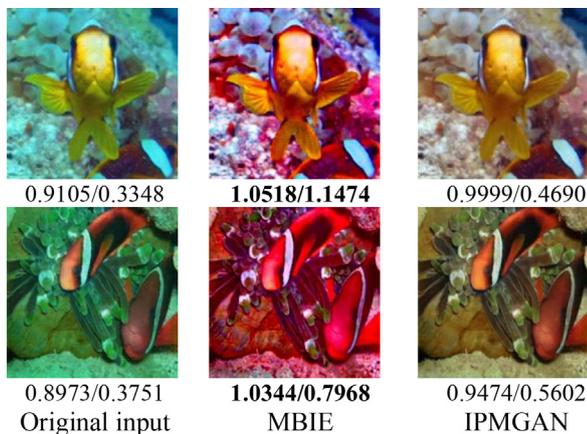


Fig. 5. Samples of UCIQE*/UIQM* evaluation on MBIE [12] and IPMGAN. The numbers under each image indicate UCIQE*/UIQM*, respectively.

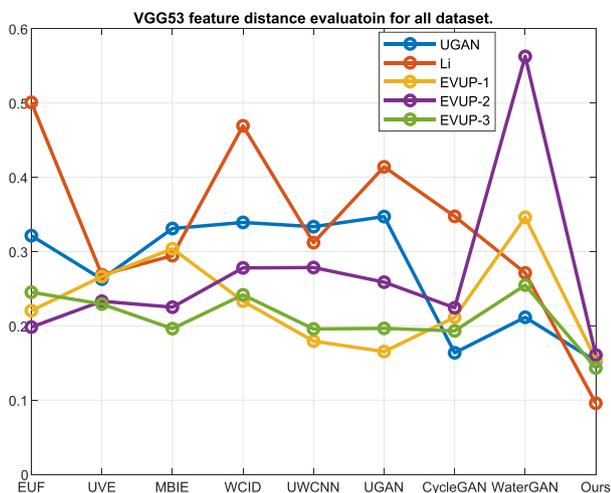


Fig. 6. Evaluation for VGG feature distance for all methods on all datasets.

This dataset consists of 890 image pairs in total, and we randomly select 780 images as the training set and the remaining 110 as test images.

Training parameters. Our network is implemented with PyTorch and trained using Adam optimizer with learning rate $lr = 0.0005$ and $\beta_1 = 0.5$. $\lambda_{GP} = 10$, $\lambda_G = 0.1$, $\lambda_1 = 10$, $\lambda_2 = 1$. These hyper-parameters are tuned and defined by cross validation. The network is trained for 50 epochs on one NVIDIA 1080Ti graphic card. The learning rate is kept the same for the first 25 epochs, and linearly decayed to zero over the next 25 epochs. All the training and test images are resized to 256×256 due to memory limitation.

4.1. Evaluation

To sufficiently evaluate our method, we make comparisons against state-of-the-art methods, and evaluate on the existing datasets both quantitatively and qualitatively. Afterwards, we conduct the ablation study to demonstrate the effectiveness and necessity of the integration of the Akkaynak-Treibitz model, and the effectiveness of the loss function.

4.1.1. Evaluation metrics

There are currently two sets of metrics for image quality evaluation: full-reference metrics and non-reference metrics. To sufficiently evaluate our method, we adopt both metrics for evaluation. In addition, we also utilize the VGG feature distance as a metric for performance evaluation.

Full-reference metrics. Underwater images with ground-truth images can be evaluated based on full-reference metrics. For example, the peak signal-to-noise ratio (PSNR) and the structural similarity index (SSIM), which are commonly utilized by this community, can effectively measure the image quality of the restored image against the ground-truth. For both metrics, a higher value indicates a more desirable outcome.

Non-reference metrics. On the other hand, there are cases where the ground-truth images are not available. Under this condition, it is possible to evaluate the underwater image with non-reference underwater image quality evaluation metrics, such as underwater image quality metric (UIQM) [34], and underwater color image quality evaluation metric (UCIQE) [35]. UIQM measures image colorfulness (UICM), sharpness (UISM) and contrast (UIConM), and UIQM is calculated via a linear combination of these sub-image attribute measures. On the other hand, UCIQE is designed to measure image chroma, saturation and contrast, which is a linear combination of those image attribute measure. We made comparisons based on UIQM and UCIQE metrics, and reported their limitations in the evaluation part, which are found to be consistent with [33].

VGG feature distance. In addition, we also include the VGG feature distance for evaluation. The VGG feature can be extracted by the pre-trained VGG net and the feature distance l_{VGG} can be calculated as:

$$l_{VGG} = \frac{1}{W_{ij}H_{ij}} \sum_{m=1}^{W_{ij}} \sum_{n=1}^{H_{ij}} (\phi_{ij}(J_c) - \phi_{ij}(G(I_c)))^2 \quad (13)$$

where ϕ_{ij} denote the operation to extract the intermediate feature map from the j^{th} convolution layer (after activation layer) before the i^{th} max-pooling layer from the pre-trained VGG16 network. W_{ij}, H_{ij} denote the dimensions of the respective feature maps within VGG network, $G(I_c)$ and J_c refer to the restored image and the ground-truth. Considering CNNs learn feature representation hierarchically, and feature map from the deeper layers encodes more semantic content of the image [36], we choose $\phi_{5,3}$ to extract content features for comparison.

4.1.2. Comparisons with state-of-the-art methods

In this section, we make comparisons with state-of-the-art methods on current existing datasets both quantitatively and qualitatively.

Methods for comparisons. Comparisons are made with the following state-of-the-art methods: 1) *Non-learning-based methods*: 1-a) enhancing underwater images by fusion (EUF) [37]. 1-b) multi-band image enhancement (MBIE) [12]. 1-c) underwater visibility enhancement (UVE) [11]. 1-d) wavelength compensation and dehazing for underwater scenes (WCID) [9]. These methods are implemented and evaluated with the source code provided by their authors. 2) *Learning-based methods*: 2-a) Enhancing the underwater image with GANs (UGAN) [21]. 2-b) Prior inspired underwater image enhancement (UWCNN) [33]. 2-c) In addition, as CycleGAN [22] sets a new baseline for image-to-image translation problems, we also include CycleGAN [22] for comparisons. 2-d) Color correction network in WaterGAN [20]. UGAN [21], CycleGAN [22] and are trained with the parameters reported in their papers. For the color correction network in WaterGAN, we only train the one-stage net based on RGB images from scratch.

