Pattern Recognition 98 (2020) 107038

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/patcog

Underwater scene prior inspired deep underwater image and video enhancement

Chongyi Li[a,1,*], Saeed Anwar[b,c,1], Fatih Porikli[d]

[a] Department of Computer Science, City University of Hong Kong (CityU), Hong Kong
[b] Data61, CSIRO, ACT 2601, Australia
[c] Australian National University, Canberra ACT 2600, Australia
[d] Research School of Engineering, The Australian National University, Canberra, ACT 0200, Australia

ARTICLE INFO

Article history:
Received 17 April 2019
Revised 23 August 2019
Accepted 4 September 2019
Available online 5 September 2019

Keywords:
Underwater image and video enhancement and restoration
Underwater image synthesis
Pattern recognition
Deep learning

ABSTRACT

In underwater scenes, wavelength-dependent light absorption and scattering degrade the visibility of images and videos. The degraded underwater images and videos affect the accuracy of pattern recognition, visual understanding, and key feature extraction in underwater scenes. In this paper, we propose an underwater image enhancement convolutional neural network (CNN) model based on underwater scene prior, called UWCNN. Instead of estimating the parameters of underwater imaging model, the proposed UWCNN model directly reconstructs the clear latent underwater image, which benefits from the underwater scene prior which can be used to synthesize underwater image training data. Besides, based on the light-weight network structure and effective training data, our UWCNN model can be easily extended to underwater videos for frame-by-frame enhancement. Specifically, combining an underwater imaging physical model with optical properties of underwater scenes, we first synthesize underwater image degradation datasets which cover a diverse set of water types and degradation levels. Then, a light-weight CNN model is designed for enhancing each underwater scene type, which is trained by the corresponding training data. At last, this UWCNN model is directly extended to underwater video enhancement. Experiments on real-world and synthetic underwater images and videos demonstrate that our method generalizes well to different underwater scenes.

1. Introduction

Acquisition of clear underwater images and videos is of great importance for underwater scene perception and understanding where autonomous and remotely operated underwater vehicles are widely used to explore, recognition, and interact with marine environments. However, raw underwater images and videos seldom meet the expectations concerning the visual quality and further challenge the performance of pattern recognition, object detection, key feature extraction, to name a few. This is because most deep networks are trained by high-quality images or the algorithms assume the inputs are clear images. Naturally, underwater images are degraded by the adverse effects of light absorption and scattering due to particles in the water, including micro phytoplankton, colored dissolved organic matter, and non-algal particles. Additionally, when the light propagates in the underwater scenario, it has the characteristic of selective attenuation with respect to the wavelength of light [1]. Fig. 1 presents a diagram of light attenuation with respect to the wavelength of light.

These absorption and scattering problems hinder the performance of underwater scene understanding and recognition, such as aquatic robot inspection and marine environmental surveillance. Moreover, traditional image enhancement methods [2,3] show limitations when they are used to process underwater image and video. Additionally, lacking sufficient and effective training data, the performance of deep learning-based underwater image and video enhancement methods does not match the success of recent deep learning-based solutions such as classification [4], analysis [5] segmentation [6], super-resolution [7], recognition [8], etc. It is necessary to develop underwater image synthesis and enhancement methods for superior underwater visual quality and improve the performance of high-level vision tasks.

In recent years, more and more deep learning-based methods [10,11] have been proposed. The deep models have some advantages than the traditional non-learning-based methods: (i) deep learning provides a strong modeling capability of the distortions and facilitates discriminative prior learning, and (ii) the inference of deep models can be made efficiently by exploiting the parallel

* Corresponding author.
E-mail address: lichongyi@tju.edu.cn (C. Li).
[1] Chongyi Li and Saeed Anwar contribute equally to this work.

https://doi.org/10.1016/j.patcog.2019.107038
0031-3203/

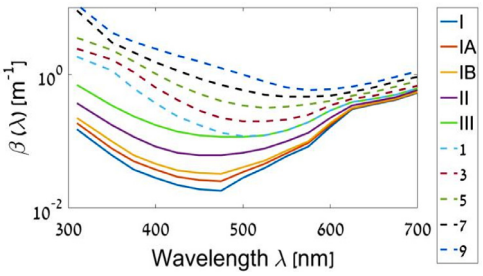


Fig. 1. Wavelength-dependent light attenuation coefficients β of Jerlov water types from [9]. Solid lines mark open ocean water types while dashed lines mark coastal water types. The Jerlov water types are I, IA, IB, II and III for open ocean waters, and 1 through 9 for coastal waters. Type-I is the clearest and Type-III is the most turbid open ocean water. Likewise, for coastal waters, Type-1 is clearest and Type-9 is the most turbid.

processing platforms. Inspired by the recent success of deep learning in pattern recognition and visual understanding [12], we propose a new underwater image synthesis algorithm using underwater scene prior, and then design to offer a robust and data-driven solution to underwater image and video enhancement. The proposed method is shown to have superior robustness, accuracy, and flexibility for diverse water types.

Contributions: We design an end-to-end solution for the complex and nonlinear underwater image formation model using a novel CNN architecture trained on the underwater scene prior based underwater images. Our model robustly restores the degraded underwater images and accurately reconstructs underlying colors and appearance. Besides, the proposed model can be easily extended to underwater video thanks to the light-weight network structure. To summarize,

- We propose a new underwater image synthesis algorithm based on underwater scene prior that is capable of simulating a diverse set of degraded underwater images. To our best knowledge, it is the first underwater image synthesis algorithm that can simulate different underwater types and degradation levels. Our image synthesis can be used as a guide for network training and full-reference image quality assessments.
- We propose a novel CNN model to reconstruct the clear latent underwater image while preserving the original structure and texture by jointly optimizing multi-term loss. Benefiting from the light-weight network design and effective training data, the proposed model can be extended to underwater video for frame-by-frame enhancement.
- Our method generalizes well both to synthetic and real-world underwater images and videos with diverse color and visibility characteristics. In addition, a lightweight network structure also can achieve decent results when effective prior information is embedded into network, which encourages the related designs for pattern recognition, visual understanding, etc.

2. Related work

From different views, the existing underwater image enhancement and restoration methods can be classified into different categories. In this paper, we classified these methods into one of three broad categories: underwater image enhancement method, underwater image restoration method, and supplementary-information specific method. Since there is little work for underwater video enhancement and restoration, we mainly introduce image processing methods in this section.

