



# Full scene underwater imaging with polarization and an untrained network

YANMIN ZHU,<sup>1,3</sup>  TIANJIAO ZENG,<sup>1,3</sup>  KEWEI LIU,<sup>2</sup> ZHENBO REN,<sup>2</sup>  AND EDMUND Y. LAM<sup>1,\*</sup> 

<sup>1</sup>*Department of Electrical and Electronic Engineering, The University of Hong Kong, Pokfulam, Hong Kong SAR, China*

<sup>2</sup>*MOE Key Laboratory of Material Physics and Chemistry under Extraordinary Conditions, and Shaanxi Key Laboratory of Optical Information Technology, School of Physical Science and Technology, Northwestern Polytechnical University, Xi'an, China*

<sup>3</sup>*These authors contributed equally to this work*

\**elam@eee.hku.hk*

**Abstract:** The veiling effect caused by the scattering and absorption of suspending particles is a critical challenge of underwater imaging. It is possible to combine the image formation model (IFM) with the optical polarization characteristics underwater to effectively remove the veiling effect and recover a clear image. The performance of such methods, to a great extent, depends on the settings of the global parameters in the application scenarios. Meanwhile, learning-based methods can fit the underwater image information degradation process nonlinearly to restore the images from scattering. Here, we propose for the first time a method for full scene underwater imaging that synergistically makes use of an untrained network and polarization imaging. By mounting a Stokes mask polarizer on the CMOS camera, we can simultaneously obtain images with different polarization states for IFM calculation and optimize the imaging automatically by an untrained network without requiring extra training data. This method makes full use of the nonlinear fitting ability of a neural network and corrects the undesirable imaging effect caused by imperfect parameter settings of the classical IFM in different scenes. It shows good performance in removing the impact of water scattering and preserving the object information, making it possible to achieve clear full scene underwater imaging.

© 2021 Optica Publishing Group under the terms of the [Optica Open Access Publishing Agreement](#)

## 1. Introduction

It is well understood that in underwater imaging, the presence of complex biological and abiotic particles, such as soil particles and the floating excrement of marine animals, can change the direction of the light path in a complex manner [1,2]. The background light scattered in the water will be superimposed with the imaging light path, reducing the contrast and resolution of the imaging.

At present, to address the problem of underwater image quality reduction due to scattering, the main approaches can be divided into two categories. The first one includes the physical model-based image restoration methods [3,4], such as inverse imaging with the underwater image formation model (IFM) [5] prior, maximum intensity prior (MIP) [6], and dark channel prior (DCP) [7]. The second one uses digital image restoration and enhancement technologies [8,9], such as gamma correction (GC) [10] and histogram enhancement (HE) [11]. The former tackles the problem by establishing the light scattering model in the process of underwater propagation. However, such physical models are mostly based on the assumption of a single or special environment, which requires setting some prior information and cannot fully simulate the changeable underwater environment realistically. For example, the degree of polarization of background scattered light is assumed constant, the degree of polarization of target information light is often ignored, and the selective scattering and absorption of water body to different

wavelengths are not taken into account. These assumptions about the environment greatly limit the application scenarios of the model. At the same time, how the hypothesis matches the real environment also determines the quality of the descattering. The digital image restoration method usually improves the visual effect of the image independent of the physical model. However, many statistical and learning-based algorithms depend on the quality of the dataset. Without considering the essential reasons for the reduction in image quality and physical principle, the extent of the enhancement of image details is limited.

Polarization-based descattering methods analyze the difference in the polarization characteristics between the object and the background light [12,13] and estimate their irradiance changes. Such methods can effectively remove the background scattered light and further carry out target recognition and 3D reconstruction [14,15]. The ability to restore multi-dimensional information, such as contrast, information details, and color, based only on polarization information makes it a promising research direction of underwater imaging technology. However, current methods require manually rotating the polarizer in front of the camera [13,16] to obtain multiple images with different polarization states for the model calculation. Moreover, the descattering effect is sensitive to the value of the parameters used to characterize the polarization state of the background light [1], which requires manual parameters tuning.

Deep learning networks, such as convolutional neural networks (CNNs), extract the features of target objects based on the massive learning from datasets, and conduct further processing such as classification and reconstruction of newly input data through conversion between low-level features and high-level features [17]. In recent years, untrained networks have been applied to computational imaging problems, such as phase retrieval [18], phase image reconstruction [19] and automatic aberration compensation [20] by using prior knowledge rather than large datasets to guide network convergence. However, as far as we know, this is the first time to deeply explore the method of untrained network for underwater imaging. In addition, the polarization state, as a unique feature different from light irradiance information, broadens the feature space dimension available for extraction in a deep learning network. This is conducive to the network's extraction of the polarization difference between the object information light and the background light, so as to perform underwater image restoration.

In this study, we present a polarimetric imaging method with an untrained network to recover images from scattering in all underwater scenarios. Without requiring additional datasets for network training, this IFM-guided network can identify the structure of the water body sensitively, based only on the input scattered images, and remove the scattering effects of ripples. This process can not only overcome the limitations caused by the assumptions on the polarization characteristics of the environment and objects, but also minimize the mismatch between the model and the real scene when the environmental priors are unknown. In addition, the input images with different polarization states used for the network model calculation are obtained simultaneously by a polarization camera mounted with specially designed Stokes mask polarizer (SMP), which eliminates the need to manually adjust the polarizer.

This paper is organized as follows. In Sec. 2, we introduce the underwater IFM, the polarization-based underwater clear imaging method, the IFM-based untrained network, and the polarization imaging system with SMP. In Sec. 3, we demonstrate the experimental results in three different scenarios, namely, homogeneous background where the object is in camera's partial field of view, composite water and glass background, where the object is in camera's partial and full field of view, respectively. We also explore the performance of our method with different loss functions and show the results at the end of this section. Finally, we discuss the limitations, advantages, and application directions of our work in Sec. 4.

## 2. Methods

### 2.1. Underwater image formation model

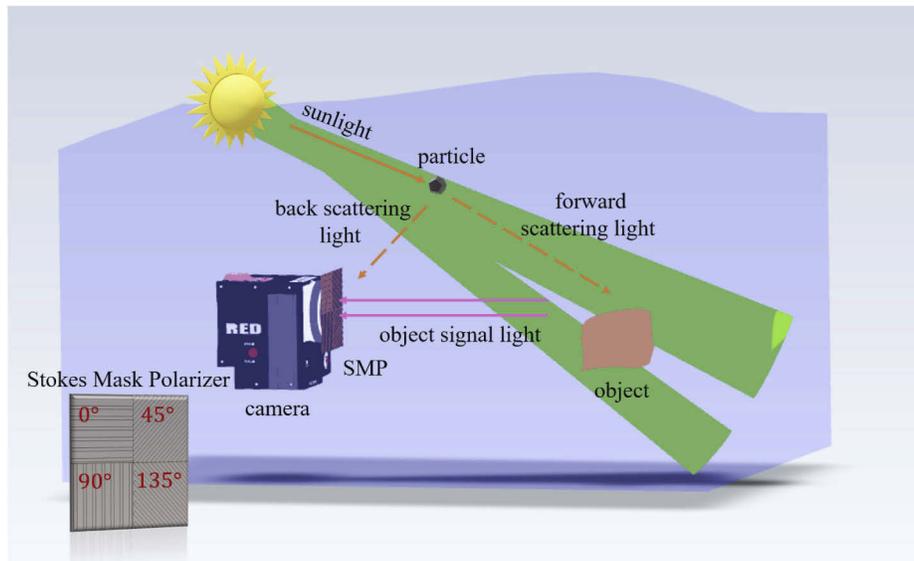
We describe the underwater image formation in Fig. 1 and elaborate on the image descattering method. In underwater IFM, the natural light comes from the converging rays within a limited cone above the water surface, which is caused by total internal reflection [21]. The light incident upon the photosensitive surface of the polarization camera is not only the object light carrying the sample information, but also the stray light generated by water scattering. We use  $S(x, y)$  to represent the irradiance of the scattering light, which can be expressed as

$$S(x, y) = [1 - T(x, y)] \cdot S_{\infty}, \quad (1)$$

where  $S_{\infty}$  denotes the irradiance of the scattering light at an infinite distance, and  $T(x, y)$  is the transmittance of the scattering light. This transmittance will decay exponentially with the rate governed by the extinction coefficient  $\beta$ , given by

$$T(x, y) = \exp[-\beta \cdot D(x, y)], \quad (2)$$

where  $D(x, y)$  is the distance between the object and the camera. This model is based on the single illumination source assumption [5], where at any given moment, the illumination of a scattering particle is coming from only one direction.