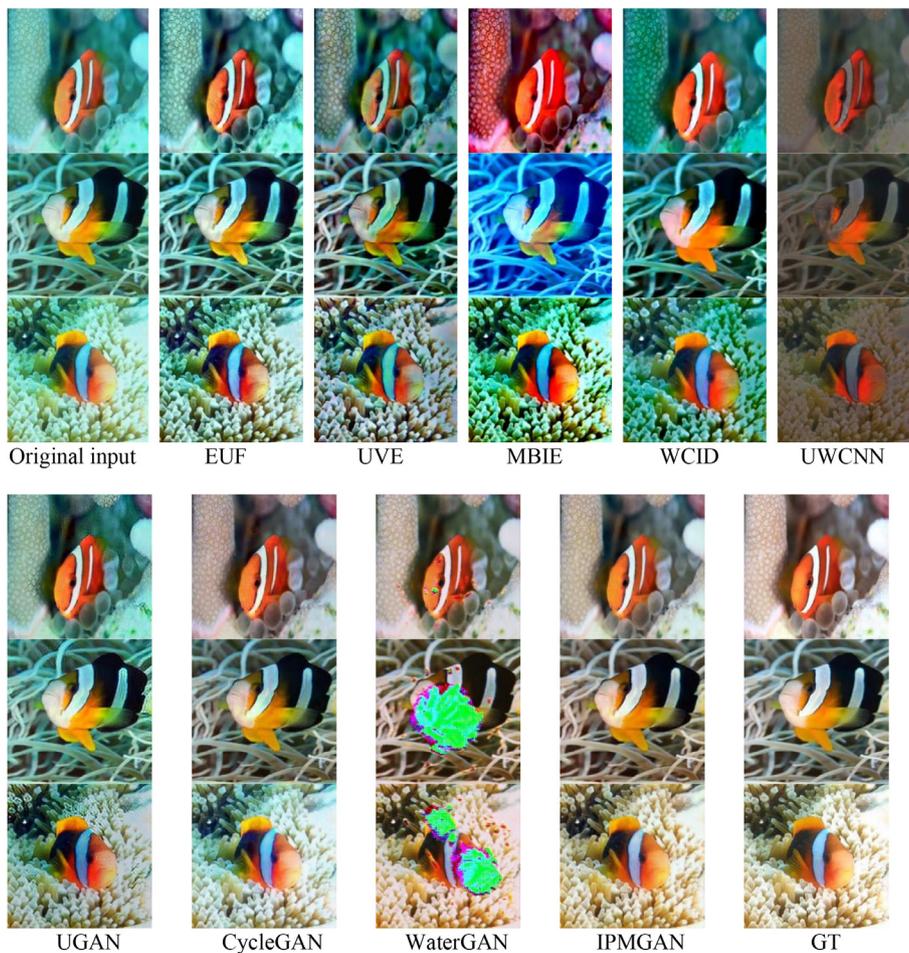


Fig. 7. Evaluation on UGAN dataset. Three samples are shown with the original inputs, and the enhancement results are EUF [37], UVE [11], MBIE [12], WCID [9], UWCNN [33], UGAN [21], CycleGAN [22], WaterGAN [20], ours and the ground-truth images.

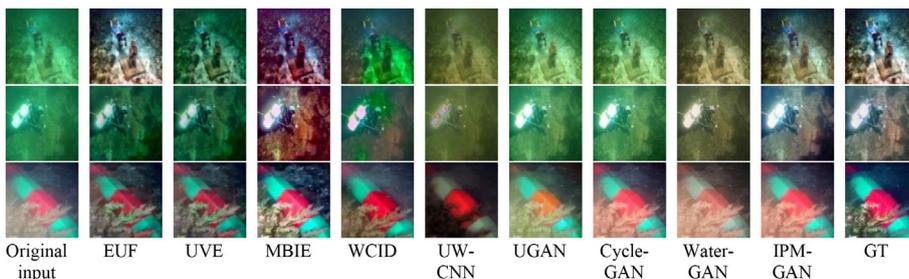


Fig. 8. Evaluation on the Li dataset. From the left to right are the original inputs, and the enhancement results are EUF [37], UVE [11], MBIE [12], WCID [9], UWCNN [33], UGAN [21], CycleGAN [22], WaterGAN [20], our enhancement results and the corresponding ground-truth images.

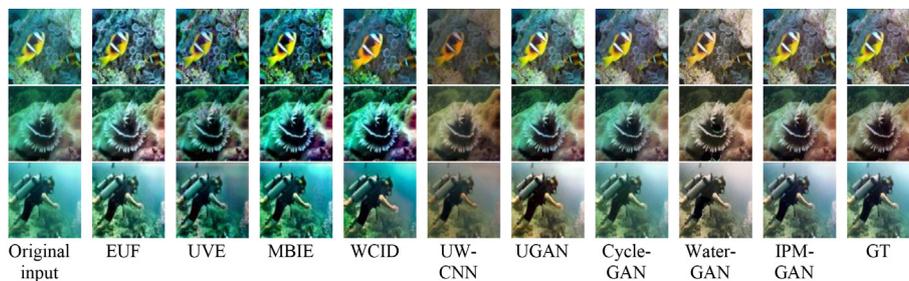


Fig. 9. Evaluation on EUVP underwater scenes dataset. From the left to right are the original inputs, and the enhancement results are EUF [37], UVE [11], MBIE [12], WCID [9], UWCNN [33], UGAN [21], CycleGAN [22], WaterGAN [20], our enhancement results and the ground-truth images.

