Underwater Image Enhancement via Medium Transmission-Guided Multi-Color Space Embedding

Chongyi Li, Saeed Anwar, *Member, IEEE*, Junhui Hou, *Senior Member, IEEE*, Runmin Cong, *Member, IEEE*, Chunle Guo, and Wenqi Ren, *Member, IEEE*

***Abstract*— Underwater images suffer from color casts and low contrast due to wavelength- and distance-dependent attenuation and scattering. To solve these two degradation issues, we present an underwater image enhancement network via medium transmission-guided multi-color space embedding, called *Ucolor*. Concretely, we first propose a multi-color space encoder network, which enriches the diversity of feature representations by incorporating the characteristics of different color spaces into a unified structure. Coupled with an attention mechanism, the most discriminative features extracted from multiple color spaces are adaptively integrated and highlighted. Inspired by underwater imaging physical models, we design a medium transmission (indicating the percentage of the scene radiance reaching the camera)-guided decoder network to enhance the response of network towards quality-degraded regions. As a result, our network can effectively improve the visual quality of underwater images by exploiting multiple color spaces embedding and the advantages of both physical model-based and learning-based methods. Extensive experiments demonstrate that our *Ucolor* achieves superior performance against state-of-the-art methods in terms of both visual quality and quantitative metrics. The code is publicly available at: https://li-chongyi.github.io/Proj_Ucolor.html.**

***Index Terms*— Underwater imaging, image enhancement, color correction, scattering removal.**

Manuscript received August 17, 2020; revised February 2, 2021 and March 28, 2021; accepted April 15, 2021. Date of publication May 7, 2021; date of current version May 17, 2021. This work was supported in part by the Hong Kong RGC under Grant CityU 21211518, Grant 11219019, and Grant 11202320; in part by the Hong Kong GRF-RGC General Research Fund under Grant 9042958 (CityU 11203820) and Grant 9042816 (CityU 11209819); in part by the Beijing Nova Program under Grant Z201100006820016; in part by the National Natural Science Foundation of China under Grant 62002014; and in part by the Young Elite Scientist Sponsorship Program by the China Association for Science and Technology under Grant 2020QNRC001. The work of Chunle Guo was supported by the CAAI-Huawei MindSpore Open Fund. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Abdesselam S. Bouzerdoum. *(Corresponding author: Junhui Hou.)*

Chongyi Li and Junhui Hou are with the Department of Computer Science, City University of Hong Kong, Hong Kong (e-mail: lichongyi25@gmail.com; jh.hou@cityu.edu.hk).

Saeed Anwar is with the Data61, Commonwealth Scientific and Industrial Research Organization (CSIRO), Clayton South, VIC 3169, Australia, and also with the Research School of Engineering, The Australian National University (ANU), Canberra, ACT 2600, Australia (e-mail: saeed.anwar@csiro.au).

Runmin Cong is with the Institute of Information Science, Beijing Jiaotong University, Beijing 100044, China, and also with the Beijing Key Laboratory of Advanced Information Science and Network Technology, Beijing Jiaotong University, Beijing 100044, China (e-mail: rmcong@bjtu.edu.cn).

Chunle Guo is with the College of Computer Science, Nankai University, Tianjin 300071, China (e-mail: guochunle@nankai.edu.cn).

Wenqi Ren is with the State Key Laboratory of Information Security, Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100093, China (e-mail: rwq.renwenqi@gmail.com).

This article has supplementary downloadable material available at https://doi.org/10.1109/TIP.2021.3076367, provided by the authors.

Digital Object Identifier 10.1109/TIP.2021.3076367

I. Introduction

UNDERWATER images inevitably suffer from quality degradation issues caused by wavelength- and distance-dependent attenuation and scattering [1]. Typically, when the light propagates through water, it suffers from selective attenuation that results in various degrees of color deviations. Besides, the light is scattered by suspending particles such as micro phytoplankton and non-algal particulate in water, which causes low contrast. An effective solution to recover underlying clean images is of great significance for improving the visual quality of images captured in water and accurately understanding underwater world.

The quality degradation degrees of underwater images can be implicitly reflected by the medium transmission that represents the percentage of the scene radiance reaching the camera. Hence, physical model-based underwater image enhancement methods [2]–[8] mainly focus on the accurate estimation of medium transmission. With the estimated medium transmission and other key underwater imaging parameters such as the homogeneous background light, a clean image can be obtained by reversing an underwater imaging physical model. Though physical model-based methods can achieve promising performance in some cases, they tend to produce unstable and sensitive results when facing challenging underwater scenarios. This is because 1) estimating the medium transmission is fundamentally ill-posed, 2) estimating multiple underwater imaging parameters is knotty for traditional methods, and 3) the assumed underwater imaging models do not always hold.

Recently, deep learning technology has shown impressive performance on underwater image enhancement [10]–[12]. These deep learning-based methods often apply the networks that were originally designed for other visual tasks to underwater images. Thus, their performance is still far behind when compared with current deep visual models [13]–[16]. The main reason is that the design of current deep underwater

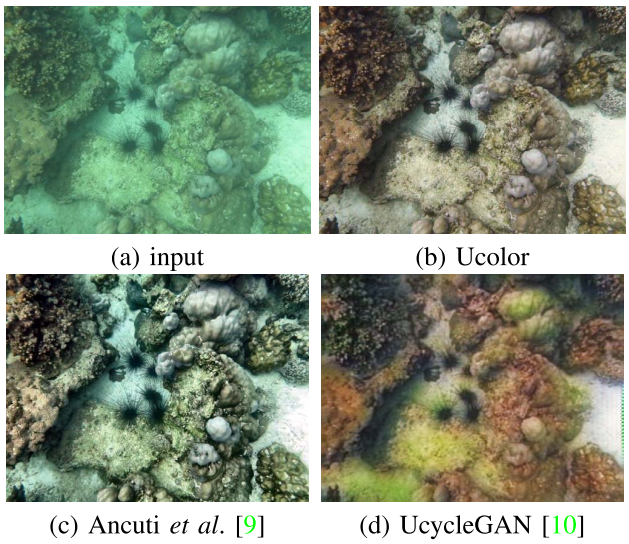


Fig. 1. Visual comparisons on a real underwater image. Our Ucolor removes both the greenish color deviation and the effect of scattering. In contrast, the compared methods either remain the color deviation or introduce extra color artifacts.

image enhancement models neglects the domain knowledge of underwater imaging.

In this work, we propose to solve the issues of color casts and low contrast of underwater images by leveraging rich encoder features and exploiting the advantages of physical model-based and learning-based methods. Unlike previous deep models [10], [17], [18] that only employ the features extracted from RGB color space, we examine the feature representations through a multi-color space encoder network, then highlight the most representative features via an attention mechanism. Such a manner effectively improves the generalization capability of deep networks, and also incorporates the characteristics of different color spaces into a unified structure. This is rarely studied in the context of underwater image enhancement. Inspired by the conclusion that the quality degradation of underwater images can be reflected by the medium transmission [19], we propose a medium transmission-guided decoder network to enhance the response of our network towards quality-degraded regions. The introduction of medium transmission allows us to incorporate the advantage of physical model-based methods into deep networks, which accelerates network optimization and improves enhancement performance. Since our method is purely data-driven, it can tolerate the errors caused by inaccurate medium transmission estimation.

In Fig. 1, we present a representative example by the proposed Ucolor against two underwater image enhancement methods. As shown, both the classical fusion-based method [9] (Fig. 1(c)) and the deep learning-based method [10] (Fig. 1(d)) fail to cope with the challenging underwater image with greenish tone and low contrast well. In contrast, our Ucolor (Fig. 1(b)) achieves the visually pleasing result in terms of color, contrast, and naturalness. The main contributions of this paper are highlighted as follows.

- We propose a multi-color space encoder network coupled with an attention mechanism for incorporating the characteristics of different color spaces into a unified structure and adaptively selecting the most representative features.
- We propose a medium transmission-guided decoder network to enforce the network to pay more attention to quality-degraded regions. It explores the complementary merits between domain knowledge of underwater imaging and deep neural networks.
- Our Ucolor achieves state-of-the-art performance on several recent benchmarks in terms of both visual quality and quantitative metrics.

II. Related Work

In addition to extra information [20], [21] and specialized hardware devices [22], underwater image enhancement can be roughly classified into two groups: traditional methods and deep learning-based methods.

Traditional Methods. Early attempts aim to adjust the pixel values for visual quality improvement, such as dynamic pixel range stretching [23], pixel distribution adjustment [24], and image fusion [9], [25], [26]. For example, Ancuti *et al*. [9] first obtained the color-corrected and contrast-enhanced versions of an underwater image, then computed the corresponding weight maps, finally combined the advantages of different versions. Ancuti *et al*. [25] further improved the fusion-based underwater image enhancement strategy and proposed to blend two versions that are derived from a white-balancing algorithm based on a multiscale fusion strategy. Most recently, based on the observation that the information contained in at least one color channel is close to completely lost under adverse conditions such as hazy nighttime, underwater, and non-uniform artificial illumination, Ancuti *et al*. [27] proposed a color channel compensation (3C) pre-processing method. As a pre-processing step, the 3C operator can improve traditional restoration methods.

Although these physical model-free methods can improve the visual quality to some extent, they omit the underwater imaging mechanism and thus tend to produce either over-/under-enhanced results or introduce artificial colors. For example, the color correction algorithm in [9] is not always reliable when encountering diverse and challenging underwater scenes. Compared with these methods, our method takes underwater image formation models into account and employs the powerful learning capability of deep networks, making the enhanced images look more natural and visually pleasing.

The widely used underwater image enhancement methods are physical model-based, which estimate the parameters of underwater imaging models based on prior information. These priors include red channel prior [28], underwater dark channel prior [3], minimum information prior [4], blurriness prior [29], general dark channel prior [30], *etc*. For example, Peng and Cosman [29] proposed an underwater image depth estimation algorithm based on image blurriness and light absorption. With the estimated depth, the clear underwater image can be restored based on an underwater imaging model. Peng *et al*. [30] further proposed a generalization of the dark channel prior to deal with diverse images captured under severe weather. A new underwater image formation model was proposed in [19]. Based on this model, an underwater

image color correction method was presented using underwater RGB-D images [21].

These physical model-based methods are either time-consuming or sensitive to the types of underwater images [31]. Moreover, the accurate estimation of complex underwater imaging parameters challenges current physical model-based methods [3], [4], [28]–[30]. For example, the blurriness prior used in [29] does not always hold, especially for clear underwater images. In contrast, our method can more accurately restore underwater images by exploiting the advantages of both physical model-based and data driven-based methods.

Deep Learning Models. The emergence of deep learning has led to considerable improvements in low-level visual tasks [32]–[36]. There are several attempts made to improve the performance of underwater image enhancement through deep learning strategy [37]. As a pioneering work, Li *et al.* [11] employed a Generative Adversarial Network (GAN) and an image formation model to synthesize degraded/clean image pairs for supervised learning. To avoid the requirement of paired training data, a weakly supervised underwater color correction network (UCycleGAN) was proposed in [10]. Furthermore, Guo *et al.* [17] introduced a multi-scale dense GAN for robust underwater image enhancement. Li *et al*. [12] proposed to simulate the realistic underwater images according to different water types and an underwater imaging physical model. With ten types of synthesized underwater images, ten underwater image enhancement (UWCNN) models were trained, in which each UWCNN model was used to enhance the corresponding type of underwater images. Recently, Li *et al.* [18] collected a real paired underwater image dataset for training deep networks and proposed a gated fusion network to enhance underwater images. This proposed gated deep model requires three preprocessing images including a Gamma correction image, a contrast improved image, and a white-balancing image as the inputs of the gated network. A wavelet corrected transformation was proposed for underwater image enhancement in [38]. Yang *et al*. [39] proposed a conditional generative adversarial network to improve the perceptual quality of underwater images.