2.1. Underwater image enhancement method

In this line of research, Li et al. [13] treated the problem of underwater image enhancement as an image dehazing step and a color correction step. Ancuti et al. [14] fused a contrast improved underwater image and a color corrected underwater image obtained from an input. In the process of mulit-scale fusion, four weights are used to determine which pixel is advantaged to appear in the final image. A hybrid method based on color correction and underwater image dehazing was proposed in [15], which corrects the color casts of the underwater image using image color prior and improves the visibility by a modified image dehazing algorithm. Li et al. [16] proposed an underwater image color correction method based on weakly supervised color transfer, which learns a cross-domain mapping function between underwater images and air images. Inspired by the generative adversarial networks (GANs), Guo et al. [17] proposed a multiscale dense GAN for underwater image enhancement, which boots the performance of underwater image enhancement by introducing multiscale, dense concatenation, and residual learning strategies. Ancuti et al. [18] modified their previous work [14] to reduce the effects of the over-enhancement and over-exposure. More recently, Li et al. [19] proposed a deep baseline model trained on the paired underwater images and the corresponding reference images. These reference images are subjectively selected from the enhanced results by different methods.

2.2. Underwater image restoration method

Underwater image restoration methods usually consider the challenge at hand as an inverse problem, and then construct physical models of the degradation, at last estimate the model parameters. Chiang and Chen [20] combined an image dehazing algorithm with a wavelength dependent compensation algorithm to restore underwater image, which can remove the bluish tone of underwater images and the effects of artificial light. A Red Channel method [21] recovered the lost contrast of the underwater image by restoring the colors associated with short wavelengths. Drews et al. [22] proposed an underwater dark-channel prior called UDCP which modifies the previous dark channel prior [23]. With the proposed UDCP, the medium transmission can be estimated in some cases; however, the UDCP does not always hold when there are white objects or artificial light in the underwater scenes. Li et al. [24,25] combined an underwater image dehazing algorithm with a contrast enhancement algorithm. Peng et al. [26] restored underwater images based on image blurriness and light absorption, which is a prior-based method. Li et al. [27] proposed a CNN based underwater image color correction model based on synthetic underwater images generated in a weakly supervised learning manner.

2.3. Supplementary-information specific method

Supplementary-information specific methods usually take advantage of the additional information obtained from polarization filters, stereo images, rough depth of the scene, etc [28].

3. Underwater image formulation model

We follow the underwater image formulation model proposed in [20]. This underwater image degradation model has been widely used in traditional underwater image restoration methods and can be expressed as:

$$\mathbf{U}_\lambda(x) = \mathbf{I}_\lambda(x) \cdot T_\lambda(x) + B_\lambda \cdot \big(1 - T_\lambda(x)\big), \tag{1}$$

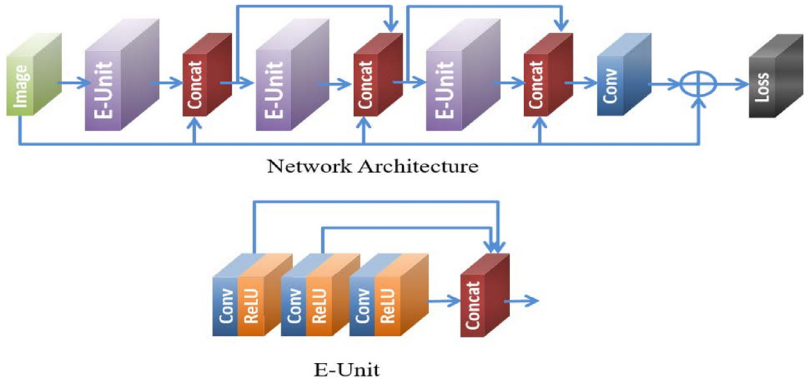


Fig. 2. Our UWCNN model where 'Conv' represents the convolutional layer, 'Concat' represents the stacked convolutional layers, and 'ReLU' represents the rectified linear unit.

where $\mathbf{U}_\lambda(x)$ is the captured underwater image; $\mathbf{I}_\lambda(x)$ is the clear latent image, also called as the scene radiance, that we aim to recover; B_λ is the homogeneous global background light; λ is the wavelength of the light for the red, green and blue channels; and x is a point in the underwater scene (for clarity, images are denoted in bold capital letters). The medium energy ratio $T_\lambda(x)$ represents the percentage of the scene radiance reaching at the camera after reflecting from the point x in the underwater scene, which thereby causes color cast and contrast degradation. In other words, $T_\lambda(x)$ is a function of the wavelength of light λ and the distance $d(x)$ from scene point x to the camera:

$$T_\lambda(x) = 10^{-\beta_\lambda d(x)} = \frac{E_\lambda\big(x, d(x)\big)}{E_\lambda(x, 0)} = N_\lambda\big(d(x)\big), \tag{2}$$

where β_λ is the wavelength-depended medium attenuation coefficient as shown in Fig. 1. Assuming the energy of a light beam emanated from x before and after it passes through a transmission medium at a distance of $d(x)$ is $E_\lambda(x, 0)$ and $E_\lambda(x, d(x))$, respectively. The normalized residual energy ratio N_λ corresponds to the ratio of residual energy to initial energy for every unit of distance propagated. Its value varies in water depending on the light wavelength. For example, red light possesses a longer wavelength; thus, it attenuates faster and gets absorbed more than other wavelengths in open water, which results in a bluish tone of most underwater images. More details can be found in Ref. [20].

The underwater image degradation model [20] is different from the image degradation models which have been widely used in image dehazing [29], image deblurring [30], and image super-resolution [31]. Specifically, the underwater image degradation model is similar to the image dehazing model; however, it is more complex due to the characteristics of wavelength-dependent light absorption and scattering. Compared with the image deblurring model which simulates the blurred image by doing convolutional operation between clear image and blur kernels and image super-resolution model which simulates the low-resolution image by down-sampling the original image, the underwater image degradation model mainly focuses on the degradation of color and visibility.

4. Proposed UWCNN model

Here, we discuss the details of the proposed UWCNN model and then present a post-processing stage to further improve our enhanced results.

4.1. Network architecture

Inspired by the recent success of deep network architectures in pattern recognition [12,32], we proposed a lightweight network for underwater image and video enhancement. Fig. 2 shows the architecture of our UWCNN model, which is a densely connected FCNN. As follows, we present its basic building blocks and hyperparameters. The input to our network is an RGB image $\mathbf{U}$.

4.1.1. Residuals

Unlike the traditional end-to-end approaches such as [3] that directly predict the clean latent image $\mathbf{I}$ by learning the mapping function $\mathbf{I} = f^{-1}(\mathbf{U})$, we allow our network to learn the difference between the synthetic underwater image and its clean counterpart. **Note that such a synthetic image generation task is a nontrivial objective for underwater image enhancement and restoration field, and it will be discussed in detail in** Section 5. As underwater image and its feature maps in the subsequent layers are processed through many convolutional filters before reaching the final loss layer. Although our network is not intentionally very deep, there is still a possibility of vanishing or exploding gradients. To avoid such issues during the training iterations, we enforce learning the residual by adding the input of the network, i.e., $\mathbf{U}$ to the output of the network i.e., $\Delta(\mathbf{U}, \theta)$ (see below) before loss function as:

$$\mathbf{I} = \mathbf{U} + \Delta(\mathbf{U}, \theta), \tag{3}$$

where '+' is the element-wise addition operation.