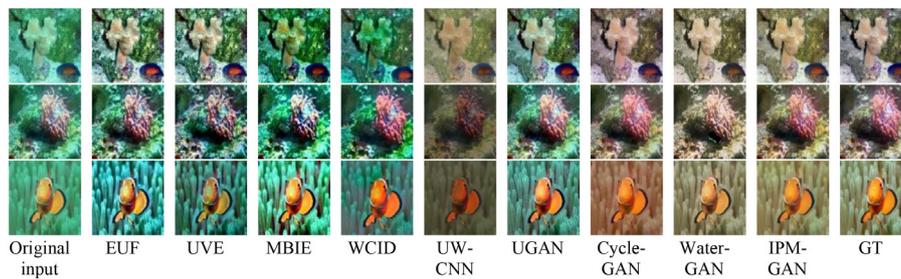


Fig. 10. Evaluation on EUVP ImageNet dataset. From the left to right are the original inputs, and the enhancement results are EUF [37], UVE [11], MBIE [12], WCID [9], UWCNN [33], UGAN [21], CycleGAN [22], WaterGAN [20], our enhancement results and the corresponding ground-truth images.

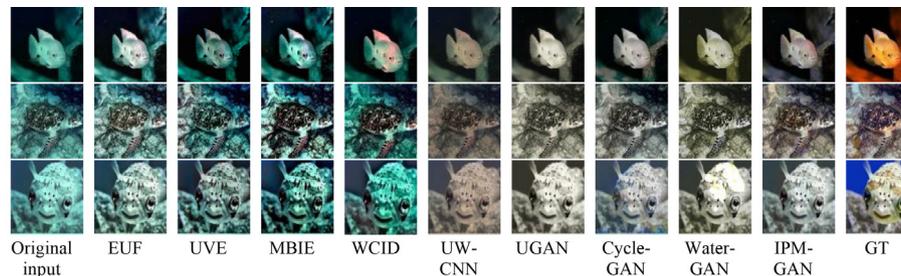


Fig. 11. Evaluation on EUVP underwater dark dataset. From the left to right are the original inputs, and the enhancement results are EUF [37], UVE [11], MBIE [12], WCID [9], UWCNN [33], UGAN [21], CycleGAN [22], WaterGAN [20], our enhancement results and the ground-truth images.

Since UWCNN [33] only provides the pre-trained model, it is evaluated on the pre-trained model.

Quantitative evaluation on PSNR and SSIM

The evaluation is made on all datasets and the average PSNR and SSIM for the test images of each dataset are reported in Tables 2 and 3. From Table 2 and Table 3, we find our method nearly achieve the best results for all datasets except the Li dataset, where EUF [38] is the best numerically. Quantitatively, our method achieves comparable results with EUF [38] on the Li dataset. More specifically, the reason why EUF [38] achieves the highest PSNR and SSIM score is that the nearly 25% ground-truth images in the Li dataset are exactly the enhancement results achieved by EUF [38] as reported in [33].

Quantitative evaluation on UIQM and UCIQE

For non-reference image quality metrics, the average of UCIQE and UIQM for the test images of each dataset are normalized and reported in Table 4 and Table 5.

Remark: As the big variance of UCIQE/UIQM exists for different methods within one dataset, we normalize all the reported UCIQE and UIQM score into (0, 1) within each dataset for better method comparisons, where the highest UCIQE/UIQM scores are set as 1, also as the baseline for each dataset. The normalized values are denoted as UCIQE*/UIQM*. Since we take the average UCIQE/UIQM for normalization, UCIQE*/UIQM* for a single image may exceed the mean value, which could result in UCIQE*/UIQM* score to be larger than 1 as the case shown in Fig. 5.

From Table 2 and Table 3, we could find our method, IPM-GAN, nearly achieve the best PSNR and SSIM for all the datasets. Interestingly, however, for UCIQE and UIQM metrics, MBIE [12] achieves the best as shown in Table 4 and Table 5 instead of IPM-GAN. Samples from MBIE [12] and IPM-GAN are shown in Fig. 5. In Fig. 5, compared with the original inputs, images processed by IPM-GAN are recovered with visually pleasing white balance achieving improvements in UCIQE*/UIQM*. On the other hand, although MBIE [12] introduces the red color bias and over-saturation, it achieves better UCIQE*/UIQM* score. This also explains why MBIE [12] achieves the best UCIQE/UIQM for all datasets. Thus, we

may report some limitations of UCIQE and UIQM for underwater image enhancement evaluation. In theory, UCIQE and UIQM try to measure image colorfulness; therefore, these metrics favor colorful objects in the image. However, these metrics lack penalization on over-saturation and would fail for these cases. The limitations of UCIQE and UIQM are also reported in [33].

Quantitative evaluation on VGG feature distance

We calculate the content feature distance between the restored images and the ground-truth images for all methods on all datasets. The results are depicted in Fig. 6, where the smaller distance to the ground-truth indicates a better visual restoration. For all datasets involved, IPM-GAN achieves the lowest distance. This shows that IPM-GAN can restore images closest to the ground-truth in VGG content feature space, indicating the best perceptual restoration among all methods.

Qualitative evaluation

In the following, we will present the restoration results for each dataset sequentially.

Evaluation on the UGAN dataset.

This dataset mainly contains images suffering from the color cast where the degraded images appear greenish or bluish. Three samples are chosen from the test results due to space limitations in Fig. 7. The original input images contain intense green hue, and EUF [37], UVE [11], WCID [9], however, fails to remove the color shift. MBIE [12], on the other hand, introduces unwanted color bias, for example, the red and blue hue in the first and second image receptively. Results obtained by UWCNN [33] appear dim. UGAN [21] and CycleGAN [22] achieved relatively better performance, however, there still remain unwanted green hue on the image in the third row. Results obtained from WaterGAN contain unexpected artifacts, which may be due to the limited learning ability of the one-stage color correction net. By contrast, our enhancement results achieve clear results with good white balance, and visually quite close to the ground-truth images. Quantitatively, our results achieve the best PSNR and SSIM scores, and outperform other methods, as shown in Tables 2 and 3 with a large margin.

