IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received May 26, 2020, accepted June 7, 2020, date of publication June 15, 2020, date of current version June 24, 2020.

*Digital Object Identifier 10.1109/ACCESS.2020.3002593*

# NUICNet: Non-Uniform Illumination Correction for Underwater Image Using Fully Convolutional Network

**XUETING CAO, SHENGHUI RONG, (Member, IEEE), YONGBIN LIU, TENGYUE LI, QI WANG, AND BO HE, (Member, IEEE)**
Underwater Vehicle Laboratory, School of Information Science and Engineering, Ocean University of China, Qingdao 266100, China

Corresponding authors: Shenghui Rong (rsh@ouc.edu.cn) and Bo He (bhe@ouc.edu.cn)

This work was supported in part by the Project Funded by the China Postdoctoral Science Foundation under Grant 2019M652472, and in part by the Fundamental Research Funds for the Central Universities under Grant 201813019 and Grant 201861009.

**ABSTRACT** Absorption and scattering in aqueous media would attenuate light and make imaging difficult. Therefore, an artificial light source is usually utilized to assist imaging in the deep ocean. However, the artificial light source typically alters the light conditions to a large extent, resulting in the non-uniform illumination of images. To solve this problem, we propose a non-uniform illumination correction algorithm based on a fully convolutional network for underwater images. The proposed algorithm model the original image as the addition of the ideal image and a non-uniform light layer. We replace the traditional pooling layer with dilated convolution to expand the receptive field and achieve higher accuracy in non-uniform illumination recognition. To improve the perception ability of the network effectively, the original image and parameters which pre-trained on the ImageNet are concentrated. The concentrated information is used as input to the network. Due to the color shift and blurred details of the underwater image, we design the novel loss function, which includes three parts of feature loss, smooth loss, and adversarial loss. Moreover, we built a dataset of the underwater image with non-uniform illumination. Experiments show that our method performs better in subjective assessment and objective assessment than some traditional methods.

**INDEX TERMS** Underwater image enhancement, illumination correction, deep learning, fully convolutional network, dilated convolution.

## I. INTRODUCTION

Underwater optical camera is an essential sensor for detecting the ocean. Absorption and scattering of light in aqueous media cause the exponential decay that light suffers as it travels [1]. Therefore, there is low underwater visibility. Adding artificial light sources is a common way to improve underwater visibility [2]. However, the artificial light sources will cause the change of lighting conditions sharply during the short distance process of capturing underwater images, which brings enormous challenges to underwater imaging [3]. Due to the artificial light sources, underwater images are usually bright in the middle and dark in the surroundings. Illumination unevenness will display the highlights and hidden information incorrectly in dark areas, and the accuracy of underwater tasks will be affected, such as underwater target detection and underwater image semantic segmentation. Thus, the correction of non-uniform illumination (NUI) to improve the image quality is of great significance for underwater tasks.

The non-uniform illumination correction (NUIC) method with image processing techniques has been studied in various ways. Many traditional NUIC algorithms were proposed, such as Retinex algorithm [4], homomorphic filtering algorithm [5], and MASK algorithm [6]. Although most methods have made better results, there are several deficiencies in traditional algorithms, like weak adaptive effect and low precision. In recent years, deep learning technologies have successfully solved some underlying problems in the field of computer vision with the improvement of computing performance, such as super-resolution [7] and deblurring [8]. The deep learning methods achieve excellent performance through a large number of data, and the pre-trained models are universal for many tasks. Moreover, the appropriate parameters of deep learning methods are not required to choose manually when processing images as a result of the

The associate editor coordinating the review of this manuscript and approving it for publication was Sudhakar Radhakrishnan.

objectively good self-adaptability. Therefore, deep learning methods are superior to traditional methods in terms of effectiveness, adaptability, and versatility. In particular, convolutional neural networks (CNN) are used in various fields. In [9], Long *et al.* proposed a Fully Convolutional Network (FCN), which extended the original CNN structure into pixel-wise prediction without a fully connected layer. Therefore, this structure can accept input of any size and process the image more accurately.

Inspired by FCN, we propose a network structure with a novel loss function to correct the NUI of underwater images. On the grounds of the characteristics of underwater imaging, the NUI image is modeled as the addition of the ideal image and the light layer. The proposed network is based on a fully convolutional network that separates the light layer and the ideal image. FCN achieves the pixel-wise prediction of the illumination field accurately. Our end-to-end network takes a single image as input and separates the light layer and the ideal image from the original image directly. The dilated convolution was used to increase the receptive field of the proposed network for precise global illumination. To further improve the perception ability of our network, we concatenate the input image with the features which trained on the benchmark dataset (ImageNet). Furthermore, a novel loss function is designed to get the optimal performance of our network in terms of image details and authenticity. Due to the rare public datasets of NUI underwater images, we built a new dataset containing the synthetic and real data for training and testing. Our main contributions are summarized as follows:

(1) A fully convolutional network structure is proposed for the NUIC of underwater optical images.

(2) We design a novel loss function for illumination correction, which improves the performance of our proposed network in terms of the details and authenticity of restored images.

(3) To boost underwater imaging processing, we have built a common suitable dataset of NUI optical images that can be used for qualitative and quantitative assessment of our method and other existing algorithms.

(4) This method has an excellent performance in subjective assessment and objective assessment, and the pre-trained model can be utilized for the NUIC of other types of images.

The remainder of this paper is organized as follows. In Section II, we review the related work. Then we give a detailed explanation of the image enhancement model and our network architecture in Section III. In Section IV, we introduce the details of network training. Section V shows the experimental results and analysis in which we evaluate the different methods for NUIC, followed by a conclusion in Section VI.

## II. RELATED WORK

The quality of the acquired image directly affects the accuracy of the next step, such as image classification, target recognition, and image semantic segmentation. Improving image quality through NUIC is an essential task. We will obtain more useful information from the corrected image. Many researchers have done a lot of work for illumination correction and achieved positive results. NUIC is typically applied to remote sensing images [10], optical microscope images [11], aerial images [12], and scanned images [13]. Also, there is a need to remove vignetting in optical photography [14], [15]. According to the image enhancement model, the traditional NUIC algorithms are mainly included three categories: incident-reflection multiplicative model, mathematical-statistical model, and light additive model.

Algorithms based on the incident-reflection multiplicative model mainly include Retinex-inspired algorithms [16] and homomorphic filtering algorithms [5]. The image is modeled as the product of incident component and reflected component, where the reflected component means the high-frequency information of the image, that is, the part with uniform illumination [17]. The Retinex algorithm achieves the purpose of correcting illumination by extracting reflection components from images with uneven illumination. Single-scale Retinex (SSR) [18], multi-scale Retinex (MSR) [19], and Multi-Scale Retinex with Color Restoration (MSCR) [20] are pioneering research in the field of illumination correction. These algorithms do not retain much detail when separating light, resulting in blurry image details and color distortion. In [21], a new light-pass filter was proposed, replacing the Gaussian filter in the traditional Retinex algorithm, which enhanced the image details while maintaining the naturalness of the image as much as possible. However, due to the lack of constraints on reflectivity, noise in images with uneven illumination will be amplified. Seow and Asari [22] proposed a homomorphic filtering algorithm based on neural network learning based on the Ratio rule. This method can obtain color images from the color information of the original image during image processing, but there is details distortion.

Algorithms based on the mathematical-statistical model believe that its mean represents the hue and brightness of an image, and its standard deviation represents the sharpness and contrast of an image. These methods achieve eliminating the inconsistency between the brightness and contrast of the image by changing mean and variance of the image [23]. The major algorithms based on statistical methods are Histogram matching [24] and Wallis algorithm [25]. Histogram matching adjusts the mean and variance by directly changing the histogram shape to fit the reference image and the target image. If the different internal features of the image differ greatly, histogram matching may change the relative distance between the gray levels, resulting in the color deviation of the image with different internal features. Due to block processing, the images processed by the traditional Wallis algorithm usually have obvious artificial blocks. Moreover, the parameters of the existing Wallis algorithm usually depend on the standard image selected from the original image, which has many different standards and is not robust.

The NUIC algorithms based on the light additive model believe that an image with non-uniform illumination is composed of an image with uniform illumination and the undesired noise. The MASK-inspired algorithms are mainstream algorithms based on the noise additive model. MASK algorithm uses a Gaussian filter to remove uneven illumination in the frequency domain and obtain an image with uniform illumination. It has a simple structure and works well. Nevertheless, the traditional MASK algorithm has poor self-adaptability and requires manual selection of filtering parameters, which is labor-intensive and highly subjective. Aiming at the problem of poor adaptability of traditional MASK, Yao *et al.* [26] proposed an adaptive method. After the image is divided into blocks, different filtering parameters are selected for each image to improve the algorithm efficiency. Although adaptive, this method may lose some image details.