4.1.2. Enhancement units

The UWCNN has a modular architecture composed of enhancement units (E-Units) having same structure and components. Suppose r and c are the notation for ReLU and convolution, then the first operation of convolution and ReLU pair, in the l-th block, is given by

$$z_{l,0} = r\big(c(\mathbf{U}); \theta_{l,0}\big), \tag{4}$$

where $z_{l,0}$ is the output of the first convolution-ReLU pairs of l-th residual enhancement unit and $\theta_{l,0}$ is a set of weights and biases associated with it. By composing the series of convolution-ReLU pairs, we obtain

$$z_{l,n} = r\big(c(\ldots r(c(\mathbf{U}; \theta_{l,0})) \ \ldots); \theta_{l,n}\big). \tag{5}$$

The output of the l-th block is obtained by concatenating along third dimension of each individual convolution-ReLU pairs output z and input image $\mathbf{U}$ as:

$$b_l = h(z_{l,0};\ \ldots;\ z_{l,n};\ \mathbf{U}). \tag{6}$$

The output of the $(l+1)$-th enhancement unit is obtained by:

$$b_{l+1} = h(z_{l+1,0};\ \ldots;\ z_{l+1,n};\ \mathbf{U};\ b_l). \tag{7}$$

Finally, we chain all enhancement unit and the output of this chain is convolved with a final convolution layer with parameters $\theta_{l+m,n}$ to predict the component as $\Delta(\mathbf{U}, \theta) = c(b_{l+m}, \theta_{l+m,n})$.

Fig. 3. Sample results for qualitative assessment. (a) Raw real-world underwater images. (b) Results of UWCNN. (c) Results of UWCNN+. As visible, our methods (i.e., UWCNN and UWCNN+) remove the greenish tone while reconstructing accurate and vivid latent images.

Fig. 4. Ten types of synthesized underwater images from the NYU-v2 RGB-D dataset [37] using a sample image and its depth map.

4.1.3. Network layers

Our network consists of three different layers indicated by different color as shown in Fig. 2. The first type is the convolutional layer represented by 'Conv', which consists of 16 convolutional kernels of size $3 \times 3 \times 3$ to produce 16 output feature maps for the first layer, while subsequent convolutional layers produce 16 maps each using $3 \times 3 \times 16$ filters. The second type is the activation layer 'ReLU' for introducing the nonlinearity. The third type is the 'Concat' layer, which is used to concatenate all the convolutional layers after each block. The last convolutional layer estimates the final output of the network.

4.1.4. Dense concatenation

We stack all convolutional layers at the end of each block. This technique is different from DenseNet provided in [33], where each convolutional layer is connected with other convolutional layers in the same block. Furthermore, we do not use any fully connected layers or batch normalization steps, which makes our network memory efficient and fast. In addition, we feed the input image to every block. The stacking of the convolutional layers with input data reduces the need for a very deep network. In summary, our network is unique since (i) the input image is applied to all enhancement unit, and (ii) it contains only the fully-convolutional layers without any batch-normalization steps.

4.1.5. Network depth

Our network is of modular structure and consists of three enhancement units where each unit is again composed of three convolutional layers. We have a single convolutional layer at the end of the network; hence, making the full depth of our network only ten layers. This makes our model computationally inexpensive and highly practical in training and inference. Besides, such a light-weight network structure can be easily extended to underwater videos for frame-by-frame enhancement, which is desired in practical applications. Such a lightweight network structure mainly benefits from the embedded prior which boosts the training and inference of networks, which encourages the designs of similar networks for pattern recognition, object detection, and visual understanding.

4.1.6. Reducing boundary artifacts

In low-level vision tasks, the output size of the system is needed to be equal to the input. This requirement sometimes results in boundary artifacts. To avoid this phenomenon, we enforce two strategies: (i) we do not use any pooling layers in our network, and (ii) we add zeros before each convolutional layer. As a consequence, the final output image of UWCNN network is almost artifacts-free around the boundaries and is of the same size as the input image.

Fig. 5. Qualitative comparisons for samples from test set. (a) Raw underwater images. (b) Results of RED [21]. (c) Results of UDCP [22]. (d) Results of ODM [25]. (e) Results of UIBLA [26]. (f) Our results. (g) Ground truth. The types of underwater images in the first column from top to bottom are Type-1, Type-3, Type-5, Type-7, Type-9, Type-I, Type-II, and Type-III. Our method removes the light absorption effects and recovers the original colors without any artifacts.

4.2. Network loss

To reconstruct an image, we use the ℓ_2 loss as in our observations it can well preserve the sharpness of edges and details, because blurring edges results in large errors. We add the estimated residual to the input underwater image, then compute the ℓ_2 loss as:

$$\ell_2 = \frac{1}{M}\sum_{i=1}^{M}\left|\left[\mathbf{U}(x_i) + \Delta\Big(\mathbf{U}(x_i), \theta(x_i)\Big)\right] - \mathbf{I}^*(x_i)\right|^2, \tag{8}$$

where $\mathbf{U}(x_i) + \Delta(\mathbf{U}(x_i), \theta(x_i)) = \mathbf{I}(x_i)$ is the estimated latent image pixel value at x_i, $i = 1, ., M$ as described in Eq. (3) and $\mathbf{I}_i^*$ is the ground truth.

In addition, we include the SSIM loss in our objective function to impose the structure and texture similarity on the latent image. We use gray images to compute SSIM scores. For each pixel x, the SSIM value is computed within a 13×13 image patch around the pixel as:

$$SSIM(x) = \frac{2\mu_{I*}(x)\mu_I(x) + c_1}{\mu_{I*}^2(x) + \mu_I^2(x) + c_1} \cdot \frac{2\sigma_{I*I}(x) + c_2}{\sigma_{I*}^2(x) + \sigma_I^2(x) + c_2}, \tag{9}$$

where $\mu_I(x)$ and $\sigma_I(x)$ correspond to the mean and standard deviation of the image patch from the latent image $\mathbf{I}$, similarly, $\mu_{I^*}(x)$ and $\sigma_{I^*}(x)$ are for the patch from the ground truth image $\mathbf{I}^*$. The cross-covariance $\sigma_{I^*I}(x)$ is computed between the patches from $\mathbf{I}$ and $\mathbf{I}^*$ for the pixel x. We set constants $c_1 = 0.02$ and $c_2 = 0.03$ based on the default in SSIM loss. Our model is insensitive to these defaults. Still, we fix them for a fair comparison. The SSIM loss is expressed as:

$$L_{SSIM} = 1 - \frac{1}{M}\sum_{i=1}^{M} SSIM(x_i). \tag{10}$$

The final loss function L is the aggregation of MSE and SSIM losses:

$$L = \ell_2 + L_{SSIM}. \tag{11}$$

4.3. Post-processing

UWCNN generates enhanced images without color casts and exceptional visibility. However, due to the limitation of our training data pairs (an indoors image as the latent image and a synthesized image from the indoors image using the aforementioned underwater image formation model as the corresponding underwater