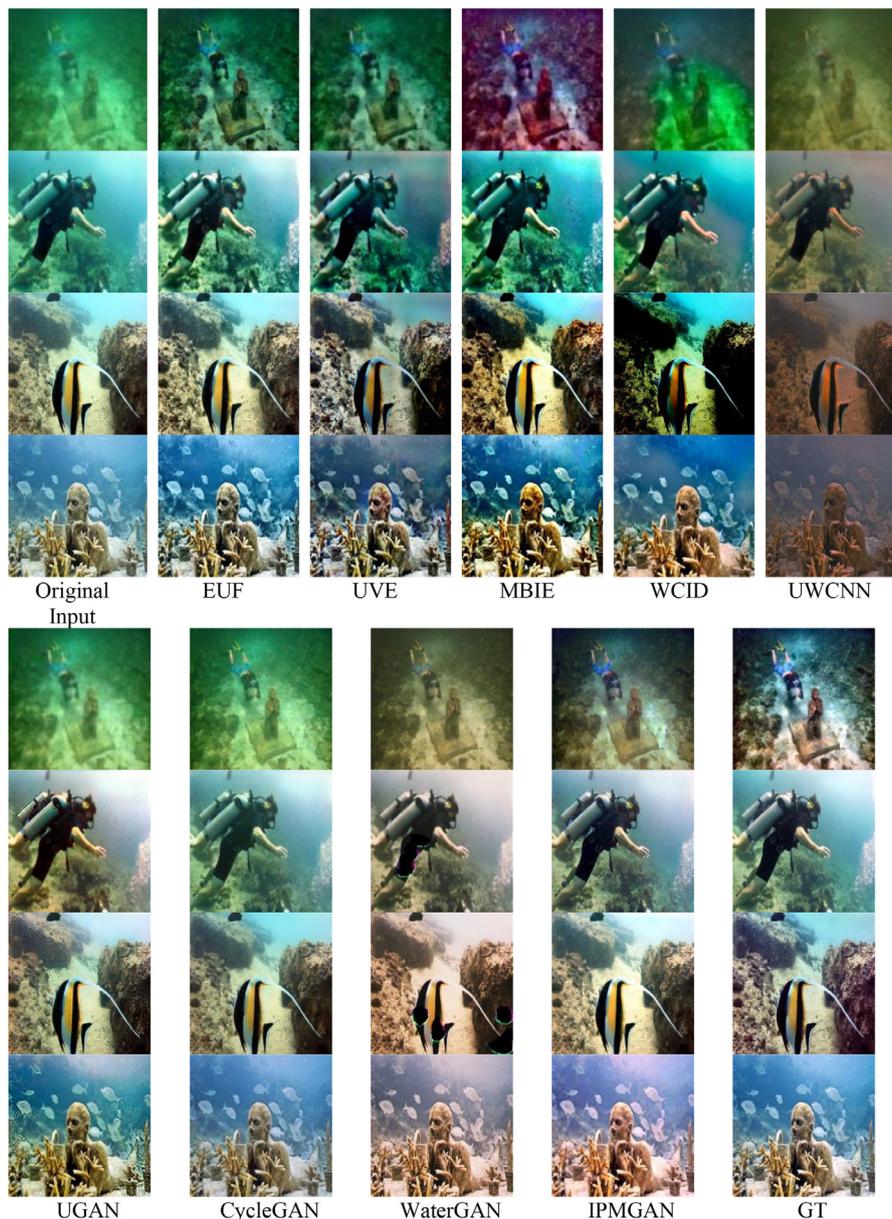


Fig. 12. Zoom-in details for images suffer severe backscatter. From the left to right are original inputs, the enhancement results from EUF [37], UVE [11], MBIE [12], WCID [9], UWCNN [33], UGAN [21], CycleGAN [22], WaterGAN [20], ours and the ground-truth images.

Evaluation on the Li dataset.

This dataset is more challenging than others as: first, it contains images with much poorer quality with limited visibility; second, the number of training images is also limited (less than 1000 images). The evaluation on the Li dataset is shown in Fig. 8, where our method can achieve better image restoration quality compared to others.

Evaluation on the EUVP dataset.

First, the evaluation on the underwater scenes and underwater ImageNet dataset are shown in Figs. 9 and 10 since these two datasets share some similarities. As shown in Figs. 9, 10, our method can effectively remove the color cast and achieve good white balance. More interestingly, as shown in the third row of the test image in Fig. 9 of the human diver, our result is more visually pleasing than the simulated ground-truth with better white balance. Quantitatively, the average PSNR and SSIM values are 28.1232 dB/0.9197 and 23.3435 dB/0.8029 for underwater scenes and underwater ImageNet dataset, respectively, which greatly outperform other methods.

Evaluation on the dark underwater dataset is shown in Fig. 11. To restore the dark underwater dataset is a bit more challenging, and nearly all methods can not guarantee to achieve acceptable colorful results, especially the restoration of the background with great diversity and variation. As shown in Fig. 11, the background for the first and the third images are red and blue respectively, causing it challenging to be precisely restored.

One possible reason for this failure may be that the number of training images is limited. Compared to the restoration of the blue or green hue dominated scenes, in a dark scenario, the majority of information in all RGB channels is lost, hence more information needs to be learned and recovered by the network. Therefore, data augmentation or more training images are required for improvement.

In addition, zoom-in details for images that suffer from severe backscatter are evaluated and presented in Fig. 12. Four samples of real underwater images are from the Li dataset and the EUVP scenes dataset. Fig. 12 demonstrates that when the images, which suffer heavier backscatter, are challenging for a perfect restoration.