**Fig. 1.** Schematic diagram of underwater image formation model (IFM) and detailed structure of Stokes mask polarizer (SMP). In IFM, the natural light at one point underwater comes from the converging light of a limited cone above the water surface and is scattered by the particles in the seawater. Part of the scattering light goes forward to the object, while another part of the back scattering light is recorded by the camera. During the underwater imaging process, the object signal light is transmitted to the photosensitive surface of the camera. A Stokes mask polarizer is mounted in front of the camera lens to obtain images with different polarization states simultaneously.

The object signal light irradiance  $O(x, y)$  that reaches the photosensitive surface of the camera is denoted as  $C(x, y)$ . Similar to the above, it will decay exponentially with distance, such that

$$\begin{aligned} C(x, y) &= \alpha O(x, y) \cdot \exp[-\beta \cdot D(x, y)] \\ &= \alpha O(x, y) \cdot T(x, y), \end{aligned} \quad (3)$$

where  $\alpha$  is a compensation factor for the water absorption.

The total irradiance is given by

$$I(x, y) = C(x, y) + S(x, y). \quad (4)$$

Combining the equations above, the object signal light can be derived as

$$O(x, y) = \frac{I(x, y) - S(x, y)}{\alpha(1 - S(x, y)/S_\infty)} \triangleq \mathcal{M}^{-1} \{I(x, y)\}. \quad (5)$$

We define  $\mathcal{M}\{\cdot\}$  as the mapping from the object to the scattered image recorded by the camera, and therefore the above equation is its inverse process. Since  $I(x, y)$  can be measured by the photosensitive surface of the camera, a clear object  $O(x, y)$  can be computed provided that the values of the scattered light irradiance  $S(x, y)$ , the global parameter  $\alpha$ , and the scattered light radiation at an infinite distance  $S_\infty$  are accurately estimated. Furthermore, we would like to point out that the global parameters  $\alpha$  and  $S_\infty$  are constant for the whole image, which can be estimated from the brightest region of the sky [5]. Therefore, as will be explained next, the key step for the descattering is to estimate  $S(x, y)$  by means of polarization.

## 2.2. Polarization method for clear underwater vision

In principle, both the scattering light and the object signal light are partially polarized. However, as distance  $D(x, y)$  gets larger, according to (1) and (3), the object signal light attenuates in the process of transmission to the camera, while the scattering light increases. Therefore, the latter is dominant in the measured polarization [5]. If we use the Stokes vectors to describe the polarization state of the incident light, then the relationships between the Stokes vectors  $(S_0, S_1, S_2)$  and the light irradiance at the polarization directions of  $0^\circ$ ,  $45^\circ$  and  $90^\circ$  ( $I_{0^\circ}$ ,  $I_{45^\circ}$ ,  $I_{90^\circ}$ ) are given by [22]

$$S_0(x, y) = I_{0^\circ}(x, y) + I_{90^\circ}(x, y), \quad (6)$$

$$S_1(x, y) = I_{0^\circ}(x, y) - I_{90^\circ}(x, y), \quad (7)$$

$$S_2(x, y) = 2I_{45^\circ}(x, y) - S_0(x, y) \quad (8)$$

where  $S_0(x, y)$  represents the total light irradiance and has the same value as  $I(x, y)$ . Also,  $S_1(x, y)$  and  $S_2(x, y)$  represent the irradiance difference between  $0^\circ$  and  $90^\circ$ , and between  $45^\circ$  and  $135^\circ$  polarization directions, respectively. Here, we define two parameters to describe the state of polarization, namely, the polarization angle  $\theta$  and the degree of the linear polarization ( $p^\circ$ ) over the whole image. They are given by

$$\theta(x, y) = \frac{1}{2} \arctan \frac{S_2(x, y)}{S_1(x, y)}, \quad (9)$$

$$p^\circ(x, y) = \frac{\sqrt{[S_1(x, y)]^2 + [S_2(x, y)]^2}}{S_0(x, y)}. \quad (10)$$

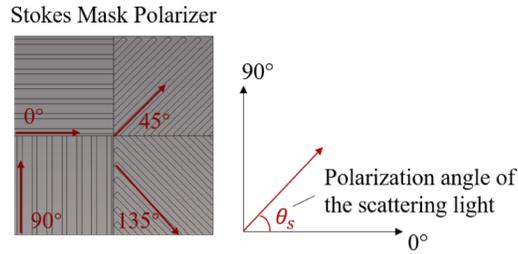
The angle and degree of the scattering light are denoted as  $\theta_S$  and  $p_S^\circ$ , respectively, and are calculated by the mean irradiance values of the region without the object, which can be identified by the sky region identification algorithm [23]. For an image captured by the camera at a

certain moment, the polarization directions of  $0^\circ$  and  $90^\circ$  are parallel to the x-axis and y-axis, respectively, as shown in Fig. 2. The irradiance of the polarized scattering light in the direction of  $\theta_S$ , which is denoted as  $I^{\parallel}(x, y)$ , is given by

$$I^{\parallel}(x, y) = \frac{I_{0^\circ}(x, y)}{\cos^2 \theta_S}. \quad (11)$$

The light irradiance, which is perpendicular to the direction of  $\theta_S$ , can be calculated as

$$I^\perp(x, y) = S_0(x, y) - I^{\parallel}(x, y). \quad (12)$$



**Fig. 2.** The polarization coordinates in the captured image. The  $0^\circ$  and  $90^\circ$  polarization directions are parallel to x- and y-axes, respectively.  $\theta_S$  is the polarization angle of the scattering light.

We further assume that the atmosphere is homogeneous for the whole image. Therefore,  $I^{\parallel}(x, y)$  contains all the polarized scattering light, half of the unpolarized scattering light, and half of the object signal light, while  $I^\perp(x, y)$  only contains half of the unpolarized scattering light and half of the object signal light [24]. The irradiance of the polarized scattering light  $S_p(x, y)$  can be estimated by the difference between  $I^{\parallel}(x, y)$  and  $I^\perp(x, y)$ , i.e.,

$$S_p(x, y) = \frac{2I_{0^\circ}(x, y)}{\cos^2 \theta_S} - S_0(x, y). \quad (13)$$

