

Research Article

Underwater Imaging Formation Model-Embedded Multiscale Deep Neural Network for Underwater Image Enhancement

Fucui Li ¹, Difei Lu ¹, ChengLang Lu ¹, and Qiuping Jiang ²

¹*School of Modern Information Technology, Zhejiang Institute of Mechanical and Electrical Engineering, Hangzhou, Zhejiang, China*

²*School of Information Science and Engineering, Ningbo University, Ningbo, Zhejiang, China*

Correspondence should be addressed to Qiuping Jiang; jiangqiuping@nbu.edu.cn

Received 15 June 2021; Revised 16 January 2022; Accepted 29 March 2022; Published 20 May 2022

Academic Editor: Abílio De Jesus

Copyright © 2022 Fucui Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Underwater image enhancement (UIE) plays an essential role in improving the quality of raw images captured in an underwater environment. Existing UIE methods can be categorized into two types: handcraft-designed and deep learning-based methods. Generally, the handcraft-designed methods are more explainable due to the leverage of knowledge-based image priors, while the deep learning-based methods are usually criticized for their weak interpretability. In this study, we address this issue by integrating the merits of both handcraft-designed and deep learning-based methods. Specifically, a physical underwater imaging model-inspired deep CNN for UIE is designed. Instead of estimating a global background light magnitude and a transmission matrix separately in traditional image restoration-based UIE methods, we directly generate a single variable as the joint estimation of these two parameters within a deep CNN and directly recover the enhanced image as an output according to a reformulated physical underwater imaging model. The whole network is trained in an end-to-end manner and more importantly has good interpretability. The proposed method has been validated for the UIE task on a real-world underwater image dataset and the experimental results well demonstrate the superiority of our method over the existing ones for UIE.

1. Introduction

Due to the effects of light absorption and scattering, the raw images captured in the underwater environment usually suffer from severe quality degradations. The suffered quality degradations greatly limit the utility of raw underwater images in practical autonomous underwater vehicles (AUVs), remotely operated vehicles (ROVs), and vessel detection in maritime video surveillance [1]. Therefore, underwater image enhancement (UIE) that aims to automatically improve the underwater image quality is a highly desired technique not only for generating visually appealing enhanced underwater images but also for boosting the performance of underwater vision tasks. Compared with in-air image enhancement, UIE is especially challenging due to the unknown and significantly varying ambient light in underwater environments. As shown in Figure 1, the raw underwater images captured in different environments can

have different degradation types, for example, color cast, reduced contrast, blurring details, and low luminance. To improve the quality of raw underwater images, during the past years, many UIE algorithms have been proposed and can be roughly categorized into two types: handcraft-designed and deep learning-based methods.

Traditional handcraft-designed UIE methods based on image enhancement and image restoration pipelines have long been explored during the past several years. Image enhancement-based UIE methods [2–7] directly manipulate image pixel intensities to enhance specific image characteristics, for example, color, contrast, sharpness, and brightness. However, these methods may not be effective for all underwater images due to their diverse and complicated degradation types. Image restoration-based UIE methods [16–31] treat image quality amelioration as an inverse problem where the physical underwater optical model is utilized and diverse image priors are explored as constraints

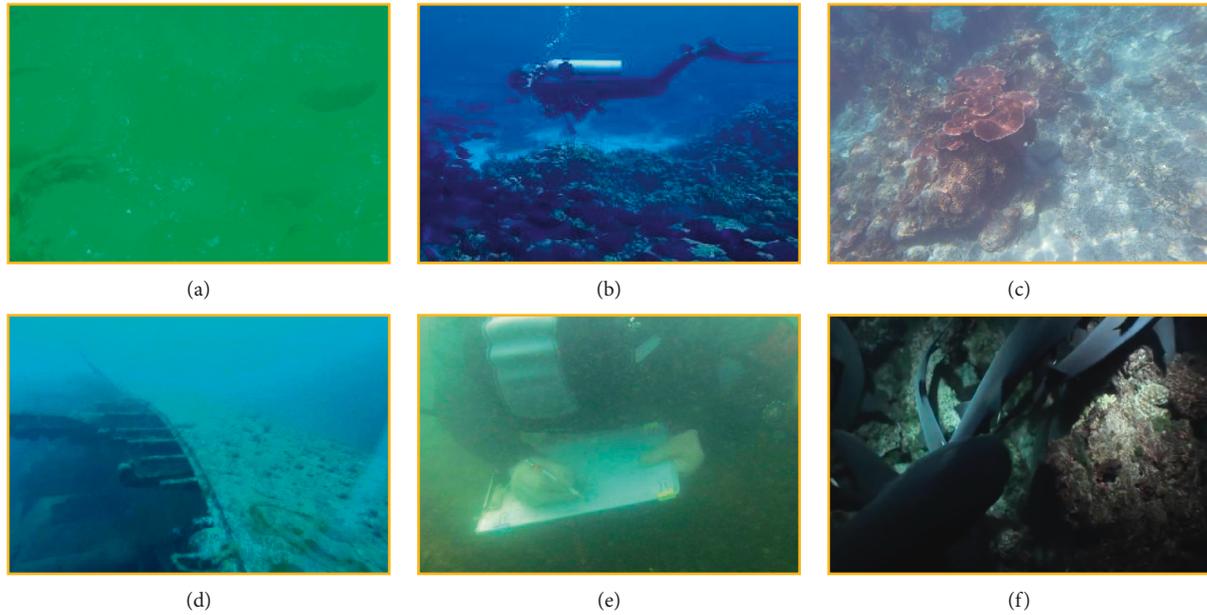


FIGURE 1: Degradation types of raw underwater images. (a), (b) Color casts (green and blue); (c) reduced contrast; (d), (e) blurring details; (f) low luminance.

to estimate a global background light magnitude and a transmission matrix involved in the physical underwater optical model. However, one inevitable limitation of these prior-based restoration methods for UIE is that these priors are invalid to some specific underwater environments and/or severe color casts. Therefore, the estimations of global background light and transmission matrix are not always accurate. Moreover, the separate/nonjoint estimation of these two parameters may even further amplify the error when putting them together.

During the past several years, the remarkable performance of deep convolutional neural networks (CNNs) in both low-level and high-level vision tasks has shed new light on UIE [8–15]. Typically, deep CNN training requires a large amount of data, either labeled or paired with ground truth. However, degraded underwater images are usually lacking of ground truth, which is a major hindrance to the same success of deep CNN-based methods for other low-level vision problems. To circumvent the problem of insufficient paired data, Shin et al. [8] synthesize hazed patches from haze-free image patches based on the classical atmospheric scattering model for underwater image dehazing. Similarly, an underwater CNN (UWCNN) model proposed in reference [9] is trained on a simulated underwater image dataset containing ten types of synthetic underwater images. Hu et al. [16] proposed a two-branch CNN, which is capable of separately removing the color cast and enhancing image contrast by fully leveraging useful properties of the HSV color space in disentangling chrominance and intensity. With the development of generative adversarial network (GAN), Li et al. [10] propose a WaterGAN to generate underwater images, taking in-air RGB-D images and a sample set of underwater images as inputs to train a generative network in an adversarial fashion. With the generated data, a two-stage color correction network is trained.