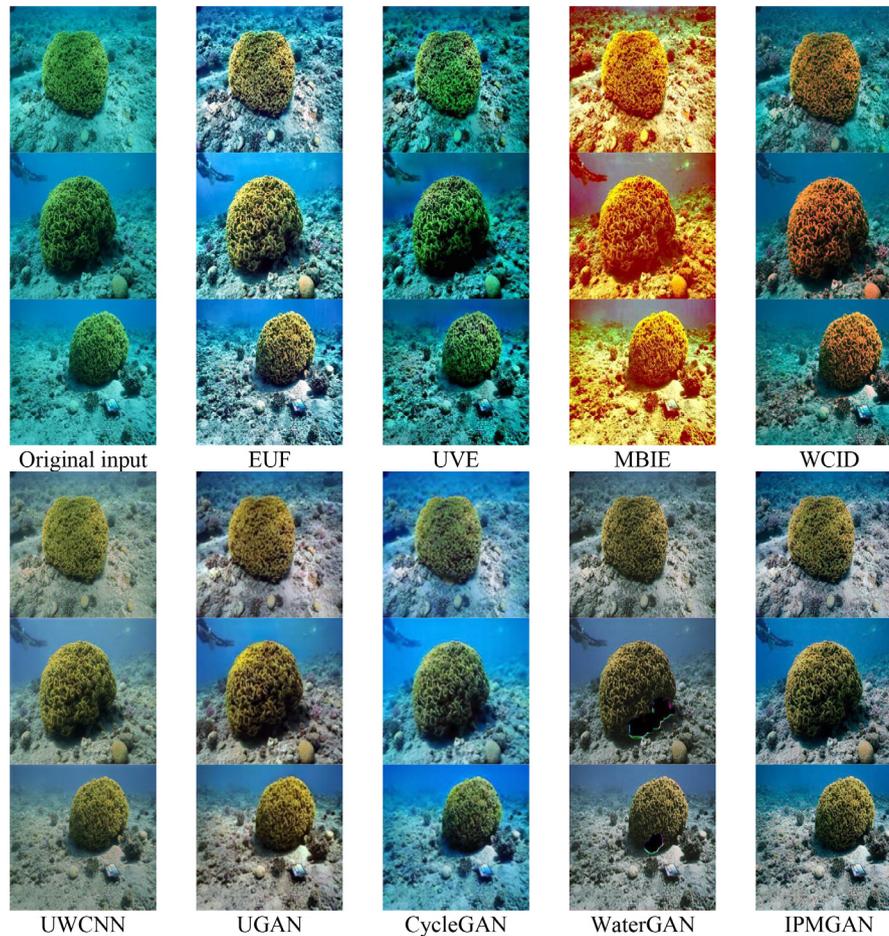


Fig. 13. Evaluation on the Akkaynak-Treibitz dataset. From the left to right are original inputs, the enhancement results are EUF [37], UVE [11], MBIE [12], WCID [9], UWCNN [33], UGAN [21], CycleGAN [22], WaterGAN [20], and ours.

Table 6 UIQM for methods on the Akkaynak-Treibitz dataset. UICM, UIQM, UIConM refer to image colorfulness measure, sharpness measure and contrast measure, respectively.

	UICM	UIQM	UIConM	UIQM
EUF	-59.2150	7.1388	0.6089	2.6153
UVE	-81.5150	7.2574	0.4820	1.5677
MBIE	54.2868	7.4361	0.5325	5.6307
WCID	-52.5412	7.1239	0.5215	2.4867
UWCNN	-24.0956	6.9938	0.8342	4.3682
UGAN	-13.5590	6.9812	0.8156	4.5951
CycleGAN	-66.2694	7.0944	0.8412	3.2339
WaterGAN	-14.6831	7.1984	0.7872	4.5259
IPMGAN	-3.6700	7.0987	0.8567	5.0559

Table 7 PSNR and SSIM comparisons against different design approaches.

	Simplified model-based [31]	End-to-end network [24]	IPMGAN
PSNR	18.3934	22.9651	23.5439
SSIM	0.5801	0.7946	0.8142

Most methods could not obtain the perceptually pleasing results, especially for patches in the distance. Besides, their corresponding "ground-truth" images are actually the reference images instead of the real ground-truth. Despite that the reference images are not clear and satisfied, our method achieves comparatively the best performance for color restoration and hazy removal. The results are visually closest to the reference images.

Evaluation on the Akkaynak-Treibitz Dataset.

The evaluation on the Akkaynak-Treibitz dataset is shown in Fig. 13. As there are no ground-truth images available for the dataset, UIQM is adopted for quantitative performance evaluation, and the results are listed in Table 6. From Fig. 13 and Table 6, we could have the following observations. EUF [37], UVE [11] and CycleGAN [22] fail to remove the color shift for these samples. MBIE [12] generates images with high saturation, together with unwanted color bias. WCID [9] could improve the contrast but fail to remove the color shift. Visually, UWCNN [33] could remove the color shift to some extent, however, images still appear dim. WaterGAN [20] would introduce unwanted artifacts like the black holes in the restored images. UGAN [21] and IPMGAN can successfully achieve good enhancement for the foreground, and ours can generate images with better details (for example, the color chart) when

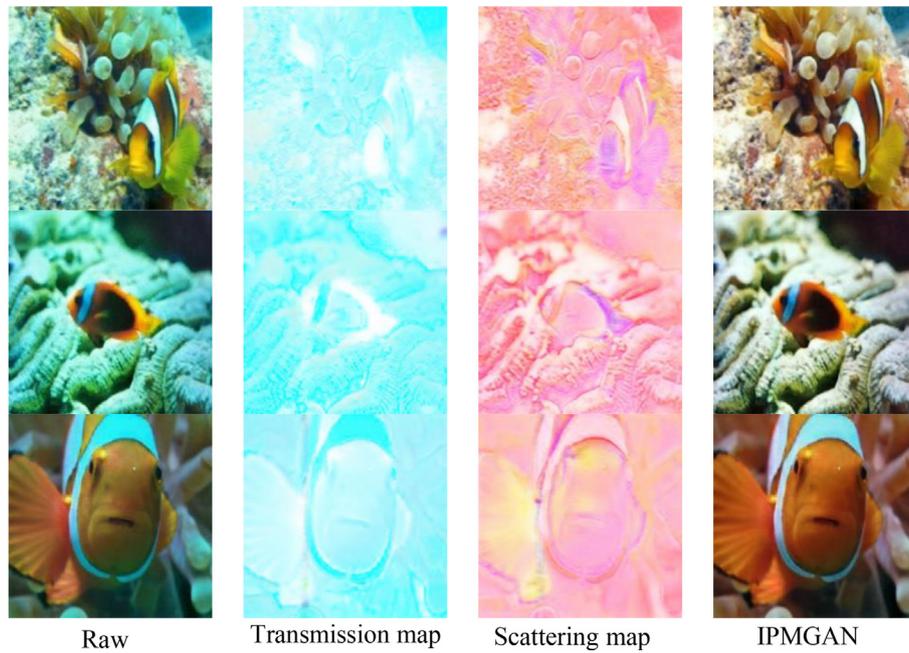


Fig. 14. From the left to right are the raw underwater input, the estimated transmission map, the estimated scattering map and the reconstructed images by IPMGAN.