From the above, the traditional NUIC algorithm has disadvantages such as poor adaptability, low accuracy, and poor robustness. In addition, the traditional algorithm is effective for images with low illumination unevenness and has poor processing effects for images with severe illumination unevenness.

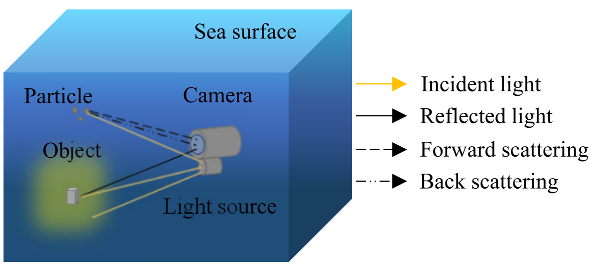



**FIGURE 1.** **Underwater imaging model.**

## III. METHODS

### A. IMAGE ENHANCEMENT MODEL

A typical underwater imaging system is shown in Figure 1. The light entering the imaging system is mainly contained the direct irradiation component and the scattered components [27]. Among them, the direct irradiation component refers to the light that is not scattered or absorbed by the water, and surface into the camera. The scattering component is divided into forward scattering and backward scattering. The forward scattered light is reflected by the target surface or the light that is scattered by suspended particles in the water and enters the imaging system. It has a marginal impact on the imaging result that is usually ignored. Therefore, the underwater imaging model can be simplified as the addition of the light reflected by the object and the backscattered components. Inspired by the underwater imaging model, the degraded image with uneven illumination is considered as an additive model of illumination. Hence, we approximately model the uneven illumination image $U \in R^{m\times n\times 3}$ as

$$U = I + L, \tag{1}$$

where $I$ is the ideal image which is uniform, and $L$ is the light layer. Our goal is to predict the light layer from the original image using a single neural network and obtain the ideal image with uniform illumination.

### B. NETWORK ARCHITECTURE

Given a non-uniform illumination image $U\epsilon[0-1]^{m\times n\times 3}$, according to the image enhancement model, we use a single neural network $f(U;\theta)$ to separate the ideal image $I$ and the light layer $L$ from the non-uniform illumination image. The objective of the network is

$$f(U;\theta) = (f_I(U;\theta), \quad f_L(U;\theta)). \tag{2}$$

Figure 2 shows the pipeline of our network. The proposed network consists of two consecutive operations, feature fusion and illumination layer separation. The former operation extracts the parameters trained on the benchmark dataset (ImageNet) [28], then combines with the input image as hyper-column features [29]. The latter operation is a fully convolutional network built sequentially by stacking nine blocks, which of eight perform feature extraction, the other two blocks are used for information fusion. In the training phase, we will optimize the result of light correction through the loss function. We calculate the L1 distance between the ground truth and the ideal image predicted by the proposed network to further improve accuracy. There is no fully connected layer in the whole network. Hence, it can accept the inputs of any size without being affected by a fully connected layer. At the same time, training the end-to-end network can predict the illuminance of an image more accurately and ensure the large receptive field.

The function of the feature fusion operation is to improve correction accuracy by fusing vital information while enhancing the perception ability of the network. The input of the network is a NUI image combined with the features extracted from ‘conv1_2’, ‘conv2_2’, ‘conv3_2’, ‘conv4_2’, ‘conv5_2’ of VGG-19. The extracted features have a total of 1472 dimensions, so the whole input dimension of the network is 1475 (1472+3). Since the extracted feature maps have different sizes, we use bilinear interpolation to scale the feature maps to the same size. Multi-dimensional features fusion can be mapped to the high-dimensional space, thereby enhancing the understanding of images by learning non-linear relationships of our network.

Details of illumination layer separation operation are shown in Figure 3. The information fusion block is composed of a convolution layer with a convolutional kernel size of $1 \times 1$, a ReLU layer, and a batch normalization layer. By changing the number of input channels, it realizes the multi-information fusion of inputs The feature extraction block consists of a dilated convolution layer with different dilated rate, a ReLU layer, and a batch normalization layer, where the convolutional kernel size is $3 \times 3$ [30]. The corresponding dilated rate $n$ of each feature extraction block is in the range of 2 to 128. The size of the receptive field of the proposed structure is shown in Table 1, and the maximum can

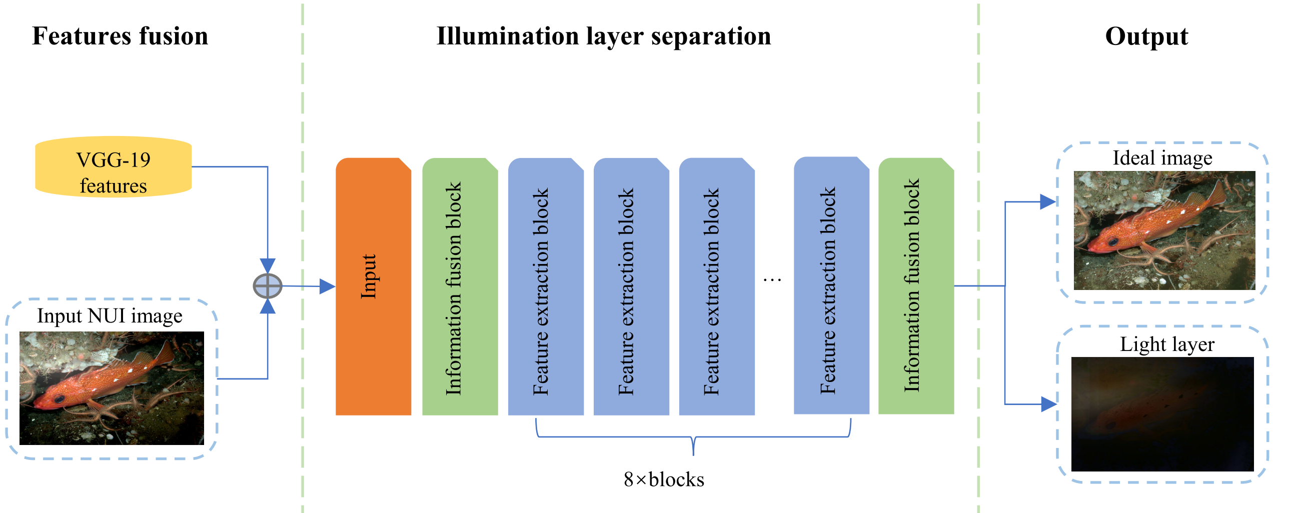



**FIGURE 2.** **The pipeline of the proposed network.**

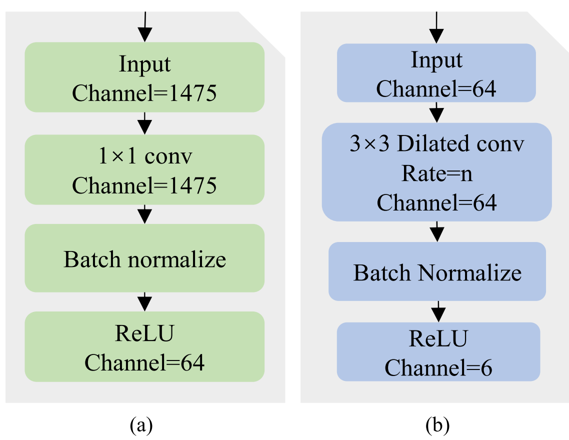



**FIGURE 3.** **From left to right: (a) is the details of information fusion block 1, and (b) is the details of feature extraction block.**

reach 513 × 513. Some critical information can be extracted by the feature extraction block like edges, colors, etc.

In the case of inputting a NUI image, the output will contain the ideal image and light layer of the input image. The first information fusion block fuses the input information of 1475 channels and compresses it to 64 channels. We set the number of information channels of the subsequent seven feature extraction blocks to 64, which using the dilated convolution with the different dilated rates in them. The last block fuses the 64-dimensional information into 6-dimensional, including the ideal images and illumination layers with uniform illumination, which are all RGB three-channel images.

### C. MORE ANALYSES OF NETWORK ARCHITECTURE

#### 1) HYPER-COLUMN FEATURE

In many works, the fusion of features across different scales is an essential way to improve network performance. Low-level features are precise in localization and contain detailed information. However, the few convolutional layers will result in less sensitive to semantics and high noises of low-level features. While the high-level features have multiple layers of information extraction and rich semantic information, the resolution is low and the ability to perceive details is weak. This characteristic suggests that reasoning at combining low-level features with high-level features has proven benefits of the network. According to the order of fusion and prediction, feature fusion is divided into early fusion and late fusion. Early fusion first fuses multiple layers of features and then trains predictors on the fused features. Late fusion improves performance by combining the prediction results of different layers.