These underwater image enhancement models usually apply the existing deep network structures for general purposes to underwater images and neglect the unique characteristics of underwater imaging. For example, [10] directly uses the CycleGAN [34] network structure, and [18] adopts a simple multi-scale convolutional network. For unsupervised models [10], [17], they still inherit the disadvantage of GAN-based models, which produces unstable enhancement results. In [12], facing an input underwater image, how to select the corresponding UWCNN model is challenging. Consequently, the robustness and generalization capability of current deep learning-based underwater image enhancement models are limited and unsatisfactory.

In contrast to existing deep learning-based underwater image enhancement methods, our method has the following unique characteristics: 1) the multi-color space encoder network coupled with an attention mechanism that enables the diverse feature representations from multi-color space and adaptively selects the most representative information; 2) the medium transmission-guided decoder network that incorporations the domain knowledge of underwater imaging into deep structures by tailoring the attention mechanism for emphasizing the quality-degraded regions; 3) our method does not require any pre-processing steps and adopts supervised learning, thus producing more stable results; 4) our method adopts end-to-end training and is able to handle most underwater scenes in a unified structure; and 5) our method achieves outstanding performance on various underwater image datasets. These innovations provide new ideas for exploring the complementary merits between domain knowledge of underwater imaging and deep learning strategy and the advantages of multi-color space encoder.

III. Proposed Method

We present the overview architecture of Ucolor in Fig. 2. **In the multi-color space encoder network**, an underwater image first goes through color space transformation. Three encoder paths named HSV path, RGB path, and Lab path are formed. In each path, the input is forwarded to three serial redisual-enhancement modules, thus obtaining three levels of feature representations using a $2\times$ downsampling operation (noted Fig. 2 as 1, 2, and 3). Simultaneously, we enhance the RGB path by densely connecting the features of the RGB path with the corresponding features of the HSV path and the Lab path. We then concatenate the same level features of these three parallel paths to form three sets of multi-color space encoder features. At last, we separately feed these three sets of features to the corresponding channel-attention module that serves as a tool to spotlight the most representative and informative features. **In the medium transmission-guided decoder network**, the selected encoder features by channel-attention modules and the same sizes of reverse medium transmission (RMT) map are forwarded to the medium transmission guidance module for emphasizing quality-degraded regions. Here, we employ the max pooling operation to achieve different sizes of RMT maps. Then, the outputs of the medium transmission guidance modules are fed to the corresponding residual-enhancement module. After three serial residual-enhancement modules and two $2\times$ upsampling operations, the decoder features are forwarded to a convolution layer for reconstructing the result.

In what follows, we detail the key components of our method, including the multi-color space encoder (Sec. III-A), the residual-enhancement module (Sec. III-B), the channel-attention module (Sec. III-C), the medium transmission guidance module (Sec. III-D), and the loss function (Sec. III-E).

A. Multi-Color Space Encoder

Compared with terrestrial scene images, the color deviations of underwater images cover more comprehensive ranges, differing from the bluish or greenish tone to a yellowish one. The diversity in color casts severely limits the traditional network

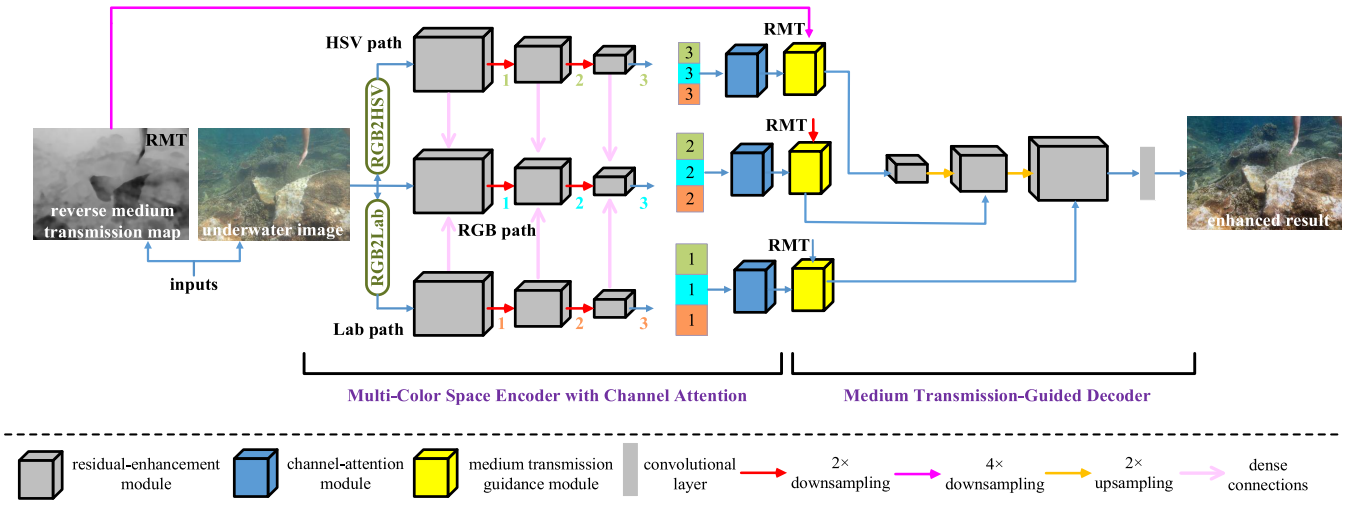


Fig. 2. Overview of the architecture of Ucolor. Our Ucolor consists of a multi-color space encoder network and a medium transmission-guided decoder network. In our method, we normalize the values of the medium transmission map to [0,1] and feed the reverse medium transmission map (denoted as RMT) to the medium transmission guidance module. 'downsampling' is implemented by max pooling, while 'upsampling' is implemented by bilinear interpolation. 'dense connections' represents the concatenation operation along the channel dimension for each set of features from the corresponding convolutional layer in different color-space encoder paths. 'convolutional layer' has the kernel of size 3×3 and stride 1. In the Ucolor, all convolutional layers adopt kernels of size 3×3 and stride 1. A detailed network structure with the hyper-parameters can be found in the Supplementary Material.

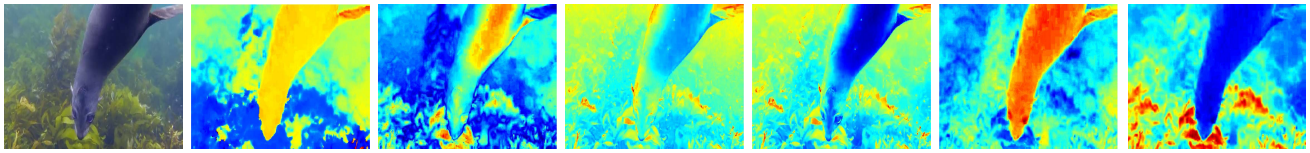


Fig. 3. A visual example of the channels extracted from different color spaces. For visualization, we normalize the values to the range of [0,1]. The channels in (b)-(g) are represented by heatmaps, where the color ranging from blue to red represents the value from small to large.

architectures [40], [41]. Inspired by the traditional enhancement algorithms that operate in various color spaces [23], [42], we extract features in three color spaces (RGB, HSV, and Lab) where the same image has different visual representations in various color systems as shown in Fig. 3.

Concretely, the image is easy to store and display in RGB color space because of its strong color physical meaning. However, the three components (R, G, and B) are highly correlated, which are easy to be affected by the changes of luminance, occlusion, shadow, and other factors. By contrast, HSV color space can intuitively reflect the hue, saturation, brightness, and contrast of the image. Lab color space makes the colors better distributed, which is able to express all the colors that the human eye can perceive.

These color spaces have obvious differences and advantages. To combine their properties for underwater image enhancement, we incorporate the characteristics of different color spaces into a unified deep structure, where all the image degradation related components (color, hue, saturation, intensity, and luminance) can be taken into account. Moreover, the color difference of two points with a small distance in one color space may be large in other color spaces. Thus, the multiple color spaces embedding can facilitate the measurement of color deviations of underwater images. Additionally, the multi-color space encoder brings more nonlinear operations during color space transformation. It is known that the nonlinear transformation generally improves the performance of deep models [43]. In the ablation study, we will analyze the contribution of each color space.

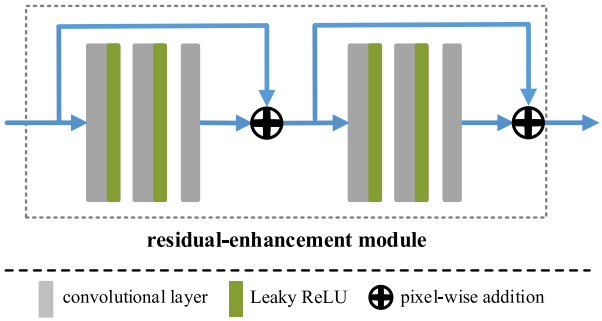


Fig. 4. The schematic illustration of the residual-enhancement module. Each residual-enhancement module is composed of two residual blocks, where each block is built by three stacked convolutions followed by the Leaky ReLU activation function, except for the last one. After each residual block, a pixel-wise addition is used as an identity connection.

B. Residual-Enhancement Module

Fig. 4 presents the details of the residual-enhancement module. This residual-enhancement module aims to preserve the data fidelity and address gradient vanishing [44]. In each residual-enhancement module, the convolutional layers have

an identical number of filters. The numbers of filters are progressively increased from 128 to 512 by a factor 2 in the encoder network while they are decreased from 512 to 128 by a factor 2 in the decoder network. All the convolutional layers have the same kernel sizes of 3×3 and stride 1.

C. Channel-Attention Module

In view of the specific definition of each color space, these features extracted from three color spaces should have different contributions. Therefore, we employ a channel-attention module to explicitly exploit the interdependencies between the channel features extracted from different color spaces. The details of the channel-attention module are depicted in Fig. 5.