Table 8
PSNR and SSIM for different loss functions.

Loss function	\mathcal{L}_{L_1}	$\mathcal{L}_{L_1} + \mathcal{L}_{SSIM}$	\mathcal{L}^*
PSNR	20.9421	21.7193	22.2804
SSIM	0.7076	0.7673	0.7713

zooming in for visual inspection. As for the background with severe haze effects, it is still challenging for restoration. This is also reported challenging even for the depth estimation-based or range-map dependent image restoration methods in [39]. To correctly estimate the depth, the areas far from the camera are quite difficult due to the low signal-to-noise-ratio [39].

Quantitatively, as discussed earlier that UIQM favors colorfulness objects, MBIE [12] again achieves the best UIQM scores on this dataset. If we exclude the "outliers" of MBIE [12], our method achieves the best reasonable UIQM.

The learning-based method is deeply dependent on the training dataset. Therefore, to further effectively restore the areas with low signal-to-noise-ratio, datasets containing severely degraded images which are paired with high-quality ground-truth or reference images are greatly expected. Possible choices would be the

Akkaynak-Treibitz dataset (when the reference images are ready and publicly available) or the simulated dataset with the Akkaynak-Treibitz model.

4.2. Ablation study

In this section, we first compare the different network design philosophies; then demonstrate the effectiveness of the loss functions.

Comparisons against other design philosophies. Considering our design is based on the revised underwater image degradation model, we make comparisons with the other two different network design philosophies:

- 1) The network based on the simplified underwater image degradation model as described in Eq. (1), which ignores the difference between the absorbing coefficients and scattering coefficients (as used in [31]).
- 2) The end-to-end network training scheme without the integration of the physical model. We take the state-of-the-art underwater image enhancement [24] for comparison.

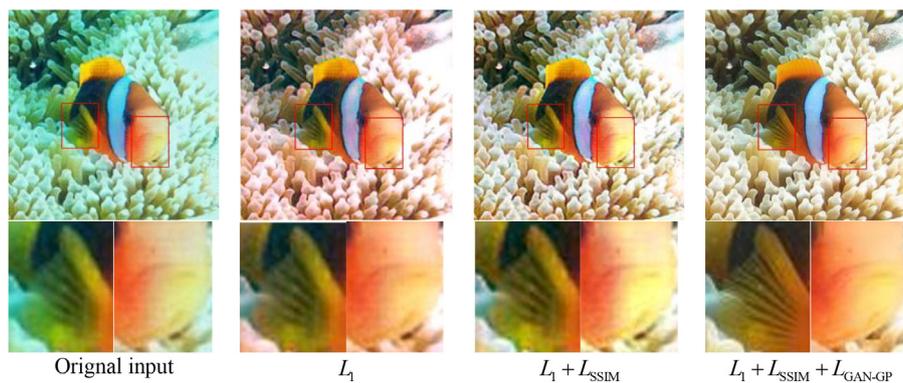


Fig. 15. Examples from the results with different loss functions.

These networks are trained on the UGAN dataset with the same training parameters. The evaluation is made based on the 628 test images, where the average PSNR and SSIM are reported in Table 7.

As shown in Table 7, our method achieves the best score in both PSNR and SSIM metrics, which shows that based on the Akkaynak-Treibitz underwater image formation model, the network is empowered with better learning ability. Compared to the simplified model-based network design, the Akkaynak-Treibitz model could capture more information from the degradation process. On the other hand, compared to the end-to-end training strategy, integrating the revised physical model further boosts the restoration ability.

In addition, we show the estimated transmission map \hat{T} and the scattering map \hat{S} in Fig. 14. Different from the in-air captured images, the transmission coefficients β_c^d and the scattering coefficients β_c^b are not equal in RGB channels due to the wavelength-dependent absorption and scattering in water. Therefore, \hat{T} and \hat{S} are shown in color. Clearly, \hat{T} and \hat{S} are different from each other.

Nevertheless, this significant difference between \hat{T} and \hat{S} were ignored in the simplified image formation model, and the over-simplification may also be the cause for the lower PSNR and SSIM scores on the evaluation (in Table 7). This also demonstrates the Akkaynak-Treibitz model's correctness and verifies the necessity of integrating the revised model into the network design, which greatly improves the performance in underwater image restoration.

Effectiveness of the loss function.

Here we would give more illustration about the effectiveness of the loss functions. The network is trained with three loss functions as a) only L_1 distance loss: \mathcal{L}_{L_1} ; b) L_1 distance loss and SSIM loss: $\mathcal{L}_{L_1} + \mathcal{L}_{SSIM}$; and c) the overall loss function $\mathcal{L}^* : \mathcal{L}_{L_1} + \mathcal{L}_{SSIM} + \mathcal{L}_{GAN-GP}$ (with weight factors). The results are evaluated in terms of PSNR and SSIM as shown in Table 8. The examples together with their zoom-in details are visualized in Fig. 15.

Table 8 shows the results of PSNR and SSIM, where our objective function \mathcal{L}^* achieves the best score. Besides, Fig. 15 provides the visual effects exerted by the losses. The result from L_1 loss appears with color cast, and relatively poor white balance. That may be due to the Euclidean distance loss is likely to push the network to achieve an average of all plausible results [36]. Incorporated with SSIM loss, more structure constraint is introduced and learned by the network, thus the background with more textures is better recovered with good white balance. Besides, from Table 8, we could notice there is a huge improvement of SSIM score when SSIM loss is introduced. Furthermore, the adversarial loss is beneficial to generate finer details [36] [22] [21], which are presented in the fin and the face of the Nemo with comparatively better clearness and smoothness.

5. Conclusion

We have presented an integration of the current two main streams underwater image restoration methods, i.e., the underwater image degradation model-based and the neural-network-based methods, aiming to make full advantages of the merits of these two approaches to mutually benefits with each other. The revised image degradation model guides of the network design and learning, on the other hand, the model-free conditional GAN framework also improves the underwater image restoration performance. Besides, to our knowledge, we are the first to incorporate the revised model into the network design and the performance is evaluated nearly on all publicly available datasets. Extensive experiment results demonstrate that our method can effectively

restore color of underwater images with fine details and alleviate the unwanted artifacts, which outperform state-of-the-arts approaches both subjectively and objectively.