We hope that our network is not limited to the application of the underwater image while it can also be extended to more scenes for its robustness, such as illumination correction of remote sensing images. Considering the better generalization of our network that the illumination corrections in other scenes are not only caused by point light sources. Inspired by the feature fusion, we make use of the extracted features of VGG-19 for tackling complex tasks and improving accuracy. Therefore, the feature fusion enables our network to recognize more different lighting scenes, not just scenes with point light sources.

#### 2) DILATED CONVOLUTION

To make the proposed network focus on the global rather than local illumination information, the utility way for expanding the receptive field of the network was applied without losing any information during operations. Generally, the function of pooling layer is to increase the receptive field of the network, thereby improving the accuracy of the network. However, the pooling layer will reduce the size of the feature map. This operation reduces the amount of calculations while losing certain information about features. The dilated convolution is instead to replace the traditional pooling layer.

**TABLE 1.** **Details of the proposed network.**

| Block | Output size | Channel | Kernel size | Dilated rate(n) | Equivalent convolution kernel size | Receptive field |
|---|---|---|---|---|---|---|
| Input | 640×640 | 1475 | - | - | - | - |
| Information fusion block 1 | 640×640 | 64 | 1×1 | 1 | 1×1 | 1×1 |
| Feature extraction block 1 | 640×640 | 64 | 3×3 | 1 | 3×3 | 3×3 |
| Feature extraction block 2 | 640×640 | 64 | 3×3 | 2 | 5×5 | 7×7 |
| Feature extraction block 3 | 640×640 | 64 | 3×3 | 4 | 9×9 | 15×15 |
| Feature extraction block 4 | 640×640 | 64 | 3×3 | 8 | 17×17 | 31×31 |
| Feature extraction block 5 | 640×640 | 64 | 3×3 | 16 | 33×33 | 63×63 |
| Feature extraction block 6 | 640×640 | 64 | 3×3 | 32 | 65×65 | 127×127 |
| Feature extraction block 7 | 640×640 | 64 | 3×3 | 64 | 129×129 | 255×255 |
| Feature extraction block 8 | 640×640 | 64 | 3×3 | 128 | 257×257 | 511×511 |
| Information fusion block 2 | 640×640 | 6 | 1×1 | 1 | 3×3 | 513×513 |
| Output | 640×640 | (3+3) | - | - | - | - |

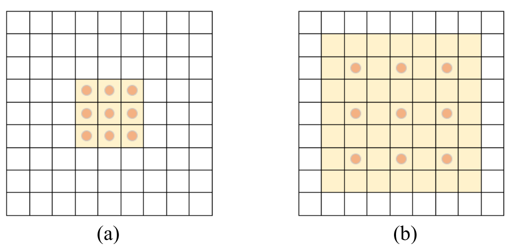


**FIGURE 4.** **Explanation of the principle of dilated convolution. (a) is traditional convolution with 3 × 3 kernel size, (b) is a 3 × 3 kernel size convolution with an expansion rate of 2. The receptive field is the same as the 5 × 5 convolution kernel, and only 9 parameters are required.**

Dilated convolution is a novel convolution method. The interval of each parameter of its convolution kernel is empty, forming some ‘holes’, as shown in Figure 4. Making ‘holes’ is not to directly padding 0, but to skip some existing pixels and empty the convolution kernel. The first benefit of dilated convolution is that it can reduce parameters and prevent overfitting. Secondly, the dilated convolution increases the receptive field without reducing the feature map, which means that each convolution output contains a more extensive range of information without losing information. Under the same calculation conditions, dilated convolution is suitable for our task. Thus, dilated convolution is adopted in all information extraction blocks of our network.

#### 3) BATCH NORMALIZATION

During network training, the problem of Internal Covariate Shift (ICS) will occur [31]. As the training progresses, the parameters in the network are updated continuously. On the one hand, when the parameters in the underlying network change slightly, the existence of linear transformation and non-linear activation mapping in each layer will lead to these weak changes amplified as the deeper network; on the other hand, the parameter’s change leads to the change of input distribution of each layer. The upper-layer network needs to constantly adapt to these distribution changes, resulting in a decrease in the network learning speed, and the network’s training process quickly falls into the gradient saturation region. Batch normalization processes each feature individually so that each feature has a distribution with a mean of 0 and a variance of 1. If batch normalization is not performed, problems such as slow learning and gradient dispersion may occur. We have added the batch normalization layer into our network.

Furthermore, we set the mini-batch mean and variance as the estimate of the overall training samples. Although each batch of data from a sampling of the overall, the mean and variance of different mini-batch will have a little discrepancy. Moreover, the batch normalization layer adds random noise during the learning process, a model the effect of the generalization ability of ascension. Insufficient data for training may lead to weak generalization ability of the model. The batch normalization layer was used to improve the generalization capability of the network.

## IV. DETAILS OF TRAINING

### A. LOSS FUNCTION

Local adjustment of the image (such as contrast, local exposure is too strong or too weak) and global information of the image (such as image color, overall light intensity) need to be learned to correct unevenly illuminated images effectively. To this end, we designed a novel loss function which contains three parts, feature loss $L_f$, smooth loss $L_s$, and adversarial loss $L_a$. Underwater images usually have color casts and blurred details, which brings great challenges to the restoration effect close to the original image. To better restore the colors, edges and other features of the separated ideal image and make it closer to the original image, we define the feature loss as

$$L_f = \sum_{(I,U)\in D} \sum_l \lambda_l \left\| \varphi_l(I) - \varphi_l(f_I(U;\theta)) \right\|_1 , \quad (3)$$

here, $\lambda_l$ is a hyperparameter used to balance the loss terms, $f_I(U;\theta)$ is the uniform illumination image which predicted by our network, $I$ is the ground truth, $\varphi_l$ represents layer $l$ in the VGG-19 network. We feed the uniform illumination image $f_I(U;\theta)$ which predicted by our network and the ground truth $I$ into VGG-19, and compute the L1 distance between $\varphi_l(I)$ and $\varphi_l(f_I(U;\theta))$ in the selected feature layers. We select the layers 'conv1-2', 'conv2-2', 'conv3-2', 'conv4-2', and 'conv5-2' in the VGG-19 network. Through training, the loss value is continuously decreasing and the prediction of the ideal image more accurate. Moreover, feature loss also combines the features of the pre-trained model on a large dataset (ImageNet) to enhance the generalization of the network, making our model suitable for a variety of scenes with uneven lighting.

According to previous work [32]–[34], natural illumination in images is usually smooth locally. After we separate the light layer from the NUI image, the illumination of the ideal image obtained may not be smooth enough locally and cannot be restored to the state of natural illumination. Thus, we use the smooth loss to correct the lighting in the ideal image so that the lighting will be closer to natural lighting and the result can be restored better. Smooth loss can not only make the lighting of an ideal image more natural but also enhance the contrast of the image. Smooth loss $L_s$ is defined as

$$L_s = \sum_p \sum_c \omega^p_{x,c}\left(\alpha_x L_p\right)^2_c + \omega^p_{y,c}\left(\alpha_y L_p\right)^2_c. \quad (4)$$

Among them, we sum each channel $c$ and each pixel $p$. $\alpha_x$ and $\alpha_y$ are the first-order partial derivatives in the $x$ and $y$ directions, respectively; $\omega^p_{x,c}$ and $\omega^p_{y,c}$ are the weights of local light smoothness, calculated as

$$\omega^p_{x,c} = \left(\left|\alpha_x G^p_i\right|^\theta_c + \epsilon\right)^{-1}, \quad (5)$$

$$\omega^p_{y,c} = \left(\left|\alpha_y G^p_i\right|^\theta_c + \epsilon\right)^{-1}, \quad (6)$$

where $G^p_i$ is the logarithmic image of the input image $U_i$; $\theta$ is a constant about the sensitivity of the image gradient; $\epsilon$ is a constant to prevent the weight equal to zero, and it often set to 0.0001.

Since after the light layer and the ideal image were separated, problems such as detailed distortion may be exposed in the ideal image. We use conditional GAN [35] to correct the error. In addition, to further correct the authenticity of the generated ideal image, an adversarial loss function is proposed with the discriminator of conditional GAN and which is expressed as

$$L_a = \sum\nolimits_{I,U\in D} \log D\left(I, f_I(U;\theta)\right) - logD\left(U, I\right), \quad (7)$$

where $D(I, x)$ is the probability distribution of the output, $I$ is the input image, and $x$ is the ideal image predicted by the proposed network. In conclusion, the complete proposed loss function is expressed as

$$L = w_1 L_f + w_2 L_s + w_3 L_a, \quad (8)$$

here $w_1$, $w_2$, and $w_3$ are the weights of each loss; we set $w_1$ as 1, $w_2$ as 0.1, and $w_3$ as 0.1.