Assume the input features $\mathcal{F} = \mathsf{Cat}(\mathbf{F}_1, \mathbf{F}_2, \cdots, \mathbf{F}_N) \in \mathbb{R}^{N \times H \times W}$, where $\mathbf{F}$ is a feature map from one path at a specific level (the level is denoted as 1, 2, and 3 in Fig. 2), N is the number of feature maps, Cat represents the feature concatenation; and H and W are the height and width of input image, respectively. We first perform the global average pooling on input features $\mathcal{F}$, leading to a channel descriptor $\mathbf{z} \in \mathbb{R}^{N \times 1}$, which is an embedded global distribution of channel-wise feature responses. The k-th entry of $\mathbf{z}$ can be expressed as:

$$z_k = \frac{1}{H \times W} \sum_{i}^{H} \sum_{j}^{W} \mathbf{F}_k(i, j), \tag{1}$$

where $k \in [1, N]$. To fully capture channel-wise dependencies, a self-gating mechanism [45] is used to produce a collection of per-channel modulation weights $\mathbf{s} \in \mathbb{R}^{N \times 1}$:

$$\mathbf{s} = \sigma(\mathbf{W}_2 * (\delta(\mathbf{W}_1 * \mathbf{z}))), \tag{2}$$

where $\sigma(\cdot)$ represents the Sigmoid activation function, $\delta(\cdot)$ represents the ReLU activation function, $*$ denotes the convolution operation, and $\mathbf{W}_1$ and $\mathbf{W}_2$ are the weights of two fully-connected layers with the numbers of their output channels equal to $\frac{N}{r}$ and N, respectively, where r is set to 16 for reducing the computational costs. At last, these weights are applied to input features $\mathcal{F}$ to generate rescaled features $\mathcal{U} \in \mathbb{R}^{N \times H \times W}$. Moreover, to avoid gradient vanishing problem and keep good properties of original features, we treat the channel-attention weights in an identical mapping fashion:

$$\mathcal{U} = \mathcal{F} \oplus \mathcal{F} \otimes \mathbf{s}, \tag{3}$$

where $\oplus$ and $\otimes$ denote the pixel-wise addition and pixel-wise multiplication, respectively.

D. Medium Transmission Guidance Module

According to the image formation model in bad weather [46], [47], which is widely used in image dehazing and underwater image restoration algorithms [2], [29], [30], the quality-degraded image can be expressed as:

$$I^c(x) = J^c(x) \otimes T(x) \oplus A^c(x) \otimes (1 - T(x)), c \in \{r, g, b\}, \tag{4}$$

where x indicates the pixel index, I is the observed image, J is the clear image, A is the homogeneous background light, and T is the medium transmission that represents the percentage of scene radiance reaching the camera after reflecting in the medium, indicating the degrees of quality degradation in different regions.

We incorporate the medium transmission map into the decoder network via the proposed medium transmission guidance module. Specifically, we use the reverse medium transmission (RMT) map (denoted as $\overline{T} \in \mathbb{R}^{H \times W}$) as the pixel-wise attention map. The RMT map $\overline{T}$ is obtained by $\mathbf{1}$-T ($T \in \mathbb{R}^{H \times W}$ is the medium transmission map in the range of [0,1], and $\mathbf{1} \in \mathbb{R}^{H \times W}$ is the matrix with all elements equal to 1), which indicates that the higher quality degradation pixels should be assigned larger attention weights.

Since the corresponding ground truth medium transmission map of an input underwater image is not available in practice, it is difficult to train a deep neural network for the estimation of medium transmission map. To solve this issue, we employ prior-based estimation algorithms to obtain the medium transmission map. Inspired by the robust general dark channel prior [30], we estimate the medium transmission map as:

$$\tilde{T}(x) = \max_{c, y \in \Omega(x)} \left(\frac{A^c - I^c(y)}{\max(A^c, 1 - A^c)}\right), \tag{5}$$

where $\tilde{T}$ is the estimated medium transmission map, $\Omega(x)$ represents a local patch of size 15×15 centered at x, and c denotes the color channel. As shown, the medium transmission estimation is related to the homogeneous background light A. In [30], the homogeneous background light is estimated based on the depth-dependent color change. Due to the limited space, we refer the readers to [30] for more details. We will compare and analyze the effects of the medium transmission maps estimated by different algorithms in the ablation study.

With the RMT map, the schematic illustration of the proposed medium transmission guidance module is shown in Fig. 6. As shown, we utilize the RMT map as a feature selector to weight the importance of different spatial positions of the features. The high-quality degradation pixels (the pixels with larger RMT values) are assigned higher weights, which can be expressed as:

$$\mathcal{V} = \mathcal{U} \oplus \mathcal{U} \otimes \overline{T}, \tag{6}$$

where $\mathcal{V} \in \mathbb{R}^{M \times H \times W}$ and $\mathcal{U} \in \mathbb{R}^{M \times H \times W}$ respectively represent the output features after the medium transmission guidance module and the input feature. We treat the RMT weights as an identity connection to avoid gradient vanishing and tolerate the errors caused by inaccurate medium transmission estimation. Besides, our purely data-driven framework also tolerates the inaccuracy of medium transmission maps.

E. Loss Function

Following previous works [48], [49], to achieve a good balance between visual quality and quantitative scores, we use the linear combination of the ℓ_2 loss L_{ℓ_2} and the perceptual loss L_{per}, and the final loss L_f for training our network is expressed as:

$$L_f = L_{\ell_2} + \lambda L_{per}, \tag{7}$$

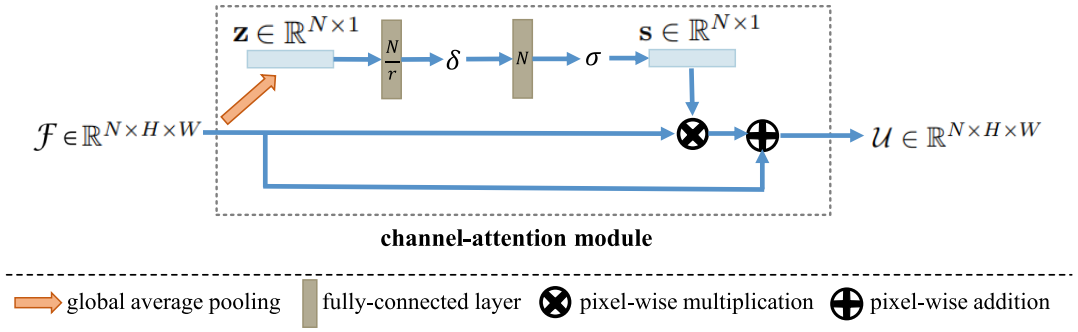


Fig. 5. The schematic illustration of the channel-attention module. The channel-attention module performs feature recalibration using global information. After going through global average pooling and fully-connected layers, the informative features are emphasized and the less usefull features are suppressed in the input features $\mathcal{F}$, thus obtaining rescaled features $\mathcal{U}$.

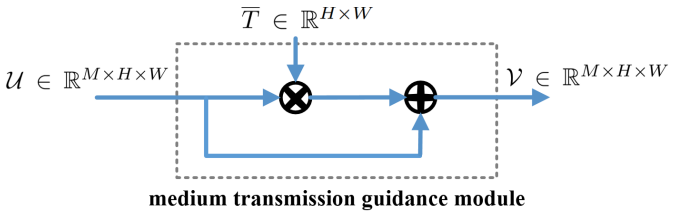


Fig. 6. The schematic illustration of the medium transmission guidance module. The RMT map $\overline{T}$ as a feature selector is used to weight the importance of different spatial positions of the input features $\mathcal{U}$, thus obtaining highlighted output features $\mathcal{V}$.

where λ is empirically set to 0.01 for balancing the scales of different losses. Specifically, the ℓ_2 loss measures the difference between the reconstructed result $\hat{J}$ and corresponding ground truth J as:

$$L_{\ell_2} = \sum_{m=1}^{H}\sum_{n=1}^{W}(\hat{J}(m,n) - J(m,n))^2. \tag{8}$$

The perceptual loss is computed based on the VGG-19 network ϕ [43] pre-trained on the ImageNet dataset [50]. Let $\phi_j(\cdot)$ be the jth convolutional layer. We measure the distance between the feature representations of the reconstructed result $\hat{J}$ and ground truth image J as:

$$L_{per} = \sum_{m=1}^{H}\sum_{n=1}^{W}|\phi_j(\hat{J})(m,n) - \phi_j(J)(m,n)|. \tag{9}$$

We compute the perceptual loss at layer relu5_4 of the VGG-19 network. An ablation study towards the loss function will be presented.

IV. Experiments

In this section, we first describe the implementation details, then introduce the experiment settings. We compare our method with representative methods and provide a series of ablation studies to verify each component of Ucolor. We show the failure case of our method at the end of this section. Due to the limited space, more experimental results can be found in the supplementary material.

A. Implementation Details

To train the Ucolor, we randomly selected 800 pairs of underwater images from UIEB [18] underwater image enhancement dataset. The UIEB dataset includes 890 real underwater images with corresponding reference images. Each reference image was selected by 50 volunteers from 12 enhanced results. It covers diverse underwater scenes, different characteristics of quality degradation, and a broad range of image content, but the number of underwater images is inadequate to train our network. Thus, we incorporated 1,250 synthetic underwater images selected from a synthesized underwater image dataset [12], which includes 10 subsets denoted by ten types of water (I, IA, IB, II, and III for open ocean water and 1, 3, 5, 7, and 9 for coastal water). To augment the training data, we randomly cropped image patches of size 128×128.

We implemented the Ucolor using the MindSpore Lite tool [51]. A batch-mode learning method with a batch size of 16 was applied. The filter weights of each layer were initialized by Gaussian distribution, and the bias was initially set as a constant. We used ADAM for network optimization and fixed the learning rate to $1e^{-4}$.

B. Experiment Settings

Benchmarks. For testing, we used the rest 90 pairs of real data of the UIEB dataset, denoted as **Test-R90**, while 100 pairs of synthetic data from each subset of [12] forming a total of 1k pairs, denoted as **Test-S1000**. We also conducted comprehensive experiments on three more benchmarks, *i.e.*, **Test-C60** [18], **SQUID** [52], and **Color-Check7** [25]. Test-C60 contains 60 real underwater images without reference images provided in the UIEB dataset [18]. Different from Test-R90, the images in Test-C60 are more challenging, which fail current methods. The SQUID [52] dataset contains 57 underwater stereo pairs taken from four different dive sites in Israel. We used the 16 representative examples presented in the project page of SQUID[1] for testing. Specifically, for four dive sites (Katzaa, Michmoret, Nachsholim, Satil), four representative samples were selected from each dive site. Each

[1]http://csms.haifa.ac.il/profiles/tTreibitz/datasets/ambient_forwardlooking/index.html

image has a resolution of 1827×2737. Color-Check7 contains 7 underwater Color Checker images taken with different cameras provided in [25], which are employed to evaluate the robustness and accuracy of underwater color correction. The cameras used to take the Color Checker pictures are Canon D10, Fuji Z33, Olympus Tough 6000, Olympus Tough 8000, Pentax W60, Pentax W80, and Panasonic TS1, denoted as Can D10, Fuj Z33, Oly T6000, Oly T8000, Pen W60, Pen W80, and Pan TS1 in this paper.

Compared Methods. We compared our Ucolor with ten methods, including one physical model-free method (Ancuti *et al*. [9]), three physical model-based methods (Li *et al*. [4], Peng and Cosman [29], GDCP [30]), four deep learning-based methods (UcycleGAN [10], Guo *et al*. [17], Water-Net [18], UWCNN [12]), and two baseline deep models (denoted as Unet-U [40] and Unet-RMT [40]) that are trained using the same training data and loss functions as our Ucolor. Different from Ucolor, Unet-U and Unet-RMT employ the structure of Unet [40]. In addition, the inputs of Unet-U and Unet-RMT are an underwater image and the concatenation of an underwater image and its RTM map that is estimated using the same algorithm as our Ucolor, respectively. The comparisons with the two baseline deep models aim at demonstrating the advantages of our network architecture and supplementing the compared deep learning-based methods.