The total scattering light irradiance  $S(x, y)$  can be expressed as

$$S(x, y) = \frac{S_p(x, y)}{\eta \cdot p_S^\circ}, \quad (14)$$

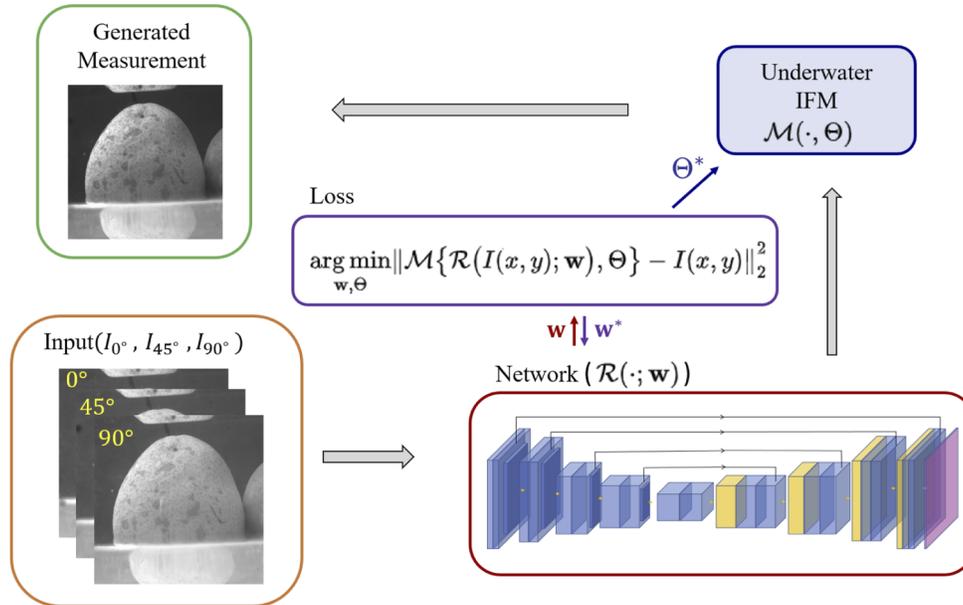
where  $\eta$  is a bias factor to adjust the relationship between  $S(x, y)$  and the first item of the Stokes parameter  $S_0(x, y)$ , and guarantees that  $S(x, y)$  is always smaller than  $S_0(x, y)$ . The range of  $\eta$  is from 1 to  $1/p_S^\circ$  [5]. By substituting (6) and (14) into (5), the object signal  $O(x, y)$  can be precisely estimated.

As  $I(x, y)$ ,  $S_0(x, y)$  and  $S_p(x, y)$  are the variables related to  $I_{0^\circ}(x, y)$ ,  $I_{45^\circ}(x, y)$  and  $I_{90^\circ}(x, y)$ , we can obtain  $O(x, y)$  when these irradiance values are known and the global parameters such as  $\alpha$ ,  $\eta$ ,  $p_S^\circ$  and  $S_\infty$  are well set. Under normal circumstances, these global variables can be estimated by the intensities of the brightest region of the sky [5], and then manually adjusted to make it more suitable for the specific applications. However, such an approach often falls short in fitting the real situations, and the resulting errors often have a significant impact on the desattering result. In the next section, we present an alternative approach using a neural network.

### 2.3. IFM-based untrained network

The goal of our neural network is to compute the critical parameters described in the previous section and compensate for the mismatch automatically. Our design is shown in Fig. 3. The

network, denoted by  $\mathcal{R}(\cdot)$ , is based on U-Net [25], which is a convolutional network with good performance in image reconstruction due to its encoder-decoder structure and skip connection blocks. The network takes only a  $512 \times 512$  image as input, and obtains the output image through a sequence of five downsampling and then five upsampling steps. The downsampling consists of a convolution layer and a maxpooling layer, whereas the upsampling is a combination of deconvolution and convolution layers. At the same time, we make use of the batch normalization layer and the ReLU layer to normalize and activate the network, respectively.



**Fig. 3.** The workflow of our method. The images with  $0^\circ$ ,  $45^\circ$  and  $90^\circ$  polarization states taken by the camera are used as input to generate descattered object images with the network  $\mathcal{R}(\cdot; \mathbf{w})$ . The descattered image is mapped back to the estimated total irradiance image by IFM  $\mathcal{M}(\cdot, \Theta)$ . The convergence of the network is guided by minimizing the Euclidean distance between the estimated image and the input images.

The original U-Net used a dataset consisting of 30 images, where each image was then augmented sevenfold for the network training. After about 5000 iterations [25], the image reconstruction task was considered complete. The whole process required a significant amount of computational time. Inspired by [18–20], we set out to eliminate the training process. The network  $\mathcal{R}(\cdot; \mathbf{w})$  is randomly initialized with the weights  $\mathbf{w}$ . We manually set the initial values for parameters  $\{\alpha, \eta, p_S^\circ, S_\infty\} \in \Theta$  of IFM (represented by  $\mathcal{M}(\cdot)$ ) according to [5,26] as well as our experimental measurements, and then adjust them by solving the optimization

$$[\mathcal{R}(\cdot; \mathbf{w}^*), \Theta^*] = \arg \min_{\mathbf{w}, \Theta} \|\mathcal{M}\{\mathcal{R}(I(x, y); \mathbf{w}), \Theta\} - I(x, y)\|_2^2. \quad (15)$$

In this work, the initial value of  $\alpha$  was set to 0.99 as the subject was placed in a comparably shallow position with pure water in the underwater environment, where the influence on the absorption of natural light could be ignored. The measurement value of  $p_S^\circ$  is set to 0.8333 and we use  $\eta = 1.13$  in the range of  $1 \leq \eta \leq 1/p_S^\circ$ . Once the weights  $\mathbf{w}$  and parameters  $\Theta$  are optimized, the resulting untrained network  $\mathcal{R}(\cdot; \mathbf{w}^*)$  can be used to generate the descattered object  $O_{\mathbf{w}^*}(x, y)$

and map to the estimated total irradiance  $\tilde{I}(x, y)$ , given by

$$O_{\mathbf{w}^*}(x, y) = \mathcal{R}(I(x, y); \mathbf{w}^*) \quad (16)$$

and

$$\tilde{I}(x, y) = \mathcal{M}\{O_{\mathbf{w}^*}(x, y), \Theta^*\}. \quad (17)$$

To summarize, the descattering only relies on the input scattered images  $I(x, y)$  and the underwater IFM, without any requirement of additional training sets and ground truth data.

#### 2.4. Polarization imaging system with SMP

In order to obtain the required images with three different polarization states simultaneously in a single shot, a Stokes Mask Polarizer (SMP) is specially designed and mounted in front of a CMOS camera. The SMP integrates four different polarizers together, as shown in Fig. 1. From top to bottom and from left to right, as displayed in Fig. 2, the polarization states are  $0^\circ$ ,  $45^\circ$ ,  $90^\circ$  and  $135^\circ$ , respectively. According to (6–8), we only use the polarization images of  $0^\circ$ ,  $45^\circ$  and  $90^\circ$  to calculate the descattering IFM. Compared with traditional polarizing cameras, the use of SMP eliminates the need to manually rotate the polarizer during the imaging process. This automates the process and simplifies the operation.

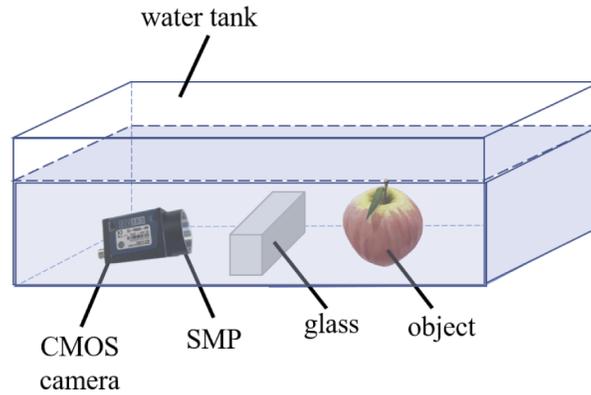
### 3. Experiments and results

The system setup is shown in Fig. 4. An object, a piece of glass, and a CMOS camera mounted with an SMP are placed in a water tank. The distance between the object and the camera is 20 cm, and the glass is placed 10 cm in front of the object. We use a CREVIS MG-A500P-22 CMOS camera with a resolution of  $2464 \times 2056$  pixels to record the degraded grayscale images. The thickness of the glass is about 2 cm. In order to better evaluate the performance of the proposed method in removing the water scattering, we have collected images under various imaging scenarios. They can be divided into the following two categories, namely, imaging with a homogeneous background (only affected by water) and imaging with a composite background (affected by water and glass at the same time). In addition, with a composite background, we further divide the scenarios into two categories based on the proportion of the object in the field of view. One is that the object is in part of the field of view, i. e., there exists a pure background area in the image. The other is that the object exists in the full field of view, i. e., there is no pure background area. In this scenario, there are no objects in the water tank, and the experimental system is placed outdoors for image recording.