However, due to the complicated underwater imaging environment, the synthetic data are not able to cover all possible underwater conditions and the network trained on such synthetic data cannot generalize well to the real-world case. To circumvent this issue, recent works have taken the advantage of CycleGAN [11] and domain adaption to relax the need for one-to-one paired data for training [12, 13, 17]. However, these CycleGAN-based UIE methods tend to produce inauthentic results in some cases. Most recently, Li et al. [14] build a real-world Underwater Image Enhancement Benchmark (UIEB), containing 890 real-world underwater images and their pseudo reference images to enable training fully supervised deep CNN models, which are out of the constraints of specific underwater scenes. Based on the constructed UIEB, an end-to-end deep Water-Net is trained for UIE. Fu et al. [15] designed a two-branch network (i.e., GL-Net) to handle the globally distorted color and locally reduced contrast, respectively. Then, a compressed-histogram equalization is implemented for further improvement. Although the existing works have achieved significant progress, the UIE models based on deep learning always focus on learning a nonlinear mapping function from the degraded underwater image to its clean version and work as black boxes, which are usually criticized for their weak interpretability.

We claim that handcraft-designed and deep learning-based UIE algorithms have their own pros and cons; that is, handcraft-designed models are more explainable while deep learning-based models have much better nonlinear mapping capability. Therefore, it is of great interest to ask can we design a model-inspired deep CNN for UIE so that the merits of hand-craft designed and deep learning-based techniques are both inherited. With this consideration and inspired by the traditional image restoration-based UIE methods, a physical underwater imaging model-inspired

deep CNN for UIE is designed in this study. Instead of estimating a global background light magnitude and a transmission matrix separately (the common practice in traditional image restoration-based UIE methods), we directly generate a single variable as the joint estimation of these two parameters within a deep CNN and directly recover the enhanced image as an output, according to a reformulated physical underwater imaging model. The whole network is trained in an end-to-end manner and has good interpretability. The most prominent feature of this work is that data-driven deep learning and handcrafted image enhancement are integrated for improving the performance of UIE. The proposed method has been validated on two real-world underwater image datasets, that is, UIEB [14] and RUIE [18], and the experimental results demonstrate the superiority of our method both qualitatively and quantitatively.

2. Related Work

In recent years, with the rapid increase in underwater research, the demand for underwater image enhancement techniques has also been significantly increased, and a large number of underwater image enhancement methods have been invented. Generally, all the existing underwater image enhancement methods can be roughly divided into three categories: image enhancement-based, image restoration-based, and deep learning-based models. We will briefly review some of them in the next.

2.1. Image Enhancement-Based Model. The purpose of image enhancement technology is to produce visually pleasing results based on certain assumptions about high-quality images. Generally speaking, these methods usually focus on enhancing specific image attributes and directly manipulate the pixel intensity of the image, while ignoring the physical degradation model. In reference [2], this study introduces a fusion-based UIE method and shows its good performance in underwater image and video enhancement. First, the color-corrected image and contrast-enhanced image are estimated from the original underwater image. Then, a multiscale fusion scheme is proposed to fuse the two intermediate images into the final enhancement result. Experimental results show that the method can improve the global contrast and visibility and has high efficiency. Then, the method is further improved by introducing a new white balance scheme and an improved fusion strategy, as described in reference [3]. Based on the method proposed in reference [4], a new adaptive histogram method [5] is proposed to enhance the underwater image. The number of underenhanced and overenhanced pixels may be significantly limited. Another important branch of UIE based on image enhancement is to process underwater images, according to the simplified Retinex model. In reference [6], a UIE method based on variational retinoic acid is introduced, which includes three main steps: color correction, layer decomposition, and postenhancement. In reference [7], an extended

multiscale UIE method based on retina is introduced, and the mixture of milk and juice is used to simulate the underwater low visibility environment.

2.2. Image Restoration-Based Model. Different from the method based on image enhancement, the method based on image restoration regards underwater image enhancement as a typical ill-posed inverse problem. In this method, physical imaging models and various prior knowledge are explored to estimate the expected results from the degraded inputs. At present, the most widely used underwater imaging model is the so-called Jaffe–Mcglary underwater optical imaging model [19, 20], which is a simplified radiative transfer model. In fact, there are two variables to be determined: the global background light and the transmission matrix. In order to solve this inverse problem, many researchers are committed to exploring effective image priors to constrain this inverse problem. In reference [21], the authors use the fog line [22] as the constraint to derive the transmission matrix. In reference [23], the authors propose a new prior method to estimate the transmission matrix by using the attenuation difference between RGB channels, which can eliminate light scattering very well. By extending this prior, the author combines the fuzzy prior in reference [24] with the proposed prior in reference [23] to estimate the transmission. Because the Jaffe–Mcglary underwater optical imaging model is very similar to the hazy imaging model in reference [25], the popular dark channel prior (DCP) was originally proposed to solve the problem of single image deblurring. After modification and adjustment, the underwater image can be restored seamlessly, which has impressive performance. In reference [26], a wavelength-dependent compensation scheme based on DCP is proposed to restore clear images. In reference [27], the underwater dark channel prior is specially designed for the underwater environment. Compared with the traditional DCP, it can estimate the transmission more conveniently and accurately. According to the phenomenon that the observed red light component increases with distance, a short wavelength color recovery method based on a red light channel is proposed. In references [28, 29], the author uses the color lines [30] to derive the backscattered light and the modified DCP to estimate the transmittance, respectively. In reference [31], the contrast enhancement algorithm and image deblurring algorithm are combined to produce two kinds of enhancement results: one is bright color and the other is high contrast to reveal image details. Recently, a generalized dark channel prior is proposed in reference [32] to estimate the transmission and restoration of degraded images.

However, one of the unavoidable limitations of these prior-based UIE restoration methods is that they are ineffective for certain underwater environments and/or severe color transformations. Therefore, the estimation of global background light and transmission matrix is not always accurate. In addition, the separate/nonjoint estimation of these two parameters may further amplify the error of putting them together.