Since the source code of Ancuti *et al.* [9] is not publicly available, we used the code[2] implemented by other researchers to realize code of [9]. For Li *et al*. [4], Peng and Cosman [29], GDCP [30], UcycleGAN [10], and Water-Net [18], we used the released codes to produce their results. The results of Guo *et al.* [17] were provided by the authors. Note that UcycleGAN [10] is an unsupervised methods, *i.e*., training with unpaired data, and thus there is no need to retrain it with our training data. Same as our Ucolor, Water-Net [18] randomly selected the same number of training data from the UIEB dataset for training. For UWCNN [12], we used the original UWCNN models, in which each UWCNN model was trained using the underwater images synthesized by one type of water. We discarded the UWCNN_typeIA and UWCNN_typeIB models because their results are similar to those of UWCNN_typeI. Besides, we also retrained the UWCNN model (denoted as UWCNN_retrain) using the same training data as our Ucolor.

Evaluation Metrics. For Test-R90 and Test-S1000, we conducted full-reference evaluations using PSNR and MSE metrics. Following [18], we treated the reference images of Test-R90 as the ground-truth images to compute the PSNR and MSE scores. A higher PSNR or a lower MSE score denotes that the result is closer to the reference image in terms of image content.

For Test-C60 and SQUID that do not have corresponding ground truth images, we employ the no-reference evaluation metrics UCIQE [53] and UIQM [54] to measure the performance of different methods. A higher UCIQE or UIQM score suggests a better human visual perception. Please note that the scores of UCIQE and UIQM cannot accurately reflect the performance of underwater image enhancement methods in some cases. We refer the readers to [18] and [52] for the discussions and visual examples. In our study, we only provide the scores of UCIQE and UIQM as the reference for the following research. In addition, we also provide the scores of NIQE [55] of different methods as the reference though it was not originally devised for underwater images. A lower NIQE score suggests better image quality. Although SQUID provides a script to evaluate color reproduction and transmission map estimation for the underwater image, this script requires the estimated transmission map or the results having a resolution of 1827×2737. However, the compared deep learning-based methods do not need to estimate the transmission maps. Moreover, our current GPU cannot process the input image with such a high resolution. Besides, the sizes of the color checker in the images of SQUID are too small to be cropped for full-reference evaluations. Instead, we resized the images in the SQUID testing dataset to a size of 512×512 and processed them by different methods. We conducted a user study to measure the perceptual quality of results on Test-C60 and SQUID. Specifically, we invited 20 human subjects to score the perceptual quality of the enhanced images independently. The scores of perceptual quality range from 1 to 5 (worst to best quality). These subjects were trained by observing the results from 1) whether the results introduce color deviations; 2) whether the results contain artifacts; 3) whether the results look natural; and 4) whether the results have good contrast and visibility.

For Color-Check7, we measured the dissimilarity of color between the ground-truth Macbeth Color Checker and the corresponding enhanced results. To be specific, we extracted the color of 24 color patches from the ground-truth Macbeth Color Checker. Then, we respectively cropped the 24 color patches for each enhanced result and computed the average color values of each color patch. At last, we followed the previous method [25] to employ CIEDE2000 [56] to measure the relative perceptual differences between the corresponding color patches of ground-truth Macbeth Color Checker and those of enhanced results. The smaller the CIEDE2000 value, the better.

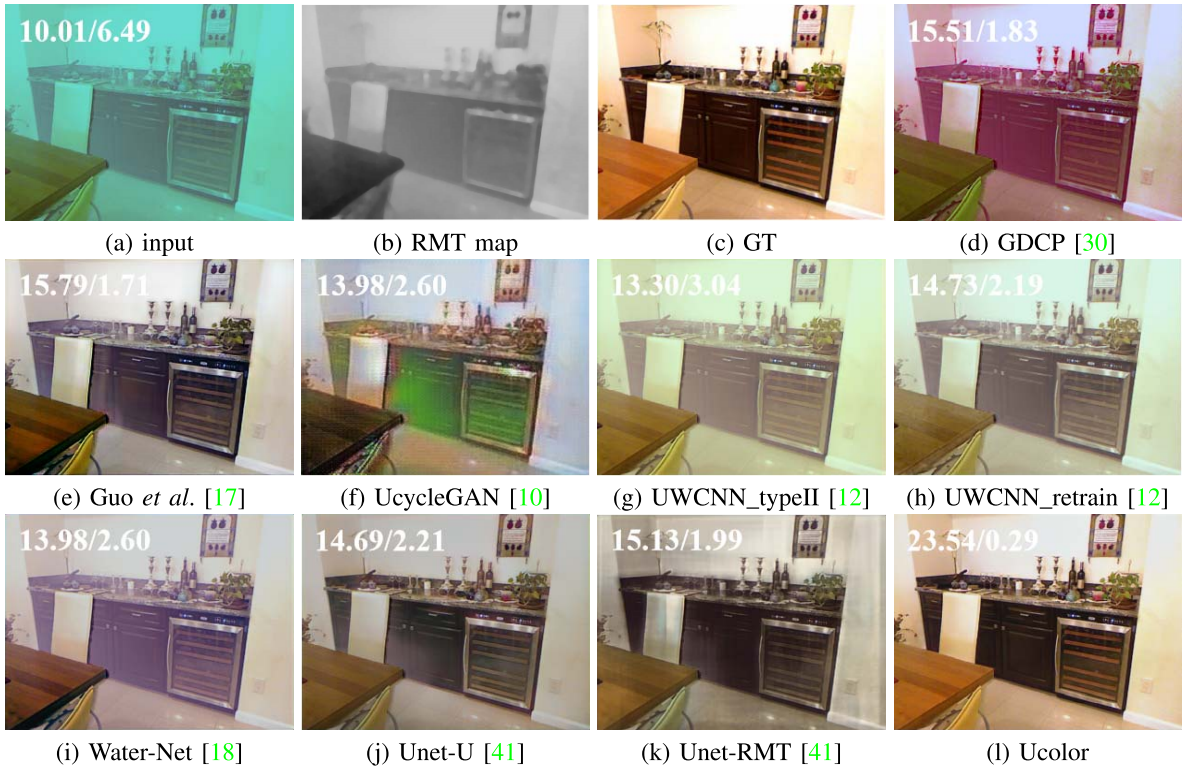
C. Visual Comparisons

In this section, we conduct visual comparisons on diverse testing datasets. We refer readers to the Google Drive link[3] for all results of different methods (about 7.4 GB).

We first show the comparisons on a synthetic image in Fig. 7. The competing methods either fail to dehaze the input image or they introduce undesirable color artifacts. All the methods under comparison fail to recover the complete scene structure. Our result (Fig. 7(l)) is closest to the ground-truth images and obtains the best PSNR/MSE scores. Furthermore, the RMT map shown in Fig. 7(b) indicates that the high degradation regions have large pixel values. With the RMT map, our method highlights these regions and enhances them well.

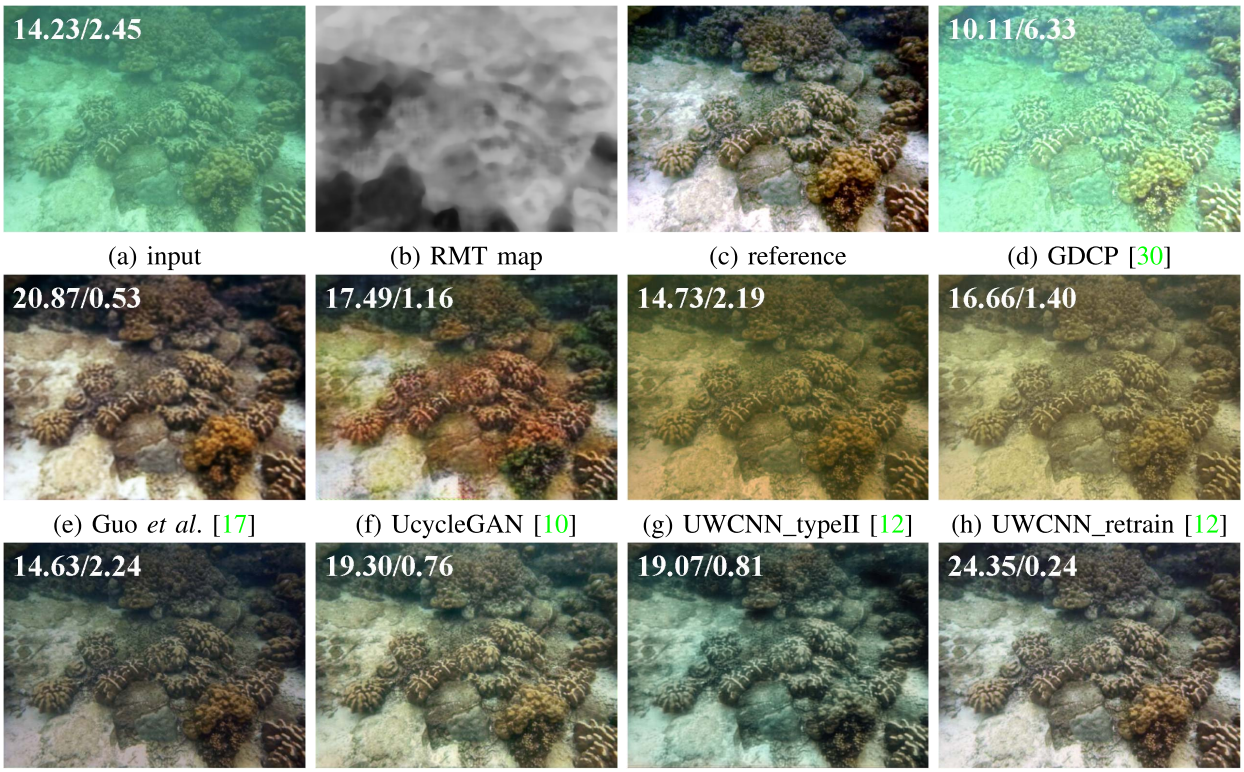
[2]https://github.com/bilityniu/underimage-fusion-enhancement

[3]https://drive.google.com/file/d/1zrynw05ZgkVMAybGo9Nhqg2mmIYIMVOT/view?usp=sharing



(a) input (b) RMT map (c) GT (d) GDCP [30]
(e) Guo *et al*. [17] (f) UcycleGAN [10] (g) UWCNN_typeII [12] (h) UWCNN_retrain [12]
(i) Water-Net [18] (j) Unet-U [41] (k) Unet-RMT [41] (l) Ucolor

Fig. 7. Visual comparisons on a synthetic underwater image sampled from **Test-S1000**. The numbers on the top-left corner of each image refer to its PSNR (dB)/MES ($\times 10^3$).



(a) input (b) RMT map (c) reference (d) GDCP [30]
(e) Guo *et al*. [17] (f) UcycleGAN [10] (g) UWCNN_typeII [12] (h) UWCNN_retrain [12]
(i) Water-Net [18] (j) Unet-U [41] (k) Unet-RMT [41] (l) Ucolor

Fig. 8. Visual comparisons on a typical real underwater image with obvious greenish color deviation and low-contrast sampled from **Test-R90**. The numbers on the top-left corner of each image refer to its PSNR (dB)/MES ($\times 10^3$).