CRedit authorship contribution statement

Xiaodong Liu: Data curation, Writing - review & editing. **Zhi Gao:** Writing - review & editing, Funding acquisition. **Ben M. Chen:** Supervision, Writing - review & editing, Funding acquisition.

Declaration of Competing Interest

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

Appendix A. Table of Notation Symbols

Symbol	Description
I_c	The captured color underwater image.
J_c	The unattenuated scene radiance (goal of image restoration).
β_c	Attenuation coefficients.
z	Distance between the camera and the objects.
B_c^∞	Veiling light.
c	Color channel (red,green,blue).
β_c^d	Attenuation coefficients for direct transmission.
β_c^b	Backscatter coefficients.
v_b	Dependencies of β_c^b .
v_d	Dependencies of β_c^d .
ρ	Scene reflectance.
R	Spectrum of ambient light.
S_c	Sensor spectral response.
s	Physical scattering coefficients of water body.
β	Beam attenuation coefficients of water body.
T	Transmission map.
S	Scattering map.
G	Generator network.
D	Discriminator network.
E	Statistical expectation value.
p	Pixel in the image.
μ_a, μ_b	Mean value for image patch a and patch b.
σ_a, σ_b	Variance for image patch a and image patch b.
C_1, C_2	Constant value in SSIM calculation.
\hat{T}	Samples along the lines between $G(I_c)$ and J_c
l_{VGG}	VGG feature distance.
$\phi_{i,j}$	Operation to extract feature map from j^{th} convolution layer before the i^{th} maxpooling layer.
$\lambda_{GP}, \lambda_G, \lambda_1, \lambda_2$	Weight factor for gradient penalty, GAN loss, L_1 loss and SSIM loss.

References

- [1] P. Ozog, N. Carlevaris-Bianco, A. Kim, R.M. Eustice, Long-term mapping techniques for ship hull inspection and surveillance using an autonomous underwater vehicle, *J. Field Robot.* 33 (3) (2016) 265–289.
- [2] F. Shkurti, A. Xu, Meghjani, et al., Multi-domain monitoring of marine environments using a heterogeneous robot team, in: *Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on IEEE, 2012*, pp. 1747–1753.
- [3] A. Khan, S.S.A. Ali, F. Meriaudeau, A.S. Malik, L.S. Soon, T.N. Seng, Visual feedback-based heading control of autonomous underwater vehicle for pipeline corrosion inspection, *Int. J. Adv. Rob. Syst.* 14 (3) (2017), 1729881416658171.
- [4] L. Paull, M. Seto, J.J. Leonard, H. Li, Probabilistic cooperative mobile robot area coverage and its application to autonomous seabed mapping, *Int. J. Robot. Res.* 37 (1) (2018) 21–45.
- [5] N. Abreu, A. Matos, Minehunting mission planning for autonomous underwater systems using evolutionary algorithms, *Unmanned Syst.* 2 (04) (2014) 323–349.
- [6] C.D. Mobley, *Light and water: radiative transfer in natural waters*, Academic Press, 1994.

