Underwater image enhancement with global–local networks and compressed-histogram equalization

Xueyang Fu [a], Xiangyong Cao [b],*

[a] School of Information Science and Technology, University of Science and Technology of China, Hefei, China
[b] School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an, China

ARTICLE INFO

ABSTRACT

Due to the light absorption and scattering, captured underwater images usually contain severe color distortion and contrast reduction. To address the above problems, we combine the merits of deep learning and conventional image enhancement technology to improve the underwater image quality. We first propose a two-branch network to compensate the global distorted color and local reduced contrast, respectively. Adopting this global–local network can greatly ease the learning problem, so that it can be handled by using a lightweight network architecture. To cope with the complex and changeable underwater environment, we then design a compressed-histogram equalization to complement the data-driven deep learning, in which the parameters are fixed after training. The proposed compression strategy is able to generate vivid results without introducing over-enhancement and extra computing burden. Experiments demonstrate that our method significantly outperforms several state-of-the-arts in both qualitative and quantitative qualities.

1. Introduction

Particles suspended in the water lead to light absorption and scattering [1], which severely degrade captured underwater image quality, e.g., color distortion and contrast reduction. Therefore, obtaining high quality underwater images is of great importance for consumer underwater photography, ocean engineer [2], underwater archeology [3] and aquatic robot inspection [4].

To improve the quality of single underwater images, various methods based on image enhancement, image restoration and deep learning are well explored. Image enhancement-based methods [5,6] aim to directly process image pixel values to enhance specific image characteristics, e.g., color, contrast and brightness. While image restoration-based methods [7,8] treat image quality improvement as an inverse problem. This kind of methods usually utilize physical imaging models and explore image prior as constraints to restore the clear images. Recently, due to the powerful modeling capabilities and rich feature representations learned from massive training data, deep learning has achieved remarkable performance on both high-level vision tasks [9–12] and image processing [13–16]. Several methods based on deep learning [17–19] are also proposed to extract effective features from synthetic data for improving the underwater image quality.

Despite achieving significant progress, addressing the light absorption and scattering problem in underwater images improvement is still challenging. On the one hand, since the underwater imaging environment is complex and affected by too many factors, it is impractical to

find a universal method by using only hand-crafted algorithms which usually adopt simplified models and have small model capacities. On the other hand, since the parameters of deep network will be fixed after training, deep learning-based methods lack sufficient flexibility to handle the changeable underwater environment. Once the type of new underwater images is different from the training set, the deep network will not be able to adjust according to the input data. In addition, existing deep learning-based methods requires large numbers of parameters to learn the complex mapping functions. This limits the potential value of these deep learning methods in practical applications.

To effectively and efficiently improve the quality of underwater images, this paper proposes a novel model by combining the merits of deep learning and classical histogram equalization. Specifically, instead of directly learning an image-to-image mapping function, we first design a two-branch network to separately handle global color distortion and local contrast reduction. This strategy is motivated by the observation that most underwater images are dominated by a relative single and uniform color distribution [5,20]. Adopting this global–local network architecture can significantly simplify the learning problem. However, as mentioned above, a single and trained deep network cannot cover all types of underwater images. Therefore, to deal with the complex and changeable underwater imaging environment, we further propose a compressed-histogram equalization to complement the global–local network. As shown in Fig. 1, by taking advantages of both data-driven deep leaning and hand-crafted image enhancement,

Fig. 1. Examples of real-world underwater images (top) and our results (bottom). Our single model is able to handle various types of underwater images while only contains 7292 trainable parameters.

our single, lightweight and easy-to-implement model is able to handle various types of underwater images.

Our global–local networks can be treated as a compensation process which play a role similar to the image restoration. The network first learns prior from training data to generates an intermediate image, in which the lost information caused by light absorption and scattering is compensated or restored to some extent. Then a compressed-histogram equalization is utilized as a post-processing for obtaining a higher image quality. After enhancing the intermediate image, the final result contains more vivid colors and higher contrast. In fact, due to the effective information compensation, the proposed global–local networks can provide an initial solution and benefit other existing methods, especially in avoiding over-enhancement.

The contributions of this work are summarized as follows:

1. We propose a novel approach by integrating data-driven deep learning and hand-crafted image enhancement for single underwater images enhancement. We argue that it is impractical to adopt only one kind of method to face such a complex underwater imaging environment. By combining the deep learning and image enhancement technology, our single model is able to handle a wide variety of underwater images without extra parameters and computational burden.

2. We propose a two-branch network to compensate both the global distorted color and local reduced contrast. To effectively connect the two network paths, a global residual-guide biases is further designed to interact global and local information. The proposed global–local networks help ease the learning process and support effective enhancement in the subsequent module.

3. We propose a compressed-histogram equalization to further improve the image quality. This algorithm, which is in the classic histogram equalization framework, is efficient and able to generate vivid visual results without introducing over-enhanced artifacts.

4. To our knowledge, the resulting easy-to-implement model is lightweight while achieves the state-of-the-art performance on both synthetic and real-world underwater images. We also show how our method can benefit downstream applications, such as diver detection and keypoints matching.

2. Related work

In general, existing methods of improving single underwater images can be roughly divided into three categories: enhancement-based methods, restoration-based methods and deep learning-based methods. The first two kinds of methods are hand-crafted designed, while the deep learning-based methods aim to automatically learn a nonlinear mapping function from training data.

2.1. Image enhancement-based methods

Image enhancement technology aims to produce visually pleasing results based on some assumptions on high-quality images. This kind of method usually focuses on enhancing specific objectives and directly process image pixel values, regardless of the physical degradation model. In [5], a fusion-based method is presented and achieve impressive performance on underwater images and videos enhancement. This method first derives a color corrected image and a contrast enhanced version from the underwater image. Then a multi-scale fusion strategy is adopted to fuse the two derived images into the final result. This method is able to improve the global contrast and visibility with a fast computational time. Recently, this method is further improved in [6] by introducing a novel white balancing strategy and a revised fusing implementation. By modifying [21], a Rayleigh-stretched contrast-limited adaptive histogram method [22] is proposed to enhance underwater images. The number of under-enhanced and over-enhanced regions can be effectively constrained. Another line of enhancement tries to process underwater images based on the simplified Retinex model. In [23], a variational Retinex-based method is proposed for underwater image enhancement. This method contains three steps, i.e., color correction, layer decomposition and post-enhancement. In [24], the authors introduce an extended multi-scale Retinex-based method and simulate the underwater turbidity conditions by using mixture of milk and fruit juice.

2.2. Image restoration-based methods

Image restoration-based methods treat the underwater image enhancement as an ill-posed inverse problem. In this line of methods, the physical imaging model and various prior are explored to estimate the desired results from degraded inputs. To the underwater community, the most widely used imaging model is the Jaffe–McGlamery underwater optical model [25,26], which is a simplified radiative transfer model. Mathematically, this model is often expressed as:

$$\mathbf{I} = \mathbf{J}e^{-\eta\mathbf{d}} + \mathbf{A}(1 - e^{-\eta\mathbf{d}}), \tag{1}$$

where $\mathbf{I}$ is the captured underwater image, $\mathbf{J}$ is the clear image, $\mathbf{A}$ is a back-scattered light, $\mathbf{d}$ is the distance between the camera and the object and η is the attenuation coefficient. The exponential term $e^{-\eta\mathbf{d}}$ is also known as the transmission $\mathbf{t}$ through the underwater medium. To invert Eq. (1) to estimate $\mathbf{J}$, many researchers focus on exploring effective image prior to constrain this inverse problem [27–30]. In [8], the authors employ a haze-lines prior [31] to estimate the transmission. In [32], the transmission is estimated by proposing a prior which exploited the attenuation difference among the RGB channels. The light scattering can be removed by using the predicted transmission. In [33],

the authors estimate the transmission by combining the proposed blurriness prior with the method [32]. Since Eq. (1) has a similar form to the hazy imaging model [34], the popular dark channel prior (DCP), which is initially proposed for single image de-hazing [35], has been widely modified and utilized to restore underwater images. In [36], based on DCP, a wavelength dependent compensation algorithm is explored to restore the clear image. In [37], an underwater dark channel prior is specifically designed for underwater scenarios, which helps estimate more accurate transmission than the conventional DCP. Based on the observation that the red component reciprocal increases with the distance to the camera, a red channel prior [38] is introduced to restore the colors with short wavelengths. In [39,40], the authors utilize the color-lines prior [41] and a modified DCP to estimate the back-scattered light and transmission, respectively. In [7,42], a contrast enhancement algorithm is combined with an image de-hazing algorithm to generate two enhanced results. One contains vivid color and the other has high contrast to reveal image details. Very recently, a generalized dark channel prior [43] is proposed to estimate the transmission and recover the degraded images.

However, due to the complex and changeable underwater environment, hand-crafted prior does not always hold. For example, when large white objects or artificial light exist in underwater images, the underwater dark channel prior will fail. Additionally, in underwater environment, color distortion strongly depends on the different light wavelengths [44–46]. Directly using the simplified model (1) cannot accurately describe the process that the selective absorption of colors. Therefore, in this work, we do not resort to an explicit physical model to compensate and restore the lost information.

2.3. Deep learning-based methods

Different with hand-crafted 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. In [47], the authors synthesize underwater images by adopting Eq. (1) for network training. However, light absorption along horizontal direction is not taken into consideration, which makes the synthetic data unable to accurately simulate the real underwater environment. To address the above problem, [19,48] utilize the generative adversarial networks (GANs) [49] to generate underwater images in the weakly supervised learning manner. By using the generated data, a real-time color correction network is further trained to improve the underwater image quality. In [50], a residual multi-scale dense block is designed and deployed in the generator to perform underwater images enhancement. In [17], the GANs are also adopted to generate a underwater dataset for underwater image restoration. In [18], a pixels disrupting strategy is introduced to improve the convergence speed and accuracy of network training. In [51], a weakly supervised model for underwater image color correction is introduced. In this method, a nonlinear cross domain mapping function of color transfer is learned, which relaxes the need for explicit one-to-one paired labels. In [52], an underwater residual convolutional neural network is proposed to predict the transmission. This deep model is trained on synthetic underwater images generated by using wavelength-dependent attenuation prior. Recently, by integrating domain-knowledge priors and information from training data, two deep prior ensemble framework [53,54] are proposed for underwater images enhancement.