We then show the results of different methods on a real underwater image with obvious greenish color deviation in Fig. 8. In Fig. 8(a), the greenish color deviation significantly hides the structural details of the underwater scene. In terms of color, GDCP [30], UcycleGAN [10], UWCNN_typeII [12] and UWCNN_retrain [12] introduce

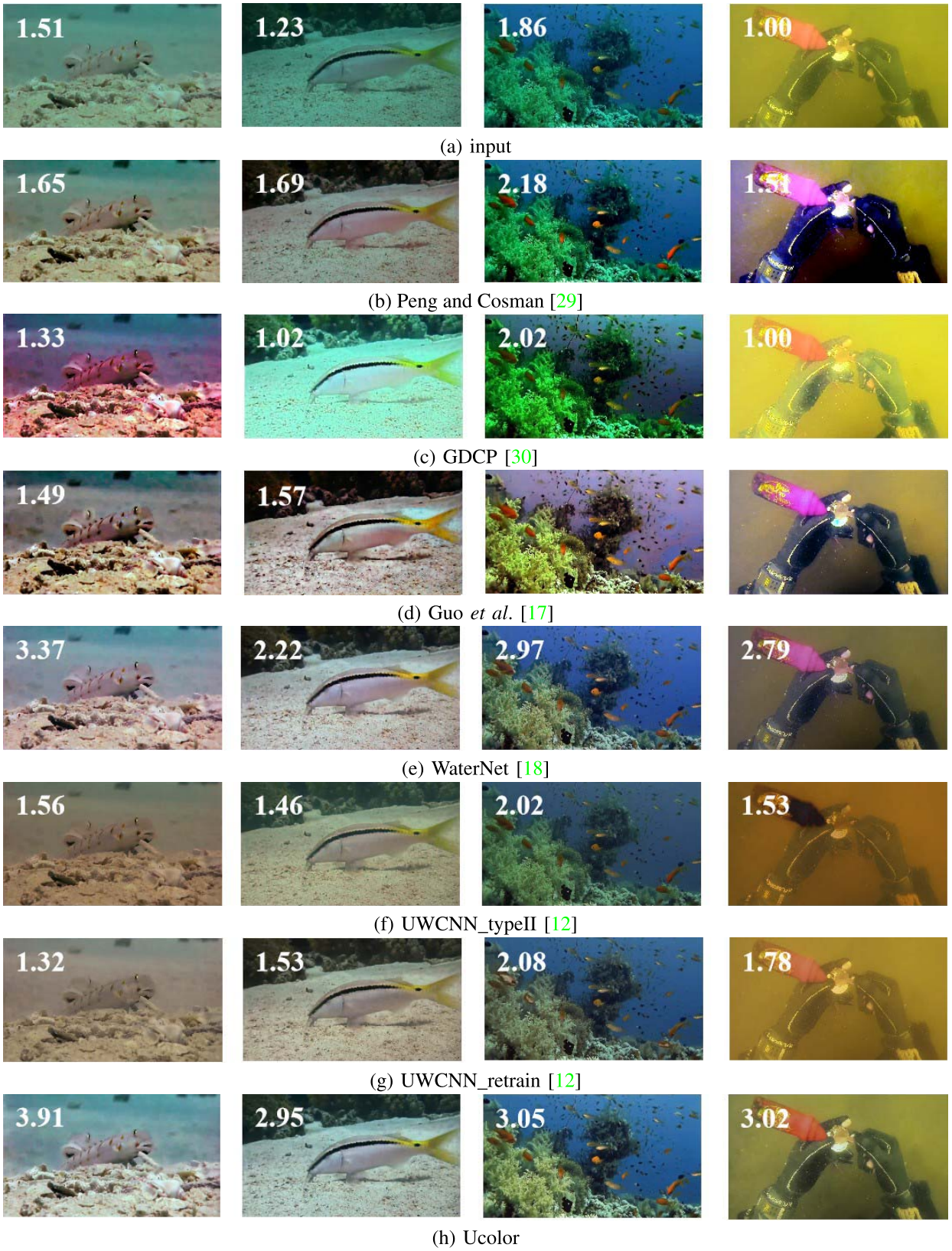


Fig. 9. Visual comparisons on challenging underwater images sampled from **Test-C60**. The number on the top-left corner of each image refers to its perceptual score (the larger, the better).

extra color artifacts. All the compared methods under-enhance the image or introduce the over-saturation. In comparison, our Ucolor effectively removes the greenish tone and improves the contrast without obvious over-enhancement and over-saturation.Although UWCNN_retrain [12], Unet-U [40], and Unet-RMT [40] were trained with the same data as our

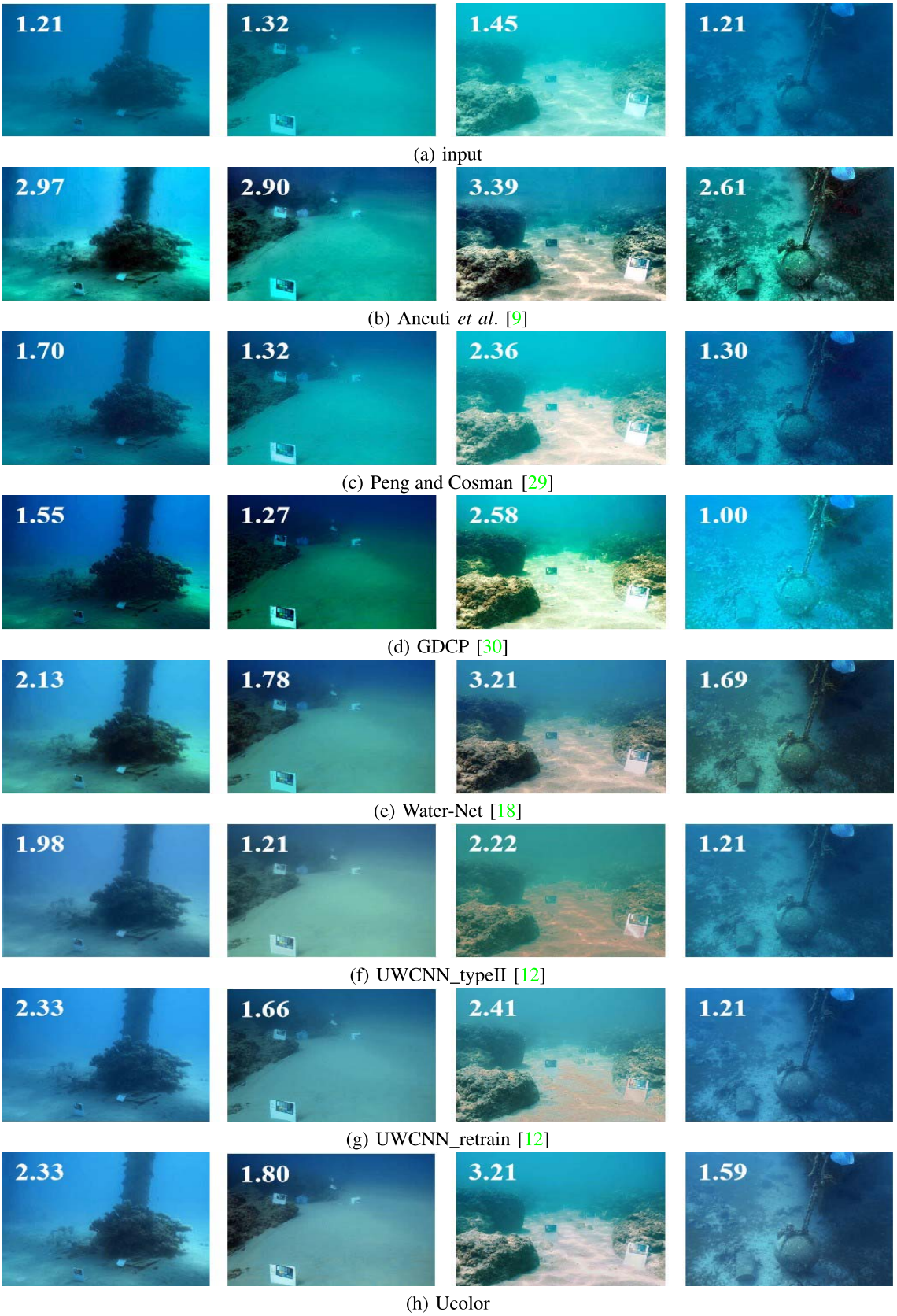


Fig. 10. Visual comparisons on challenging underwater images sampled from **SQUID**. From left to right, the images were taken from four different dive sites Katzaa, Michmoret, Nachsholim, and Satil. The number on the top-left corner of each image refers to its perceptual score (the larger, the better).

Ucolor, their performance is not as good as our Ucolor, which demonstrates the advantage of our specially designed network structure for underwater image enhancement.

We also show comparisons on challenging underwater images sampled from Test-C60 in Fig. 9. These underwater images suffer from high backscattering and color deviations

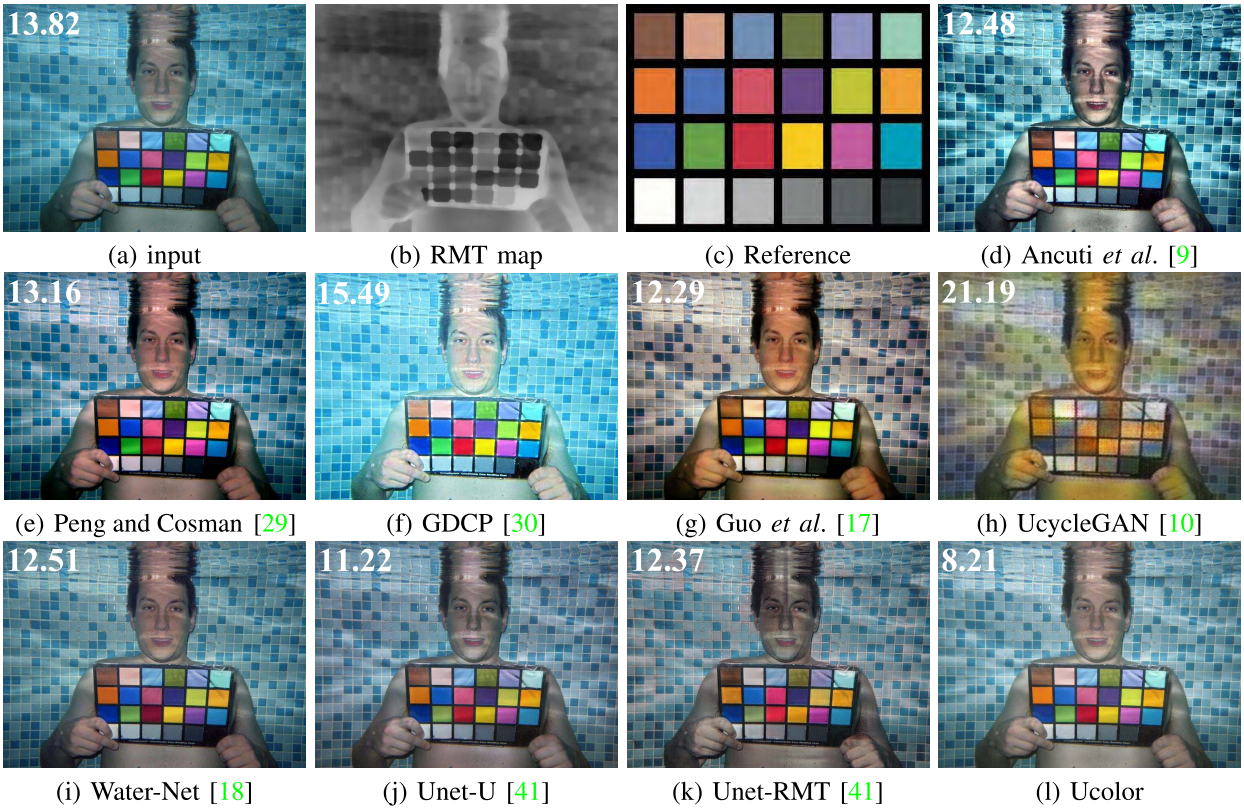


Fig. 11. Visual comparisons on a color checker image taken by a Pentax W60 camera sampled from **Color-Checker7**. The values of CIEDE2000 metric for the regions of Color Checker are reported on the top-left corner of the images (the smaller, the better).

as shown in Fig. 9(a). For these images, all competing methods cannot achieve satisfactory results. Some of them even introduce artifacts, such as GDCP [30], Guo *et al.* [17], and UWCNN_typeII [12]. Additionally, some methods introduce artificial colors. For example, the red object in the fourth image becomes purple and the sand in the first two images becomes reddish. In contrast, our Ucolor not only recovers the relatively realistic color but also enhances details, which is credited to the effective designs of multiple color spaces embedding and the introduction of medium transmission-guided decoder structure. The perceptual scores of our results also suggest the visually pleasing quality of our results.