To investigate the feasibility of the proposed method, we compared the resulting images in all three scenarios with other leading underwater image descattering and restoration methods, including based on maximum intensity prior (MIP) [6], histogram enhancement (HE) [11], relative global histogram stretching (RGHS) [27], dark channel prior (DCP) [7], gamma correction (GC) [10] and the method of removal of water scattering (RoWS) [28]. Since the CMOS camera can only record the light intensity, only grayscale images can be obtained and used for the experimental result evaluations. Following the method proposed in [29], we make the following adjustments on DCP such that it can estimate the transmission map without taking into account the color information and be applicable to grayscale images. In experiments based on DCP, the input image  $I(x, y)$  is first multiplied by a low-pass Gaussian filter. Then, the standard deviation  $\sigma_p$  of their product is subtracted from the filtered image. With such adjustments, the resulting input image  $I_G(x, y)$  can be denoted as

$$I_G(x, y) = G_\sigma \cdot I(x, y) - \sigma_p, \quad (18)$$

where  $G_\sigma$  is the Gaussian function of the standard deviation  $\sigma_p$ .



**Fig. 4.** Diagram for the experimental setup. The object, glass, and a CMOS camera mounted with an SMP are placed in a water tank successively. According to the different experimental scenario settings, the images are recorded by the camera either only through the water body in a homogeneous background or both through the water body and glass in a composite background. The water in the tank is uniformly purified water. The glass wall of the water tank is clean and thin. The impact of it on the imaging system can be neglected.

All the experiments are implemented using python on Intel Xeon CPU with 12 GB RAM, while the proposed learning-based method is evaluated on a Tesla K80 GPU. All images are resized to  $512 \times 512$ . The red, green, and blue channels have the same input values, which are equal to the intensity value of the grayscale image. An Adam optimizer with a 0.01 learning rate is adopted to optimize the network. The number of iteration is set to 600 to ensure that the network fully converges.

To numerically compare the different methods, we measure the contrast of the descattered images. The contrast of an image  $I(x, y)$  with a size of  $M \times N$  is calculated by

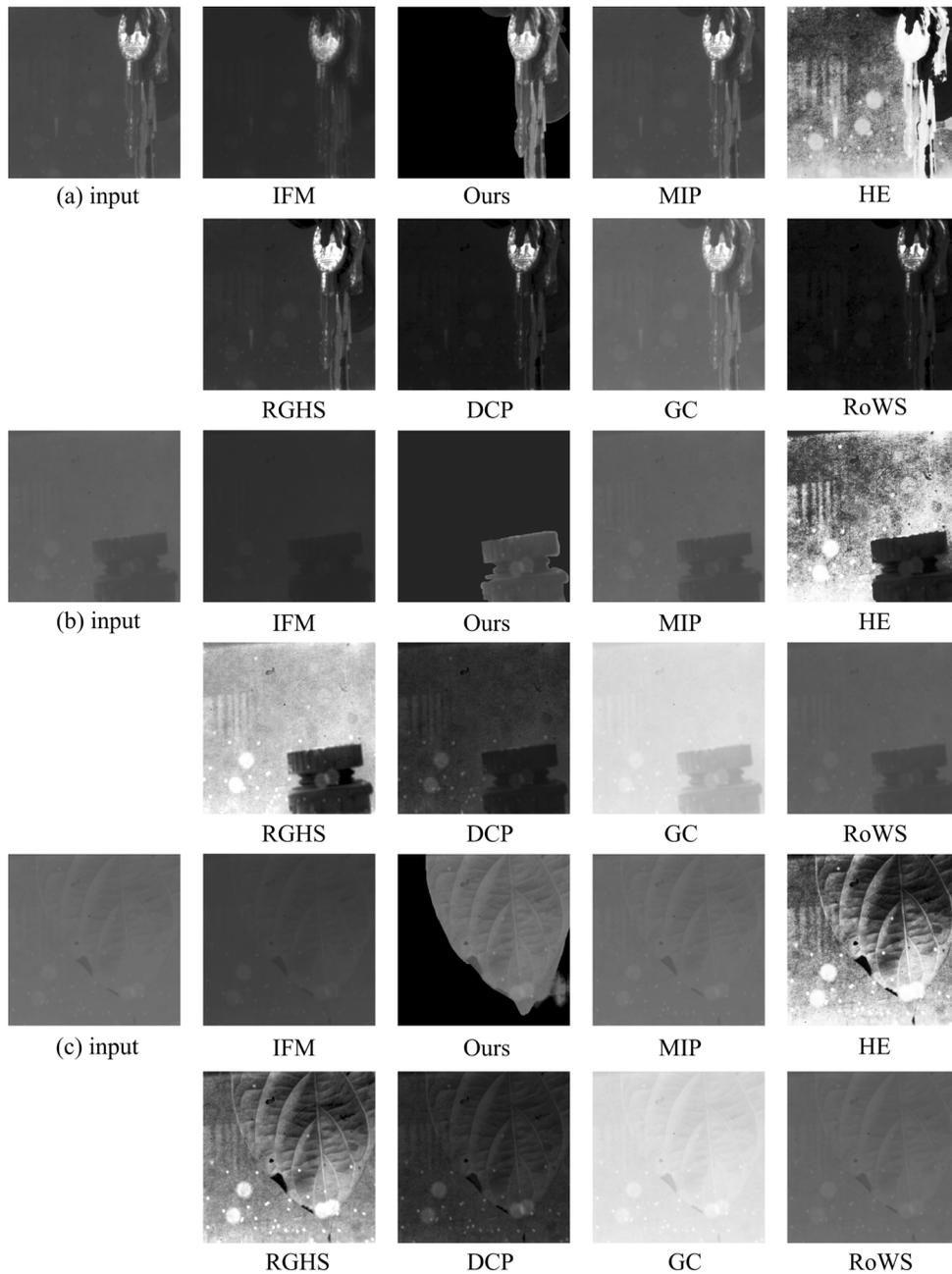
$$C = \frac{\sqrt{\frac{1}{MN} \sum_{x=0}^{N-1} \sum_{y=0}^{M-1} (I(x, y) - \bar{I})^2}}{\bar{I}}, \quad (19)$$

where  $\bar{I}$  is the average intensity of all image pixels. A higher value of C indicates a better image quality [30].

### 3.1. For a homogeneous background

Three different objects with rich details are placed in the camera's field of view for image acquisition, as shown in Fig. 5. These three sets of images have a homogeneous background, i.e., only affected by the water. The target area of interest is smaller than the camera's field of view, so there is a pure background area in the image.

We observe from Fig. 5 that, with our method, the contour of the object is detected automatically. The scattering effect of water is completely removed in the pure background area, while the details of the object are enhanced. With IFM and HE, image details are substantially lost while HE and RGHS also introduce many artifacts. In addition, GC, DCP and RoWS change the intensity of the entire image as a whole, as they cannot automatically detect the object area and the background area. MIP introduces fewer artifacts in the image. However, according to the contrast values shown in Table 1, it only gives a slight improvement on the image quality. It can be concluded that our method has a significant improvement in the image quality. Although HE may have higher image contrast values, the artifacts introduced cannot be ignored.



**Fig. 5.** Visual inspection on the descattering performance for the images with a homogeneous background using our method, IFM, MIP [6], HE [11], RGHS [27], DCP [7], GC [10] and RoWS [28].