Despite the powerful nonlinear modeling capacity, the performance of existing deep learning-based underwater image enhancement methods do not match the success of recent deep learning-based low-level vision problems [13–16,55]. This is because collecting sufficient and effective real-world underwater/clear image pairs for training deep networks is very difficult. Meanwhile, due to the complex underwater imaging environment, the synthetic underwater images are unable to cover all possible underwater conditions. Since the parameters of deep networks are fixed after training, the prior and features learned from synthetic data cannot well generalize to the changeable real-world scenarios, which limits the practical values of deep learning-based methods.

3. Methodology

Based on the above analysis, we argue that utilizing only hand-crafted or deep learning method cannot well handle the tough underwater environment. Therefore, we choose to take advantages of both kinds of methods for underwater images enhancement. We illustrate the proposed framework in Fig. 2. As discussed in more detail below, to the problem of global distorted color and local reduced contrast, we first utilize a lightweight two-branch network to compensate the lost information. We then introduce a compressed-histogram equalization to complement the global–local networks to further improve visual quality. The final enhanced image thus contains more vivid color and higher contrast.

3.1. Global-local networks

3.1.1. Motivation

Since deep learning has powerful modeling capabilities, an intuitive way to enhance underwater images quality is to directly train a deep network that transforms the input $\mathbf{I}$ to output $\mathbf{J}$. However, different with other image restoration problems, e.g., de-noising [14] and super-resolution [13], the underwater image contains not only local details lost but also global color distortion over all image pixels. Therefore, to underwater community, most existing deep learning-based [17,48,56] methods usually adopt the U-Net architecture [57] to capture both global and local features, at the cost of parameters burden. However, the spatial information, e.g., textures and structures, contained in the down-sampled features still introduce interference for color correction. Therefore, it is difficult to simultaneously compensate global distorted color and local reduced contrast by directly learning an image-to-image mapping function, even with the help of GANs [49] to extract the underlying distribution.

To effectively address the above issues, we instead use a priori knowledge to separately process the global average and local centered image. The motivation for designing this global–local network is that most underwater images are dominated by a relative single and uniform color distribution. In other words, the global average of an underwater image is able to represent its overall color information. This inspires us to design our global–local networks to separately process the global average $\mu_{\mathbf{I}}$ and the centered image $\mathbf{I}_{centered} = \mathbf{I} - \mu_{\mathbf{I}}$.

3.1.2. Network architecture

As shown in Fig. 2, the top branch is used to compensate global color information while the bottom one aims to improve local contrast. We name the two sub-networks as Network-G and Network-L, respectively. In the Network-G, we input both average $\mu_{\mathbf{I}}$ and standard deviation $\sigma_{\mathbf{I}}$ to provide not only first-order color measurement but also second-order dispersion information. To color images, both $\mu_{\mathbf{I}}$ and $\sigma_{\mathbf{I}}$ are 1×3 vectors. We concatenate $\mu_{\mathbf{I}}$ and $\sigma_{\mathbf{I}}$ and adopt a fully connected network to construct the top branch network. Inspired by the densely connected networks [11], we concatenate all hidden features to predict the residual $\hat{\Delta\mu}$ and use it to generate the compensated average $\hat{\mu}_{\mathbf{J}}$. The Network-G contains four layers and can be expressed as:

$$\mathbf{h}_1 = ReLU(\mathbf{w}_1 \cdot concat(\mu_{\mathbf{I}}, \sigma_{\mathbf{I}}) + \mathbf{b}_1),$$
$$\mathbf{h}_2 = ReLU(\mathbf{w}_2 \cdot \mathbf{h}_1 + \mathbf{b}_2),$$
$$\mathbf{h}_3 = ReLU(\mathbf{w}_3 \cdot \mathbf{h}_2 + \mathbf{b}_3),$$
$$\hat{\Delta\mu} = \mathbf{w}_4 \cdot concat(\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3) + \mathbf{b}_4,$$
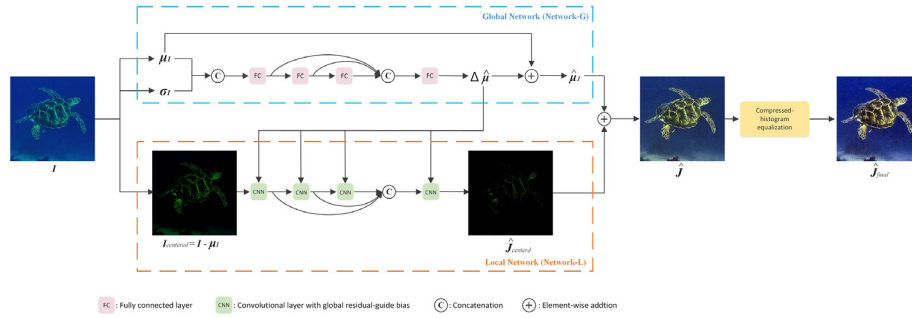$$\hat{\mu}_{\mathbf{J}} = Sigmoid(\mu_{\mathbf{I}} + \hat{\Delta\mu}), \tag{2}$$

Fig. 2. Our proposed model for underwater image enhancement. The global–local network contains two branches, i.e., Network-G for global average and Network-L for local contrast. The compensated image $\hat{\mathbf{J}}$ is further enhanced by our compressed-histogram equalization to obtain the final result $\hat{\mathbf{J}}_{final}$.
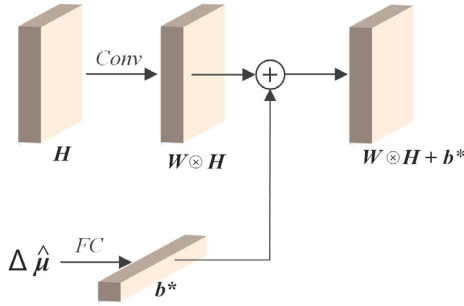

Fig. 3. Global residual-guide bias in our CNN unit. The $\Delta\hat{\mu}$ is fully connected to produce the vector $\mathbf{b}^*$. The new feature map is generated with common convolutional operations, except for using $\Delta\hat{\mu}$ to generate the bias.

where $\mathbf{h}$ is the hidden features, $\mathbf{w}$ and $\mathbf{b}$ are learnable weights and biases, $ReLU(\cdot)$ and $Sigmoid(\cdot)$ are the rectified linear units and sigmoid functions, respectively. $concat(\cdot)$ is the concatenation. The sigmoid activation is utilized to constrain the range of $\hat{\mathbf{J}}_{\mathbf{J}}$ between 0 and 1.

The Network-L has a similar architecture with the top one, except for using the convolutional operations to process the input matrix $\mathbf{I}_{centered}$. It is worth noting that $\mathbf{I}_{centered}$ still contains color distortion even the average is subtracted. Simply stacking convolution layers tends to remove distorted color information, which is not good for accurate local contrast compensation. To address this issue, we propose a global residual-guide bias, as shown in Fig. 3, to replace the initialized bias with value zero used in common deep CNNs. We use the symbol $\mathbf{b}^*$ to denote this new bias that depends on the predicted residual $\Delta\hat{\mu}$ and varies with respect to this global compensated information. We generate $\mathbf{b}^*$ by using a one-layer fully connected network:

$$\mathbf{b}^* = \mathbf{w} \cdot \Delta\hat{\mu}, \tag{3}$$

where $\mathbf{w}$ is the learnable weight to perform a fully connection without bias. The Network-L structure also contains four layers can be expressed as:

$$\mathbf{H}_1 = ReLU(\mathbf{W}_1 \otimes \mathbf{I}_{centered} + \mathbf{b}_1^*),$$
$$\mathbf{H}_2 = ReLU(\mathbf{W}_2 \otimes \mathbf{H}_1 + \mathbf{b}_2^*),$$
$$\mathbf{H}_3 = ReLU(\mathbf{W}_3 \otimes \mathbf{H}_2 + \mathbf{b}_3^*),$$
$$\mathbf{H}_3 = ReLU(\mathbf{W} \otimes \mathbf{H} + \mathbf{b}^*),$$
$$\hat{\mathbf{J}}_{centered} = \mathbf{W}_4 \otimes concat(\mathbf{H}_1, \mathbf{H}_2, \mathbf{H}_3) + \mathbf{b}_4^*, \tag{4}$$

where $\mathbf{W}$ is the learnable convolutional kernels, $\mathbf{H}$ is the hidden feature maps and $\otimes$ is the convolutional operation. The convolutional operation used in our global–local networks is the same with standard CNNs, except for using $\Delta\hat{\mu}$ to generate the bias $\mathbf{b}^*$. The kernels $\mathbf{W}$ are used to extract local features while the $\Delta\hat{\mu}$ guide the global direction of contrast

compensation. In addition, during the back-propagation process, the gradient of the Network-L also flows through the Network-G, which provides more supervised information for compensating distorted colors. In this way, the two branch networks can be effectively connected and interacted. The intermediate image with compensated information can be obtained by:

$$\hat{\mathbf{J}} = \hat{\mathbf{J}}_{centered} + \hat{\mu}_{\mathbf{J}}. \tag{5}$$

3.1.3. Loss function

Given a training set $\{\mathbf{I}, \mathbf{J}_{GT}\}$, where $\mathbf{I}$ are the underwater images and $\mathbf{J}_{GT}$ are the corresponding clear images, the most widely used loss function for training a network is ℓ_2 loss, i.e., the mean squared error. However, ℓ_2 usually generates over-smoothed results due to the squared penalty that works poorly at edges in an image. Following [58], we choose the SSIM [59] as our loss function $\mathcal{L}_{SSIM}$ for reconstructing underwater images:

$$\mathcal{L}_{SSIM} = 1 - SSIM(\mathbf{J}_{GT}, \hat{\mathbf{J}}), \tag{6}$$

where $SSIM(\cdot)$ is the SSIM and the detailed definition can be found in [59].