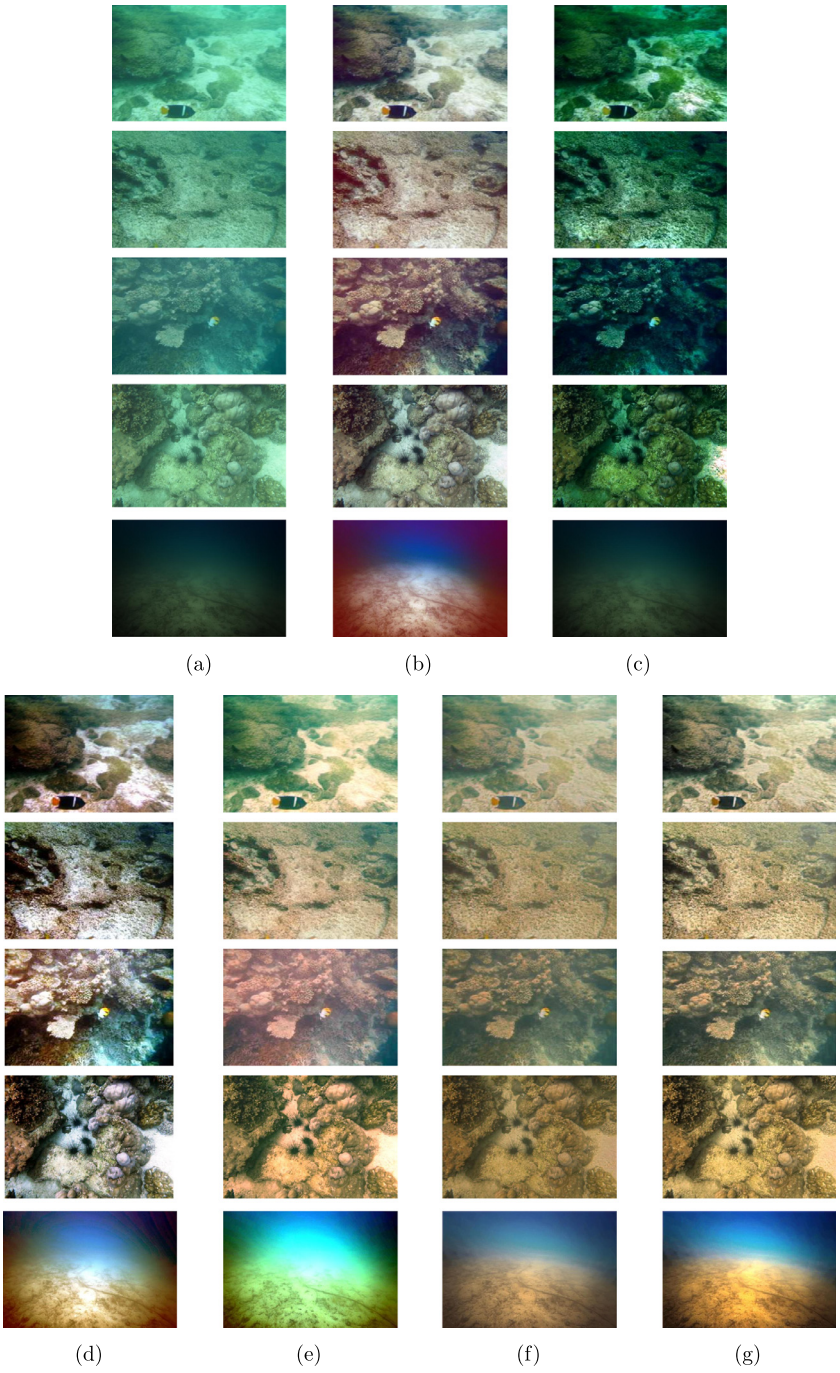


Fig. 6. Qualitative comparisons on real-world underwater images. (a) Real-world underwater images. (b) Results of RED [21]. (c) Results of UDCP [22]. (d) Results of ODM [25]. (e) Results of UIBLA [26]. (f) Results of our UWCNN. (g) Results of our UWCNN+. Our method (i.e., UWCNN and UWCNN+) produces the results without any visual artifacts, color deviations, and over-saturations. It also unveils spatial motifs and details.

image), the enhanced images have lower dynamic range. In practice, one would expect the enhanced results to have vivid colors and higher contrast.

To solve this issue, we employ a simple yet effective adjustment as a post-processing stage. We denote UWCNN with post-processing as UWCNN+. The image is first transformed to HSI color space. Then, the ranges of its saturation and intensity components in the HSI color space are normalized to [0,1] as:

$$y_{out} = \frac{y_{in} - y_{min}}{y_{max} - y_{min}}, \tag{12}$$

where $y_{\max}$ and $y_{\min}$ are the maximum and minimum saturation or intensity values in the UWCNN image. After this simple saturation and intensity normalization, we transform the modified result back to RGB color space.

Sample results are shown in Fig. 3. As visible, UWCNN effectively removes the dominant greenish color distortion in these real-world underwater images and significantly improves the contrast while preserving the natural look and authenticity of the images. Compared to UWCNN, the saturation and intensity normalization in UWCNN+ further improves the contrast and brightness, unveiling more details.

Table 1
N_λ values for synthesizing ten underwater image types.

Types	I	IA	IB	II	III
Blue	0.982	0.975	0.968	0.940	0.89
Green	0.961	0.955	0.950	0.925	0.885
Red	0.805	0.804	0.830	0.800	0.750
Types	1	3	5	7	9
Blue	0.875	0.800	0.670	0.500	0.290
Green	0.885	0.820	0.730	0.610	0.460
Red	0.750	0.710	0.670	0.620	0.550

5. Proposed underwater image synthesis algorithm

Unlike the high-level visual tasks [34–36] where large training datasets are often available, lacking underwater image dataset with corresponding ground truth constrains the development of deep learning-based underwater image enhancement and quality evaluation. To fill the gap, we propose an underwater image synthesis algorithm based on an underwater imaging physical model and the optical properties of underwater scenes. To the best of our knowledge, this is the first physical model based underwater image synthesis algorithm that can simulate a diverse set of water types and degradation levels, which is a significant contribution for the development of underwater image and video enhancement.