2.3. Deep Learning-Based Model. Different from handcrafted methods, deep learning technology aims to automatically extract representations and learn nonlinear mapping functions from training data. To the underwater community, several deep learning-based methods are also proposed to estimate the clear image. Typically, deep CNN training requires a large amount of data, either labeled or paired with ground truth. However, degraded underwater images are usually lacking of ground truth, which is a major hindrance to the same success of deep CNN-based methods for other low-level vision problems. To circumvent the problem of insufficient paired data, Shin et al. [8] synthesize hazed patches from haze-free image patches based on the classical atmospheric scattering model for underwater image dehazing. Similarly, an underwater CNN (UWCNN) model proposed in reference [9] is trained on a simulated underwater image dataset containing ten types of synthetic underwater images. With the development of generative adversarial network (GAN), Li et al. [10] proposed a WaterGAN to generate underwater images, taking in-air RGB-D images and a sample set of underwater images as inputs to train a generative network in an adversarial fashion. With the generated data, a two-stage color correction network is trained. However, due to the complicated underwater imaging environment, the synthetic data are not able to cover all possible underwater conditions and the network trained on such synthetic data cannot generalize well to the real-world case. To circumvent this issue, recent works have taken the advantage of CycleGAN [11] to relax the need for one-to-one paired data for training [12, 13]. However, these CycleGAN-based UIE methods tend to produce inauthentic results in some cases. Most recently, Li et al. [14] build a real-world Underwater Image Enhancement Benchmark (UIEB), containing 890 real-world underwater images and their pseudo reference images to enable training fully supervised deep CNN models, which are out of the constraints of specific underwater scenes. Based on the constructed UIEB, an end-to-end deep Water-Net is trained for UIE. Fu et al. [15] designed a two-branch network (i.e., GL-Net) to respectively handle the globally distorted color and locally reduced contrast, respectively. Then, a compressed-histogram equalization is implemented for further improvement. Although the existing works have achieved significant progress, the UIE models based on deep learning always focus on learning a nonlinear mapping function from the degraded underwater image to its clean version and work as black boxes, which are usually criticized for their weak interpretability.

3. Methodology

This section first introduces the widely used physical underwater imaging model. Then, the details of our proposed model including network architecture and implementations will be illustrated.

3.1. Physical Underwater Imaging Model. Prior works on image restoration-based UIE have focused on explicitly modeling the underwater degradation model and estimating

the model parameters from the observation and various priors to restore the clean underwater images. The physical underwater imaging model derived from the classical Jaffe-McGlamery model [19, 20] is expressed as follows:

$$I^c(p) = J^c(p)T^c(p) + A^c(1 - T^c(p)), \quad c \in \{r, g, b\}, \quad (1)$$

where $I^c(p)$ represents a pixel in the raw image, $J^c(p)$ represents a pixel in the clear image to be recovered, and c denotes the color channel. In the above model, there are two parameters to be determined for restoring J^c from I^c , that is, the global background light A^c and the transmission matrix T^c . Previous studies resort to various prior assumptions, for example, underwater dark channel prior [27], red channel prior [33], and adaptive attenuation-curve prior [34], as constraints to solve the two parameters separately. Nevertheless, these used priors are invalid to some specific underwater environments and/or severe color casts. Therefore, the accuracy of the estimated global background light and transmission matrix cannot be always guaranteed. Moreover, the separate estimation of these two parameters may even further amplify the error when combining them together during the recovery process.

In order to enable a joint estimation of the global background light A^c and the transmission matrix T^c , the above model can be reformulated by merging A^c and T^c into a single variable K^c [35], such that

$$J^c(p) = K^c(p)I^c(p) - K^c(p) + 1, \quad (2)$$

where

$$K^c(p) = \frac{(1/T^c(p))(I^c(p) - A^c) + (A^c - 1)}{I^c(p) - 1}. \quad (3)$$

According to the above-reformulated model, we only need to estimate K such that J can be directly recovered from I according to equation (2). Specifically, in this study, we resort to a multiscale densely connected deep CNN model to address this problem. Details regarding the proposed network are given below.

3.2. Network Architecture

3.2.1. Architecture Overview. The proposed network architecture is depicted in Figure 2. The ultimate goal of our network is to enhance the quality of an input raw underwater image I with certain quality degradation. Towards this end, we propose to first estimate a single intermediate variable K via a generator and recover a clear version J from I according to equation (2). Specifically, we design a Gaussian pyramid architecture to implement the generator at multiscales and the estimated K_s values in different scales are fused into a single one for recovering the final clear version according to equation (2). In addition, we also leverage the widely-adopted deep supervision strategy to add supervision on the recovered image at each scale to facilitate multiscale feature learning by allowing error information back-propagation from multiple paths and alleviating the problem of vanishing gradients in deep neural networks. With the

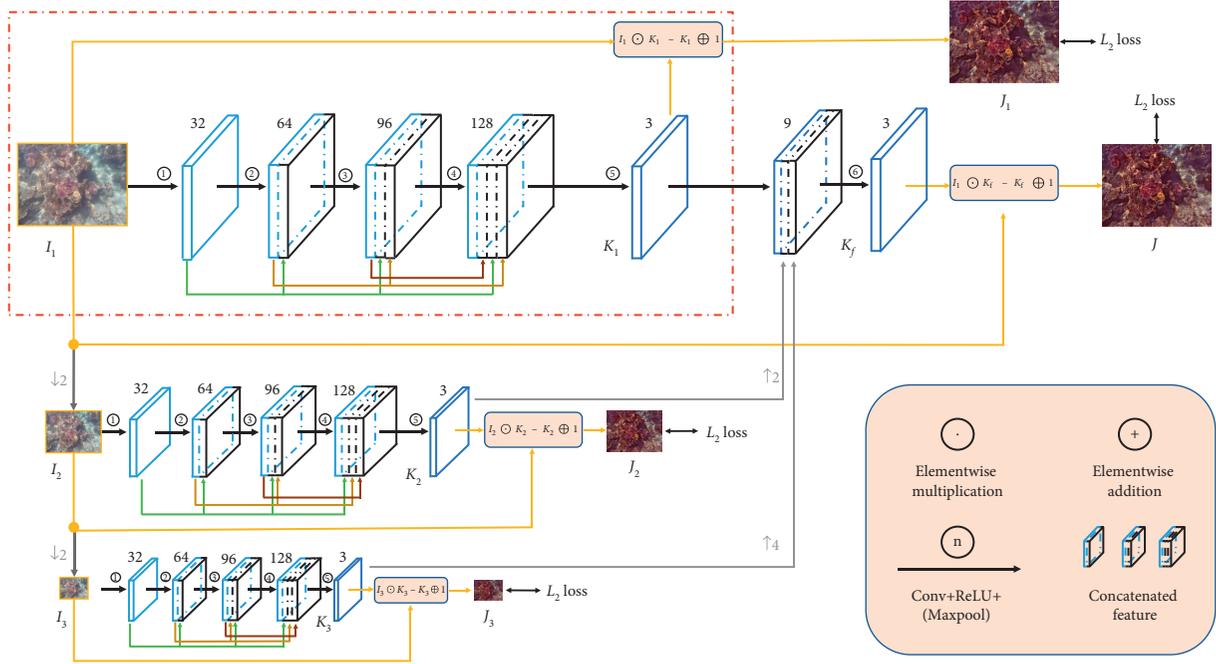


FIGURE 2: Architecture of our proposed model-inspired multiscale densely connected deep CNN model for UIE.

help of Gaussian pyramid architecture and multiscale supervisions, our proposed method is able to well handle the challenging UIE task. In the following sections, we will illustrate the details regarding the proposed deep neural network and its practical implementations and configurations.