Since the proposed global–local network has two branches, it may cause instability if only the SSIM loss is used. To obtain a stable solution, we also formulate an average loss $\mathcal{L}_{avg}$ to intentionally guide the Network-G training. This loss is defined based on the cosine distance of two 1×3 vectors:

$$\mathcal{L}_{avg} = 1 - \cos(\mu_{GT}, \hat{\mu}_{\mathbf{J}}), \tag{7}$$

where $\cos(\mathbf{a}, \mathbf{b}) = \frac{\mathbf{a}^{\mathrm{T}}\mathbf{b}}{\|\mathbf{a}\|\|\mathbf{b}\|}$, $(\cdot)^T$ is transpose operation and $\|\cdot\|$ is magnitude of the vector. Since the ℓ_2 loss only numerically measures the values difference, it cannot ensure that the vectors have the same direction. Moreover, the calculating process of cosine, which is simple and fast for computation, has already implicitly measured the ℓ_2 difference.

The overall loss function consists of two components and is minimized during the network training. It is expressed as:

$$\mathcal{L}_{total} = \mathcal{L}_{SSIM} + \alpha\mathcal{L}_{avg}, \tag{8}$$

where α is the corresponding weight.

3.2. Compressed-histogram equalization

After network training, the parameters of proposed global–local networks will be fixed. Since the underwater imaging environment is complex and changeable, it is impractical to use single trained network to cover all kinds of underwater images. Therefore, we further introduce a compressed-histogram equalization to complement the trained global–local networks and generate results with more vivid colors and higher global contrast.

Due to the simplicity and quickness, histogram equalization (HE) [60] is the most widely used global contrast enhancement technique.

HE is designed based on the assumption that the desired histogram should meet a uniform distribution. The enhanced image is obtained by matching the cumulative input histogram to the uniform histogram. However, one main drawback of HE is that it often generates over-enhanced artifacts when large peaks exist in the input histogram. In this work, we propose a simple yet effective compressed-histogram equalization in the classical HE framework. Specifically, we adopt a compression operation to modify the input histogram. By adding this simple operation into the classic HE procedure, the compressed-histogram equalization can generate naturally looking results without introducing artifacts.

Given the compensated image $\hat{\mathbf{J}}$, the goal is to generate the final enhanced image $\hat{\mathbf{J}}_{final}$ with a pixel value range of $\hat{J}_{min}$ and $\hat{J}_{max}$. To avoid the over-enhancement problem, an intuitive solution is to compress peaks existed in the input histogram. Due to the character of compression and monotonically increasing, we adopt the logarithmic operation to modify the input histogram. Utilizing the log-operation can effectively compress large peak values while preserves the order of the input histogram. Additionally, as described by Weber's law [61], the log-operation accurately characterizes human visual perception, which implicitly modify the input histogram according to human visual preferences. Other effective compression methods can also be used to compress the histogram.

First, the compressed histogram $\mathbf{h}_{comp}$ is simply generated by:

$$\mathbf{h}_{comp}(l) = \log(\mathbf{h}(l)), \tag{9}$$

where $\mathbf{h}$ is the input histogram, $l = 1, \ldots, L$ and L is the number of the grayscale levels that exist in $\hat{\mathbf{J}}$. Then, the distribution function f is obtained by normalizing $\mathbf{h}_{comp}$:

$$f(l) = \mathbf{h}_{comp}(l) / \sum_{t=1}^{L} \mathbf{h}_{comp}(t). \tag{10}$$

The uniform distribution function F is computed as:

$$F(l) = \sum_{t=1}^{l} f(t). \tag{11}$$

The final enhanced image $\hat{\mathbf{J}}_{final}$ can be obtained by using F, $\hat{J}_{min}$ and $\hat{J}_{max}$ in a standard lookup table-based HE procedure. We point out that, by removing Eq. (9), this algorithm equals the classic HE processing. Adding the log-operation is easy to implement and does not introduce a large computational burden. In this work, each RGB channel is separately enhanced. In Fig. 4, we use a gray-scale image to demonstrate the effect of our proposed compressed-histogram equalization. When peaks exist in the input histogram, the classical HE tends to produce over-enhanced results as shown in Fig. 4(b). By adding the simple log-operation, the global enhanced result has a significant improvement on artifacts suppression and naturalness preservation. Meanwhile, the histogram shape is well retained shown in Fig. 4(c).

One result on a real-world underwater image is shown in Fig. 5. It can be seen that the global–local network first compensates the lost information. Both global color and local contrast are significantly improved, shown in Fig. 5(b). Then, by adding the compressed-histogram equalization, the enhanced result contains more vivid color and higher global contrast, shown in Fig. 5(d). Moreover, compared to the classical HE in Fig. 5(c), our method achieves a good trade-off between image enhancement and naturalness preservation.

3.3. Discussion

Our model is composed of two parts: global–local networks and compressed-histogram equalization. The former is used to compensate lost information which can be seen as an intermediate result. The later aims to further improve the image quality and can be seen as a post-processing stage. We first utilize our domain specific knowledge to compensate the distorted color by predicting the global average. Compared to learning an image-to-image transformation, mapping one 1×6 vector (i.e., concatenated $\mu_{\mathbf{I}}$ and $\sigma_{\mathbf{I}}$) to a 1×3 vector (i.e., $\hat{\mu}_{\mathbf{J}}$) is significantly easier for a deep learning model. This is because the mapping range has been significantly decreased. By using this divide-and-conquer network architecture, the problem of underwater image information compensation can be easily handled.

Since our compressed-histogram equalization does not require training data, it is able to complement global–local networks to handle new kinds of underwater images. Meanwhile, the output of global–local networks also provides a good initial result for the compressed-histogram equalization.

Moreover, our method is easy to implement, all network modules are constructed by using standard deep learning operations, e.g., fully connected layers and convolutional layers. While the proposed compressed-histogram equalization is in the classical HE framework. This is because we decompose the tough underwater image problem into different easy sub-problems, which are separately handled by each part of our model. Since the sub-problems are simplified, the corresponding parts can well solve these problems with relative simple operations. In this work, we mainly focuses on how to well handle underwater images from the perspective of signal processing. Other advanced methods, such as squeeze-and-excitation networks blocks [62], non-local networks [63] and 2D histogram equalization [64], can also be incorporated into our framework to further improve the image quality.

4. Experimental results

In this section, we first present the experimental settings and then conduct several groups of experiments to verify the effectiveness of the proposed method.

4.1. Experimental settings

Parameters setting. For the global–local networks, all kernel size for convolutions are 3×3 and the numbers of hidden layers (both Network-G and Network-L) are 16. The total numbers of trainable parameters are 7292, far fewer than the hundreds of thousands often encountered in deep learning. The parameter α is empirically set as 0.5. For the compressed-histogram equalization, the entire dynamic range is used, i.e., $\hat{J}_{min} = 0$ and $\hat{J}_{max} = 2^8 - 1 = 255$ for 8-bit images.

Network training setting. We use synthetic underwater images from [17] and [65] as our training and testing data. The dataset of [17] contains 6128 synthetic images, of which the first 5000 images used for training and the rest 1128 images for testing. While the dataset of [65] contains 890 synthetic images, of which the first 700 images used for training and the rest 190 images for testing. Note that we only train a single model to handle both datasets. We use TensorFlow [66] to train our networks using the Adam solver [67] with a mini-batch size of 1. We set the learning rate as 0.001. The network is trained in an end-to-end fashion and finished after 40 epochs.

Compared methods. We compare our model with five state-of-the-art methods, which include one image enhancement-based approach (fusion-based [5]), three image restoration-based approaches (histogram distribution prior [7], haze-line prior [8] and blurriness-based [33]), and one deep learning-based approach (DL-based [50]).

4.2. Experiments with ground truth

In this section, two datasets with ground truth are chosen for comparison. One is from [17] and contains 1128 testing images. This dataset is generated by using the underwater imaging model. The other is from [65] which contains 190 synthetic images. Note that this dataset from [65] is constructed by using real-world underwater images while
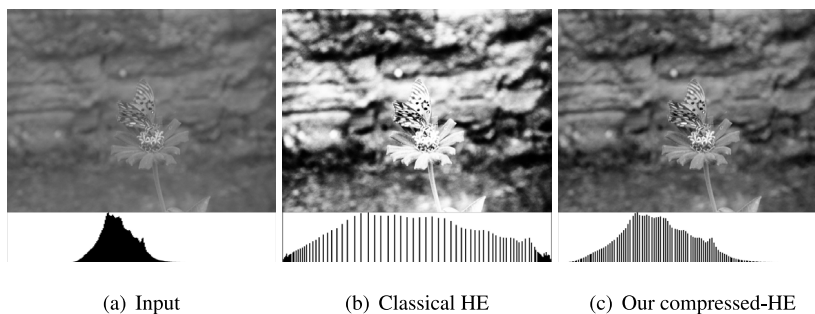
(a) Input   (b) Classical HE   (c) Our compressed-HE

Fig. 4. An example of our proposed compressed-HE algorithm on a gray-scale image. Bottom shows corresponding histograms.