### B. DATASET

#### 1) SYNTHETIC DATA

A large number of training data pairs that contain the non-uniform illumination image and the ideal image is very essential to train our model. Given that there are no publicly available datasets and the acquisition of real underwater images is difficult because of the unstable condition. Therefore, we made synthetic data to train and test the network. During the synthesization, we only consider the effect of illumination and ignore the effects of underwater forward and backward scattering, which allows us to focus on the purpose of NUIC. In practical underwater applications, the light source is relatively small due to the limited size of the underwater vehicle. Compared with the imaging distance, the light source usually can be considered as a point one. Thus, the point light source is adopted as the main form of artificial light sources in synthetic data experiments.

The point artificial light source usually casts a light spot in the center of the illuminated area, which will cause NUI on the image. Apparently, the brightness is the strongest in the spot center and gradually attenuates in the radial direction. Besides, the brightness of the spot depends on the power of the artificial light source, and the position of light spot depends on the angle of the light source. According to the above analysis, we can synthesize NUI on images with different intensity, size, and position.

We selected images with uniform light from some public underwater image data sets SUN [36], fish4knowledge [37] as ideal images, including 4,816 underwater images and 1,000 other natural scene images, and modeled uneven underwater light images according to its characteristics. The gray value of uneven illumination can be modeled as

$$f(x, y) = k \times \left(1 - \frac{\sqrt{(x-x_0)^2 + (y-y_0)^2}}{r}\right), \quad (9)$$

where $k$ is the illumination coefficient, $r$ is the radius of the artificial light source and the point $(x_0, y_0)$ is selected as the artificial light source center point, and $f(x, y)$ is the gray value of each pixel of the artificial light source. A point light source with a radius $r$ is simulated, and the farther away from the center of the bright point, the smaller the gray value will be. The illumination coefficient $k$, the radius $r$, and the artificial light source point $(x_0, y_0)$ are all random values which follow Gaussian distribution. We add simulated illumination to the ground truth to get samples with uneven illumination. Figure 5 shows the synthetic samples in the dataset.

#### 2) REAL DATA

Training data is critical to the performance of the model. The amount of data, data type, and data quality will directly affect the quality of training. In order to enlarge the dataset and enrich the type of data, we obtained the real data through the interior and outside experiments. The real data consists of two parts, the NUI image and the ideal image. The real

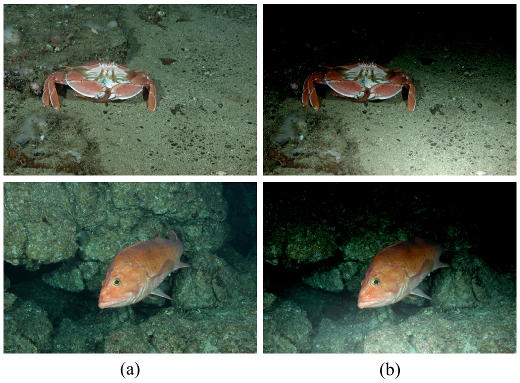



**FIGURE 5.** Samples in dataset. (a) the ground truth, (b) the synthetic image with non-uniform illumination.

data was taken with the Bumblebee XB3 optical camera, and we made a watertight capsule for the camera to adapt to the underwater environment. Our experimental environment consisted of indoor pools (filled with seawater) and artificial lakes, including a darker indoor environment and an overcast outdoor environment. First, we fixed the camera with a watertight cabin and the target in the pool was taken into the ideal image. Then, the position of the camera and the target was kept unchanged, the uneven illumination image was shot by adding an artificial light source, in which the angle and intensity of the artificial light source are random. We made a total of 100 pairs of real data and expanded the data volume again through flipping and cropping. The real data were randomly divided into train set and test set. The dataset we built is public at *https://github.com/caoxueting555/uneven-illumination-correction*.

### C. OPTIMIZATION METHOD

We trained our network using Adam optimization method [38]. Adam is a first-order gradient-based optimization of the stochastic objective function, based on adaptive estimates of lower-order moments. The learning rate can be adaptively adjusted for different parameters based on the first and second moment estimates of the gradient, so the calculation efficiency is high, and the memory requirement is small. It combines the advantages of two recently popular methods: AdaGrad and RMSProp, which are faster than the RMSprop method when gradients become sparse [39]. Moreover, it has invariance for diagonal gradient scaling, which solves the problem of large data and difficult training.

Adam updates the parameters as follows

$$g_t = \frac{1}{n}\nabla\theta \sum_i L\left(F\left(Y_i, \theta\right), X_i\right), \tag{10}$$

$$m_t = u \times m_{t-1} + (1-u) \times g_t, \tag{11}$$

$$n_t = v \times n_{t-1} + (1-v) \times g_t^2, \tag{12}$$

$$\hat{m}_t = \frac{m_t}{1-u^t}, \tag{13}$$

$$\hat{n}_t = \frac{n_t}{1-v^t}, \tag{14}$$

$$\nabla\theta_t = -\eta \times \frac{\hat{m}_t}{\sqrt{\hat{n}_t + \varepsilon}}, \tag{15}$$

$$\theta_{t+1} = \theta_t + \nabla\theta_t, \tag{16}$$

where $g_t$ is the gradient of the mean square error function $L(\theta)$ versus $\theta$, $m_t$ is the first-order moment meter for the gradient, $n_t$ is the second-order moment estimate for the gradient, $\hat{m}_t$ is the deviation correction for $m_t$, and $\hat{n}_t$ is deviation correction of $n_t$, the exponential decay rate $u$ of the moment estimation is 0.9, $v$ is 0.99, and the step size $\eta$ is 0.001. The numerically stable small constant $\varepsilon$ is $10^{-8}$, $\nabla\theta_t$ is the calculated $\theta_t$ update value, $\theta_{t+1}$ is the $\theta$ value at time $t+1$, that is, the sum of the values of $\theta_t$ and $\nabla\theta_t$ is applied to $\theta_{t+1}$. We train the variants until the training loss converges. After then we select the model which performs the highest accuracy on our training datasets.

### D. TRAINING SETTINGS

In the training processing, 5816 pairs of synthetic data and 80 pairs of real data were selected as training data, all the data was cut into $640 \times 640$ pixels for training. The training image pair includes the NUI image and ground truth. The network inputs the NUI image to predict the light layer and the ideal image. The loss function calculates the distance between the ideal image and ground truth and continuously optimizes it. We trained our network for 200 epochs on an NVIDIA GeForce RTX 2080 GPU, using the Tensorflow framework [40] to minimize loss functions. The optimization method is Adam optimization method, the exponential decay rate $\boldsymbol{u}$ of the moment estimation is 0.9, $\boldsymbol{v}$ is 0.99, and the step size is 0.001.

## V. EXPERIMENTAL RESULTS AND ANALYSES

In this section, we evaluate the proposed algorithm. To effectively evaluate the proposed algorithm, we compare the proposed algorithm with seven algorithms including classic algorithms and excellent algorithms in recent years. The compared algorithms are BPDHE [41], FEA [42], ALT [16], EFF [43], ROBUST [44], JED [45], and MLE [46].

### A. EXPERIMENT SETTINGS

We provided different types of data for our pre-trained model. The first of which contains the real data without ground truth collected in the public dataset (Figure 6) [46], the other part consists of the real data we captured and the synthetic data with ground truth (Figure 7). There is a certain difference in the size and scene of test data, which can fully confirm the effectiveness of the algorithm.