3.2.2. Single-Scale Version. We first introduce the single-scale version. The architecture is shown in the rectangle region marked with red dashed lines (Figure 2). Specifically, a densely connected deep CNN is designed as the encoder to generate K , which is then fused with the raw underwater image I according to equation (2) to achieve UIE. Note that the explicit model-based operational formula (equation (2)) is embedded into the network architecture and the whole network is trained end to end. Therefore, compared with the traditional studies where the global background light and transmission matrix are estimated based on diverse image priors separately, the optimization of these two parameters is coupled and more importantly the parameter estimation and restoration process are jointly optimized in a data-driven manner. The utilization of dense connection aims to avoid gradient vanishing, strengthen the transmission of features, increase the utilization of features, and decrease the amount of parameters to some extent. Mathematically, the network architecture can be described as follows:

$$\begin{aligned} f^{(1)} &= \mathcal{O}^{(1)}(I; \theta^{(1)}), \\ f^{(l+1)} &= \mathcal{O}^{(l+1)}(\text{concat}(f^{(1)}, f^{(2)}, \dots, f^{(l)}); \theta^{(l+1)}), \end{aligned} \quad (4)$$

where $f^{(l)}$ denotes the output feature maps of the l -th block, $\mathcal{O}^{(l)}$ denotes the mapping function parameterized by $\theta^{(l)}$ of the l -th block built by a combination of convolutional layer,

ReLU layer, and max-pooling layer, and $\text{concat}(\cdot)$ means concatenation of all the features from $f^{(1)}$ to $f^{(l)}$. The output feature maps of the 5-th block are used as the estimated K , which is then fused with the raw underwater image I according to equation (2) to generate the final-enhanced underwater image J .

3.2.3. Multiscale Extension. In addition, on considering the scale variance problem in underwater images, we adopted a Gaussian pyramid architecture for multiscale fusion to handle this issue. The Gaussian pyramid architectures have been widely demonstrated to be effective in many deep learning-based low-level and high-level vision tasks. Leveraging this classical idea, CNN produces a feature pyramid through stacked convolutional layers and spatial pooling layers. Specifically, as shown in Figure 2, three different scales are considered in our framework, that is, the input underwater image is downsampled into two smaller scales, that is, 1/2 scale and 1/4 scale, respectively. For each scale, we use the same densely connected deep CNN encoder but the weights are not shared. Taking three-scale images as inputs, the estimated K maps corresponding to the three scales are denoted as K_1 , K_2 , and K_3 , respectively. Then, K_1 , K_2 , and K_3 are upsampled and concatenated followed by a convolution layer to generate the fused K map, that is, K_f . This process is formulated as follows:

$$\begin{aligned} K_C &= \text{concat}(K_1, K_2^{\times 2}, K_3^{\times 4}), \\ K_f &= \mathcal{O}^{(6)}(K_C; \theta^{(6)}), \end{aligned} \quad (5)$$

where $K_2^{\times 2}$ denotes the $\times 2$ upsampled version of K_2 and $K_3^{\times 4}$ denotes the $\times 4$ up-sampled version of K_3 . We implement downsampling and upsampling by bicubic interpolation. The

details of the multiscale network architecture have been listed in Table 1. Note that we only give the details of the first scale for brevity. The other two scales are almost the same with the first scale. The only difference is that the input sizes of the other two scales are changed into 64×64 and 32×32 , respectively.

Finally, we use K_f to generate the final-enhanced underwater image J_f according to equation (2), such that

$$J_f = K_f I - K_f + 1. \quad (6)$$

3.3. Loss Function. As previously illustrated, the whole network is trained and optimized with multiscale supervisions, that is, the supervisions are induced into each scale. Therefore, we in total have four objectives to minimize as follows: supervising three outputs at three different scales and the final output. For supervising the three-scale outputs, we have the following loss function:

$$L_s = \sum_{s=1}^3 \hat{J}_s - J_s(I_s; \theta_s)^2, \quad (7)$$

where I_s denotes the input image in the s -th scale, \hat{J}_s is the downsampled versions of the ground truth, θ_s denotes the parameters to be optimized, and J_s is the recovered image in the s -th scale. For the final output, we have

$$L_f = \hat{J}_1 - J_f(I; \theta)^2, \quad (8)$$

where J_f is the final-enhanced underwater image as an output. The final loss function can be expressed as follows:

$$L_{\text{total}} = L_f + \lambda_s L_s, \quad (9)$$

where λ_s controls the trade-off between the two terms. In our implementation, we empirically set $\lambda_s = 1$.

4. Experiments

In this section, we will conduct experiments to demonstrate the effectiveness of the proposed method. First, we will introduce the used benchmark datasets and the implementation details regarding the proposed deep neural network. Then, we conduct both qualitative and quantitative performance comparisons with state-of-the-art UIE algorithms. Finally, we also perform an application test to demonstrate whether the enhanced result by the proposed method can facilitate the application test performance, such as salient object detection and key point matching.

4.1. Dataset and Implementation Details

4.1.1. Datasets. For network training, following the strategy in reference [14], a random set of 800 images and their corresponding reference images collected from the UIEB dataset [14] are used to generate the original training set. Flipping and rotation are used for original training data augmentation. Each training image is resized into 128×128 . The remaining 90 images are used for testing. In addition, for

TABLE 1: Details of the multiscale network architecture.

Block	Input size	Operation	Configuration
1	$128 \times 128 \times 3$	Convolution	K3S1P1C32
	$128 \times 128 \times 32$	ReLU	—
2	$128 \times 128 \times 32$	Convolution	K3S1P1C32
	$128 \times 128 \times 32$	ReLU	—
	$128 \times 128 \times 32$	Maxpool	K3S1P1
3	$128 \times 128 \times 64$	Convolution	K3S1P1C32
	$128 \times 128 \times 32$	ReLU	—
	$128 \times 128 \times 32$	Maxpool	K5S1P1
4	$128 \times 128 \times 96$	Convolution	K3S1P1C32
	$128 \times 128 \times 32$	ReLU	—
	$128 \times 128 \times 32$	Maxpool	K7S1P1
5	$128 \times 128 \times 128$	Convolution	K3S1P1C32
	$128 \times 128 \times 3$	ReLU	—
6	$128 \times 128 \times 9$	Convolution	K3S1P1C32
	$128 \times 128 \times 3$	ReLU	—

Note. K, kernel size; S, stride; P, padding; C, channel.

more comprehensive performance validation, we also conduct experiments on the RUIE dataset [18], which is also a real-world dataset of underwater images. Note that we directly apply the model trained on the UIEB dataset to test the RUIE dataset. Such a strategy can well validate the robustness of UIE methods.

4.1.2. Implementation Details. We implemented our model with Pytorch [36] on a PC with an NVIDIA GeForce RTX 2080 Ti GPU. During training, a batch-mode learning strategy is used with a batch size of 48. The learning rate and momentum were set as 0.0001 and 0.9, respectively. All the parameters were optimized using ADAM [37] with default parameters. During testing, the test image is directly fed into the network without any resizing operation.