We present the results of different methods on challenging underwater images sampled from SQUID in Fig. 10. As presented, the input underwater images challenge all underwater image enhancement methods. Ancuti *et al.* [9] achieves better contrast than the other methods but produces color deviations, *e.g.*, the reddish tone in the third image and the greenish tone in the fourth image. Our Ucolor dehazes the input image, thus improving the contrast of input images. Moreover, our method does not produce obvious artificial colors on the third image. According to the quantitative scores, we can find that the artificial colors significantly affect the perceptual scores given by subjects.

To analyze the robustness and accuracy of color correction, we conduct the comparisons on the underwater Color Checker image in Fig. 11. As shown in Fig. 11(a), the professional underwater camera (Pentax W60) also inevitably introduces various color casts. Both traditional and learning-based methods change the colors of input from an overall perspective. As indicated by the CIEDE2000 values on the results, our Ucolor achieves the best performance (8.21 under CIEDE2000 metric) in terms of the accuracy of color correction.

TABLE I

THE EVALUATIONS OF DIFFERENT METHODS ON **TEST-S1000** AND **TEST-R90** IN TERMS OF AVERAGE PSNR (DB) AND MSE ($\times 10^3$) VALUES. THE BEST RESULT IS IN RED UNDER EACH CASE

Methods	Test-S1000		Test-R90	
	PSNR↑	MSE↓	PSNR↑	MSE↓
input	12.96	4.60	16.11	2.03
Ancuti *et al.* [9]	13.27	5.15	19.19	0.78
Li *et al.* [4]	14.29	3.64	16.73	1.38
Peng and Cosman [29]	13.04	4.53	15.77	1.72
GDCP [30]	11.67	5.98	13.85	3.40
Guo *et al.* [17]	15.78	2.57	18.05	1.18
UcycleGAN [10]	14.73	3.13	16.61	1.65
Water-Net [18]	15.47	3.26	19.81	1.02
UWCNN_type1 [12]	16.27	2.68	13.62	3.52
UWCNN_type3 [12]	15.70	2.87	12.84	4.23
UWCNN_type5 [12]	14.78	2.94	13.26	3.65
UWCNN_type7 [12]	12.38	4.35	13.02	3.67
UWCNN_type9 [12]	12.83	3.85	12.79	3.89
UWCNN_typeI [12]	10.44	6.42	10.57	6.24
UWCNN_typeII [12]	17.51	2.59	14.75	2.57
UWCNN_typeIII [12]	17.41	2.39	13.26	3.40
UWCNN_retrain [12]	15.87	2.74	16.69	1.71
Unet-U [41]	19.14	1.22	18.14	1.32
Unet-RMT [41]	17.93	1.43	16.89	1.71
Ucolor	23.05	0.50	20.63	0.77

TABLE II

THE AVERAGE PERCEPTUAL SCORES (PS), UIQM [54] SCORES, UCIQE [53] SCORSE, AND NIQE [55] SCORES OF DIFFERENT METHODS ON **TEST-C60** AND **SQUID**. THE BEST RESULT IS IN RED UNDER EACH CASE. "-" REPRESENTS THE RESULTS ARE NOT AVAILABLE

Methods	Test-C60				SQUID			
	PS↑	UIQM↑	UCIQE↑	NIQE↓	PS↑	UIQM↑	UCIQE↑	NIQE↓
input	1.34	0.84	0.48	7.14	1.21	0.82	0.42	4.93
Ancuti *et al.* [9]	2.11	1.22	0.62	4.94	2.93	1.30	0.62	5.01
Li *et al.* [4]	1.22	1.27	0.65	5.32	1.00	1.34	0.66	4.81
Peng and Cosman [29]	2.07	1.13	0.58	6.01	2.34	0.99	0.50	4.39
GDCP [30]	1.98	1.07	0.56	5.92	2.47	1.11	0.52	4.48
Guo *et al.* [17]	2.63	1.11	0.60	5.71	-	-	-	-
UcycleGAN [10]	1.01	0.91	0.58	7.67	1.16	1.11	0.56	5.93
Water-Net [18]	3.52	0.97	0.56	6.04	2.78	1.03	0.54	4.72
UWCNN_typeII [12]	2.19	0.77	0.47	6.76	2.72	0.69	0.44	4.60
UWCNN_retrain [12]	2.91	0.84	0.49	6.66	2.67	0.77	0.46	4.38
Unet-U [41]	3.37	0.94	0.50	6.12	2.61	0.82	0.50	4.38
Unet-RMT [41]	3.04	1.03	0.52	6.12	2.53	0.82	0.49	5.16
Ucolor	3.74	0.88	0.53	6.21	2.82	0.82	0.51	4.29

All the visual comparisons demonstrate that our Ucolor not only renders visually pleasing results but also generalizes well to different underwater scenes.

D. Quantitative Comparisons

We first perform quantitative comparisons on Test-S1000 and Test-R90. The average scores of PSNR and MSE of different methods are reported in Table I. As presented in Table I, our Ucolor outperforms all competing methods on Test-S1000 and Test-R90. Compared with the second-best performer, our Ucolor achieves the percentage gain of 20%/59% and 4.1%/1.3% in terms of PSNR/MSE on Test-S1000 and Test-R90, respectively. There are two interesting findings from the quantitative comparisons. 1) Although the medium transmission map used in our Ucolor is the same as the traditional GDCP [30], the performance is significantly different. Such a result suggests that the performance of underwater image enhancement can be improved by the effective combination of domain knowledge with deep neural networks. 2) The performance of Unet-U [40] is better than Unet-RMT [40], which suggests that simple concatenation of input image and its reverse medium transmission map cannot improve the underwater image enhancement performance of deep model, and even decreases the performance. 3) The generalization capability of UWCNN models [12] is limited because they require the images to be taken in the accurate type of water as inputs. Due to the limited space, we only present the results of the original UWCNN model that performs the best in Table I in the following experiments, *i.e.*, UWCNN_typeII [12].

Next, we conduct a user study on Test-C60 and SQUID. The average perceptual scores of the results by different methods are reported in Table II, where it can be observed that these two challenging testing datasets fail most underwater image enhancement methods in terms of perceptual quality. Some methods such as Li *et al.* [4] and UcycleGAN [10] even achieve lower perceptual scores than inputs. For the Test-60 testing dataset, the deep learning-based methods achieve relatively higher perceptual scores. Among them, our Ucolor is superior to the other competing methods. For the SQUID testing dataset, the traditional fusion-based method [9] obtains the highest perceptual score while our Ucolor ranks the second best. Other deep learning-based methods achieve similar perceptual scores. As shown in Fig. 10, all deep learning-based methods cannot handle the color deviations of images in SQUID well, while the haze can be removed. In contrast, the fusion-based method [9] achieves better contrast, thus obtaining a higher perceptual score. Observing the scores of non-reference image quality assessment metrics, we can see that Li *et al.* [4] obtains the best performance in terms of UIQM and UCIQE scores. For the NIQE scores, our Ucolor achieves the lowest NIQE score on the SQUID testing set while Ancuti *et al.* [9] obtains the best performance on the Test-C60 testing set.

To demonstrate the robustness to different cameras and the accuracy of color restoration, we report the average CIEDE2000 scores on Color-Checker7 in Table III. For the cameras of Pentax W60, Pentax W80, Cannon D10, Fuji Z33, and Olympus T6000, our Ucolor obtains the lowest color dissimilarity. Moreover, our Ucolor achieves the best average score across seven cameras. Such results demonstrate the superiority of our method for underwater color correction. It is interesting that some methods achieve worse performance in terms of the average CIEDE2000 score than the original input, which suggests that some competing methods cannot recover the real color and even break the inherent color.

E. Ablation Study

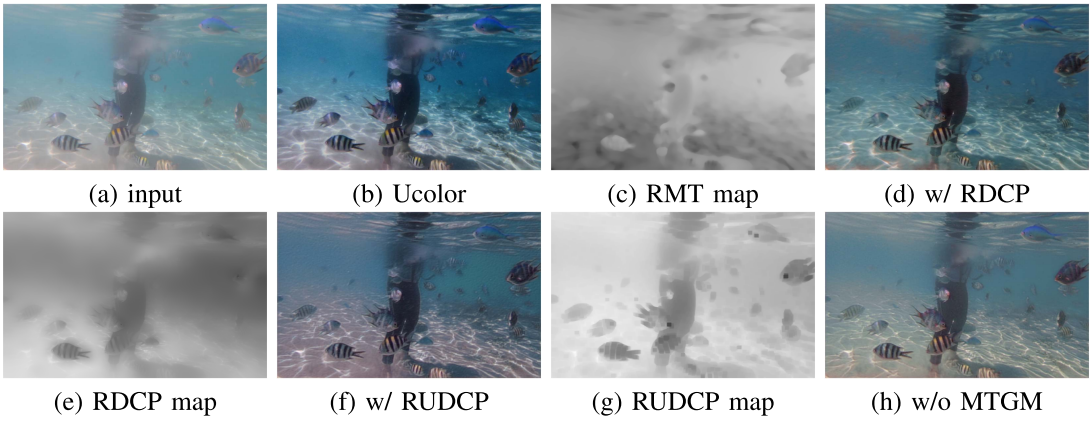
We perform extensive ablation studies to analyze the core components of our method, including the multi-color space encoder (MCSE)), the medium transmission guidance module (MTGM), and the channel-attention module (CAM). Additionally, we analyze the combination of the ℓ_2 loss and the perceptual loss. More specifically,

- w/o HSV, w/o Lab, and w/o HSV+Lab stand for Ucolor without the HSV, Lab, and both HSV and Lab color spaces encoder paths, respectively.
- w/ 3-RGB means that all inputs of three encoder paths are RGB images.
- w/o MTGM refers to the Ucolor without the medium transmission guidance module.
- w/ RDCP and w/ RUDCP are the models by replacing the medium transmission map estimated via [30] with the algorithms in [57] and [3], respectively.