(a) Input   (b) Our $\hat{\mathbf{J}}$ (c) HE on $\hat{\mathbf{J}}$ (d) Our $\hat{\mathbf{J}}_{final}$
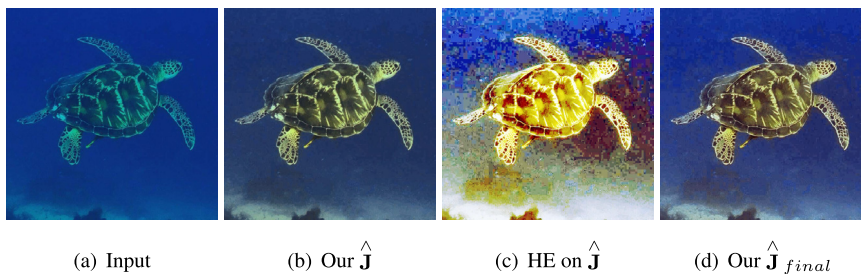
Fig. 5. Effect of each part of our model. The global–local network compensates the lost information shown in (b). Using the classical HE to enhance $\hat{\mathbf{J}}$ introduces obvious over-enhanced artifacts shown in (c). While utilizing our compressed-histogram equalization generates a natural looking result with more vivid color and higher contrast, as shown in (d).
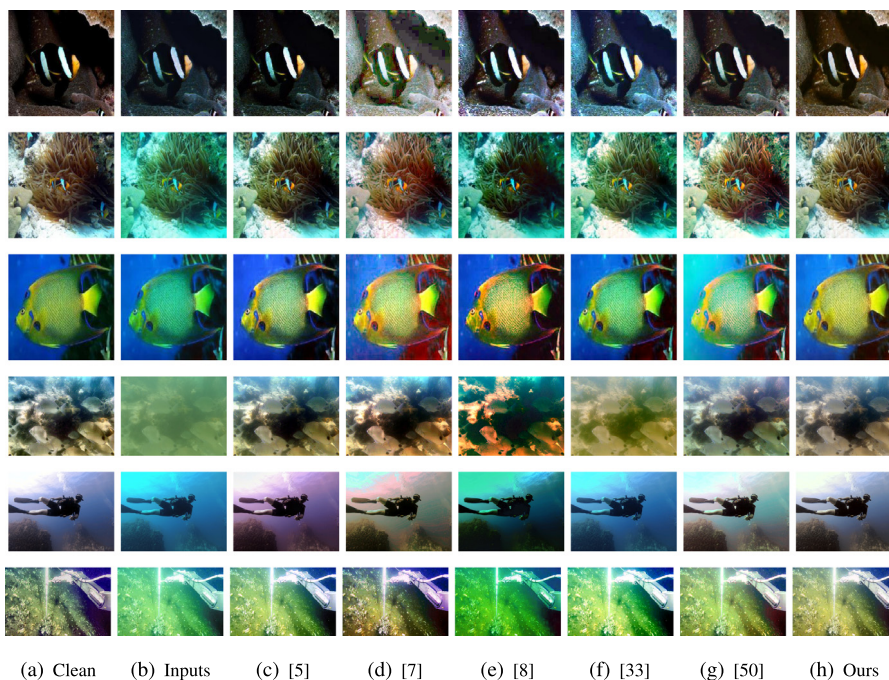

(a) Clean   (b) Inputs   (c) [5]   (d) [7]   (e) [8]   (f) [33]   (g) [50]   (h) Ours

Fig. 6. Visual comparisons on synthetic underwater images. The testing images in the first three rows are from the dataset [17] while the rest are from [65].

the corresponding clear images are provided according to laborious pairwise comparisons.

Fig. 6 shows visual results from the two dataset. As can be seen, histogram distribution prior [7] tends to generate obvious reddish color deviation (2nd and 3rd rows) while haze-line prior [8] and blurriness-based [33] fail to correct distorted colors (4th and 5th rows). This because the above three methods are designed based on hand-crafted prior, which may not be strong enough to handle all types of underwater imaging environment. The similar problem also exist in the DL-based method [50]. Since the back-scattered light and piece-wise constant transmission are jointly obtained through down-sampled feature map, local textures and details will lead to inaccurate feature extraction and image restoration. The fusion-based method [5] has relatively good performance on color correction and contrast enhancement. However, the performance is limited when facing underwater images under harsh conditions, e.g., insufficient illumination (1st row) and extreme bluish appearance (5th row). This is due to the limitation of this method which directly maps pixels without considering the imaging process. Our model has comparable visual results with fusion-based
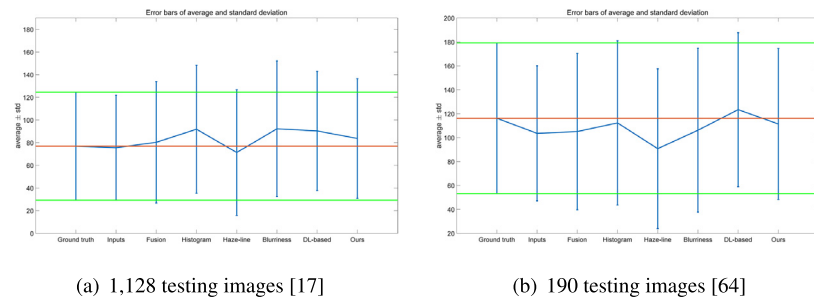
(a) 1,128 testing images [17]   (b) 190 testing images [64]

Fig. 7. Error bars of statistical analysis of different methods. The values of ground truth are indicated by red line and green line, respectively.
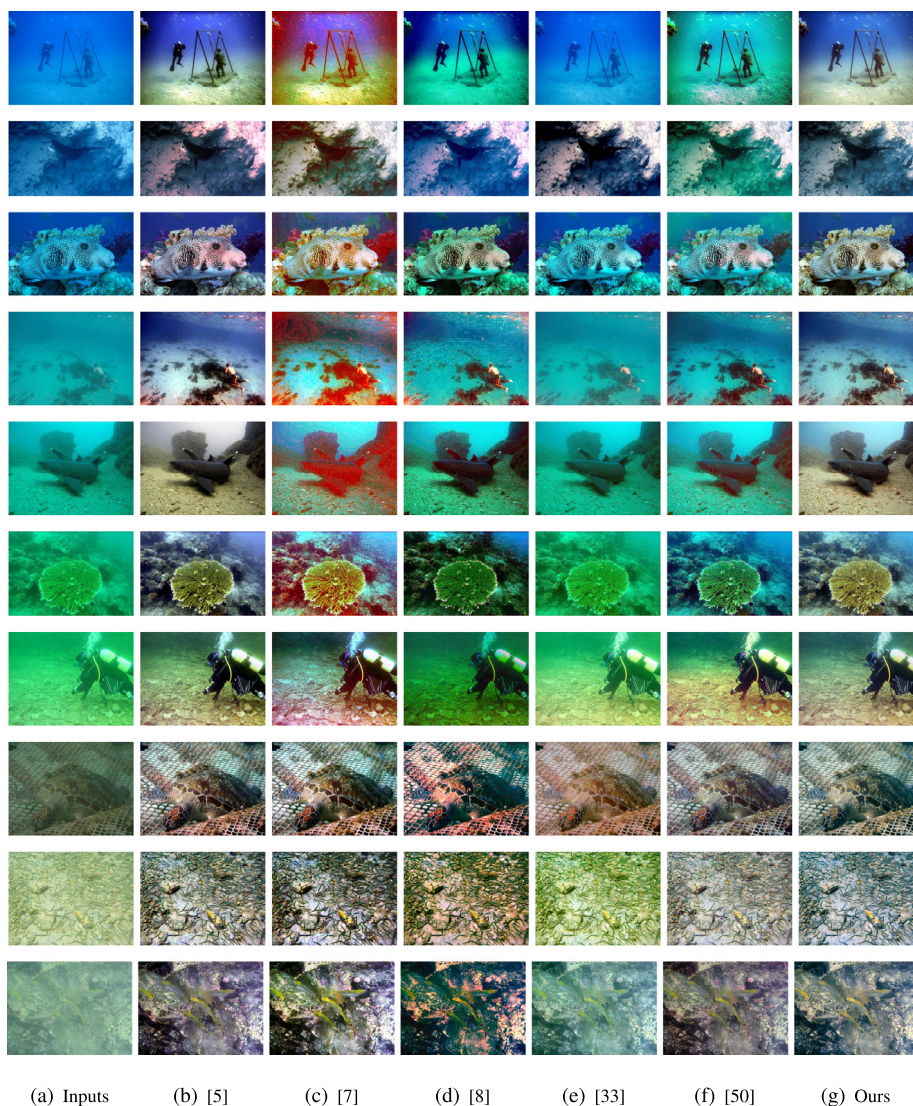

(a) Inputs   (b) [5]   (c) [7]   (d) [8]   (e) [33]   (f) [50]   (g) Ours

Fig. 8. Visual comparisons on the UCCS dataset [68]. The proposed method is able to handle all the three degrees of color cast.

method and outperforms other methods. Moreover, our single model is able to handle different kinds of underwater images. The reason is two-fold: first, we design our network to explicitly and separately process the global average and local contrast. Without interference from other components, the proposed global–local networks is able to extract more accurate features for more accurate image compensation. Second, we utilize the HE technology to further improve the image quality. The defect of fixed parameters of trained network is thus complemented. Therefore, although we only train a single network with relatively few parameters, these additional aspects enables our whole model to well handle different kinds of underwater images.

We then show a statistical analysis of different methods in Fig. 7. On the dataset [17], the average and standard deviation of inputs are very close to the ground truth. This is because this dataset adopt a relatively simple operations to synthesize underwater images. While on the dataset [65], which built on real-world underwater images, the mean of global averages of our methods has the smallest distance from ground truth. Meanwhile, our standard deviation is relatively narrow, which implies that the over-enhancement is well suppressed since the standard deviation represents the intensity of changes in image content.

We also adopt PSNR and SSIM [59] to perform quantitative evaluations in Table 1. Our method has the best PSNR and SSIM values on the two datasets. Though we do not use the ℓ_2 loss, which is also used

| (a) Inputs | (b) [5] | (c) [7] | (d) [8] | (e) [33] | (f) [50] | (g) Ours |

Fig. 9. Visual comparisons on real-world underwater images. The proposed method is able to well enhance different kinds of real-world underwater images with naturalness preservation.

Table 1
Average SSIM and PSNR comparisons with **first** and *second* best performances indicated.

Dataset	Underwater		Fusion [5]		Histogram [7]		Haze-line [8]		Blurriness [33]		DL-based [50]		Ours	
	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR
1128 images [17]	0.61	15.83	0.63	16.81	0.40	12.26	0.40	11.93	0.53	13.70	0.63	16.55	**0.66**	**18.10**
190 images [65]	0.79	19.40	0.86	21.33	0.67	15.08	0.80	17.09	0.77	16.03	0.79	18.03	**0.88**	**21.98**
Ideal value	1.00	+∞	1.00	+∞	1.00	+∞	1.00	+∞	1.00	+∞	1.00	+∞	1.00	+∞

for calculating PSNR, our model still achieves the best PSNR. Note that the overall color of the image has the greatest impact on PSNR since PSNR measures global pixel errors. This demonstrates the validity of using the global average to compensate the global distorted color.