### 3.2. For a composite background

#### 3.2.1. Object exists in the camera's partial field of view

To better evaluate the ability of our method on detecting and processing the target area, we also carry out experiments with a composite background, consisting of both water and glass. As

**Table 1. Contrast value comparisons of different methods in scenarios with a homogeneous water background. Numbers in red indicate the best performance and the blue ones give the second best performance.**

Methods	input	IFM	Ours	MIP	HE	RGHS	DCP	GC	RoWS
Fig. 5(a)	0.0909	0.0875	0.1956	0.0992	0.1286	0.1329	0.1057	0.0806	0.0983
Fig. 5(b)	0.0337	0.0173	0.0735	0.0374	0.0951	0.0464	0.0551	0.0559	0.0378
Fig. 5(c)	0.0211	0.0205	0.2141	0.0234	0.2895	0.2115	0.0482	0.0306	0.0255

shown in Fig. 6, the upper part, where the object is located in a homogeneous water background, and the lower part, with glass in front of the object, form a composite background of water and glass. A bright border line can be observed between these two areas. In this set of experiments, we select the objects with rich details as the observation objects. For example, pears with dense spots in Fig. 6(a) and Fig. 6(c), and apples with fine texture in Fig. 6(b), are used to better evaluate the ability of each image processing method to enhance the detailed information of the objects.

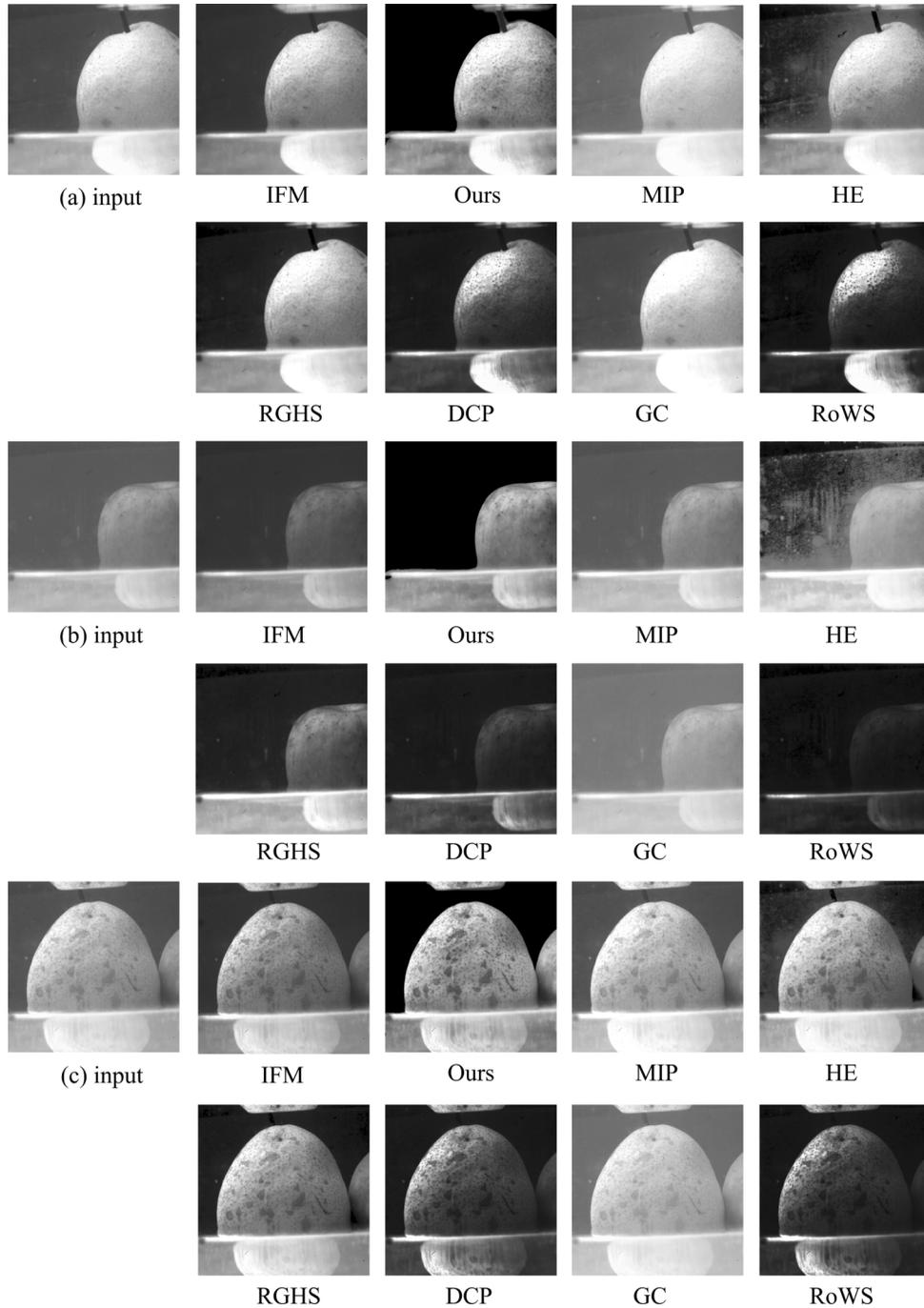
We observe from Fig. 6 that, with the processing of our method, only the background area with the homogeneous scattering effects in the upper part of the image is completely removed, while the object information is preserved. However, in the lower part of the composite background, the image effect caused by the glass has not been preserved. This reflects the specificity and effectiveness of our method for removing water scattering. By making full use of the principle of IFM, as well as the polarization information, the proposed method can precisely target the artifacts caused by water scattering and suppress it specifically, instead of simply adjusting the intensity of the entire image. In addition, the introduction of the physical prior-based untrained network further strengthens this ability of our method. We can see that the boundary between the object and the background is accurately detected. The background information is removed, while the object information is retained and enhanced through the iterative process of the network and automatic parameter adjustment. At the same time, the obvious artifacts introduced by HE are not observed in the result of our method. For other IFM-based methods such as MIP, DCP, RGHS, and RoWS, they will cause the loss of some object information details (such as RGHS), image overexposure (such as MIP), and image dark (such as DCP and RoWS). The contrast values are given in Table 2. Our method delivers good performances in all three sets of images.

**Table 2. Contrast value comparisons of different methods in scenarios with a composite background where the object covers camera's partial field of view. Numbers in red indicate the best performance and the blue ones give the second best performance.**

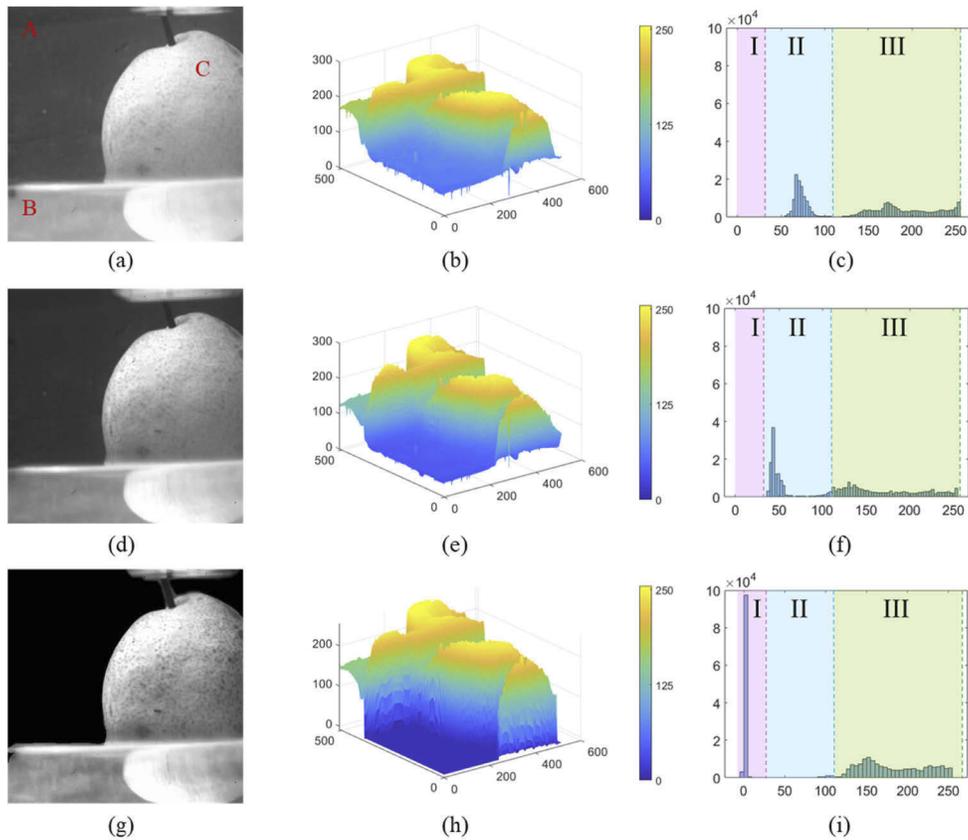
Methods	input	IFM	Ours	MIP	HE	RGHS	DCP	GC	RoWS
Fig. 6(a)	0.2571	0.2730	0.3536	0.2785	0.2863	0.3346	0.2625	0.2164	0.2521
Fig. 6(b)	0.1733	0.1929	0.3116	0.1924	0.2792	0.2951	0.1595	0.1457	0.0865
Fig. 6(c)	0.1931	0.2365	0.3160	0.2358	0.2863	0.2965	0.2171	0.1785	0.1964