TABLE III

THE COLOR DISSIMILARITY COMPARISONS OF DIFFERENT METHODS ON **COLOR-CHECK7** IN TERMS OF THE CIEDE2000. THE BEST RESULT IS IN RED UNDER EACH CASE

Methods	Pen W60	Pen W80	Can D10	Fuj Z33	Oly T6000	Oly T8000	Pan TS1	Avg
input	13.82	17.26	16.13	16.37	14.89	23.14	19.06	17.24
Ancuti *et al.* [9]	12.48	13.30	14.28	11.43	11.57	12.58	10.63	12.32
Li *et al.* [4]	15.41	17.56	18.52	25.01	16.01	17.12	12.03	17.38
Peng and Cosman [29]	13.16	16.01	14.78	14.09	12.24	14.79	19.59	14.95
GDCP [30]	15.49	24.32	16.89	13.73	12.76	16.82	12.93	16.13
Guo *et al.* [17]	12.29	15.50	14.58	16.65	39.71	15.14	12.40	18.04
UcycleGAN [10]	21.19	21.23	22.96	26.28	20.88	23.42	19.02	22.14
Water-Net [18]	12.51	19.57	15.44	12.91	17.55	21.73	18.84	16.94
UWCNN_typeII [12]	16.73	20.55	17.73	17.20	16.31	17.94	20.97	18.20
UWCNN_retrain [12]	13.64	20.33	14.91	13.38	14.72	18.11	20.19	16.47
Unet-U [41]	11.22	15.17	13.32	11.91	10.87	15.12	17.31	13.56
Unet-RMT [41]	12.37	19.01	15.57	14.80	13.26	16.47	19.55	15.86
Ucolor	8.21	10.59	12.27	8.11	7.22	14.42	14.54	10.77



(a) input (b) Ucolor (c) RMT map (d) w/ RDCP

(e) RDCP map (f) w/ RUDCP (g) RUDCP map (h) w/o MTGM

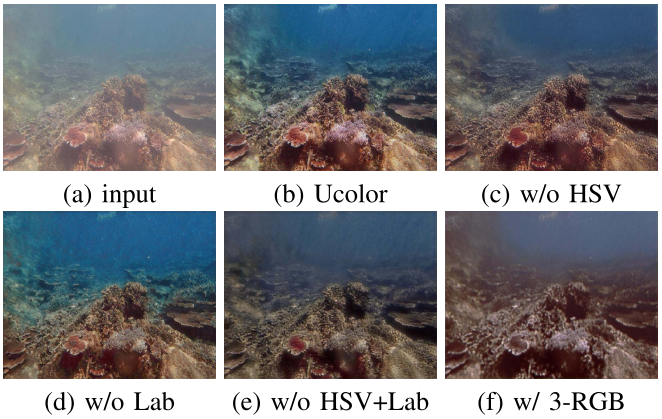
Fig. 12. Ablation study of the effects of the reverse medium transmission map. RDCP map represents the reverse DCP map, where the DCP map was estimated by [57]. RUDCP map represents the reverse UDCP map, where the UDCP map was estimated by [3]. Compared with the RDCP and RUDCP maps, the RMT map can more accurately indicate the degradation of underwater image, thus leading to better enhancement performance of Ucolor.

- w/o CAM stands for the Ucolor without the channel-attention module.
- w/o perc loss means that the Ucolor is trained only with the constraint of ℓ_2 loss.

The quantitative PSNR (dB) and MSE ($\times 10^3$) values on Test-S1000 and Test-R90 are presented in Table IV. The visual comparisons of the effects of reverse medium transmission maps, the contributions of each color space encoder path, the effectiveness of channle-attention module, and the effect of perceptual loss are shown in Figs. 12, 13, 14, and 15, respectively. The conclusions drawn from the ablation studies are listed as follows.

1) As presented in Table IV, our full model achieves the best quantitative performance across two testing datasets when compared with the ablated models, which implies the effectiveness of the combinations of MCSE, MTGM, and CAM modules.

2) Our RMT map can relatively accurately assign the high-quality degradation regions a higher weight than the RDCP map and the RUDCP map, thus achieving better visual quality, especially for high scattering regions as shown in Fig. 12. Additionally, the quantitative performance of the Ucolor is better than the ablated models w/ RDCP, w/ RUDCP, and w/o MTGM, which suggests the importance of an accurate medium transmission map.



(a) input (b) Ucolor (c) w/o HSV

(d) w/o Lab (e) w/o HSV+Lab (f) w/ 3-RGB

Fig. 13. Ablation study of the contributions of each color space encoder path. Ucolor (three color spaces encoder) can achieve vivid color, high contrast, and clear details.

3) The ablated models w/o HSV+Lab and w/ 3-RGB produce comparable performance as shown in Table IV and Fig. 13. The results indicate that aimlessly adding more parameters or the same color encoder will not bring extra representational power to better enhance underwater images. In contrast, a well-design multi-color space embedding helps to learn more powerful representations to improve enhancement performance. In addition, removing any one of the three

(a) input (b) Ucolor (c) w/o CAM

Fig. 14. Ablation study of the effectiveness of the channel-attention module for highlighting the multi-color space features. More realistic result is achieved by Ucolor (with channle-attention module) than the ablated model w/o CAM.

(a) input (b) Ucolor (c) w/o perc loss

Fig. 15. Ablation study towards the perceptual loss. By adding a perceptual loss to the ℓ_2 loss, the visual quality of final result is improved.

TABLE IV

QUANTITATIVE RESULTS OF THE ABLATION STUDY IN TERMS OF AVERAGE PSNR (DB) AND MSE ($\times 10^3$) VALUES

Modules	Baselines	Test-S1000		Test-R90	
		PSNR↑	MSE ↓	PSNR↑	MSE ↓
	full model	23.05	0.50	20.63	0.77
MCSE	w/o HSV	16.52	1.83	16.08	1.97
	w/o Lab	18.33	1.37	17.54	1.50
	w/o HSV+Lab	16.62	1.88	15.91	2.10
	w/ 3-RGB	16.59	2.06	15.84	2.11
MTGM	w/o MTGM	17.02	1.91	17.37	1.59
	w/ RDCP	18.74	0.87	18.09	1.01
	w/ RUDCP	18.94	0.83	17.56	1.14
CAM	w/o CAM	16.36	1.88	16.02	2.02
loss function	w/o perc loss	23.11	0.49	18.29	0.98

encoder paths (w/o HSV and w/o Lab) will decrease the performance as shown in Table IV.

4) The ablated model w/o CAM produces an under-saturated result as shown in Fig. 14. This may be induced by removing the CAM module that integrates and highlights the features extracted from multiple color spaces.

5) The quantitative results in Table IV show that only using the ℓ_2 loss can slightly improve the quantitative performance on Test-S1000 in terms of PNSR and MSE values. However, from the visual results in Fig. 15, it can be observed that the enhanced image by Ucolor trained with the full loss function (*i.e.*, the combination between the ℓ_2 loss and the perceptual loss) is better than that by Ucolor trained without the perceptual loss. Thus, it is necessary to add the perceptual loss for improving the visual quality of final results. Note that only using the perceptual loss for training does not make sense, and thus we did not conduct such an ablation study.

F. Failure Case

When facing underwater images with limited lighting, our Ucolor, as well as other state-of-the-arts might not work well. Fig. 16 presents an example where both our Ucolor and latest deep learning-based Water-Net [18] fail to produce a visually compelling result when processing an underwater image with limited lighting. The potential reason lies in few such images

(a) input (b) Water-Net [18] (c) Ucolor

Fig. 16. Failure case. The input underwater image has limited lighting. Although our Ucolor cannot effectively enhance this image, it does not introduce color casts like water-net.

in the training data sets. Thus, it is difficult for supervised learning-based networks such as Water-Net and Ucolor to handle such underwater images. The stronger ability and more diverse training data that handle such kinds of underwater images will be our future goal.

V. CONCLUSION

We have presented a deep underwater image enhancement model. The proposed model learns the feature representations from diverse color spaces and highlights the most discriminative features by the channel-attention module. Besides, the domain knowledge is incorporated into the network by employing the reverse medium transmission map as the attention weights. Extensive experiments on diverse benchmarks have demonstrated the superiority of our solution and the effectiveness of multi-color space embedding and the reverse medium transmission guided decoder network structure. The effectiveness of the key components of our method has been verified in the ablation studies.

ACKNOWLEDGMENT

The authors would like to thank Codruta O. Ancuti, Cosmin Ancuti, Christophe De Vleeschouwer, and Philippe Bekaert for providing the underwater Color Checker images [25]. They would also like to thank Yecai Guo, Hanyu Li, and Peixiain Zhuang for providing their results [17].

REFERENCES

[1] D. Akkaynak, T. Treibitz, T. Shlesinger, R. Tamir, Y. Loya, and D. Iluz, "What is the space of attenuation coefficients in underwater computer vision?" in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4931–4940.

[2] J. Y. Chiang and Y.-C. Chen, "Underwater image enhancement by wavelength compensation and dehazing," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1756–1769, Apr. 2012.

[3] 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 Comput. Graph. Appl.*, vol. 36, no. 2, pp. 24–35, Mar. 2016.

[4] C.-Y. Li, J.-C. Guo, R.-M. Cong, Y.-W. Pang, and B. Wang, "Underwater image enhancement by dehazing with minimum information loss and histogram distribution prior," *IEEE Trans. Image Process.*, vol. 25, no. 12, pp. 5664–5677, Dec. 2016.

[5] C. Li, J. Guo, S. Chen, Y. Tang, Y. Pang, and J. Wang, "Underwater image restoration based on minimum information loss principle and optical properties of underwater imaging," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2016, pp. 1993–1997.

[6] D. Berman, T. Treibitz, and S. Avidan, "Diving into haze-lines: Color restoration of underwater images," in *Proc. Brit. Mach. Vis. Conf. (BMVC)*, 2017, pp. 1–12.

[7] C. Li, J. Guo, C. Guo, R. Cong, and J. Gong, "A hybrid method for underwater image correction," *Pattern Recognit. Lett.*, vol. 94, pp. 62–67, Jul. 2017.

[8] P. Zhuang, C. Li, and J. Wu, "Bayesian retinex underwater image enhancement," *Eng. Appl. Artif. Intell.*, vol. 101, May 2021, Art. no. 104171.

[9] C. Ancuti, C. O. Ancuti, T. Haber, and P. Bekaert, "Enhancing underwater images and videos by fusion," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 81–88.

[10] C. Li, J. Guo, and C. Guo, "Emerging from water: Underwater image color correction based on weakly supervised color transfer," *IEEE Signal Process. Lett.*, vol. 25, no. 3, pp. 323–327, Mar. 2018.

[11] J. Li, K. A. Skinner, R. M. Eustice, and M. Johnson-Roberson, "WaterGAN: Unsupervised generative network to enable real-time color correction of monocular underwater images," *IEEE Robot. Autom. Lett.*, vol. 3, no. 1, pp. 387–394, Jan. 2018.

[12] C. Li, S. Anwar, and F. Porikli, "Underwater scene prior inspired deep underwater image and video enhancement," *Pattern Recognit.*, vol. 98, pp. 107038–107049, Feb. 2020.