4.3. Experiments without ground truth

To demonstrate the generalization ability of our model, we test our model on two real-world datasets. The first is the underwater color cast set (UCCS) which contains 300 images provided by [68]. The images of this dataset are divided into green–blue, blue and green according different color cast degrees. We also collect 200 real-world underwater images from the Internet as the second dataset. The images of this dataset contain various and severe degradations. Figs. 8 and 9 shows visual results on the two real-world datasets. As can be seen, the

compared methods cannot well handle all types of degraded images. For example, the fusion-based method [5] has obvious reddish color deviation, while other methods cannot accurately correct colors with severe distortions. On the contrary, our model is able to consistently generate natural looking results on all testing images. This is because our model combines the merits of data-driven deep learning and hand-crafted image enhancement, which helps to deal with real-world scenarios that not contained in the training data.

Since no ground truth exists, we cannot definitively say which method performs quantitatively the best. Instead, we adopt two reference-free metrics, i.e., UIQM [69] and UCIQE [70] for evaluation. These two metrics are widely used for evaluating real-world underwater image quality. A higher UIQM value indicates the image is more consistent with the human visual perception. A higher UCIQE value indicates the image has better balance among the chroma, saturation

Table 2
Average UIQM and UCIQE values on the UCCS dataset [54], with **first** and *second* best performances indicated.

	Underwater	Fusion [5]	Histogram [7]	Haze-line [8]	Blurriness [33]	DL-based [50]	Ours
UIQM	0.21	2.50	2.98	**3.79**	2.33	2.34	2.50
UCIQE	0.41	0.52	0.53	**0.67**	0.50	0.51	0.57

(a) Inputs (b) Fusion [5] (c) Histogram [7]

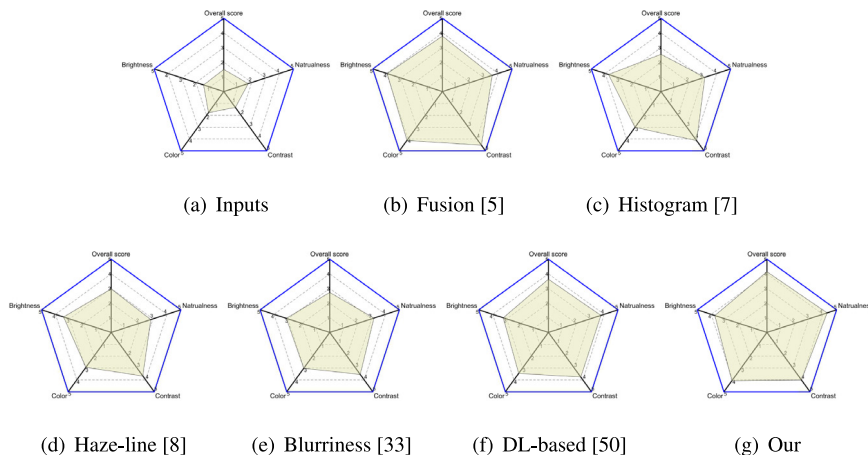(d) Haze-line [8] (e) Blurriness [33] (f) DL-based [50] (g) Our

Fig. 10. Radar charts of average user study scores on overall visual quality, naturalness preservation, contrast enhancement, color perception and brightness improvement. The score 1 represents the worst quality and 5 represents the best quality.
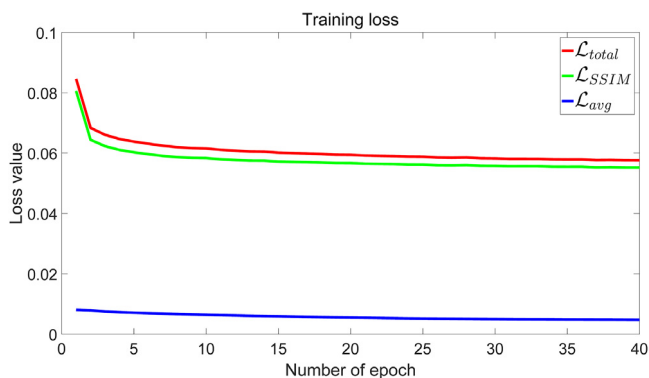

Fig. 11. Convergence curve of network training.

and contrast. The average scores of 200 testing real-world images are shown in Tables 2 and 3. As can be seen, our method achieves promising results on both UCIQE and UIQM metrics.

However, as with all reference-free image quality metrics, UCIQE and UIQM are arguably not always subjectively correct. To provide realistic feedback and quantify the subjective evaluation, we also constructed an independent user study. In this experiment, we use the results of all the 500 real-world underwater images scored with UCIQE and UIQM. For each image, we randomly order the outputs of the all algorithms, as well as the original underwater image, and display them on a screen. We then separately asked 10 participants to rank each image from 1 to 5 subjectively according to five measurements, i.e., the overall visual quality, naturalness preservation, contrast enhancement, color perception and brightness improvement. The participants are given instructions that color distortion and over-enhanced artifacts should decrease the quality, and high contrast should increase the quality. The score 1 represents the worst quality image and 5 represents the best quality image. We show the average overall scores in Table 4. As is evident, the fusion-based method [5] does clearly improve the underwater image, while our model is subjectively superior to all methods. Moreover, we also show the radar chart of the five measurements in Fig. 10. It is clear that our method has promising results on all

measurements, which gives additional support that our model improves the subjective visual qualities of real-world images.

4.4. Convergence and testing runtime

We first show the average training loss as a function of training epoch in Fig. 11. Note that the average loss $\mathcal{L}_{avg}$ has low values in the initial epochs of training, which proves that the problem of global average compensation is easy to handle. Since we simplify the problem in a divide-and-conquer way, this allows our network to converge quickly and easily fit the mapping functions of each subproblem.

To demonstrate the efficiency of our model, we further compare the average testing runtime of different methods. Three different image sizes are chosen and each one is tested over 100 images. All experiments are performed on a PC with Intel(R) i7-8700 CPU, 32 GB RAM and one NVIDIA GTX 1080Ti GPU. The deep learning-based methods are tested on both CPU and GPU. The average running time is shown in Table 5. The fusion-based method [5] is the fastest across different image sizes on CPU, while our model ranks the second fastest. This is because our network is completely feed-forward after network training. Other image restoration-based methods have relatively slow running time since complicated inferences are required to process each new image. Meanwhile, our method can be significantly accelerated by using parallel GPU implementation. Note that our global–local networks contains relatively few parameters and compact structures, which makes our model require less operations than the DL-based method. We further record the average run time of each stage on CPU for clarity. As shown in Table 6, the global–local network consumes most of the running time. This is because our network contains two branches, and the bottom branch needs to wait for the top branch to finish running. In addition, compared to other operations, convolutional operations require more calculations and runtime. Therefore, the global–local network is significantly slower than compressed-histogram equalization.

4.5. Ablation studies

Here we discuss different configurations to study their impact on performance.
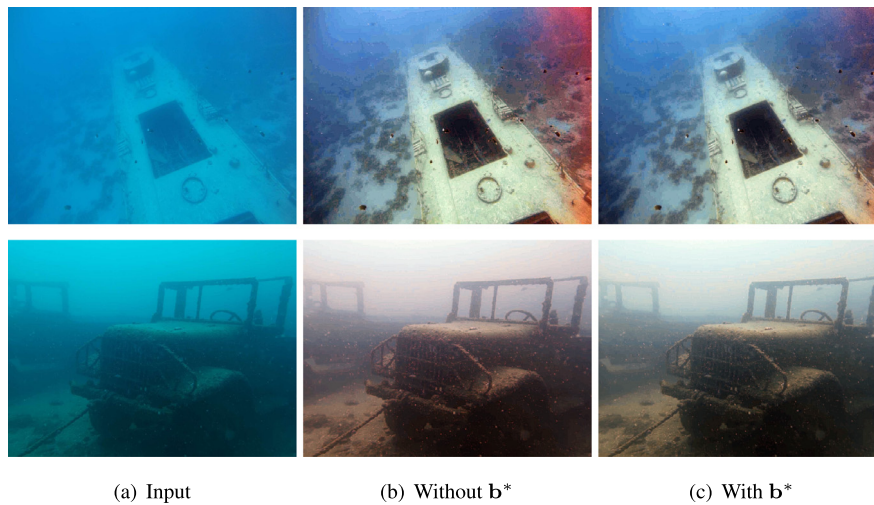
(a) Input   (b) Without $\mathbf{b}^*$ (c) With $\mathbf{b}^*$

Fig. 12. Two examples of using global residual-guide bias.

Table 3

Average UIQM and UCIQE values on our 200 real-world images, with **first** and *second* best performances indicated.

	Underwater	Fusion [5]	Histogram [7]	Haze-line [8]	Blurriness [33]	DL-based [50]	Ours
UIQM	0.98	3.48	3.13	3.22	3.12	2.27	**3.77**
UCIQE	0.46	0.52	0.56	0.52	0.54	0.48	**0.60**

Table 4

Average scores of user study on real-world images, with **first** and *second* best performances indicated.

	Underwater	Fusion [5]	Histogram [7]	Haze-line [8]	Blurriness [33]	DL-based [50]	Ours	Ideal
Scores	1.48	3.67	2.84	2.94	2.35	3.42	**4.21**	5

Table 5

Comparison of testing runtime (seconds).

Image size	Fusion [5]	Histogram [7]	Haze-line [8]	Blurriness [33]	DL-based [50]		Ours	
	CPU	CPU	CPU	CPU	CPU	GPU	CPU	GPU
512×512	0.32	1.02	20.54	22.66	1.24	0.07	0.51	0.05
1024×1024	1.10	3.49	72.54	98.74	4.25	0.17	1.95	0.08
2048×2048	4.59	13.00	292.76	471.89	11.57	0.37	7.49	0.19

Table 6

Comparison of testing runtime (seconds) of each stage on CPU.

Image size	Global–local networks	Compressed-HE	All
512×512	0.32	0.19	0.51
1024×1024	1.38	0.57	1.95
2048×2048	6.03	1.46	7.49

Table 7

Average SSIM/PSNR values on 190 images [65] using different network sizes.