4.2. Qualitative Comparison. To demonstrate the advantages of the proposed method, several state-of-the-art UIE methods, such as Retinex [6], BL-TM [38], MILHP [31], UDCP [27], UW-CNN [9], and GL-Net [15], are compared. Among these methods, the last two are deep learning based proposed in the year 2020. Several examples of the enhanced results on the two datasets by different UIE methods are shown in Figures 3 and 4, respectively. As can be seen, raw underwater images suffer from diverse degradation. Even though the BL-TM [38], GL-Net [15], and MILHP [31] models indeed enhance the raw images to a certain extent, these three methods, however, result in overexposure in specific locations with very high luminance. The enhanced images by the Retinex [6] model seem to be of insufficient saturation, thus looking to be unnatural. The enhanced images by UDCP [27] suffer from severe detail loss in those areas with very low luminance. The enhanced images by UW-CNN [9] suffer from the overenhancement issue. On the contrary, our model produces more visually pleasing images that look better in color naturalness, contrast perception, and detail preservation.

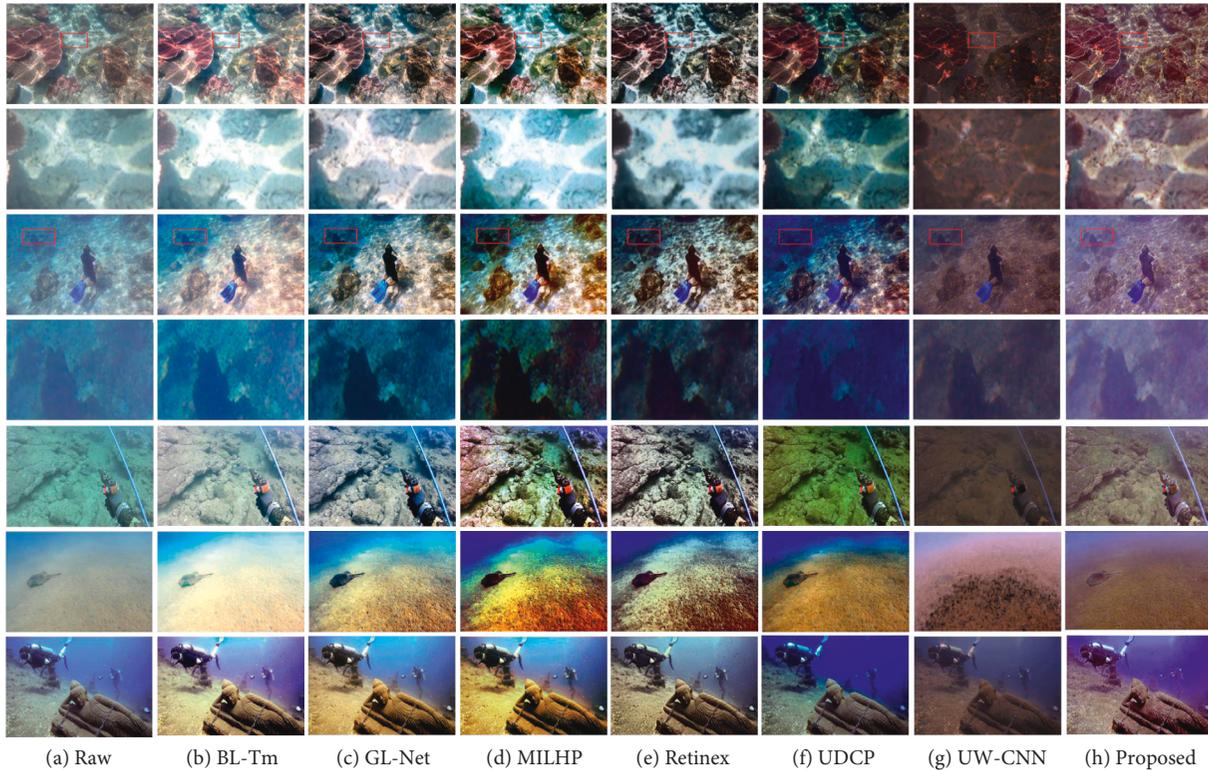


FIGURE 3: Examples of the enhanced results by different UIE methods on the UIEB dataset.

4.3. Quantitative Comparison. Besides the qualitative comparison, we also compare the performance of different UIE methods using two existing objective quality metrics dedicated designed for underwater images, such as UIQM [39] and UCIQE [40]. Note that, although there exist many full-reference image quality metrics [41–44] and no-reference image quality metrics [45–48] in the literature, they cannot be directly applied to evaluate the quality of enhanced underwater images due to the specific color distortion and contrast change suffered by underwater images. Therefore, we only use the two dedicated UIQM [39] and UCIQE [40] metrics for the quantitative comparison in this work. The average UIQM and UCIQE scores of different UIE methods are listed in Table 2. As we can see, our proposed method achieves the highest UIQM and UCIQE scores on the testing dataset, which further demonstrate the best performance of our proposed method.

4.4. User Study. For more comprehensive and reliable performance comparisons, we conduct a human subjective user study using all the testing images. Generally, there are two methodologies for subjective image quality assessment, such as absolute category rating and pairwise comparison. On considering that ACR will result in a huge bias and uncertainty when the observers do not have sufficient experience, and the pairwise comparison, which aims to provide a binary preference label between a pair of stimuli instead of rating an absolute quality level to a single stimulus, is adopted. Pairwise comparison is particularly useful when

the quality difference between stimuli is quite subtle, which is likely the case in our study. Specifically, $C_7^2 = 21$ image pairs are involved for each raw underwater image and $C_7^2 \times 90 = 1890$ pairs in total for all the testing images. Each subject is required to assign a binary label (1/0) to each presented image pair, reflecting which one is better in terms of color naturalness, contrast perception, and detail preservation. A total number of 20 subjects are invited to participate in our user study. To compute the global ranking from paired comparisons, we use the Bradley–Terry model (B-T model) [49], a probability model that predicts the outcome of the paired comparison. Finally, we evaluate the competing UIE methods by the average B-T scores, as shown in Figure 5. As seen, our method gets the highest B-T score of all the compared methods, which indicates that the enhanced images using our proposed method are of high visual quality.

4.5. Application Test. Besides visual quality comparison, we further demonstrate the effectiveness of our method for machine vision tasks. The application tests include key-point matching [50] and salient objection detection [51]. Here, we only show two examples for each test due to the limited space. As can be seen in Figure 6, the matched key points are greatly boosted after enhancement by our method. In Figure 7, the salient object detection results in the enhanced image are obviously better than that in the raw image. These two application tests demonstrate that our model works well in promoting the results of enhanced images in vision tasks.