In order to further analyze the descattering performance of our method, we randomly select a degraded input image and plot its 3D intensity image and histogram before and after processing. For the input image shown in Fig. 7(a), we divide it into three categories of regions depending on the content and mark them with red capital letters A, B and C. Specifically, region A is an example of a part of area with only the homogeneous water background and no object information. Region B is an example of composite background with both water and glass. Region C is the area containing object information. To better investigate the impact of our physical prior-based untrained network on the light intensity of different parts of the image, we process the degraded input image Fig. 7(a) using IFM and our approach, and show the results in Fig. 7(d) and Fig. 7(g). Furthermore, to emphasize the effectiveness of our method in preserving the original image

details, we also plot the 3D intensity maps in Fig. 7(b), (e) and (h) and the histograms in Fig. 7(c), (f) and (i).



**Fig. 6.** Visual inspection on the descattering performance for the images with a composite background where the object covers camera's partial field of view using our method, IFM, MIP [6], HE [11], RGHS [27], DCP [7], GC [10] and RoWS [28].



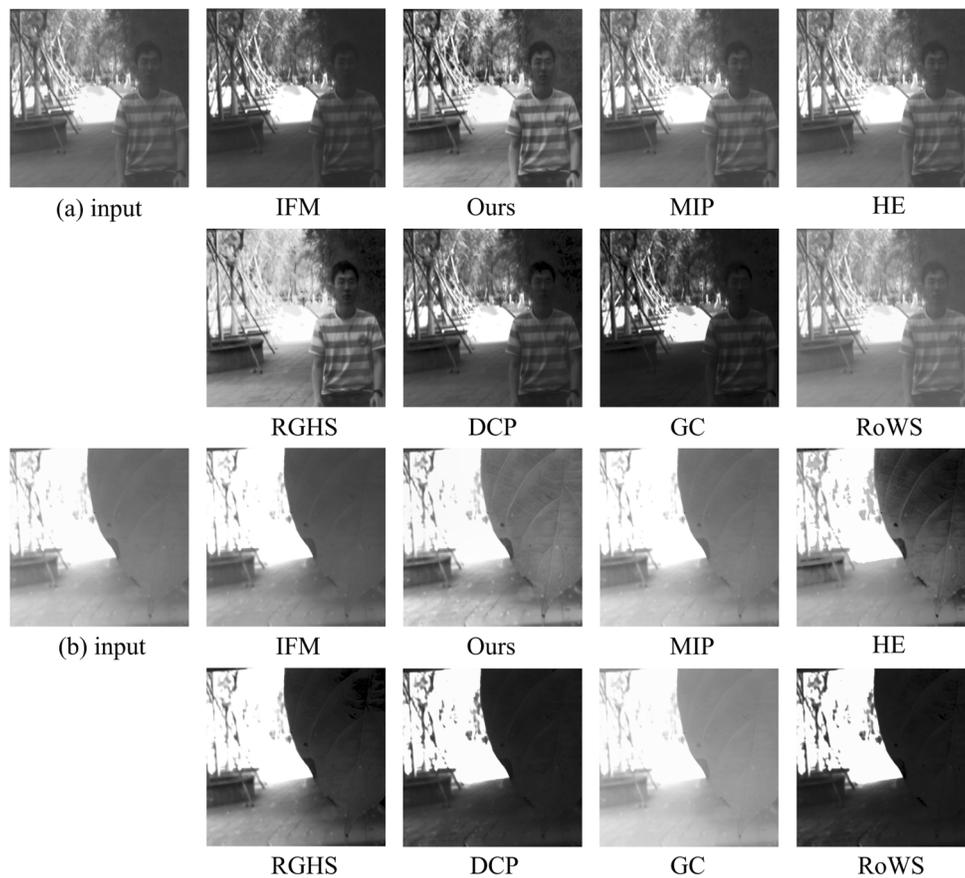
**Fig. 7.** 3D intensity images and histograms of a randomly selected degraded input image before and after the image processing of our method. (a) is the input image. The corresponding 3D intensity image and histogram are shown in (b) and (c). (d) is the image processed only with IFM. The 3D intensity image and histogram are shown in (e) and (f). (g) is the resulting image processed by our method. (h) and (i) are the corresponding 3D intensity image and histogram of (g). The red letters A, B and C in (a) refer to the area with a homogeneous background, the area with a composite background and the area containing object information. To better describe the value distribution of the histogram, we separate the region of the histogram into three parts, which are marked as I, II and III.

From the 3D intensity maps, we observe that in Fig. 7(b), there is an obvious water background with an intensity level of around 50. In (e), this background still exists at about the same level. In comparison, in (h), the intensity level significantly decreases to about 0 in the dark region, while the other areas are not affected inadvertently. In the histograms ((c), (f) and (i)), we evaluate the performance of the corresponding descattering methods by observing the distribution of pixels in regions I, II, and III. In the histogram (c) of the original image (a), due to the scattering of water, there is no completely clean background area without scattering. Thus, there are very few pixels in region I. Pixels are mainly in region II and region III. In the histogram (f), there are still only a small number of pixels in region I. In region II, the distribution of pixels has shifted more to the left. Nevertheless, compared to (i), the pixels in regions I, II, and III have not changed significantly. It can be seen that the descattering effect of the IFM is relatively limited. In contrast, in histogram (i), there are a lot more pixels in region I, with a negligible number in region II. Moreover, in region III, there are still a considerable number of pixels. Compared with

the pixel distribution in (c) and (f), we can see that our method can better preserve the object information while having a good descattering performance on the image.

### 3.2.2. Object exists in the camera's full field of view

Another set of experiments are carried out where the object is in camera's full field of view with more complicated object information and no area containing only background information. The results are shown in Fig. 8. No area in the image is recognized as a pure background region and is mistakenly removed. All the object information is preserved and enhanced even though it is complicated and varies in intensity, including trees, roads, leaves and human. In contrast, methods such as RGHS, DCP and RoWS result in object information lost, for example, the face information in Fig. 8(a) and the texture on the leaf in Fig. 8(b), which gives a good example of the ability of our method in object detail preservation. The contrast values are displayed in Table 3.



**Fig. 8.** Visual inspection on the descattering performance for the images with a composite background where the object covers camera's full field of view using our method, IFM, MIP [6], HE [11], RGHS [27], DCP [7], GC [10], and RoWS [28].