	$N = 8$	$N = 16$	$N = 32$
$D = 3$	0.836/19.941	0.839/20.057	0.841/20.121
$D = 4$	0.840/20.124	0.846/20.222	0.846/20.327
$D = 5$	0.844/20.214	0.847/20.254	0.848/20.289

4.5.1. Network breadth versus depth

In general, increasing the network capacity helps to improve the performance. This increased capacity mainly comes in two forms: one is to increase the network breadth by using more features in each hidden layer, and the other is to increase the network depth by stacking more hidden layers. In this section, we test the impact of network breadth and depth on the two synthetic datasets. Specifically, we test different feature numbers $N \in \{8, 16, 32\}$ and depth $D \in \{3, 4, 5\}$. To eliminate the influence of compressed-histogram equalization, we calculate PSNR and SSIM results on $\hat{\mathbf{J}}$ of the 190 images [65] and the results are shown in Table 7. As is clear, due to the larger nonlinear modeling capacity, adding more hidden layers achieves better results over increasing the features number per layer. However, keeping increasing the hidden layer brings only limited improvement. This is because the learning problem is well simplified by using our global–local strategy. A relatively lightweight network is able to tackle the problem. Moreover, continuously increasing the network capacity without adding new

training data may lead to an over-fitting problem. Thus, we chose depth $D = 4$ and feature numbers $N = 16$ as the default setting.

4.5.2. Global residual-guide bias

We also test the impact of the proposed global residual-guide bias $\mathbf{b}^*$. We train a network with the same structure except for using randomly initialed biases. We show PSNR and SSIM results in Table 8 and the results are very close. However, as shown in Fig. 12, without using global residual-guide bias tends to generate reddish color deviation, which cannot be reflected by PSNR and SSIM. This is because the residual $\hat{\Delta\mu}$ represents the global scale and direction that need to be compensated, which can be further utilized provides guidance information for the local contrast compensation process. In addition, the deployment of this global residual-guide bias introduces very few parameters while can produce natural looking results.
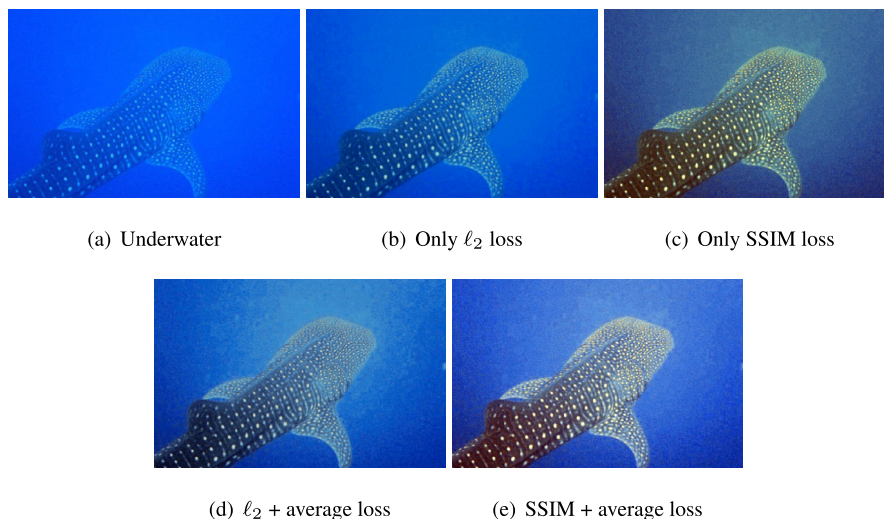
(a) Underwater   (b) Only ℓ_2 loss   (c) Only SSIM loss

(d) ℓ_2 + average loss   (e) SSIM + average loss

Fig. 13. An example by using different loss functions. SSIM + average loss achieves the best trade-off between color correction and contrast enhancement.
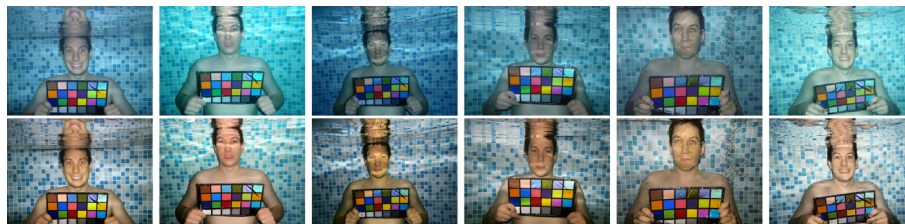

Fig. 14. Our results on the underwater images captured by different cameras. Top: underwater images, bottom: our results. These images are captured approximately one meter away from the subjects and provided by [6]. From left to right, the camera types are Canon D10, Olympus Tough 6000, Olympus Tough 8000, Pentax W60, Pentax W80 and FujiFilm Z33.

Table 8
SSIM and PSNR comparison on global residual-guide bias.

Datasets	Without $\mathbf{b}^*$		With $\mathbf{b}^*$	
	SSIM	PSNR	SSIM	PSNR
1128 images [17]	0.651	17.543	0.678	18.105
190 images [65]	0.872	20.491	0.876	21.983
Parameter #	7190		7292	

4.5.3. Loss function

We use SSIM as a part of loss function (8) for two main reasons. First, SSIM is calculated based on local image characteristics, e.g., local contrast, luminance and details, which is better than the global characteristics of ℓ_2 loss. Thus, using SSIM as the loss function is appropriate to guide the network training. Second, the human visual system is also sensitive to local image characteristics. SSIM has been motivated as generating more visually pleasing results, unlike ℓ_2 loss. We also use the average loss, i.e., Eq. (7), to intentionally guide the top branch network and stabilize the training. We have tested various combinations of loss functions and the results are shown in Fig. 13. As can be seen, using our combined loss (8) can generate better local contrast and color perception.

4.6. Robustness to different cameras

To test the robustness of our model to different cameras, we enhance the underwater images, which contain the standard Macbeth Color Checker, taken by various cameras [6]. As shown In Fig. 14, due to the different cameras settings, the underwater images contain various kinds of color distortion. While our single model is able to deal with all distorted color and make the results have consistent color perception. This test demonstrates the robustness of our method to different devices.

4.7. Potential applications

To verify our model could benefit other vision tasks, we perform three applications: pre-processing for other image processing algorithms, diver detection and local keypoints matching.

4.7.1. Pre-processing

Since our global–local networks are able to generate an intermediate image, i.e., $\hat{\mathbf{J}}$, with compensated information, it can be combined with other image processing algorithms as an effective pre-processing. Fig. 15 shows examples of using different algorithms to process an underwater image. As can be seen in the first row, due to the extreme color distortion, both restoration-based [7] and enhancement-based methods [71] are not able to accurately correct the distorted color. When adding our global–local networks as a pre-processing, the deficiencies in the first row are well addressed. This shows the flexibility of our model, i.e., it allows the user to combine other mature algorithms according to the practical needs.

4.7.2. Diver detection

Most existing deep models for high-level vision tasks are trained using high quality images. These learned models will have degraded performance when applied to severely degraded underwater images. In this case, an enhanced results can be useful for these high-level vision applications. To test whether using our model can improve the detection performance, we analyze the performances of diver detection [72] on our enhanced images. Fig. 16 shows visual results in
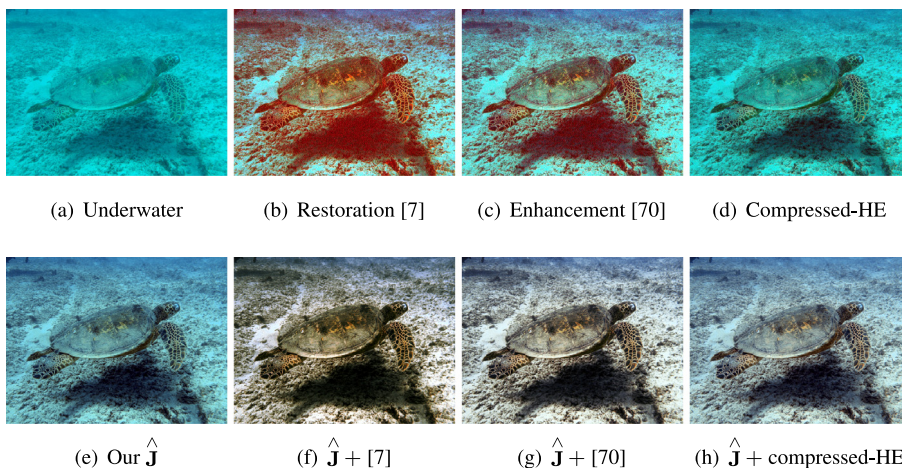
(a) Underwater   (b) Restoration [7]   (c) Enhancement [70]   (d) Compressed-HE

(e) Our $\hat{\mathbf{J}}$ (f) $\hat{\mathbf{J}}$ + [7]   (g) $\hat{\mathbf{J}}$ + [70]   (h) $\hat{\mathbf{J}}$ + compressed-HE

Fig. 15. Pre-processing for other image processing algorithms. The compensated result $\hat{\mathbf{J}}$, which generated by our global–local networks, can be used as a good initial result and benefits other image processing algorithms.
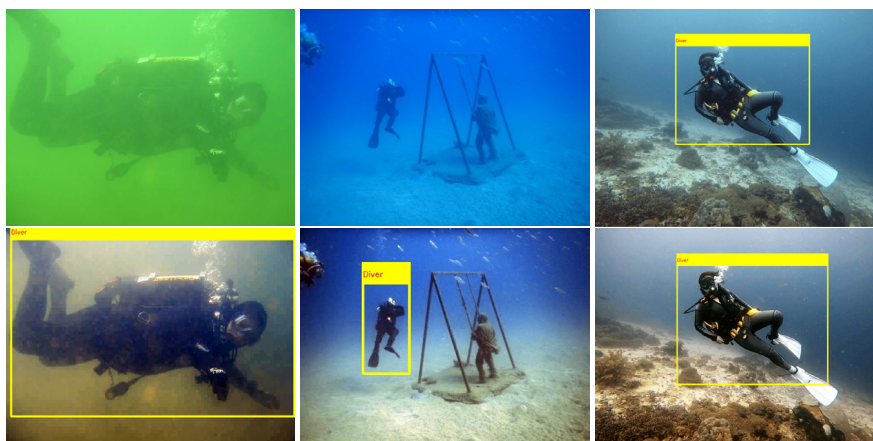

Fig. 16. Pre-processing for diver detection [72] (threshold = 0.8). Top: detection on underwater images; bottom: detection on our results.

(a) Underwater: 3 matches.   (b) Our result of (a): 60 matches.