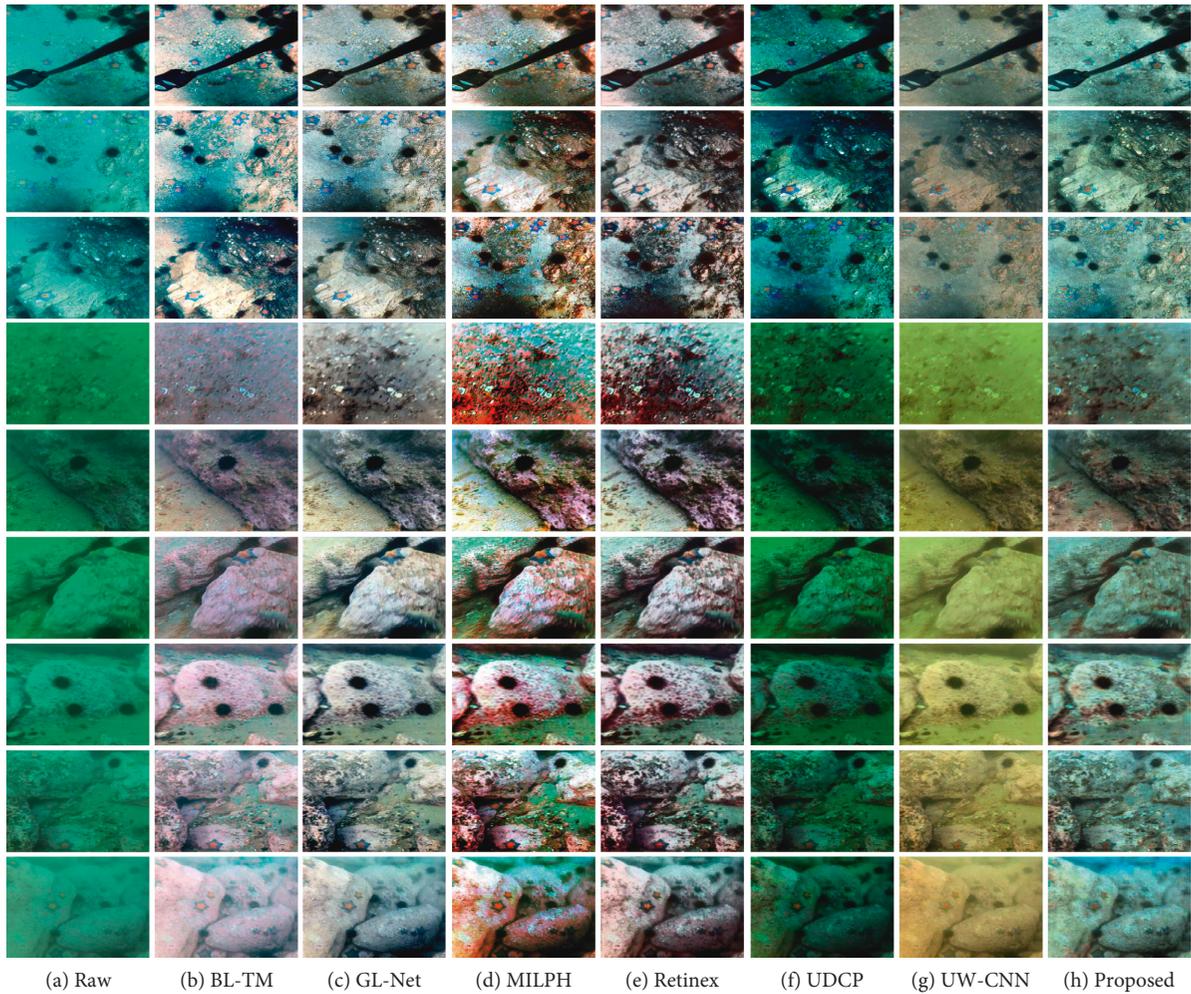


FIGURE 4: Examples of the enhanced results by different UIE methods on the RUIE dataset.

TABLE 2: Average UIQM and UCIQE scores of different UIE methods.

Datasets	UIEB		RUIE	
	UIQM (\uparrow)	UCIQE (\uparrow)	UIQM (\uparrow)	UCIQE (\uparrow)
Retinex [6]	4.8047	0.5824	4.2382	0.6055
BL-TM [38]	3.6668	0.6074	2.7331	0.5365
MILHP [31]	4.6453	0.6534	3.9178	0.6782
UDCP [27]	4.3117	0.6128	2.5631	0.5175
UW-CNN [9]	4.1060	0.4693	2.6658	0.4557
GL-Net [15]	4.4801	0.6253	3.7248	0.5780
Proposed	4.8052	0.6607	4.4432	0.6804

4.6. Influence of Scale Number. Since the proposed method utilizes a multiscale architecture, it is necessary to validate the selection of scale numbers. Here, we conducted experiments to validate the selection of scale numbers. The results are listed in Table 3. As we can find, with the increase of scale numbers,

the performance increases first and then becomes relatively stable when the scale number exceeds 3. Since the increase of scale numbers will cause heavier computational burden, we finally set the number of scales to 3 to achieve a balance between the performance and computational efficiency.

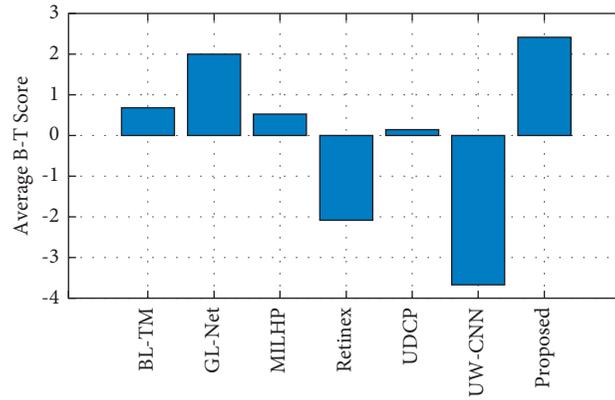


FIGURE 5: Average B-T scores of different UIE methods.

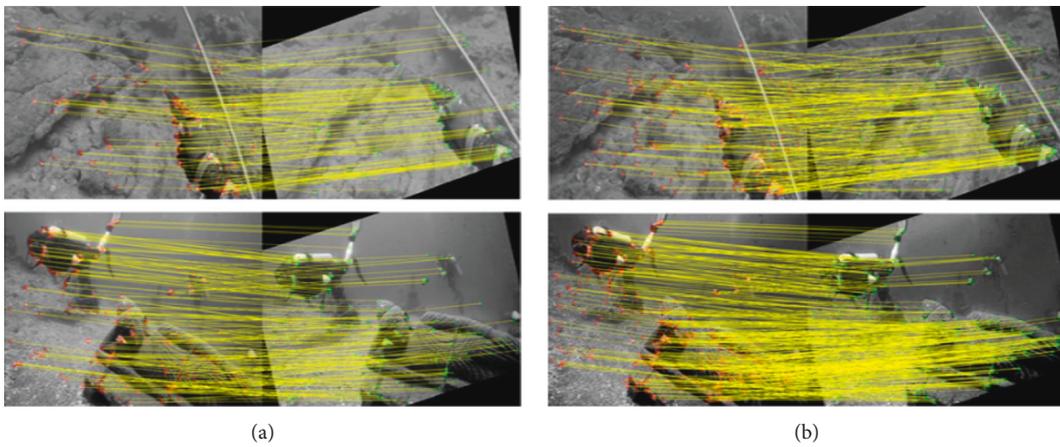


FIGURE 6: Comparisons on key-point matching results between the raw and enhanced images. (a) Raw. (b) Enhanced.