### B. SUBJECTIVE ASSESSMENT

The last row in Figure 6 and Figure 7 represents the predicted results of the light layer by the proposed algorithm. It shows the precise visual prediction of the illumination field, both in the position and edge. The penultimate row in Figure 6 and Figure 7 are the ideal images separated by our network. After processing by our model, we can intuitively observe the

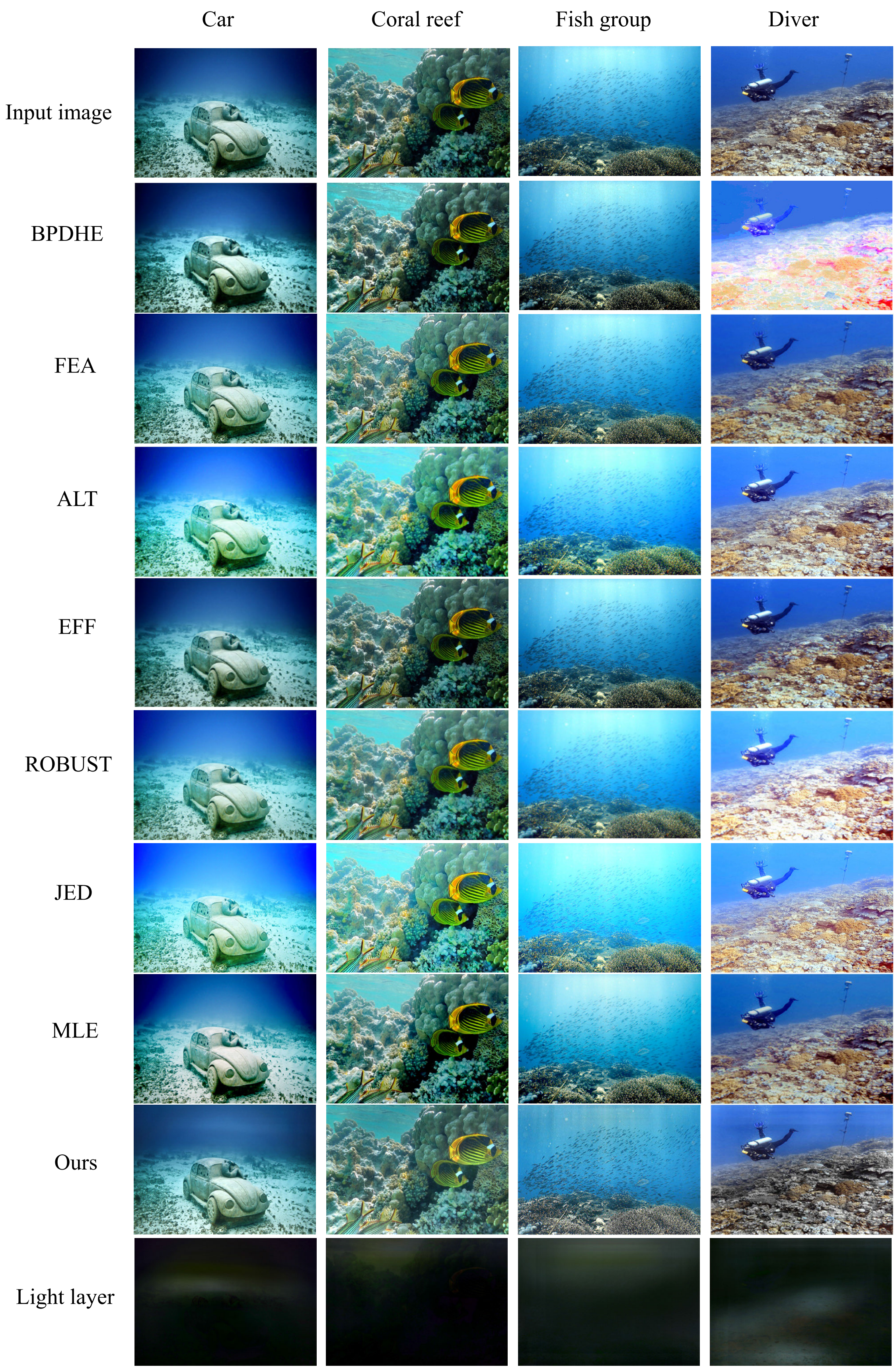



**FIGURE 6.** Results of non-uniform illumination image without ground truth.

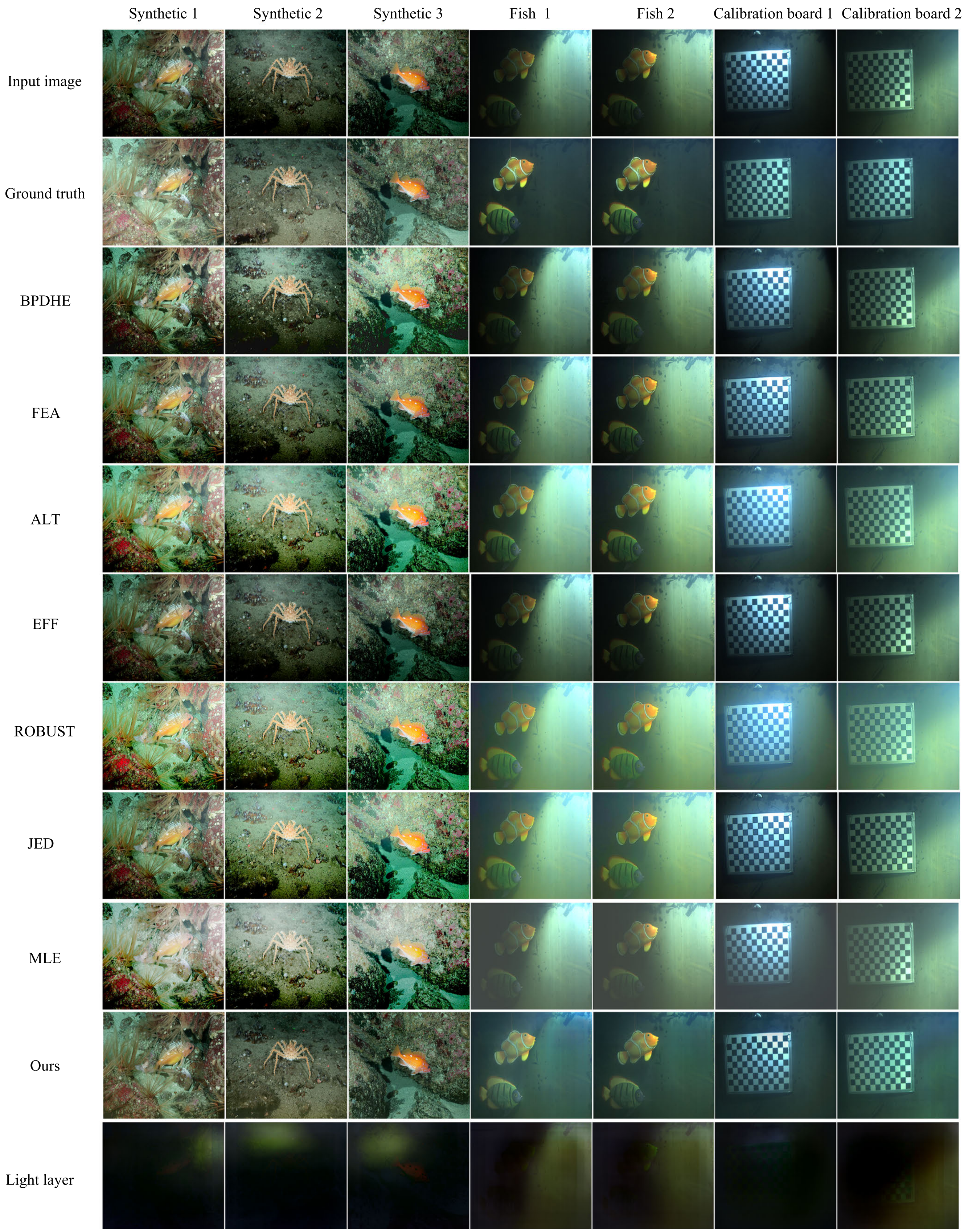



**FIGURE 7.** Results of synthetic data and real data with ground truth. From left to right: the first three columns are the experimental results of the synthetic data, and the last four columns are the experimental results of the real data captured by us.

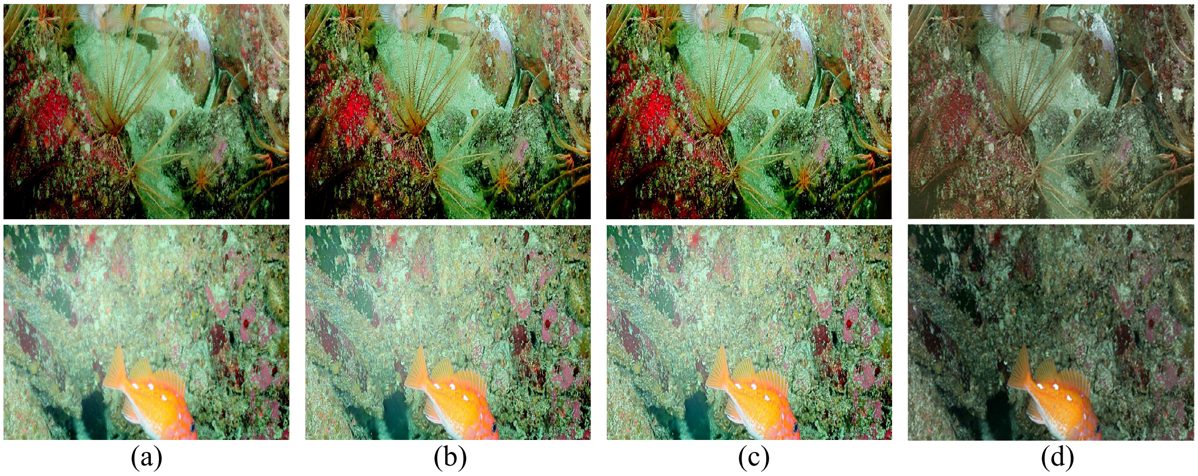


**FIGURE 8.** Detalis of processed image. From left to right: (a) ALT, (b) JED, (c) ROBUST, (d) the proposed method.

changes of the NUI image where the light becomes uniform, and the bright spots almost disappear. The ideal image has the same color as the original image, and the brightness distortion is small, which is visually comfortable.