- [7] S. Zhang, T. Wang, J. Dong, H. Yu, Underwater image enhancement via extended multi-scale retinex, *Neurocomputing* 245 (2017) 1–9.
- [8] X. Fu, P. Zhuang, Y. Huang, Y. Liao, X.-P. Zhang, X. Ding, A retinex-based enhancing approach for single underwater image, in: 2014 IEEE International Conference on Image Processing (ICIP) IEEE, 2014, pp. 4572–4576.
- [9] J.Y. Chiang, A. Ying-Ching Chen, Underwater image enhancement by wavelength compensation and dehazing (wcid), *IEEE Trans. Image Process.* 21 (4) (2012) 1756–1769.
- [10] A. Galdran, D. Pardo, A. Picón, A. Alvarez-Gila, Automatic red-channel underwater image restoration, *J. Vis. Commun. Image Rep.* 26 (2015) 132–145.
- [11] Y. Cho, A. Kim, Visibility enhancement for underwater visual slam based on underwater light scattering model, in: Robotics and Automation (ICRA), 2017 IEEE International Conference on, IEEE, 2017, pp. 710–717.
- [12] Y. Cho, J. Jeong, A. Kim, Model assisted multi-band fusion for single image enhancement and applications to robot vision, *IEEE Rob. Automat. Lett.*
- [13] D. Berman, T. Treibitz, S. Avidan, Diving into haze-lines: Color restoration of underwater images, in: Proc. British Machine Vision Conference (BMVC), vol. 1, 2017.
- [14] G. Hou, Z. Pan, G. Wang, H. Yang, J. Duan, An efficient nonlocal variational method with application to underwater image restoration, *Neurocomputing* 369 (2019) 106–121.
- [15] D. Akkaynak, T. Treibitz, A revised underwater image formation model, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 6723–6732.
- [16] D. Akkaynak, T. Treibitz, Sea-thru, A method for removing water from underwater images, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 1682–1691.
- [17] Y. Li, L. Jia, Z. Wang, Y. Qian, H. Qiao, Un-supervised and semi-supervised hand segmentation in egocentric images with noisy label learning, *Neurocomputing* 334 (2019) 11–24.
- [18] X. Zou, Z. Wang, Q. Li, W. Sheng, Integration of residual network and convolutional neural network along with various activation functions and global pooling for time series classification, *Neurocomputing* 367 (2019) 39–45.
- [19] Y. Wang, J. Zhang, Y. Cao, Z. Wang, A deep cnn method for underwater image enhancement, in: Image Processing (ICIP), 2017 IEEE International Conference on, IEEE, 2017, pp. 1382–1386.
- [20] J. Li, K.A. Skinner, R.M. Eustice, M. Johnson-Roberson, Watergan: unsupervised generative network to enable real-time color correction of monocular underwater images, *IEEE Rob. Automat. Lett.* 3 (1) (2018) 387–394.
- [21] C. Fabbri, M.J. Islam, J. Sattar, Enhancing underwater imagery using generative adversarial networks, in: 2018 IEEE International Conference on Robotics and Automation (ICRA) IEEE, 2018, pp. 7159–7165.
- [22] J.-Y. Zhu, T. Park, P. Isola, A. A. Efros, Unpaired image-to-image translation using cycle-consistent adversarial networks, *arXiv preprint*.
- [23] J. Lu, N. Li, S. Zhang, Z. Yu, H. Zheng, B. Zheng, Multi-scale adversarial network for underwater image restoration, *Opt. Laser Technol.* 110 (2019) 105–113.
- [24] X. Liu, Z. Gao, B. M. Chen, Mlfcgan: Multilevel feature fusion-based conditional gan for underwater image color correction, *IEEE Geosci. Remote Sens. Lett.*
- [25] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets, in: Advances in neural information processing systems, 2014, pp. 2672–2680.
- [26] X. Mao, Q. Li, H. Xie, R. Y. Lau, Z. Wang, S. P. Smolley, Least squares generative adversarial networks, in: Computer Vision (ICCV), 2017 IEEE International Conference on, IEEE, 2017, pp. 2813–2821.
- [27] M. Arjovsky, S. Chintala, L. Bottou, Wasserstein gan, *arXiv preprint arXiv:1701.07875*.
- [28] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, A.C. Courville, Improved training of wasserstein gans, *Adv. Neural Inform. Process. Syst.* (2017) 5767–5777.
- [29] P. Isola, J.-Y. Zhu, T. Zhou, A.A. Efros, Image-to-image translation with conditional adversarial networks, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 1125–1134.
- [30] C. Li, J. Guo, C. Guo, Emerging from water: Underwater image color correction based on weakly supervised color transfer, *IEEE Signal Process. Lett.* 25 (3) (2018) 323–327.
- [31] X. Ding, Y. Wang, Y. Yan, Z. Liang, Z. Mi, X. Fu, Jointly adversarial network to wavelength compensation and dehazing of underwater images, *arXiv preprint arXiv:1907.05595*.
- [32] M. J. Islam, Y. Xia, J. Sattar, Fast underwater image enhancement for improved visual perception, *arXiv preprint arXiv:1903.09766*.
- [33] C. Li, S. Anwar, F. Porikli, Underwater scene prior inspired deep underwater image and video enhancement, *Pattern Recogn.* 98 (2020) 107038.
- [34] Y.-T. Peng, P.C. Cosman, Underwater image restoration based on image blurriness and light absorption, *IEEE Trans. Image Process.* 26 (4) (2017) 1579–1594.
- [35] M. Yang, A. Sowmya, An underwater color image quality evaluation metric, *IEEE Trans. Image Process.* 24 (12) (2015) 6062–6071.
- [36] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. P. Aitken, A. Tejani, J. Totz, Z. Wang, et al., Photo-realistic single image super-resolution using a generative adversarial network., in: CVPR, vol. 2, 2017, p. 4.
- [37] C.O. Ancuti, C. Ancuti, C. De Vleeschouwer, P. Bekaert, Color balance and fusion for underwater image enhancement, *IEEE Trans. Image Process.* 27 (1) (2017) 379–393.
- [38] C. Ancuti, C.O. Ancuti, T. Haber, P. Bekaert, Enhancing underwater images and videos by fusion, in: 2012 IEEE Conference on Computer Vision and Pattern Recognition IEEE, 2012, pp. 81–88.
- [39] D. Berman, D. Levy, S. Avidan, T. Treibitz, Underwater single image color restoration using haze-lines and a new quantitative dataset, *IEEE Trans. Pattern Anal. Mach. Intelligence*.



Xiaodong Liu received the B.Eng. degree in automation from Shandong University, Shandong, China, in 2013, the M.Eng. degree in instrument science from Beihang University, China in 2016. Currently he is a Ph.D candidate in the Department of Electrical and Computer Engineering of National University of Singapore. His research interests include image enhancement, image processing, robotics and deep learning.



Zhi Gao received the B.Eng and the Ph.D degrees from Wuhan University, China in 2002 and 2007 respectively. Since 2008, he joined the Interactive and Digital Media Institute, National University of Singapore (NUS), as a Research Fellow (A) and project manager. In 2014, he joined Temasek Laboratories in NUS (TL@NUS) as a Research Scientist (A) and Principal Investigator. He is currently working as a full professor with the School of Remote Sensing and Information Engineering, Wuhan University. Since 2019, he has been supported by the distinguished professor program of Hubei Province, China. He has published more than 70 research papers on top journals and conferences, such as IJCV, IEEE T-PAMI, IEEE T-GRS, IEEE T-ITS, IEEE T-CSVT, IEEE GRSL, IEEEESPL, MVAP, J REAL-TIME IMAGE PR, ACM journals, ECCV, ACCV, BMVC, etc. His research interests include computer vision, machine learning, remote sensing, and their applications. In particular, he has strong interests in UAV based surveillance research and applications.



Ben M. Chen received his B.S. in mathematics and computer science from Xiamen University, China, 1983, M.S. in Electrical Engineering from Gonzaga University, Spokane, Washington, 1988, and PhD in Electrical and Computer Engineering from Washington State University, Pullman, Washington, 1991. He is currently a Professor at the Department of Electrical and Computer Engineering, National University of Singapore (NUS), and the Department of Mechanical & Automation Engineering, Chinese University of Hong Kong. His current research interests are in systems and control, unmanned aerial systems, and financial market modeling.

Prof. Chen is an IEEE Fellow. He is the author/co-author of 10 research monographs including *Optimal Control* (Prentice Hall, 1995), *Robust and Control* (Springer, 2000), *Hard Disk Drive Servo Systems* (Springer, 1st Edition, 2002; 2nd Edition, 2006), *Linear Systems Theory* (Birkhäuser, 2004), *Unmanned Rotorcraft Systems* (Springer, 2011), and *Stock Market Modelling and Forecasting* (Springer, 2013). He had served on the editorial boards of several journals including *IEEE Transactions on Automatic Control*, *Systems & Control Letters*, and *Automatica*. He currently serves as an Editor-in-Chief of *Unmanned Systems* and a Deputy Editor-in-Chief of *Control Theory & Technology*.