### 3.3. Discussion on the image processing time

We compare the required processing time for all the methods and show the results in Table 4. We randomly select an input image and the experiments for each method are repeated for ten times without any setting adjustment. The implementation hardware platform is the same for all the

**Table 3. Contrast value comparisons of different methods in scenarios with a composite background where the object covers camera's full field of view. Numbers in red indicate the best performance and the blue ones give the second best performance.**

Methods	input	IFM	Ours	MIP	HE	RGHS	DCP	GC	RoWS
Fig. 8(a)	0.2071	0.2424	0.2930	0.2224	0.2885	0.2620	0.2517	0.1762	0.2464
Fig. 8(b)	0.2214	0.2862	0.4343	0.2317	0.3149	0.3644	0.3729	0.1733	0.3876

methods. The results are the average values of the ten experiments and expressed in seconds. As shown, the physical model-based methods, such as MIP, RGHS, DCP, and RoWS, require about 10 seconds, while the digital image processing methods, such as HE and GC, require less than one second. The reason for this difference is that the physical model-based methods need to estimate the parameters of the model based on the input image, while the digital image processing methods only perform a numerical transformation on the intensity information of the overall image. In addition, the image processing time of our method is about 1.5 minutes, which mainly depends on the number of iterations required for the full convergence of the network and the implementation platform. However, for this research, we mainly focus on the underwater image quality, which means that the evaluation of the method feasibility needs to both consider the contrast of the restored image and the image processing time. In addition, the learning-based underwater image descattering method usually requires tens of hours or even days of network training in a large number of data sets to complete a competitive image processing effect [31–35]. Compared with such methods, the proposed method has higher efficiency. Based on the above considerations, our method achieves a good trade-off between the image quality and the required processing time.

**Table 4. Processing time comparisons of different methods. The result is the average value of 10 repeated experiments. The unit of the data in the table is seconds.**

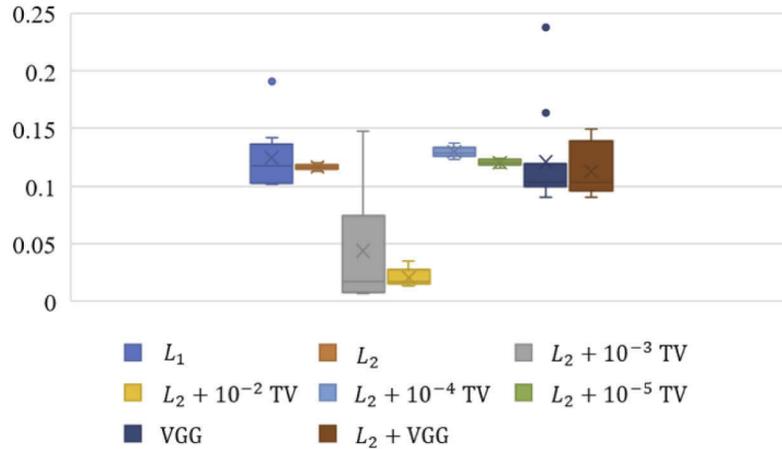
	Ours	MIP	HE	RGHS	DCP	GC	RoWS
Time (s)	104	13	≤ 1	9	12	≤ 1	13

### 3.4. Experiments for the untrained network with different loss functions

For further analysis, we use several different loss functions and compare the contrast values of the outputs. The mean value and standard deviation are calculated based on the 10 repeated experiments. As displayed in Fig. 9, the combination of  $L_2$  loss and total variation (TV) [36] loss with the weight of  $10^{-4}$  gives the highest contrast value with a relatively small standard deviation, while other combinations of  $L_2$  loss and TV loss have either lower mean contrast values or bigger standard deviations. The possible reason is that introducing TV loss can reduce some image noise, but when its weight becomes bigger, the image may also become blurry. In addition, we also demonstrate the results with a VGG [37] loss function and the combination of  $L_2$  and VGG loss, which gives a lower mean contrast and bigger standard deviation, compared to the combination of  $L_2$  and  $10^{-4}$  TV loss. This may be because the processing of the network with a VGG loss will increase the image sharpness, and may also introduce image distortion [38,39]. In this research, we follow the experimental results and use  $L_2 + 10^{-4}$  TV function as the loss function to indicate the quality of the generated images. The detailed mean and standard deviation values for each loss function can be found in [Supplement 1](#).

### 3.5. Limitations and discussion

In some practical application scenarios, natural light is too weak to have satisfactory imaging effects with the scattering and absorption of the water. Therefore, artificial light sources are



**Fig. 9.** Contrast value comparison of the images processed by the untrained network with different loss functions. The combination of  $L_2$  loss and total variation (TV) loss with a weight of  $10^{-4}$  gives the highest image contrast with a relatively small standard deviation.

added in these scenarios to meet the imaging requirements. The difference between underwater active imaging and passive imaging is that due to the absorption of water, the ambient light in the former changes from natural light to partially polarized light, and the latter changes from completely polarized light to partially polarized light [22]. The parameters of IFM should be adjusted according to the polarization state of the light source. However, in this method, the automatic adjustment of the parameters in the physical model and the weights in the network can compensate for the imaging effect and overcome the limitations caused by such changes. In addition, under normal circumstances, backscattering is the main factor leading to image degradation, not blurring. Similar conclusions have also been confirmed in the visual degradation model in the atmosphere and some psychophysical studies based on human perception [40,41]. In some special scenarios dominated by blurring [42], the imaging results of this method may be limited to a certain extent.

#### 4. Conclusion

In summary, we propose a clear underwater imaging method based on polarimetric imaging and an untrained network, which uses a CMOS camera equipped with Stokes mask polarizer to obtain intensity images with different polarization states in a single shot. This design simplifies the operation of polarization imaging and improves the performance of the polarization camera. Subsequently, we use underwater IFM as a physical model to guide network convergence and complete image descattering. This process overcomes the limitations brought by the assumptions of the IFM in the application scenarios, and automatically improves the imaging quality, while not requiring additional datasets for network training. We expect that the principle we demonstrate with this method, i.e., combining the regularity of the physical model with the flexibility of the network, can be applied to more imaging applications. In the future, we will apply this full scene underwater imaging method to the detection of underwater micro-objects, such as microplastic particles [43,44] and marine microbes [45].

**Funding.** Research Grants Council of Hong Kong (GRF 17201818, 17200019, 17201620); National Natural Science Foundation of China (61905197).

**Acknowledgements.** An image of a recognizable person in Fig. 8(a) is the face of Kewei Liu who is the co-author of this paper. The authors affirm that human research participants provided informed consent for the publication of the images in Fig. 8(a).