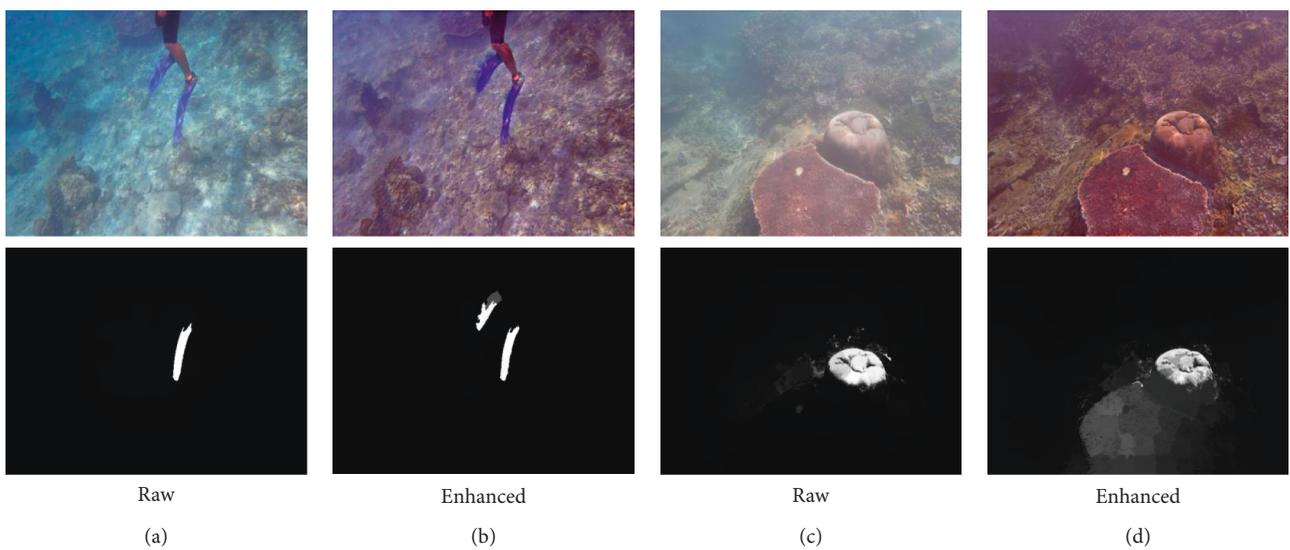


FIGURE 7: Comparisons on salient object detection results between the raw and enhanced images.

TABLE 3: Average UIQM and UCIQE scores with different scale numbers.

Datasets Scale numbers	UIEB		RUIE	
	UIQM (\uparrow)	UCIQE (\uparrow)	UIQM (\uparrow)	UCIQE (\uparrow)
1	4.4547	0.6024	4.0492	0.6162
2	4.7739	0.6381	4.2310	0.6575
3	4.8052	0.6607	4.4432	0.6804
4	4.8167	0.6728	4.4548	0.6911
5	4.8185	0.6743	4.4595	0.6989

5. Conclusion

In this study, we propose a multiscale densely connected deep CNN-based underwater image enhancement model inspired by the underwater optical imaging formulation. The most prominent feature of this work is that data-driven deep learning and handcrafted image enhancement are integrated for improving the performance of UIE. This model can not only promote the visual quality of raw underwater images but also improve the performance of the enhanced images in vision tasks. Our experiments including qualitative and quantitative comparisons, user study, and application tests demonstrate the superior performance of our model against other existing competing models.

Data Availability

UIEB dataset is available at “https://li-chongyi.github.io/proj_benchmark.html” and RUIE dataset is available at “<https://github.com/dlut-dimt/Realworld-Underwater-Image-Enhancement-RUIE-Benchmark>.”

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported in part by the Natural Science Foundation of Zhejiang Province under grant no. LR22F020002 and the Fundamental Research Funds for the Provincial Universities of Zhejiang under grant no. SJLZ2020003.

References

- [1] Y. Guo, Y. Lu, and R. W. Liu, “Lightweight deep network-enabled real-time low-visibility enhancement for promoting vessel detection in maritime video surveillance,” *Journal of Navigation*, vol. 75, no. 1, pp. 230–250, 2022.
- [2] C. Ancuti, C. O. Ancuti, T. Haber, and P. Bekaert, “Enhancing underwater images and videos by fusion,” *2012 IEEE Conference on Computer Vision and Pattern Recognition*, in *Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 81–88, RI, USA, June 2012.
- [3] C. O. Ancuti, C. Ancuti, C. De Vleeschouwer, and P. Bekaert, “Color balance and fusion for underwater image enhancement,” *IEEE Transactions on Image Processing*, vol. 27, no. 1, pp. 379–393, 2018.
- [4] K. Iqbal, M. Odetayo, A. James, R. Rosalina Abdul Salam, and A. Abdullah Zawawi Hj Talib, “Enhancing the low quality images using unsupervised colour correction method,” in *Proceedings of the 2010 IEEE International Conference on Systems, Man and Cybernetics*, pp. 1703–1709, Istanbul, Turkey, October 2010.
- [5] A. s. Abdul ghani and N. A. Mat Isa, “Underwater image quality enhancement through integrated color model with Rayleigh distribution,” *Applied Soft Computing*, vol. 27, pp. 219–230, 2015.
- [6] X. Fu, P. Zhuang, Y. Huang, Y. Liao, X.-P. Zhang, and X. Ding, “A retinex-based enhancing approach for single underwater image,” in *Proceedings of the 2014 IEEE International Conference on Image Processing (ICIP)*, pp. 4572–4576, Paris, France, October 2014.
- [7] S. Zhang, T. Wang, J. Dong, and H. Yu, “Underwater image enhancement via extended multi-scale retinex,” *Neurocomputing*, vol. 245, pp. 1–9, 2017.
- [8] Y.-S. Shin, Y. Cho, G. Pandey, and A. Kim, “Estimation of ambient light and transmission map with common convolutional architecture,” in *Proceedings of the OCEANS 2016 MTS/IEEE Monterey*, pp. 1–7, CA, USA, December 2016.
- [9] C. Li, S. Anwar, and F. Porikli, “Underwater scene prior inspired deep underwater image and video enhancement,” *Pattern Recognition*, vol. 98, Article ID 107038, 2020.
- [10] 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 Robotics and Automation Letters*, page, 2017.
- [11] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, “Unpaired image-to-image translation using cycle-consistent adversarial networks,” in *Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 2242–2251, Venice, Italy, October 2017.
- [12] C. Fabbri, M. J. Islam, and J. Sattar, “Enhancing underwater imagery using generative adversarial networks,” in *Proceedings of the 2018 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 7159–7165, Brisbane, QLD, Australia, May 2018.
- [13] C. Li, J. Guo, and C. Guo, “Emerging from water: underwater image color correction based on weakly supervised color transfer,” *IEEE Signal Processing Letters*, vol. 25, no. 3, pp. 323–327, 2018.
- [14] C. Li, C. Guo, W. Ren et al., “An underwater image enhancement benchmark dataset and beyond,” *IEEE Transactions on Image Processing*, vol. 29, pp. 4376–4389, 2020.
- [15] X. Fu and X. Cao, “Underwater image enhancement with global-local networks and compressed-histogram equalization,” *Signal Processing: Image Communication*, vol. 86, Article ID 115892, 2020.
- [16] J. Hu, Q. Jiang, R. Cong, W. Gao, and F. Shao, “Two-branch deep neural network for underwater image enhancement in hsv color space,” *IEEE Signal Processing Letters*, vol. 28, pp. 2152–2156, 2021.
- [17] Q. Jiang, Y. Zhang, F. Bao, X. Zhao, C. Zhang, and P. Liu, “Two-step domain adaptation for underwater image enhancement,” *Pattern Recognition*, vol. 122, Article ID 108324, 2022.
- [18] R. Liu, X. Fan, M. Zhu, M. Hou, and Z. Luo, “Real-world underwater enhancement: challenges, benchmarks, and solutions under natural light,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 12, pp. 4861–4875, 2020.