It can be seen from Figure 6 and Figure 7 that BPDHE algorithms amplifies the noise, and does not continue processing so that the details in the dark area are blurred. ALT, FEA, and EFF yield superior performance for slight light unevenness. It is worth noting that the result is not ideal for severe light unevenness which showed in fish 1 and fish 2 of Figure 7. The JED algorithm shows a stronger visual processing effect, in which the light is progressively corrected and the image details are clear, but the brightness is far from the original image. Although the MLE method has a good effect on the correction of bright areas, it is not ideal for the enhancement of dark areas. Besides, we notice that the ALT, JED, and ROBUST algorithms retain better details and color consistency while maintaining visual comfort. Therefore, we compared our method with them concerning small color distortion and clear details. Figure 8 shows the ideal image predicted by our model, which has clearer details and no color distortion.

Our proposed method gives a more uniform result of illumination, the only inadequacy is that the whole image is slightly dark compared to the other algorithm. This is because our algorithm mainly concerns with the NUIC process and keeps the image intact as much as possible. Besides, it is an easy task to adjust the whole brightness of the image. On the one hand, it is of vital importance for the uniform illumination and clear details of the image in our work. on the other hand, the image processed by other algorithms has the problem of brightness distortion, which is not friendly to the process of ambient light estimation for underwater image restoration. The above can be concluded that our method can practically resolve the brightness distortion, making it more similar to the original image, which means the effectiveness of subsequent tasks.

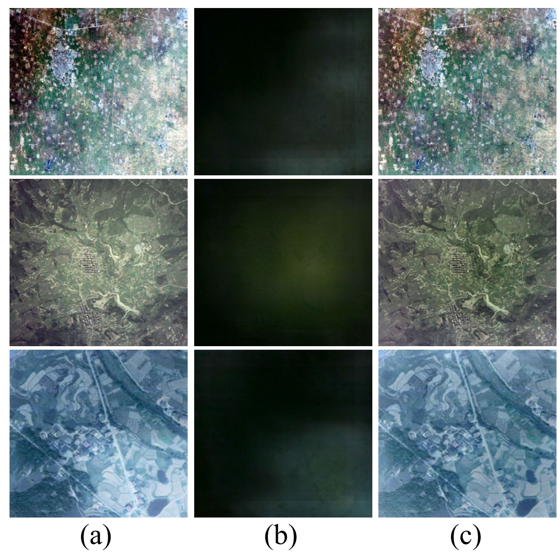


**FIGURE 9.** Results of remoting image. (a) input image, (b) light layer, (c) ideal image.

In addition, we tested the adaptability of our algorithm in remote sensing images. It can be seen from Figure 9 that the non-uniform illumination can be corrected significantly, and the prediction of the light layer is still accurate. Experiments on remote sensing images further demonstrate the robustness of the proposed algorithm.

### C. OBJECTIVE ASSESSMENT

#### 1) ALL-REFERENCE METRICS

Peak Signal to Noise Ratio (PSNR) [47] and Structural Similarity Image Metric (SSIM) [48] are all-reference objective evaluation system. Because of their simple calculation and mature evaluation criteria, it is often used as an objective evaluation index for images.

PSNR is mainly used to measure the distortion between the processed images and the ground truth. A larger PSNR value represents less distortion, which means a more accurate

**TABLE 2.** PSNR and SSIM of experiment result.

| | | BPDHE | MSR | FEA | ALT | EFF | ROBUST | JED | MLE | Ours |
|---|---|---|---|---|---|---|---|---|---|---|
| Synthetic 1 | PSNR | 10.8976 | 11.5488 | 11.5893 | 12.5559 | 10.2640 | 12.6291 | 11.5893 | 12.3593 | **14.6024** |
| | SSIM | 0.7591 | 0.5779 | 0.8166 | 0.8557 | 0.7897 | 0.7556 | 0.8166 | 0.8032 | **0.9152** |
| Synthetic 2 | PSNR | 12.5081 | 10.8430 | 13.1517 | 12.3253 | 12.6417 | 12.4653 | 12.5931 | 11.4665 | **20.1644** |
| | SSIM | 0.7128 | 0.4697 | 0.7789 | 0.7776 | 0.7771 | 0.7097 | 0.7447 | 0.7204 | **0.9246** |
| Synthetic 3 | PSNR | 13.1038 | 11.2381 | 14.2252 | 13.0475 | 13.4250 | 12.9470 | 13.1646 | 12.1795 | **20.4375** |
| | SSIM | 0.7920 | 0.5466 | 0.8197 | 0.8141 | 0.8022 | 0.7361 | 0.7634 | 0.7552 | **0.9399** |
| Fish 1 | PSNR | 11.7573 | 10.7287 | 10.8804 | 9.5202 | 12.0720 | 9.1050 | 9.1087 | 9.9333 | **17.3206** |
| | SSIM | 0.7508 | 0.5281 | 0.7374 | 0.7469 | 0.7551 | 0.7169 | 0.7174 | 0.7253 | **0.8618** |
| Fish 2 | PSNR | 12.955 | 10.8386 | 11.7955 | 10.1396 | 13.2627 | 9.5631 | 9.5681 | 10.6604 | **23.3724** |
| | SSIM | 0.7964 | 0.5454 | 0.7698 | 0.7691 | 0.7913 | 0.7219 | 0.7224 | 0.7445 | **0.9188** |
| Calibration board 1 | PSNR | 19.3826 | 11.5078 | 18.2695 | 14.3516 | 18.9712 | 12.3757 | 18.2695 | 15.8762 | **23.6316** |
| | SSIM | 0.8554 | 0.5069 | 0.8452 | 0.8278 | 0.8562 | 0.7614 | 0.8452 | 0.8374 | **0.9053** |
| Calibration board 2 | PSNR | 14.1737 | 11.0272 | 13.1709 | 11.4198 | 15.3693 | 9.9727 | 13.1709 | 12.1655 | **20.8819** |
| | SSIM | 0.8183 | 0.5567 | 0.7923 | 0.7739 | 0.8390 | 0.7015 | 0.7923 | 0.7633 | **0.8940** |

correction of non-uniform illumination. For SSIM, it evaluates the brightness, contrast, and structure of the image, which is more in line with human visual perception. From the perspective of image synthesis, SSIM defines structural information as the structural properties of the object that do not depend on brightness and contrast and model distortion as a combination of brightness, contrast, and structure. The value range of SSIM is $[0 - 1]$, the closer the SSIM value is to 1, the better the image quality.

Table 2 shows the quantitative comparison of different algorithms on the test images. We get a large gain of PSNR and SSIM values in all test images by using our algorithm. The EFF method is not good at processing severely non-uniform illumination, the processing effect of different images is unstable, and the curve fluctuates greatly, such as the Fish1 in Figure 7. BPDHE The detection of the local maximum value is affected by the setting of the Gaussian filtering process parameters, which may distort the processing result. The fluctuation of the BPDHE algorithm is small, but the overall distortion is serious, such as the Diver in Figure 6, the PSNR and SSIM value is low. Compared with other methods, the MLE algorithm flow is simple, and the separate channel processing is more targeted to make the results accurate, but among them, the histogram stretching parameters are estimated based on the overall of each channel, and the processing ability for the local is weak. The car in Figure 6 shows that the MLE method is good for images with slightly uneven lighting, and has obvious effects for overexposure of areas, but not for dark areas. The ROBUST and JED algorithms perform well in contrast, but the SSIM value is lower due to brightness distortion. In short, the experimental results show that the proposed method is more prominent in PSNR and SSIM than other methods, with low distortion and conforms to the human visual system observation standard.

### 2) NO-REFERENCE METRICS

CAF and UIQM are non-reference image quality evaluation indicators. They all contain sub-indicators in different aspects to comprehensively evaluate the image quality. We use CAF and UIQM as objective evaluation indicators to compare the performance of the proposed algorithm with other traditional methods.