To synthesize underwater image degradation datasets, we use the attenuation coefficients described in [9] for the different water types of oceanic and coastal classes (i.e., I, IA, IB, II, and III for open ocean waters, and 1, 3, 5, 7, and 9 for coastal waters). As mentioned before, Type-I is the clearest and Type-III is the most turbid open ocean water. Similarly, for coastal waters, Type-1 is the clearest and Type-9 is the most turbid. We apply Eqs. (1) and (2) to build ten types of underwater image datasets using the RGB-D NYU-v2 indoor dataset [37] which consists of 1449 images. We select the first 1000 images as the training set and the remaining 449 images as the test set.

To synthesize an underwater image, we first generate a random homogeneous global atmospheric light $0.8 < B_\lambda < 1$. Then, we modify the depth $d(x)$ from 0.5 m to 15 m, which is followed by the selection of the corresponding N_λ values of the red, green, and blue channels for the water types presented in Table 1. For each image, we generate 5 underwater images based on random B_λ and $d(x)$; therefore, we obtain a training set of 5000 and a test set of 2495 samples. For computational efficiency, we resize these images to 310×230. In total, we synthesize ten underwater image datasets according to different water types.

Fig. 4 shows these ten different types of underwater images for a sample. It is evident that the underwater images of Type-I, Type-IA, and Type-IB are similar in their physical appearance and characteristics. Thus, we select a total of eight models out of ten to display the results of synthetic underwater images.

6. Experimental evaluations

In this part, we perform qualitative and quantitative comparisons with the state-of-the-art underwater image enhancement methods on both synthetic and real-world underwater images. In addition, we also compare the performance of different methods on underwater videos. These compared methods include UDCP [22], RED [21], ODM [25], and UIBLA [26]. We run the source codes provided by the authors with the recommended parameter settings to produce the best results for an objective evaluation. For real-world images where the light-attenuation coefficients are not available, we apply each of the ten UWCNN models we learned and present the results that are visually more appealing. This process can be improved by using a classification stage to choose the best model, which leave it as future work. For synthetic data, we present the results without post-processing since the models are derived from the synthetic data thus no intensity and saturation normalization are required. At last, we conduct an ablation study to demonstrate the effect of each component in our network.

Table 2
Quantitative evaluations on test set. As seen, our method achieves the best scores in all metrics on all underwater image types.

	Types	RAW	RED	UDCP	ODM	UIBLA	Ours
MSE	1	2367.3	3489.7	2062.3	2508.6	2812.6	**587.70**
	3	2676.5	4953.2	3380.6	3130.1	3490.1	**747.50**
	5	4851.2	8385.8	6708.9	3488.9	4563.7	**1295.1**
	7	7381.1	9809.8	8591.6	5337.1	6737.9	**2974.1**
	9	9060.6	5952.3	9500.1	10634.0	8433.1	**4121.5**
	I	1449.0	936.9	1020.7	1272.0	1492.2	**209.70**
	II	941.9	851.3	1466.0	1401.9	1141.4	**251.60**
	III	1851.0	2240.0	2337.6	1701.1	1697.8	**456.40**
PSNR	1	15.535	15.596	15.757	16.085	15.079	**21.790**
	3	14.688	12.789	14.474	14.282	13.442	**20.251**
	5	12.142	11.123	10.862	14.123	12.611	**17.517**
	7	10.171	9.991	9.467	12.266	10.753	**14.219**
	9	9.502	11.620	9.317	9.302	10.090	**13.232**
	I	17.356	19.545	18.816	18.095	17.488	**25.927**
	II	20.595	20.791	17.204	17.610	18.064	**24.817**
	III	16.556	16.690	14.924	16.710	17.100	**22.633**
SSIM	1	0.7065	0.7406	0.7629	0.7240	0.6957	**0.8558**
	3	0.5788	0.6639	0.6614	0.6765	0.5765	**0.7951**
	5	0.4219	0.5934	0.4269	0.6441	0.4748	**0.7266**
	7	0.2797	0.5089	0.2628	0.5632	0.3052	**0.6070**
	9	0.1794	0.3192	0.1624	0.4178	0.2202	**0.4920**
	I	0.8621	0.8816	0.8264	0.8172	0.7449	**0.9376**
	II	0.8716	0.8837	0.8387	0.8251	0.8017	**0.9236**
	III	0.7526	0.7911	0.7587	0.7546	0.7655	**0.8795**

6.1. Network implementation and training

We train our model using ADAM and set the learning rate to 0.0002, β_1 to 0.9, β_2 to 0.999. We fix the learning rate in the entire training procedure. The batch size is set to 16. It takes around three hours to optimize a model over 20 epochs. We use TensorFlow as the deep learning framework on an Inter(R) i7-6700k CPU, 32GB RAM, and an Nvidia GTX 1080 Ti GPU.

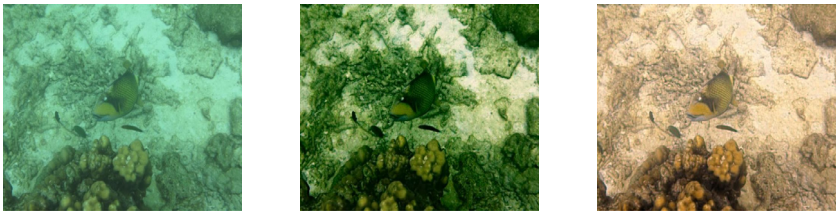
6.2. Evaluation on synthetic underwater image

We first present the results of underwater image enhancement on the synthetic underwater images from our test set. In Fig. 5(a), the synthetic underwater images accord with the measurement of [9]. The RED [21] is effective for the clear types, i.e., Type-1, Type-3, Type-5, and Type-I; however, for turbid types, i.e., Type-7, Type-9, Type-II, and Type-III, it leaves the haze on those images, moreover, it introduces color deviations. Similarly, UDCP [22] produces distinctly darkish results while ODM [25] and UIBLA [26] introduce artificial color or color deviations. On the other hand, our method not only enhances the visibility of the images but also restores an aesthetically pleasing texture and vibrant yet genuine colors. In comparison to other methods, the visual quality of our results resembles the ground-truth.

Furthermore, we quantify the accuracy of the recovered images on the synthetic test set including 2495 samples for each type. In Table 2, the accuracy is measured by three different metrics: mean square error (MSE), peak signal to noise ratio (PSNR), and the structural similarity index metric (SSIM) [38]. In the case of MSE and PSNR metrics, the lower MSE (higher PSNR) denotes the result is closer to the ground truth in terms of image content. In the case of the SSIM metric, the higher SSIM scores mean the result is more similar to the ground truth in terms of image structure

(a) Raw (b) ODM (c) UWCNN+

Fig. 7. Comparison with the ODM [25]. (a) Real-world underwater image. (b) Result produced by ODM [25] (incorrect reddish tones). It blindly introduces wrong colors, in particular, in red gamut. (c) Result produced by our UWCNN+. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)



(a) Raw (b) UIBLA (c) UWCNN+

Fig. 8. Comparison with the UIBLA [26]. (a) Real-world underwater image. (b) Result produced by UIBLA [26]. It is a failure case since only greenish tones are enhanced. (c) Result produced by our UWCNN+.

and texture. Here, the presented results are the average scores. The values in bold represent the best results.