**Disclosures.** The authors declare no conflict of interest.

**Data availability.** Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

**Supplemental document.** See [Supplement 1](#) for supporting content.

## References

1. J. N. Lythgoe and C. Hemmings, "Polarized light and underwater vision," *Nature* **213**(5079), 893–894 (1967).
2. S. Luria and J. A. S. Kinney, "Underwater vision," *Science* **167**(3924), 1454–1461 (1970).
3. E. H. Zhou, H. Ruan, C. Yang, and B. Judkewitz, "Focusing on moving targets through scattering samples," *Optica* **1**(4), 227–232 (2014).
4. X. Li, H. Hu, L. Zhao, H. Wang, Y. Yu, L. Wu, and T. Liu, "Polarimetric image recovery method combining histogram stretching for underwater imaging," *Sci. Rep.* **8**(1), 12430 (2018).
5. Y. Y. Schechner, S. G. Narasimhan, and S. K. Nayar, "Polarization-based vision through haze," *Appl. Opt.* **42**(3), 511–525 (2003).
6. N. Carlevaris-Bianco, A. Mohan, and R. M. Eustice, "Initial results in underwater single image dehazing," in *Proceedings of IEEE Oceans*, (IEEE, 2010), pp. 1–8.
7. K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE Trans. Pattern Anal. Mach. Intell.* **33**(12), 2341–2353 (2011).
8. K. Lee and Y. Park, "Exploiting the speckle-correlation scattering matrix for a compact reference-free holographic image sensor," *Nat. Commun.* **7**(1), 13359 (2016).
9. J. Yang, Q. He, L. Liu, Y. Qu, R. Shao, B. Song, and Y. Zhao, "Anti-scattering light focusing by fast wavefront shaping based on multi-pixel encoded digital-micromirror device," *Light: Sci. Appl.* **10**(1), 149 (2021).
10. S.-C. Huang, F.-C. Cheng, and Y.-S. Chiu, "Efficient contrast enhancement using adaptive gamma correction with weighting distribution," *IEEE Trans. on Image Process.* **22**(3), 1032–1041 (2013).
11. R. Hummel, "Image enhancement by histogram transformation," *Comput. Graph. Image Process.* **6**(2), 184–195 (1977).
12. Y. Ding and S. Pau, "Circularly and elliptically polarized light under water and the Umov effect," *Light: Sci. Appl.* **8**(1), 32 (2019).
13. T. Treibitz and Y. Y. Schechner, "Active polarization descattering," *IEEE Trans. Pattern Anal. Mach. Intell.* **31**(3), 385–399 (2009).
14. S. B. Powell, R. Garnett, J. Marshall, C. Rizk, and V. Gruev, "Bioinspired polarization vision enables underwater geolocalization," *Sci. Adv.* **4**(4), eaao6841 (2018).
15. N. Shashar, R. T. Hanlon, and A. d. Petz, "Polarization vision helps detect transparent prey," *Nature* **393**(6682), 222–223 (1998).
16. S. Panigrahi, J. Fade, R. Agaisse, H. Ramachandran, and M. Alouini, "An all-optical technique enables instantaneous single-shot demodulation of images at high frequency," *Nat. Commun.* **11**(1), 549 (2020).
17. Z. Ren, Z. Xu, and E. Y. Lam, "Learning-based nonparametric autofocusing for digital holography," *Optica* **5**(4), 337–344 (2018).
18. X. Zhang, F. Wang, and G. Situ, "Blindnet: An untrained learning approach toward computational imaging with model uncertainty," *J. Phys. D: Appl. Phys.* **55**(3), 034001 (2022).
19. F. Wang, Y. Bian, H. Wang, M. Lyu, G. Pedrini, W. Osten, G. Barbastathis, and G. Situ, "Phase imaging with an untrained neural network," *Light: Sci. Appl.* **9**(1), 77 (2020).
20. E. Bostan, R. Heckel, M. Chen, M. Kellman, and L. Waller, "Deep phase decoder: self-calibrating phase microscopy with an untrained deep neural network," *Optica* **7**(6), 559–562 (2020).
21. J. Gwyneth, "Total internal reflection," *Nature* **403**(6771), 707 (2000).
22. D. H. Goldstein, *Polarized Light* (Chemical Rubber Company, 2017), chap. 5, pp. 60–77, 3rd ed.
23. W. Zhang, J. Liang, L. Ren, H. Ju, E. Qu, Z. Bai, Y. Tang, and Z. Wu, "Real-time image haze removal using an aperture-division polarimetric camera," *Appl. Opt.* **56**(4), 942–947 (2017).
24. J. Liang, L. Ren, E. Qu, B. Hu, and Y. Wang, "Method for enhancing visibility of hazy images based on polarimetric imaging," *Photonics Res.* **2**(1), 38–44 (2014).
25. O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-assisted Intervention*, (Springer, 2015), pp. 234–241.
26. C. D. Mobley, *Light and Water: Radiative Transfer in Natural Waters* (Academic, 1994), chap. 3, 5.
27. D. Huang, Y. Wang, W. Song, J. Sequeira, and S. Mavromatis, "Shallow-water image enhancement using relative global histogram stretching based on adaptive parameter acquisition," in *International Conference on Multimedia Modeling*, (Springer, 2018), pp. 453–465.
28. L. Chao and M. Wang, "Removal of water scattering," in *Proceedings of IEEE Conference on Computer Engineering and Technology*, (IEEE, 2010), pp. V2–35.

29. K. O. Amer, M. Elbouz, A. Alfalou, C. Brosseau, and J. Hajjami, "Enhancing underwater optical imaging by using a low-pass polarization filter," *Opt. Express* **27**(2), 621–643 (2019).
30. J. Liang, L. Ren, H. Ju, W. Zhang, and E. Qu, "Polarimetric dehazing method for dense haze removal based on distribution analysis of angle of polarization," *Opt. Express* **23**(20), 26146–26157 (2015).
31. M. K. Moghimi and F. Mohanna, "Real-time underwater image enhancement: A systematic review," *J. Real-Time Image Process.* **18**(5), 1509–1525 (2021).
32. T. Zeng, Y. Zhu, and E. Y. Lam, "Deep learning for digital holography: a review," *Opt. Express* **29**(24), 40572–40593 (2021).
33. Y. Li, H. Lu, J. Li, X. Li, Y. Li, and S. Serikawa, "Underwater image de-scattering and classification by deep neural network," *Comput. Electr. Eng.* **54**, 68–77 (2016).
34. H. Hu, Y. Zhang, X. Li, Y. Lin, Z. Cheng, and T. Liu, "Polarimetric underwater image recovery via deep learning," *Opt. Lasers Eng.* **133**, 106152 (2020).
35. P. Liu, G. Wang, H. Qi, C. Zhang, H. Zheng, and Z. Yu, "Underwater image enhancement with a deep residual framework," *IEEE Access* **7**, 94614–94629 (2019).
36. L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys. D* **60**(1-4), 259–268 (1992).
37. J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *European Conference on Computer Vision*, (Springer, 2016), pp. 694–711.
38. T. Zeng and E. Y. Lam, "Robust reconstruction with deep learning to handle model mismatch in lensless imaging," *IEEE Transactions on Comput. Imaging* **7**, 1080–1092 (2021).
39. M. Deng, A. Goy, S. Li, K. Arthur, and G. Barbastathis, "Probing shallower: perceptual loss trained Phase Extraction Neural Network (PLT-PhENN) for artifact-free reconstruction at low photon budget," *Opt. Express* **28**(2), 2511–2535 (2020).
40. J. S. McLellan, S. Marcos, P. M. Prieto, and S. A. Burns, "Imperfect optics may be the eye's defence against chromatic blur," *Nature* **417**(6885), 174–176 (2002).
41. M. Garcia, C. Edmiston, R. Marinov, A. Vail, and V. Gruev, "Bio-inspired color-polarization imager for real-time in situ imaging," *Optica* **4**(10), 1263–1271 (2017).
42. Y.-T. Peng and P. C. Cosman, "Underwater image restoration based on image blurriness and light absorption," *IEEE Trans. on Image Process.* **26**(4), 1579–1594 (2017).
43. Y. Zhu, C. H. Yeung, and E. Y. Lam, "Digital holographic imaging and classification of microplastics using deep transfer learning," *Appl. Opt.* **60**(4), A38–A47 (2021).
44. Y. Zhu, C. H. Yeung, and E. Y. Lam, "Microplastic pollution monitoring with holographic classification and deep learning," *JPhys Photonics* **3**(2), 024013 (2021).
45. J. J. P. Karlusich, E. Pelletier, F. Lombard, M. Carsique, E. Dvorak, S. Colin, M. Picheral, F. M. Cornejo-Castillo, S. G. Acinas, R. Pepperkok, E. Karsenti, C. d. Vargas, P. Wincker, C. Bowler, and R. A. Foster, "Global distribution patterns of marine nitrogen-fixers by imaging and molecular methods," *Nat. Commun.* **12**(1), 1–18 (2021).