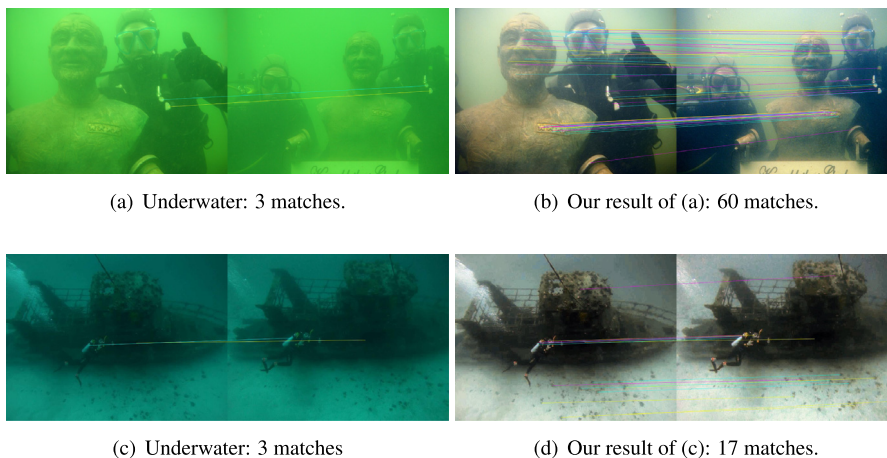(c) Underwater: 3 matches   (d) Our result of (c): 17 matches.

Fig. 17. Local keypoints matching by applying the SIFT operator [73]. Compared with the underwater images, the matching results shown in our enhanced images are improved significantly.

which the divers are not detected and the positions of the bounding box is shift. On the contrary, using our model as pre-processing the detection is improved by detecting divers, and having more accurate positions of the bounding box. In Table 9, we show the quantitative comparison results on the testing dataset [72], which contains 6K images. Two standard performance metrics, i.e., mean Average Precision (mAP) and Intersection over Union (IoU), are used for evaluations. The former measures the detection accuracy, while the later measures the object localization performance. As shown in Table 9, compared to the other methods, our algorithm achieves the best overall detection performances. This further demonstrates that our method can benefit downstream vision tasks.
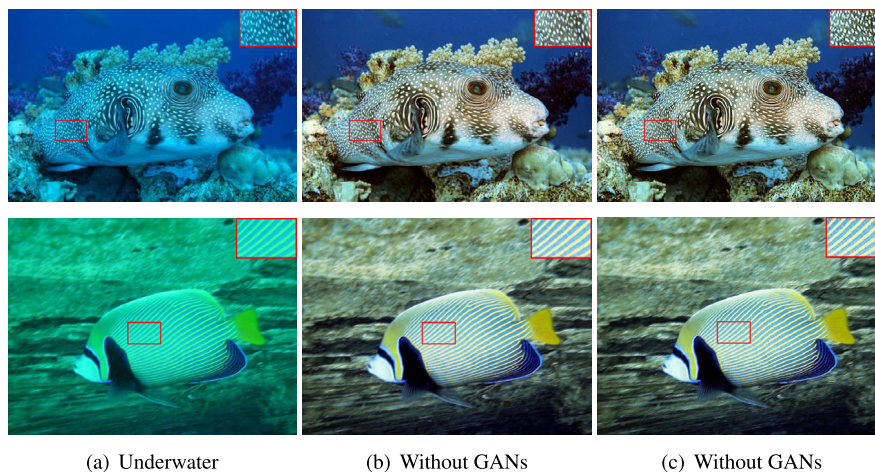
(a) Underwater   (b) Without GANs   (c) Without GANs

Fig. 18. Visual results of introducing GANs [49]. Using GANs generates sharper results.

Table 9
Comparison of diver detection on the dataset [72].

Metrics	Underwater	Fusion [5]	Histogram [7]	Haze-line [8]	Blurriness [33]	DL-based [50]	Ours
mAP (%)	53.74	54.75	53.74	50.07	51.87	49.45	**56.79**
IoU (%)	60.81	61.24	60.28	58.14	58.47	55.79	**63.41**

4.7.3. Local keypoints matching

We also adopt local keypoints matching, which aims to find correspondences between two similar scenarios, to test the effectiveness of our method. We utilize the SIFT operator [73] for a pair of real-world underwater images and as well for the corresponding enhanced images. The matching results are shown in Fig. 17. It is clear that the number of matched keypoints is significantly increased in the enhanced image pairs. This verifies that our technique can also recover local features of underwater images.

4.8. Discussion

Our global–local network belongs to supervised methods and predicts pixel values based on synthetic training data. On the other hand, the generative adversarial networks [49] are able to capture data distribution forms in an unsupervised manner. As described in the method [74], using adversarial learning generates realistic details, which makes the results more sharp. We test the GANs effect and show visual results in Fig. 18. It is clear that using adversarial loss can generate sharper results, as shown in the enlarged regions. Therefore, using GANs may help to improve the visual quality.

Our global–local network also belongs to data-driven methods and directly learns the relationship between inputs and desirable high quality outputs. The physical models reflecting the underwater imaging processes are ignored. However, our deep model can be combined with optimization-based algorithms to take fully advantages of both methods. For example, our global network branch can be used to estimate the back-scattered light, i.e. $\mathbf{A}$, in Eq. (1). The local network branch can be used to estimate the transmission $\mathbf{t}$. The compressed-histogram equalization can be used to adjust the output $\mathbf{J}$. Each of the above modules can be used as regularization terms to implicitly express the complex prior within an optimization framework. In this way, both powerful representation ability of deep networks and prior of physical models can be jointly exploited, which may further boost the performance. We leave this to our future work.

5. Conclusion

In this paper, we have introduced a lightweight model for single underwater image enhancement. Our model contains two modules, i.e., global–local networks for information compensation and

compressed-histogram equalization for further improvement. By taking advantages of both deep learning and hand-crafted enhancement, our model has fewer than 8K trainable parameters while still achieving state-of-the-art performances. Due to the generality and lightweight architecture, our networks have potential values for other vision tasks.

The goal of this work is to design an effective model for the tough underwater image enhancement. For practical applications, the computational time can be further improved by combining out network with more efficient network architectures [75,76]. In addition, other factors, such as underwater imaging process and active light sources, can also be integrated into our framework to further improve the performance. We will incorporate above issues into our future work.

CRediT authorship contribution statement

Xueyang Fu: Conceptualization, Methodology, Software, Validation, Data curation, Writing - original draft, Writing - review & editing, Visualization. **Xiangyong Cao:** Methodology, Formal analysis, Writing - review & editing, Supervision, Project administration, Funding acquisition.

Declaration of competing interest

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

Acknowledgments

References

[1] C.D. Mobley, Light and Water: Radiative Transfer in Natural Waters, Academic press, 1994.

[2] K. Lebart, C. Smith, E. Trucco, D.M. Lane, Automatic indexing of underwater survey video: algorithm and benchmarking method, IEEE J. Ocean. Eng. 28 (4) (2003) 673–686.

[3] M. Johnson-Roberson, M. Bryson, A. Friedman, O. Pizarro, G. Troni, P. Ozog, J.C. Henderson, High-resolution underwater robotic vision-based mapping and three-dimensional reconstruction for archaeology, J. Field Robotics 34 (4) (2017) 625–643.

[4] G.L. Foresti, Visual inspection of sea bottom structures by an autonomous underwater vehicle, IEEE Trans. Syst. Man Cybern. B 31 (5) (2001) 691–705.

[5] C. Ancuti, C.O. Ancuti, T. Haber, P. Bekaert, Enhancing underwater images and videos by fusion, in: CVPR, 2012, pp. 81–88.

[6] C.O. Ancuti, C. Ancuti, C. De Vleeschouwer, P. Bekaert, Color balance and fusion for underwater image enhancement, IEEE Trans. Image Process. 27 (1) (2018) 379–393.

[7] C.-Y. Li, J.-C. Guo, R.-M. Cong, Y.-W. Pang, B. Wang, Underwater image enhancement by dehazing with minimum information loss and histogram distribution prior, IEEE Trans. Image Process. 25 (12) (2016) 5664–5677.

[8] D. Berman, T. Treibitz, S. Avidan, Diving into haze-lines: Color restoration of underwater images, in: British Machine Vision Conference, Vol. 1, 2017.

[9] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, in: NIPS, 2012, pp. 1097–1105.

[10] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.

[11] G. Huang, Z. Liu, L. Van Der Maaten, K.Q. Weinberger, Densely connected convolutional networks, in: IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 4700–4708.

[12] X. Cao, F. Zhou, L. Xu, D. Meng, Z. Xu, J. Paisley, Hyperspectral image classification with markov random fields and a convolutional neural network, IEEE Trans. Image Process. 27 (5) (2018) 2354–2367.

[13] C. Dong, C.C. Loy, K. He, X. Tang, Image super-resolution using deep convolutional networks, IEEE Trans. Pattern Anal. Mach. Intell. 38 (2) (2016) 295–307.

[14] K. Zhang, W. Zuo, Y. Chen, D. Meng, L. Zhang, Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising, IEEE Trans. Image Process. 26 (7) (2017) 3142–3155.

[15] X. Fu, J. Huang, X. Ding, Y. Liao, J. Paisley, Clearing the skies: A deep network architecture for single-image rain removal, IEEE Trans. Image Process. 26 (6) (2017) 2944–2956.

[16] J. Cai, S. Gu, L. Zhang, Learning a deep single image contrast enhancer from multi-exposure images, IEEE Trans. Image Process. 27 (4) (2018) 2049–2062.

[17] C. Fabbri, M.J. Islam, J. Sattar, Enhancing underwater imagery using generative adversarial networks, in: IEEE International Conference on Robotics and Automation, 2018.

[18] Y. Wang, J. Zhang, Y. Cao, Z. Wang, A deep CNN method for underwater image enhancement, in: IEEE International Conference on Image Processing, 2017, pp. 1382–1386.

[19] M.J. Islam, Y. Xia, J. Sattar, Fast underwater image enhancement for improved visual perception, IEEE Robot. Autom. Lett. 5 (2) (2020) 3227–3234.