As visible, among all underwater image enhancement methods we tested, our method comes out as the best performer across all metrics and all degradation types, demonstrating its effectiveness and robustness. Regarding the SSIM response, our method is at least 10% better than the second-best performer. Similarly, our PSNR is higher (less erroneous as indicated by the MSE scores) than the compared methods.

6.3. Evaluation on real-world underwater image

In this part, we evaluate the proposed method on real-world underwater images. Visual comparisons with competitive methods are presented in Fig. 6. The used real-world underwater images have diverse tone, light, and contrast.

A first glance at Fig. 6 may give the impression that the results of ODM [25] and UIBLA [26] might be sharper; however, careful inspection reveals that the ODM [25] causes over-enhancement and over-saturation (besides color casts) because the histogram distribution prior used in the ODM [25] is not always valid. Similarly, the images produced by the UIBLA [26] are not natural and consist of over-enhancement, a shortcoming of this method as the robustness of the background light and the medium transmission score estimated by the prior are suboptimum. Figs. 7 and 8 show the failure cases of the ODM [25] and UIBLE [26]. The RED [21] and UDCP [22] have little effect on the inputs. In contrast, our UWCNN+ shows promising results on real-world images, without introducing any artificial colors, color casts, over- or under-enhanced areas.

Observing the failure cases in Figs. 7 and 8, one can find that the ODM [25] tends to introduce extra colors (e.g., the reddish color around the coral in Fig. 7) while our method improves the contrast, similar performance to the ODM [25], but maintains a genuine color distribution of the original underwater image. For the failure case of the UIBLA [26] in Fig. 8, it aggravates the greenish color and produces visually unpleasing results. In contrast, our method removes color casts and improves contrast and brightness, which generates better visibility and pleasant perception.

Fig. 9. Real-world underwater images with diverse tones and degradation levels.

We note that the assessments in [39,40] are slanted toward over-exposure or over-enhancement, where the histogram equalization method is regarded to yield better scores. For a more objective assessment, we conduct a user study to provide realistic feedback and quantify subjective visual quality. We collect 20 real-world underwater images from the Internet and related papers. We show samples from this dataset in Fig. 9. Some corresponding results have been presented in Fig. 6.

For the user study, we randomize the order of the results and then display them on a screen to human subjects. There are 20 participants with image processing expertise. Each subject ranks the results based on the perceived visual quality from 1 to 5 where 1 is the worst and 5 is the best. One expects that the

Table 3
User study on real-world underwater image dataset. The best result is in bold.

	RED	UDCP	ODM	UIBLA	UWCNN+
Scores	2.95	2.55	3.25	3.20	**3.35**

Table 4
Average running time of different methods. The best result is in bold.

Time	RED	UDCP	ODM	UIBLA	UWCNN-C/-G
	3.250	3.319	5.829	47.254	2.250/**0.225**

results with high contrast, good visibility, natural color, and authentic texture should receive higher ranks while the results with over-enhancement/exposure, under-enhancement/exposure, color casts, and artifacts should have lower ranks. The average subjective scores are given in Table 3. Our UWCNN+ receives the highest rankings, which indicates that our method can produce better performance on real-world underwater images from a subjective visual perspective.

6.4. Evaluation on underwater video

To validate the capability of our model for underwater video enhancement, we conduct experiments on underwater videos. Due to the limited space, we only present parts of experimental results in Fig. 10.

As shown in Fig. 10, our method can remove the color casts and improve the contrast of the underwater video. Moreover, our results between different frames are consistent and without flickering artifacts. In contrast, the compared methods produce inconsistent enhancement between different frames, which decreases their visual quality. For example, for frame 54, the ODM [25] produces visually pleasant result; however, this method introduces reddish color casts in frames 1–4. The other methods also have similar inconsistent enhancement performance. Besides, we report the running time (second) of different methods to demonstrate our model can be used for frame-by-frame video enhancement in Table 4. The average running time for an image with size 640×480 is computed on the above-mentioned machine. UWCNN-C/-G indicates that our model runs only using CPU or GPU, respectively.

In Table 4, our UWCNN-G is faster than the compared methods with a large margin, which might benefit from GPU acceleration.