CAF is an image quality assessment metric with sub-indicators. Its sub-indicators include average contrast (AC), average information entropy (AIE), and average luminance (AL).

$$CAF = AIE^{\alpha} \times AC^{\beta} \times NNF^{\gamma}, \tag{17}$$

here, $\alpha = 1$, $\beta = 1/4$, and $\gamma = 3$. The calculation formula of $NNF$ is as

$$NNF = \frac{OL - dist\,(AL, OL)}{OL} \tag{18}$$

where OL=127.5, and dist refers to the distance between the two in parentheses.

UIQM is an evaluation index for the comprehensive quality of underwater images, which includes three kinds of underwater image attribute measures: underwater image chromaticity measure (UICM), underwater image clarity measure (UISM) and underwater image contrast measure (UIConM). Among them, UICM is related to chroma, UISM is related to sharpness, and can measure the attributes of underwater images to maintain fine details and edge-related, UIConM is a measure of the contrast of underwater images. The higher the score, the better the image quality.

Table 3 shows that the proposed method stands out on UIQM, which shows that the image processed by the proposed method has a good overall quality, especially for underwater images. But it did not perform well on CAF, ranking in the top three. The main reason is that the CAF sub-indicator contains the average luminance (AL) which has a large weight as expressed in Equation (17). The brightness of the proposed method results is almost the same as the original image, but the brightness of the original image is not necessarily the best. Therefore, the proposed method may perform worse on CAF than other methods. But we think that our goal is to adjust the unevenness of the lighting. If the

**TABLE 3. CAF and UIQM of experimental results.**

| | | BPDHE | FEA | ALT | EFF | ROBUST | JED | MLE | Ours |
|---|---|---|---|---|---|---|---|---|---|
| Car | CAF | 1.1377 | 1.9590 | 1.9979 | 1.1432 | 2.0877 | **3.4024** | 1.9618 | 2.2246 |
| | UIQM | 1.6124 | 1.4533 | 0.4752 | 1.5861 | 0.7556 | 0.9687 | 1.0824 | **1.6193** |
| Coral reef | CAF | 1.8116 | 2.8959 | 3.3025 | 1.5169 | 3.1373 | **6.7417** | 1.9832 | 3.5323 |
| | UIQM | 3.2621 | 2.9810 | 1.9659 | 3.1285 | -0.0464 | 2.3471 | 2.0448 | **3.5348** |
| Fish group | CAF | 1.2795 | 1.7536 | 1.9801 | 1.4502 | 1.9532 | **3.0879** | 1.7056 | 1.9925 |
| | UIQM | 0.7172 | 0.5457 | -0.5491 | 0.6876 | 0.0730 | -0.9641 | -0.3328 | **1.6466** |
| Diver | CAF | 1.3028 | 4.0026 | **11.0003** | 4.5233 | 4.4680 | 9.0730 | 4.3392 | 4.9669 |
| | UIQM | 0.9741 | 3.1173 | 3.1285 | 3.3101 | 2.3329 | 2.8790 | 3.1721 | **3.3766** |

lighting is even, it is easy to adjust the overall brightness of the image. Moreover, keeping the brightness of the original image at all times is also conducive to subsequent image processing, such as background light estimation. As can be seen from the car in Figure 6 and Figure 7, the JED results method performs well in terms of brightness, which is significantly different from the proposed method. Therefore, Table 3 shows that the JED method is superior to the proposed method on CAF, but our method is excellent on UIQM. The BPDHE method results in distortion on individual images, such as the Diver in Figure 6, UIQM performs the worst. As can be seen in the lower part of the Diver in Figure 6, although the MLE results are not much different from the original image, its details are not clear, the sharpness and contrast are not ideal, so the CAF and UIQM scores are not dominant. The main goal of the EFF algorithm is to enhance the contrast of low-illuminance images, so it is dazzling in contrast and obtains a high score in the evaluation indicator, but it does not perform well for the correction of severe uneven illumination. In our experiment, the details of the EFF method results are relatively clear, and the color is also very close to the original image, so the EFF method gets high scores on CAF and UIQM, but the correction of uneven lighting is not obvious enough, such as the Fish1 in Figure 7.

In summary, compared with other selected comparison methods, the proposed method in this paper is more robust and achieves specific high scores on various quality indicators. It means that our method is indeed capable of correcting the non-uniform illumination of underwater images and high-quality images are obtained.

## VI. CONCLUSION AND FUTURE WORK

In this paper, we propose an end-to-end fully convolutional network to correct the non-uniform illumination of the image. The image with non-uniform illumination is regarded as an illumination additive model. When the image with non-uniform illumination was input into the proposed network, the ideal image and the light layer will be separated. In this process, we use the dilated convolution to increase the receptive field, which makes the light layer prediction more accurate. In view of the correction of non-uniform illumination, we design a novel loss function, which shows an excellent performance in the smoothness of the illumination and the authenticity of the image. Furthermore, due to the lack of publicly available datasets of images with uneven underwater illumination, we built a dataset to train and test for all algorithms. Experimental results in subject performance and objective evaluation indicators show the effectiveness and adaptability of our algorithm when compared with traditional methods. This confirmed that our network is not only suitable for correcting the illumination of underwater images with point light sources, but also has a strong ability to predict remote sensing images.

Our future work is to improve the real-time performance of our network, which can be applied to the embedded platform with limited computing resources. In addition, we believe that we can make full use of the Generative Adversarial Network to recover the dark corners of underwater images, which allow more improvements and applications.

## REFERENCES

[1] R. H. Stavn, "Lambert-Beer law in ocean waters: Optical properties of water and of dissolved/suspended material, optical energy budgets," *Appl. Opt.*, vol. 27, no. 2, pp. 222–231, 1988.

[2] H. Lu, Y. Li, T. Uemura, H. Kim, and S. Serikawa, "Low illumination underwater light field images reconstruction using deep convolutional neural networks," *Future Gener. Comput. Syst.*, vol. 82, pp. 142–148, May 2018.

[3] R. Schettini and S. Corchs, "Underwater image processing: State of the art of restoration and image enhancement methods," *EURASIP J. Adv. Signal Process.*, vol. 2010, no. 1, pp. 1–14, Dec. 2010.

[4] E. H. Land and J. J. McCann, "Lightness and retinex theory," *J. Opt. Soc. Amer.*, vol. 61, no. 1, pp. 1–11, 1971.

[5] C.-N. Fan and F.-Y. Zhang, "Homomorphic filtering based illumination normalization method for face recognition," *Pattern Recognit. Lett.*, vol. 32, no. 10, pp. 1468–1479, Jul. 2011.

[6] M. Wang, J. Pan, S. Chen, and H. Li, "A method of removing the uneven illumination phenomenon for optical remote sensing image," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, vol. 5, Jul. 2005, pp. 3243–3246.

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

[8] S. Nah, T. H. Kim, and K. M. Lee, "Deep multi-scale convolutional neural network for dynamic scene deblurring," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 3883–3891.

[9] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3431–3440.

[10] X. Lan, H. Shen, L. Zhang, and Q. Yuan, "A spatially adaptive retinex variational model for the uneven intensity correction of remote sensing images," *Signal Process.*, vol. 101, pp. 19–34, Aug. 2014.

[11] K. Smith, Y. Li, F. Piccinini, G. Csucs, C. Balazs, A. Bevilacqua, and P. Horvath, "CIDRE: An illumination-correction method for optical microscopy," *Nature Methods*, vol. 12, no. 5, p. 404, 2015.

[12] Z. Zhang and S. Zou, "An improved algorithm of mask image dodging for aerial image," *Proc. SPIE*, vol. 8006, Nov. 2011, Art. no. 80060S.

[13] G. Meng, S. Xiang, N. Zheng, and C. Pan, "Nonparametric illumination correction for scanned document images via convex hulls," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 7, pp. 1730–1743, Jul. 2013.

[14] Y. Zheng, S. Lin, S. B. Kang, R. Xiao, J. C. Gee, and C. Kambhamettu, "Single-image vignetting correction from gradient distribution symmetries," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 6, pp. 1480–1494, Jun. 2013.

[15] L. Lopez-Fuentes, G. Oliver, and S. Massanet, "Revisiting image vignetting correction by constrained minimization of log-intensity entropy," in *Proc. Int. Work-Conf. Artif. Neural Netw.*, 2015, pp. 450–463.

[16] H. Ahn, B. Keum, D. Kim, and H. S. Lee, "Adaptive local tone mapping based on retinex for high dynamic range images," in *Proc. IEEE Int. Conf. Consum. Electron. (ICCE)*, Jan. 2013, pp. 153–156.

[17] Y. Wang, H. Wang, C. Yin, and M. Dai, "Biologically inspired image enhancement based on retinex," *Neurocomputing*, vol. 177, pp. 373–384, Feb. 2016.