- [19] B. L. McGlamery, "A computer model for underwater camera systems," *SPIE Proceedings*, vol. 208, 1980.
- [20] J. S. Jaffe, "Computer modeling and the design of optimal underwater imaging systems," *IEEE Journal of Oceanic Engineering*, vol. 15, no. 2, pp. 101–111, 1990.
- [21] D. Menaker, T. Treibitz, and S. Avidan, "Color restoration of underwater images," *Proceedings of the British Machine Vision Conference 2017*, 2017.
- [22] D. Berman, T. Treibitz, and S. Avidan, "Non-local image dehazing," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, in *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1674–1682, Las Vegas, NV, USA, June 2016.
- [23] N. Carlevaris-Bianco, A. Mohan, and R. M. Eustice, "Initial results in underwater single image dehazing," in *Proceedings of the Oceans 2010 Mts/ieec Seattle*, pp. 1–8, Seattle, WA, USA, September 2010.
- [24] Y.-T. Peng and P. C. Cosman, "Underwater image restoration based on image blurriness and light absorption," *IEEE Transactions on Image Processing*, vol. 26, no. 4, pp. 1579–1594, 2017.
- [25] H. Koschmeider, "Therie der horizontalen sichtweite," *Beitr. Phys. Freien Atmosph.* vol. 12, 1924.
- [26] J. Y. Chiang and Y.-C. Ying-Ching Chen, "Underwater image enhancement by wavelength compensation and dehazing," *IEEE Transactions on Image Processing*, vol. 21, no. 4, pp. 1756–1769, 2012.
- [27] P. L. J. Drews, E. R. Nascimento, S. S. C. Botelho, and M. F. Montenegro Campos, "Underwater depth estimation and image restoration based on single images," *IEEE Computer Graphics and Applications*, vol. 36, no. 2, pp. 24–35, 2016.
- [28] H. Lu, Y. Li, L. Zhang, and S. Serikawa, "Contrast enhancement for images in turbid water," *Journal of the Optical Society of America A*, vol. 32, no. 5, p. 886, 2015.
- [29] Y. Zhou, Q. Wu, K. Yan, L. Feng, and W. Xiang, "Underwater image restoration using color-line model," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 29, no. 3, pp. 907–911, 2019.
- [30] R. Fattal, "Dehazing using color-lines," *ACM Transactions on Graphics*, vol. 34, no. 1, pp. 1–14, 2014.
- [31] C.-Y. Li, J.-C. Guo, R.-M. Cong, Y.-W. Pang, and B. Wang, "Underwater image enhancement by dehazing with minimum information loss and histogram distribution prior," *IEEE Transactions on Image Processing*, vol. 25, no. 12, pp. 5664–5677, 2016.
- [32] Y.-T. Peng, K. Cao, and P. C. Cosman, "Generalization of the dark channel prior for single image restoration," *IEEE Transactions on Image Processing*, vol. 27, no. 6, pp. 2856–2868, 2018.
- [33] A. Galdran, A. Alvarez-Gila, D. Pardo, and A. Picon, "Automatic red-channel underwater image restoration," *Journal of Visual Communication and Image representation*, vol. 26, 2015.
- [34] Y. Wang, H. Liu, and L.-P. Chau, "Single underwater image restoration using adaptive attenuation-curve prior," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 65, no. 3, pp. 992–1002, 2018.
- [35] B. Li, X. Peng, Z. Wang, J. Xu, and D. Feng, "Aod-net: all-in-one dehazing network," *2017 IEEE International Conference on Computer Vision (ICCV)*, in *Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 4780–4788, Venice, Italy, October 2017.
- [36] A. Paszke, S. Gross, F. Massa et al., "Pytorch: An Imperative Style, High-Performance Deep Learning Library," 2019, <https://arxiv.org/abs/1912.01703>.
- [37] D. Kingma and J. Ba, "A method for stochastic optimization," *International Conference on Learning Representations*, vol. 12, 2014.
- [38] 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 Transactions on Broadcasting*, vol. 66, no. 1, pp. 153–169, 2020.
- [39] K. Panetta, C. Gao, and S. Agaian, "Human-visual-system-inspired underwater image quality measures," *IEEE Journal of Oceanic Engineering*, vol. 41, no. 3, pp. 541–551, 2016.
- [40] M. Yang and A. Sowmya, "An underwater color image quality evaluation metric," *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 6062–6071, 2015.
- [41] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.
- [42] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Transactions on Image Processing*, vol. 15, no. 2, pp. 430–444, 2006.
- [43] D. M. Chandler and D. M. Chandler, "Most apparent distortion: full-reference image quality assessment and the role of strategy," *Journal of Electronic Imaging*, vol. 19, no. 1, Article ID 011006, 2010.
- [44] W. Xue, L. Zhang, X. Mou, and A. C. Bovik, "Gradient magnitude similarity deviation: a highly efficient perceptual image quality index," *IEEE Transactions on Image Processing*, vol. 23, no. 2, pp. 684–695, 2014.
- [45] Q. Jiang, W. Gao, S. Wang et al., "Blind image quality measurement by exploiting high-order statistics with deep dictionary encoding network," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 10, pp. 7398–7410, 2020.
- [46] W. Xue, X. Mou, L. Zhang, A. C. Bovik, and X. Feng, "Blind image quality assessment using joint statistics of gradient magnitude and Laplacian features," *IEEE Transactions on Image Processing*, vol. 23, no. 11, pp. 4850–4862, 2014.
- [47] P. Peng Ye, J. Kumar, L. Le Kang, and D. Doermann, "Un-supervised feature learning framework for no-reference image quality assessment," in *Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1098–1105, RI, USA, June 2012.
- [48] J. Xu, P. Ye, Q. Li, H. Du, Y. Liu, and D. Doermann, "Blind image quality assessment based on high order statistics aggregation," *IEEE Transactions on Image Processing*, vol. 25, no. 9, pp. 4444–4457, 2016.
- [49] R. A. Bradley and M. E. Terry, "Rank analysis of incomplete block designs: I. the method of paired comparisons," *Biometrika*, vol. 39, no. 3/4, p. 324, 1952.
- [50] H. Bay, T. Tuytelaars, and L. Van Gool, "SURF: speeded up robust features," *Computer Vision - ECCV 2006*, vol. 3951, pp. 404–417, 2006.
- [51] H. Peng, B. Li, H. Ling, W. Hu, W. Xiong, and S. J. Maybank, "Salient object detection via structured matrix decomposition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 4, pp. 818–832, 2017.