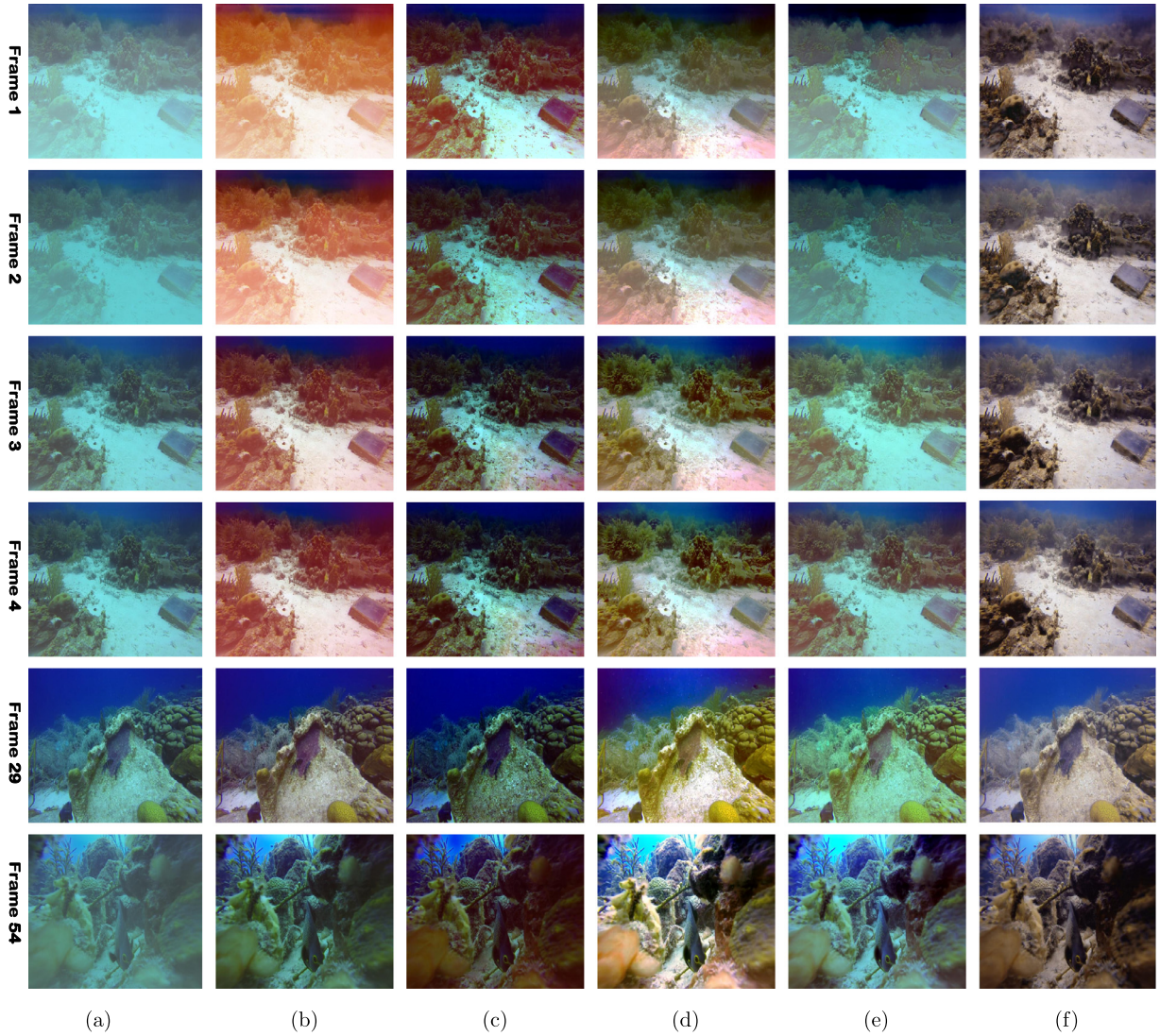


Fig. 10. Qualitative comparisons on the underwater video. (a) Raw underwater video (from top to bottom are frame 1, frame 2, frame 3, frame 4, frame 29, and frame 54 in this video). (b) Results of RED [21]. (c) Results of UDCP [22]. (d) Results of ODM [25]. (e) Results of UIBLA [26]. (f) Results of our UWCNN.

(a) Type-1 (b) -w/o SSIM (c) UWCNN (d) GT

Fig. 11. An example of the importance of SSIM loss. (a) An underwater image with Type-1 degradation. (b) Result produced by UWCNN-w/o SSIM, which is a failure case since the background is not similar to the GT. (c) Result produced by our UWCNN. (d) Ground truth.

However, our UWCNN-C ranks second fastest, which indicates our light-weight network structure also contributes to the processing speed of our method.

6.5. Ablation study

To demonstrate the effect of each component in our network, we carry out an ablation study involving the following experiments: (i) UWCNN without residual learning (UWCNN-w/o RL), (ii) UWCNN without dense concatenation (UWCNN-w/o DC), and (iii) UWCNN without SSIM loss (UWCNN-w/o SSIM). The quantitative evaluations are only performed on Type-1 and Type-III synthetic test set due to the limited space. The average scores in terms of MSE, PSNR, and SSIM are reported in Table 5.

Table 5
Quantitative results for the Type-1 and the Type-III test set. The best result for each evaluation is in bold, whereas the second best one is underlined.

	Types	-w/o RL	-w/o DC	-w/o SSIM	UWCNN
MSE	1	756.96	648.18	**398.77**	587.70
	III	542.68	789.76	**402.92**	456.40
PSNR	1	20.290	20.805	**22.902**	21.790
	III	21.556	20.289	**23.026**	22.633
SSIM	1	0.8450	0.8449	0.8214	**0.8558**
	III	0.8579	0.8359	0.8151	**0.8795**

From Table 5, one can see that replacing conventional learning strategy (i.e., UWCNN-w/o RL) with residual learning (i.e., UWCNN) could boost the performance. Comparing the UWCNN with the UWCNN-w/o DC, we observe that the dense concatenation also could improve the performance of underwater image enhancement. The use of SSIM loss (i.e., UWCNN) improves the structure and texture similarity at the cost of the decreased MSE and PSNR scores (i.e., UWCNN-w/o SSIM). However, such a sacrifice is necessary for better subjective perception. Such an example is presented in Fig. 11, which demonstrates the importance of SSIM loss. In Fig. 11, after adding SSIM loss, the result of UWCNN has a more smooth background than that of UWCNN-w/o SSIM.

7. Conclusion

We have presented an underwater image and video enhancement network inspired by underwater scene prior. Experiments on synthetic and real-world underwater images and videos demonstrate the robust and effective performance of our method. To our advantage, our method only contains ten convolutional layers and 16 feature maps at each convolutional layer, which provides fast and efficient training and testing on GPU platforms. Experimental results also demonstrate that the residual learning, dense concatenation, and SSIM loss used in our network boost the performance quantitatively and qualitatively.

In the future, we will investigate using only one single model to predict the correct output from one single blind model of UWCNN to attain further accelerating in the process of UWCNN model enhancement and also take the low contrast induced by indoor training data into consideration in a complete image degradation model. Borrowing the effective network structures and losses from the deep models designed for pattern recognition and computer vision, we will try to further improve the performance of our method.

Acknowlgedgments

This work was supported by the National Natural Science Foundation of China (61771334) and the Fundamental Research Funds for the Central Universities under Grant 2019RC039.