[20] S. Anwar, C. Li, F. Porikli, Deep underwater image enhancement, 2018, arXiv preprint arXiv:1807.03528.

[21] K. Iqbal, M. Odetayo, A. James, R.A. Salam, A.Z.H. Talib, Enhancing the low quality images using unsupervised colour correction method, in: International Conference on Systems, Man and Cybernetics, 2010, pp. 1703–1709.

[22] A.S.A. Ghani, N.A.M. Isa, Underwater image quality enhancement through integrated color model with rayleigh distribution, Appl. Soft Comput. 27 (2015) 219–230.

[23] X. Fu, P. Zhuang, Y. Huang, Y. Liao, X.-P. Zhang, X. Ding, A retinex-based enhancing approach for single underwater image, in: IEEE International Conference on Image Processing (ICIP), 2014, pp. 4572–4576.

[24] S. Zhang, T. Wang, J. Dong, H. Yu, Underwater image enhancement via extended multi-scale retinex, Neurocomputing 245 (2017) 1–9.

[25] B. McGlamery, A computer model for underwater camera systems, in: Ocean Optics VI, Vol. 208, 1980, pp. 221–232.

[26] J.S. Jaffe, Computer modeling and the design of optimal underwater imaging systems, IEEE J. Ocean. Eng. 15 (2) (1990) 101–111.

[27] R. Schettini, S. Corchs, Underwater image processing: state of the art of restoration and image enhancement methods, EURASIP J. Adv. Signal Process. 2010 (1) (2010) 746052.

[28] C. Li, J. Guo, Underwater image enhancement by dehazing and color correction, J. Electron. Imaging 24 (3) (2015) 033023.

[29] C.O. Ancuti, C. Ancuti, C. De Vleeschouwer, R. Garcia, Locally adaptive color correction for underwater image dehazing and matching, in: IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2017, pp. 1–9.

[30] C. Li, J. Guo, C. Guo, R. Cong, J. Gong, A hybrid method for underwater image correction, Pattern Recognit. Lett. 94 (2017) 62–67.

[31] D. Berman, S. Avidan, et al., Non-local image dehazing, in: IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 1674–1682.

[32] N. Carlevaris-Bianco, A. Mohan, R.M. Eustice, Initial results in underwater single image dehazing, in: OCEANS 2010 MTS/IEEE SEATTLE, 2010, pp. 1–8.

[33] Y.-T. Peng, P.C. Cosman, Underwater image restoration based on image blurriness and light absorption, IEEE Trans. Image Process. 26 (4) (2017) 1579–1594.

[34] H. Koschmieder, Theorie der horizontalen sichtweite, Beitrage Phys. Freien Atmos. (1924) 33–53.

[35] K. He, J. Sun, X. Tang, Single image haze removal using dark channel prior, IEEE Trans. Pattern Anal. Mach. Intell. 33 (12) (2011) 2341–2353.

[36] J.Y. Chiang, Y.-C. Chen, Underwater image enhancement by wavelength compensation and dehazing, IEEE Trans. Image Process. 21 (4) (2012) 1756–1769.

[37] P. Drews, E. Nascimento, F. Moraes, S. Botelho, M. Campos, Transmission estimation in underwater single images, in: IEEE International Conference on Computer Vision Workshops, 2013, pp. 825–830.

[38] A. Galdran, D. Pardo, A. Picón, A. Alvarez-Gila, Automatic red-channel underwater image restoration, J. Vis. Commun. Image Represent. 26 (2015) 132–145.

[39] H. Lu, Y. Li, L. Zhang, S. Serikawa, Contrast enhancement for images in turbid water, J. Opt. Soc. Amer. A 32 (5) (2015) 886–893.

[40] Y. Zhou, Q. Wu, K. Yan, L. Feng, W. Xiang, Underwater image restoration using color-line model, IEEE Trans. Circuits Syst. Video Technol. 29 (3) (2019) 907–911.

[41] R. Fattal, Dehazing using color-lines, ACM Trans. Graph. 34 (1) (2014) 13.

[42] C. Li, J. Guo, S. Chen, Y. Tang, Y. Pang, J. Wang, Underwater image restoration based on minimum information loss principle and optical properties of underwater imaging, in: IEEE International Conference on Image Processing, 2016, pp. 1993–1997.

[43] Y.-T. Peng, K. Cao, P.C. Cosman, Generalization of the dark channel prior for single image restoration, IEEE Trans. Image Process. 27 (6) (2018) 2856–2868.

[44] D. Akkaynak, T. Treibitz, T. Shlesinger, Y. Loya, R. Tamir, D. Iluz, What is the space of attenuation coefficients in underwater computer vision?, in: IEEE Conference on Computer Vision and Pattern Recognition, 2017.

[45] D. Akkaynak, T. Treibitz, A revised underwater image formation model, in: IEEE Conference on Computer Vision and Pattern Recognition, 2018.

[46] D. Akkaynak, T. Treibitz, Sea-thru: A method for removingwater from underwater images, in: IEEE Conference on Computer Vision and Pattern Recognition, 2019.

[47] Y.-S. Shin, Y. Cho, G. Pandey, A. Kim, Estimation of ambient light and transmission map with common convolutional architecture, in: OCEANS 2016 MTS/IEEE Monterey, 2016, pp. 1–7.

[48] J. Li, K.A. Skinner, R.M. Eustice, M. Johnson-Roberson, WaterGAN: unsupervised generative network to enable real-time color correction of monocular underwater images, IEEE Robot. Autom. Lett. 3 (1) (2018) 387–394.

[49] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets, in: Advances in Neural Information Processing Systems, 2014, pp. 2672–2680.

[50] Y. Guo, H. Li, P. Zhuang, Underwater image enhancement using a multiscale dense generative adversarial network, IEEE J. Ocean. Eng. (2019) 1–9, http://dx.doi.org/10.1109/JOE.2019.2911447.

[51] C. Li, J. Guo, C. Guo, Emerging from water: Underwater image color correction based on weakly supervised color transfer, IEEE Signal Process. Lett. 25 (3) (2018) 323–327.

[52] M. Hou, R. Liu, X. Fan, Z. Luo, Joint residual learning for underwater image enhancement, in: IEEE International Conference on Image Processing, 2018, pp. 4043–4047.

[53] R. Liu, L. Ma, Y. Wang, L. Zhang, Learning converged propagations with deep prior ensemble for image enhancement, IEEE Trans. Image Process. 28 (3) (2019) 1528–1543.

[54] C. Li, S. Anwar, F. Porikli, Underwater scene prior inspired deep underwater image and video enhancement, Pattern Recognit. 98 (2020) 107038.

[55] C. Guo, C. Li, J. Guo, C.C. Loy, J. Hou, S. Kwong, R. Cong, Zero-reference deep curve estimation for low-light image enhancement, in: IEEE Conference on Computer Vision and Pattern Recognition, 2020.

[56] Y. Guo, H. Li, P. Zhuang, Underwater image enhancement using a multi-scale dense generative adversarial network, IEEE J. Ocean. Eng. (2019).

[57] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, 2015, pp. 234–241.

[58] H. Zhao, O. Gallo, I. Frosio, J. Kautz, Loss functions for image restoration with neural networks, IEEE Trans. Comput. Imaging 3 (1) (2017) 47–57.

[59] Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, et al., Image quality assessment: from error visibility to structural similarity, IEEE Trans. Image Process. 13 (4) (2004) 600–612.

[60] R.C. Gonzalez, R.E. Woods, et al., Digital image processing, Publ. House Electron. Ind. 141 (7) (2002).

[61] W. Schreiber, R.A. Haddad, Fundamentals of electronic imaging systems, Appl. Opt. 29 (1990) 2504.

[62] J. Hu, L. Shen, G. Sun, Squeeze-and-excitation networks, in: IEEE Conference on Computer Vision and Pattern Recognition, 2018.

[63] X. Wang, R. Girshick, A. Gupta, K. He, Non-local neural networks, in: IEEE Conference on Computer Vision and Pattern Recognition, 2018.

[64] T. Celik, Spatial entropy-based global and local image contrast enhancement, IEEE Trans. Image Process. 23 (12) (2014) 5298–5308.

[65] C. Li, C. Guo, W. Ren, R. Cong, J. Hou, S. Kwong, D. Tao, An underwater image enhancement benchmark dataset and beyond, IEEE Trans. Image Process. (2019) http://dx.doi.org/10.1109/TIP.2019.2955241.

[66] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, et al., Tensorflow: A system for large-scale machine learning, in: USENIX Symposium on Operating Systems Design and Implementation, 2016.

[67] D.P. Kingma, J. Ba, Adam: A method for stochastic optimization, in: International Conference on Learning Representations, 2014.

[68] R. Liu, X. Fan, M. Zhu, M. Hou, Z. Luo, Real-world underwater enhancement: Challenges, benchmarks, and solutions under natural light, 2020, http://dx.doi.org/10.1109/TCSVT.2019.2963772.

[69] K. Panetta, C. Gao, S. Agaian, Human-visual-system-inspired underwater image quality measures, IEEE J. Ocean. Eng. 41 (3) (2015) 541–551.

[70] M. Yang, A. Sowmya, An underwater color image quality evaluation metric, IEEE Trans. Image Process. 24 (12) (2015) 6062–6071.

[71] H. Xu, G. Zhai, X. Wu, X. Yang, Generalized equalization model for image enhancement, IEEE Trans. Multimed. 16 (1) (2013) 68–82.

[72] M.J. Islam, M. Fulton, J. Sattar, Toward a generic diver-following algorithm: Balancing robustness and efficiency in deep visual detection, IEEE Robot. Autom. Lett. 4 (1) (2019) 113–120.

[73] D.G. Lowe, Distinctive image features from scale-invariant keypoints, Int. J. Comput. Vis. 60 (2) (2004) 91–110.

[74] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, et al., Photo-realistic single image super-resolution using a generative adversarial network, in: IEEE Conference on Computer Vision and Pattern Recognition, 2017.

[75] X. Zhang, X. Zhou, M. Lin, J. Sun, Shufflenet: An extremely efficient convolutional neural network for mobile devices, in: IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 6848–6856.

[76] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, L.-C. Chen, Mobilenetv2: Inverted residuals and linear bottlenecks, in: IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 4510–4520.