[13] C. Guo *et al.*, "Zero-reference deep curve estimation for low-light image enhancement," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 1780–1789.

[14] C. Li, C. Guo, and C. L. Chen, "Learning to enhance low-light image via zero-reference deep curve estimation," *IEEE Trans. Pattern Anal. Mach. Intell.*, early access, Mar. 3, 2021, doi: 10.1109/TPAMI.2021.3063604.

[15] C. Li, R. Cong, J. Hou, S. Zhang, Y. Qian, and S. Kwong, "Nested network with two-stream pyramid for salient object detection in optical remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 11, pp. 9156–9166, Nov. 2019.

[16] C. Li *et al.*, "ASIF-net: Attention steered interweave fusion network for RGB-D salient object detection," *IEEE Trans. Cybern.*, vol. 51, no. 1, pp. 88–100, Jan. 2021.

[17] Y. Guo, H. Li, and P. Zhuang, "Underwater image enhancement using a multiscale dense generative adversarial network," *IEEE J. Ocean. Eng.*, vol. 45, no. 3, pp. 862–870, Jul. 2020.

[18] C. Li *et al.*, "An underwater image enhancement benchmark dataset and beyond," *IEEE Trans. Image Process.*, vol. 29, pp. 4376–4389, 2020.

[19] D. Akkaynak and T. Treibitz, "A revised underwater image formation model," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 6723–6732.

[20] M. Bryson, M. Johnson-Roberson, O. Pizarro, and S. Williams, "Colour-consistent structure-from-motion models using underwater imagery," in *Robotics: Science and Systems VIII*. Cambridge, MA, USA: MIT Press, 2012.

[21] D. Akkaynak and T. Treibitz, "Sea-thru: A method for removing water from underwater images," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 1682–1691.

[22] Y. Y. Schechner and N. Karpel, "Clear underwater vision," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun./Jul. 2004, pp. I–I.

[23] K. Iqbal, M. Odetayo, A. James, R. Abdul Salam, and A. Zawawi Hj Talib, "Enhancing the low quality images using unsupervised colour correction method," in *Proc. IEEE Int. Conf. Syst., Man Cybern.*, Oct. 2010, pp. 1703–1709.

[24] A. Ghani and N. Isa, "Underwater image quality enhancement through integrated color model with Rayleigh distribution," *Appl. Soft Comput.*, vol. 27, pp. 219–230, Feb. 2015.

[25] C. O. Ancuti, C. Ancuti, C. De Vleeschouwer, and P. Bekaert, "Color balance and fusion for underwater image enhancement," *IEEE Trans. Image Process.*, vol. 27, no. 1, pp. 379–393, Jan. 2018.

[26] S.-B. Gao, M. Zhang, Q. Zhao, X.-S. Zhang, and Y.-J. Li, "Underwater image enhancement using adaptive retinal mechanisms," *IEEE Trans. Image Process.*, vol. 28, no. 11, pp. 5580–5595, Nov. 2019.

[27] C. O. Ancuti, C. Ancuti, C. D. Vleeschouwer, and M. Sbert, "Color channel compensation (3C): A fundamental pre-processing step for image enhancement," *IEEE Trans. Image Process.*, vol. 29, pp. 2653–2665, 2019.

[28] A. Galdran, D. Pardo, A. Picón, and A. Alvarez-Gila, "Automatic red-channel underwater image restoration," *J. Vis. Commun. Image Represent.*, vol. 26, pp. 132–145, Jan. 2015.

[29] Y.-T. Peng and P. C. Cosman, "Underwater image restoration based on image blurriness and light absorption," *IEEE Trans. Image Process.*, vol. 26, no. 4, pp. 1579–1594, Apr. 2017.

[30] Y.-T. Peng, K. Cao, and P. C. Cosman, "Generalization of the dark channel prior for single image restoration," *IEEE Trans. Image Process.*, vol. 27, no. 6, pp. 2856–2868, Jun. 2018.

[31] R. Liu, X. Fan, M. Zhu, M. Hou, and Z. Luo, "Real-world underwater enhancement: Challenges, benchmarks, and solutions under natural light," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 12, pp. 4861–4875, Dec. 2020.

[32] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295–307, Feb. 2016.

[33] J. Pan, D. Sun, H. Pfister, and M.-H. Yang, "Deblurring images via dark channel prior," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 10, pp. 2315–2328, Oct. 2018.

[34] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2223–2232.

[35] C. Guo, C. Li, J. Guo, R. Cong, H. Fu, and P. Han, "Hierarchical features driven residual learning for depth map super-resolution," *IEEE Trans. Image Process.*, vol. 28, no. 5, pp. 2545–2557, May 2019.

[36] W. Ren *et al.*, "Low-light image enhancement via a deep hybrid network," *IEEE Trans. Image Process.*, vol. 28, no. 9, pp. 4364–4375, Sep. 2019.

[37] S. Anwar and C. Li, "Diving deeper into underwater image enhancement: A survey," *Signal Process., Image Commun.*, vol. 89, Nov. 2020, Art. no. 115978.

[38] A. Jamadandi and U. Mudenagudi, "Exemplar-based underwater image enhancement augmented by wavelet corrected transforms," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR) Workshops*, Jun. 2019, pp. 11–17.

[39] M. Yang, K. Hu, Y. Du, Z. Wei, Z. Sheng, and J. Hu, "Underwater image enhancement based on conditional generative adversarial network," *Signal Process., Image Commun.*, vol. 81, Feb. 2020, Art. no. 115723.

[40] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. New York, NY, USA: Springer, 2015, pp. 234–241.

[41] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1646–1654.

[42] S. K. Naik and C. A. Murthy, "Hue-preserving color image enhancement without gamut problem," *IEEE Trans. Image Process.*, vol. 12, no. 12, pp. 1591–1598, Dec. 2003.

[43] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2015.

[44] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[45] J. Hu, L. Shen, and G. Sun, "Squeeze-and-Excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7132–7141.

[46] S. G. Narasimhan and S. K. Nayar, "Vision and the atmosphere," *Int. J. Comput. Vis.*, vol. 48, no. 3, pp. 233–254, 2002.

[47] R. T. Tan, "Visibility in bad weather from a single image," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2008, pp. 1–8.

[48] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2016, pp. 694–711.

[49] C. Ledig *et al.*, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4681–4690.

[50] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255.

[51] (2020). *Mindspore*. [Online]. Available: https://www.mindspore.cn/

[52] D. Berman, D. Levy, S. Avidan, and T. Treibitz, "Underwater single image color restoration using haze-lines and a new quantitative dataset," *IEEE Trans. Pattern Anal. Mach. Intell.*, early access, Mar. 2, 2020, doi: 10.1109/TPAMI.2020.2977624.

[53] M. Yang and A. Sowmya, "An underwater color image quality evaluation metric," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 6062–6071, Dec. 2015.

[54] K. Panetta, C. Gao, and S. Agaian, "Human-visual-system-inspired underwater image quality measures," *IEEE J. Ocean. Eng.*, vol. 41, no. 3, pp. 541–551, Jul. 2016.

[55] A. Mittal, R. Soundararajan, and A. Bovik, "Making a 'completely blind' image quality analyzer," *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 209–212, Mar. 2013.

[56] G. Sharma, W. Wu, and E. N. Dalal, "The CIEDE2000 color-difference formula: Implementation notes, supplementary test data, and mathematical observations," *Color Res. Appl.*, vol. 30, no. 1, pp. 21–30, Feb. 2005.
[57] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 12, pp. 2341–2353, Dec. 2011.

Chongyi Li received the Ph.D. degree from the School of Electrical and Information Engineering, Tianjin University, Tianjin, China, in June 2018. From 2016 to 2017, he was a joint-training Ph.D. Student with The Australian National University, Australia. He was a Postdoctoral Fellow with the Department of Computer Science, City University of Hong Kong, Hong Kong. He is currently a Research Fellow with the School of Computer Science and Engineering, Nanyang Technological University (NTU), Singapore. His current research focuses on image processing, computer vision, and deep learning, particularly in the domains of image restoration and enhancement.

Saeed Anwar (Member, IEEE) received the Ph.D. degree from the Computer Vision Research Group (CVRG), Data61, Commonwealth Scientific and Industrial Research Organization (CSIRO), and the College of Electrical and Computer Science (CECS), ANU. He is currently a Research Scientist in cyber-physical systems at Data61, CSIRO, Australia, an Adjunct Lecturer with The Australian National University (ANU), and a Visiting Fellow with the University of Technology Sydney (UTS).

Junhui Hou (Senior Member, IEEE) received the B.Eng. degree in information engineering (Talented Students Program) from the South China University of Technology, Guangzhou, China, in 2009, the M.Eng. degree in signal and information processing from Northwestern Polytechnical University, Xi'an, China, in 2012, and the Ph.D. degree in electrical and electronic engineering from the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, in 2016. He has been an Assistant Professor with the Department of Computer Science, City University of Hong Kong, since 2017. His research interests fall into the general areas of visual computing, such as image/video/3D geometry data representation, processing and analysis, semi/un-supervised data modeling, data compression, and adaptive transmission. He is a member of the Multimedia Systems and Applications Technical Committee (MSA-TC) and IEEE CAS. He was a recipient of several prestigious awards, including the Chinese Government Award for Outstanding Students Study Abroad from the China Scholarship Council in 2015 and the Early Career Award (3/381) from the Hong Kong Research Grants Council in 2018. He served as the Area Chair for ACM MM 2019 and 2020, IEEE ICME 2020, and WACV 2021. He is currently serving as an Associate Editor for IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, *The Visual Computer*, and *Signal Processing: Image Communication*, and a Guest Editor for the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING.

Runmin Cong (Member, IEEE) received the Ph.D. degree in information and communication engineering from Tianjin University, Tianjin, China, in June 2019. He was a Visiting Student/Staff with Nanyang Technological University (NTU), Singapore, and City University of Hong Kong (CityU), Hong Kong. He is currently an Associate Professor with the Institute of Information Science, Beijing Jiaotong University, Beijing, China. His research interests include computer vision, multimedia processing and understanding, visual attention perception and saliency computation, remote sensing image interpretation and analysis, and visual content enhancement in an open environment.

Chunle Guo received the Ph.D. degree from Tianjin University, China, under the supervision of Prof. Jichang Guo. As a Visiting Student, he conducted the Ph.D. research at the School of Electronic Engineering and Computer Science, Queen Mary University of London (QMUL), U.K. He continued his research as a Research Associate with the Department of Computer Science, City University of Hong Kong (CityU), from 2018 to 2019. He is currently a Postdoctoral Research Fellow working with Prof. Ming-Ming Cheng at Nankai University. His research interests lies in image processing, computer vision, and deep learning.

Wenqi Ren (Member, IEEE) received the Ph.D. degree from Tianjin University, Tianjin, China, in 2017. From 2015 to 2016, he was supported by the China Scholarship Council and working with Prof. Ming-Husan Yang as a joint-training Ph.D. Student with the Electrical Engineering and Computer Science Department, University of California at Merced. He is currently an Associate Professor with the Institute of Information Engineering, Chinese Academy of Sciences, China. His research interests include image processing and related high-level vision problems. He received the Tencent Rhino Bird Elite Graduate Program Scholarship in 2017 and the MSRA Star Track Program in 2018.