References

[1] D. Akkaynak, T. Treibitz, A revised underwater image formation, in: Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit. (CVPR), IEEE, 2018, pp. 6723–6732.
[2] S. Chikkerur, N. Cartwright, V. Govindaraju, Fingerprint enhancement using stft analysis, Pattern Recognit. 40 (1) (2007) 198–211.
[3] G. Lore, A. Akintayo, S. Sarkar, Llnet: a deep autoencoder approach to natural low-light image enhancement, Pattern Recognit. 61 (2017) 650–662.
[4] W. Wang, Y. Xu, J. Shen, S. Zhu, Attentive fashion grammar network for fashion landmark detection and clothing category classification, in: Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit. (CVPR), IEEE, 2018, pp. 4271–4280.
[5] F. Yang, K. Yan, S. Lu, H. Jia, X. Xie, W. Gao, Attentive driven person re-identification, Pattern Recognit. 86 (2019) 143–155.
[6] W. Wang, J. Shen, F. Porikli, R. Yang, Semi-supervised video object segmentation with super-trajectories, IEEE Trans. Pattern Anal. Mach. Intell. 41 (4) (2019) 985–998.
[7] C. Guo, C. Li, J. Guo, R. Cong, H. Fu, P. Han, Hierarchical features driven residual learning for depth map super-resolution, IEEE Trans. Image Process. 28 (5) (2019) 2545–2557.
[8] Z. Wu, C. Shen, A.V.D. Hengel, Wider or deeper: revisiting the resnet model for visual recognition, Pattern Recognit. 90 (2019) 119–133.
[9] D. Berman, T. Treibitz, S. Avidan, Diving into haze-lines: color restoration of underwater images, in: Proc. Brit. Mach. Vis. Conf. (BMVC), Springer, 2017, pp. 1–11.
[10] W. Wang, J. Shen, Deep visual attention prediction, IEEE Trans. Image Process. 27 (5) (2018) 2368–2378.
[11] H. Song, W. Wang, S. Zhao, J.S.K. Lam, Pyramid dilated deeper convlstm for video salient object detection, in: Proc. Eur. Conf. Comput. Vis. (ECCV), Springer, 2018, pp. 715–731.
[12] J. Gu, Z. Wang, J. Kuen, L. Ma, A. Shahroudy, B. Shuai, T. Liu, X. Wang, J. Cai, T. Chen, Recent advances in convolutional neural networks, Pattern Recognit. 77 (2018) 354–377.
[13] C. Li, J. Guo, Underwater image enhancement by dehazing and color correction, J. Electron. Imag. 24 (3) (2015) 033023–1 033023–10.
[14] C. Ancuti, C.O. Ancuti, Enhancing underwater images and videos by fusion, in: Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit. (CVPR), IEEE, 2012, pp. 81–88.
[15] C. Li, J. Guo, C. Guo, R. Cong, J. Gong, A hybrid method for underwater image correction, Pattern Recognit. Lett. 94 (2017) 62–67.
[16] 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.
[17] Y. Guo, H. Li, P. Zhuang, Underwater image enhancement using a multiscale dense generative adversarial network, IEEE J. Ocean. Engineer. (2019) 1–9.
[18] C. Ancuti, C.O. Ancuti, C. Vleeschouwer, Color balance and fusion for underwater image enhancement, IEEE Trans. Image Process. 27 (1) (2018) 379–393.
[19] C. Li, C. Guo, W. Ren, R. Cong, J. Hou, S. Kwong, An underwater image enhancement benchmark dataset and beyond (2019). arXiv: 1901.05495.
[20] J. Chiang, Y. Chen, Underwater image enhancement by wavelength compensation and dehazing, IEEE Trans. Image Process. 21 (4) (2012) 1756–1769.
[21] A. Galdran, D. Pardo, A. Picn, A. Alvarez-Gila, Automatic red-channel underwater image restoration, Vis. Commun. Image Rep. 26 (2015) 132–145.

[22] P. Drews, E. Nascimento, S. Botelho, M. Campos, Underwater depth estimation and image restoration based on single images, IEEE Comput. Graph. Appl. 36 (2) (2016) 24–35.
[23] K. He, J. Sun, X. Tang, Single image haze removal using dark channel prior, IEEE Trans. Pattern Anal. Mach. Intell. 33 (12) (2011) 2341–2343.
[24] 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: Proc. IEEE Int. Conf. Image Process. (ICIP), IEEE, 2016, pp. 1993–1997.
[25] C. Li, J. Guo, R. Cong, Y. 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.
[26] Y. Peng, P. Cosman, Underwater image restoration based on image blurriness and light absorption, IEEE Trans. Image Process. 26 (4) (2017) 1579–1594.
[27] J. Li, K. Skinner, R. Eustice, M. Roberson, Watergan: unsupervised generative network to enable real-time color correction of monocular underwater images, IEEE Robot. Autom. Lett. 3 (1) (2017) 387–394.
[28] M. Sheinin, Y. Schechner, The next best underwater view, in: Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit. (CVPR), IEEE, 2017, pp. 1436–1443.
[29] C. Li, C. Guo, J. Guo, P. Han, H. Fu, R. Cong, Pdr-net: perception-inspired single image dehazing network with refinement, IEEE Trans. Multimed. (2019) 1.
[30] Z. Shen, W.L.T. Xu, J. Kautz, M. Yang, Deep semantic face deblurring, in: Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit. (CVPR), IEEE, 2018, pp. 8260–8269.
[31] H. Huang, H. He, X. Fan, J. Zhang, Super-resolution of human face image using canonical correlation analysis, Pattern Recognit. 43 (7) (2010) 2532–2543.
[32] T. Lopes, E. de Aguiar, F.D. Souza, T. Oliveira-Santos, Facial expression recognition with convolutional neural networks: croping with few data and the training sample order, Pattern Recognit. 61 (1) (2017) 610–628.
[33] G. Huang, Z. Liu, L. van der Matten, K. Weinberger, Densely connected convolutional networks, in: Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit. (CVPR), IEEE, 2017, pp. 4700–4708.
[34] W. Wang, J.S.F. Guo, M. Cheng, A. Borji, Revisiting video saliency: a large-scale benchmark and a new model, in: Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit. (CVPR), IEEE, 2018, pp. 4894–4903.
[35] C. Zhou, J. Yuan, Multi-label learning of part detectors for occluded pedestrian detection, Pattern Recognit. 86 (2019) 99–111.
[36] C. Li, R. Cong, J. Hou, S. Zhang, Y. Qian, S. Kwong, Nested network with two-stream pyramid for salient object detection in optical remote sensing images, IEEE Trans. Geosci. Remote Sens. (2019) 1.
[37] N. Silberman, D. Hoiem, P. Kohli, R. Fergus, Indoor segmentation and support inference from rgbd images, in: Proc. Eur. Conf. Comput. Vis. (ECCV), Springer, 2012, pp. 746–760.
[38] Z. Wang, A. Bovik, H. Sherikh, E. Simoncelli, Image quality assessment: from error visibility to structural similarity, IEEE Trans. Image Process. 13 (4) (2004) 600–612.
[39] M. Yang, A. Sowmya, An underwater color image quality evaluation metric, IEEE Trans. Image Process. 24 (12) (2015) 6062–6071.
[40] K. Panetta, C. Gao, S. Agaian, Human-visual-system-inspired underwater image quality measures, IEEE J. Ocean. Eng. 41 (3) (2016) 541–551.

Chongyi Li received his Ph.D. degree from Tianjin University, China, in June 2018. From 2016 to 2017, he took one year study at the Research School of Engineering, Australian National University (ANU) as a visiting Ph.D. student. Now, he is a Postdoc Research Fellow at the Department of Computer Science, City University of Hong Kong (CityU), Hong Kong. He received Excellent Doctoral Degree Dissertation Award from Beijing Society of Image and Graphics. His research interests include image processing, computer vision, and deep learning, particularly in the domains of image dehazing, underwater image enhancement, image super-resolution, low-light image enhancement, and salient object detection.

Saeed Anwar is a Research Fellow at Data61, CSIRO (Commonwealth Scientific and Industrial Research Organization), Australia in Cyber Physical Systems. He received his Ph.D degree from the Australian National University (ANU) and Data61/CSIRO. He has been working as a Lecturer and Assistant Professor at the National University of Computer and Emerging Sciences (NUCES), Pakistan. His major research interests are low-level vision, image enhancement, image restoration, computer vision, and optimization.