[18] D. J. Jobson, Z. Rahman, and G. A. Woodell, "Properties and performance of a center/surround retinex," *IEEE Trans. Image Process.*, vol. 6, no. 3, pp. 451–462, Mar. 1997.

[19] Z.-U. Rahman, D. J. Jobson, and G. A. Woodell, "Multi-scale retinex for color image enhancement," in *Proc. 3rd IEEE Int. Conf. Image Process.*, vol. 3, Sep. 1996, pp. 1003–1006.

[20] Z.-U. Rahman, D. J. Jobson, and G. A. Woodell, "Retinex processing for automatic image enhancement," *J. Electron. Imag.*, vol. 13, no. 1, pp. 100–111, 2004.

[21] S. Wang, J. Zheng, H.-M. Hu, and B. Li, "Naturalness preserved enhancement algorithm for non-uniform illumination images," *IEEE Trans. Image Process.*, vol. 22, no. 9, pp. 3538–3548, Sep. 2013.

[22] M.-J. Seow and V. K. Asari, "Ratio rule and homomorphic filter for enhancement of digital colour image," *Neurocomputing*, vol. 69, nos. 7–9, pp. 954–958, Mar. 2006.

[23] L. Y. Zou, Z. Y. Qin, W. Shang, and J.-R. Wang, "Image contrast dodging algorithm of maintaining consistency," *J. Geomatics Sci. Technol*, vol. 1, p. 13, Jan. 2011.

[24] G. Zhang, Q. Chen, and Q. Sun, "Illumination normalization among multiple remote-sensing images," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 9, pp. 1470–1474, Sep. 2014.

[25] J. Tian, X. Li, F. Duan, J. Wang, and Y. Ou, "An efficient seam elimination method for UAV images based on Wallis dodging and Gaussian distance weight enhancement," *Sensors*, vol. 16, no. 5, p. 662, May 2016.

[26] F. Yao, H. Hu, and Y. Wan, "Research on the improved image dodging algorithm based on Mask technique," in *Proc. ISPAR*, vol. 39, 2012, pp. 519–524.

[27] B. L. McGlamery, "A computer model for underwater camera systems," *Proc. SPIE*, vol. 208, pp. 221–231, Mar. 1980.

[28] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet large scale visual recognition challenge," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, Dec. 2015.

[29] B. Hariharan, P. Arbelaez, R. Girshick, and J. Malik, "Hypercolumns for object segmentation and fine-grained localization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 447–456.

[30] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Apr. 2018.

[31] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," 2015, *arXiv:1502.03167*. [Online]. Available: http://arxiv.org/abs/1502.03167

[32] J. T. Barron and J. Malik, "Shape, illumination, and reflectance from shading," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 8, pp. 1670–1687, Aug. 2015.

[33] Y. Li and M. S. Brown, "Single image layer separation using relative smoothness," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 2752–2759.

[34] C. Rother, M. Kiefel, L. Zhang, B. Schölkopf, and P. V. Gehler, "Recovering intrinsic images with a global sparsity prior on reflectance," in *Proc. Adv. Neural Inf. Process. Syst.*, 2011, pp. 765–773.

[35] B. Dai, S. Fidler, R. Urtasun, and D. Lin, "Towards diverse and natural image descriptions via a conditional GAN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2970–2979.

[36] J. Xiao, J. Hays, K. A. Ehinger, A. Oliva, and A. Torralba, "SUN database: Large-scale scene recognition from abbey to zoo," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 3485–3492.

[37] R. B. Fisher, Y.-H. Chen-Burger, D. Giordano, L. Hardman, and F.-P. Lin, *Fish4Knowledge: Collecting and Analyzing Massive Coral Reef Fish Video Data*. Berlin, Germany: Springer, 2016.

[38] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*. [Online]. Available: http://arxiv.org/abs/1412.6980

[39] M. C. Mukkamala and M. Hein, "Variants of RMSProp and Adagrad with logarithmic regret bounds," in *Proc. 34th Int. Conf. Mach. Learn.*, vol. 70, 2017, pp. 2545–2553.

[40] M. Abadi *et al.*, "TensorFlow: A system for large-scale machine learning," in *Proc. 12th Symp. Oper. Syst. Design Implement.*, 2016, pp. 265–283.

[41] H. Ibrahim and N. P. Kong, "Brightness preserving dynamic histogram equalization for image contrast enhancement," *IEEE Trans. Consum. Electron.*, vol. 53, no. 4, pp. 1752–1758, Nov. 2007.

[42] X. Dong, G. Wang, Y. Pang, W. Li, J. Wen, W. Meng, and Y. Lu, "Fast efficient algorithm for enhancement of low lighting video," in *Proc. IEEE Int. Conf. Multimedia Expo*, Jul. 2011, pp. 1–6.

[43] Z. Ying, G. Li, Y. Ren, R. Wang, and W. Wang, "A new image contrast enhancement algorithm using exposure fusion framework," in *Proc. Int. Conf. Comput. Anal. Images Patterns*, 2017, pp. 36–46.

[44] M. Li, J. Liu, W. Yang, X. Sun, and Z. Guo, "Structure-revealing low-light image enhancement via robust retinex model," *IEEE Trans. Image Process.*, vol. 27, no. 6, pp. 2828–2841, Jun. 2018.

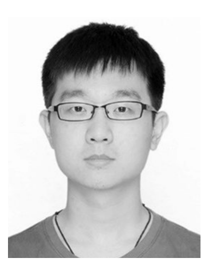
[45] X. Ren, M. Li, W.-H. Cheng, and J. Liu, "Joint enhancement and denoising method via sequential decomposition," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, May 2018, pp. 1–5.

[46] S. S. Sankpal and S. S. Deshpande, "Nonuniform illumination correction algorithm for underwater images using maximum likelihood estimation method," *J. Eng.*, vol. 2016, pp. 1–9, Jan. 2016.

[47] Q. Huynh-Thu and M. Ghanbari, "Scope of validity of PSNR in image/video quality assessment," *Electron. Lett.*, vol. 44, no. 13, pp. 800–801, Jun. 2008.

[48] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.


**XUETING CAO** was born in Ulanqab, Inner Mongolia, China, in 1996. She received the B.S. degree in electronic information science and technology from Inner Mongolia University, Inner Mongolia, in 2018. She is currently pursuing the master's degree in signal and information processing with the Ocean University of China, Qingdao, China. Her research interests include underwater image processing and underwater target recognition.



**SHENGHUI RONG** (Member, IEEE) was born in Rizhao, Shandong, China, in 1989. He received the B.S. degree in electronic science and technology and the Ph.D. degree in physical electronics from Xidian University, Xi'an, China, in 2011 and 2018, respectively. In 2016, he was funded by the China Scholarship Council (CSC) to conduct his research in the direction of 3-D image processing and recognition at Griffith University, Australia. He is currently a Lecturer with the School of Information Science and Engineering, Ocean University of China (OUC). His primary research interests include optoelectronic countermeasures, computer vision, and pattern recognition.

**YONGBIN LIU** was born in Sanmenxia, Henan, China, in 1995. He received the B.S. degree in communication engineering from the Ocean University of China (OUC), Qingdao, China, in 2017, where he is currently pursuing the master's degree in communication and information systems. His research interests include image processing and underwater object recognition.

**TENGYUE LI** was born in Chaoyang, Liaoning, China, in 1990. He received the B.S. degree in electronic information science and technology and the M.S. degree in optical engineering from the Ocean University of China (OUC), Qingdao, China, in 2013 and 2015, respectively, where he is currently pursuing the Ph.D. degree in smart information and communication system. He has published two conference papers and has applied one Chinese invention patent and one Chinese utility model patent. His research interests include image processing, underwater 3-D reconstruction, underwater object recognition, and underwater robotics technology.

**QI WANG** was born in Qingdao, Shandong, China, in 1995. She received the B.S. degree from Shandong Agricultural University, in 2017. She is currently pursuing the M.S. degree in electronic and communication engineering with the Ocean University of China (OUC), Qingdao. Her current research interests include image processing, sonar image segmentation, and deep learning.

**BO HE** (Member, IEEE) was born in Qingdao, Shandong, China, in 1971. He received the M.S. and Ph.D. degrees from the Harbin Institute of Technology, China, in 1996 and 1999, respectively. From 2000 to 2003, he worked with Nanyang Technological University, Singapore, as a Postdoctoral Fellow. His research work focused on the precise navigation and control and communication for the platform of mobile robots and unmanned vehicles. In 2004, he joined the Ocean University of China (OUC), where he is currently a Full Professor and the Deputy Head of the Department of Electronic Engineering, School of Information Science and Engineering. His current research interests include AUV design and applications, AUV SLAM, AUV control, and machine learning.