

Article

# MDNet: A Fusion Generative Adversarial Network for Underwater Image Enhancement

Song Zhang <sup>1,2,3,4</sup>, Shili Zhao <sup>1,2,3,4</sup>, Dong An <sup>1,2,3,4</sup>, Daoliang Li <sup>1,2,3,4</sup>  and Ran Zhao <sup>1,2,3,4,\*</sup>

<sup>1</sup> National Innovation Center for Digital Fishery, China Agricultural University, Beijing 100083, China; 15620503230@163.com

<sup>2</sup> Key Laboratory of Smart Farming Technologies for Aquatic Animal and Livestock, Ministry of Agriculture and Rural Affairs, China Agricultural University, Beijing 100083, China

<sup>3</sup> Beijing Engineering and Technology Research Center for Internet of Things in Agriculture, China Agricultural University, Beijing 100083, China

<sup>4</sup> College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, China

\* Correspondence: ran.zhao@cau.edu.cn

**Abstract:** Underwater images are widely used in ocean resource exploration and ocean environment surveillance. However, due to the influence of light attenuation and noise, underwater images usually display degradation phenomena such as blurring and color deviation; an enhancement method is required to make the images more visible. Currently, there are two major approaches for image enhancement: the traditional methods based on physical or non-physical models, and the deep learning method. Inspired by the fusion-based idea, this paper attempts to combine traditional methods with deep learning and proposes a multi-input dense connection generator network (MDNet) for underwater image enhancement. Raw images and processed images are input into the network together, the shallow information is fully utilized by dense connection, and the network is trained in generative and adversarial manner. We also design a multiple loss function to improve the visual quality of the generated images. We conduct both qualitative and quantitative experiments, and then compare the results with state-of-the-art approaches comprehensively using three representative datasets. Results show that the proposed method can effectively improve the perceptual and statistical quality of underwater images.



**Citation:** Zhang, S.; Zhao, S.; An, D.; Li, D.; Zhao, R. MDNet: A Fusion Generative Adversarial Network for Underwater Image Enhancement. *J. Mar. Sci. Eng.* **2023**, *11*, 1183. <https://doi.org/10.3390/jmse11061183>

Academic Editors: Mohamed Benbouzid, Anne Blavet and Nicolas Guillou

Received: 8 May 2023

Revised: 29 May 2023

Accepted: 29 May 2023

Published: 6 June 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Keywords:** underwater image enhancement; deep learning; generative adversarial network

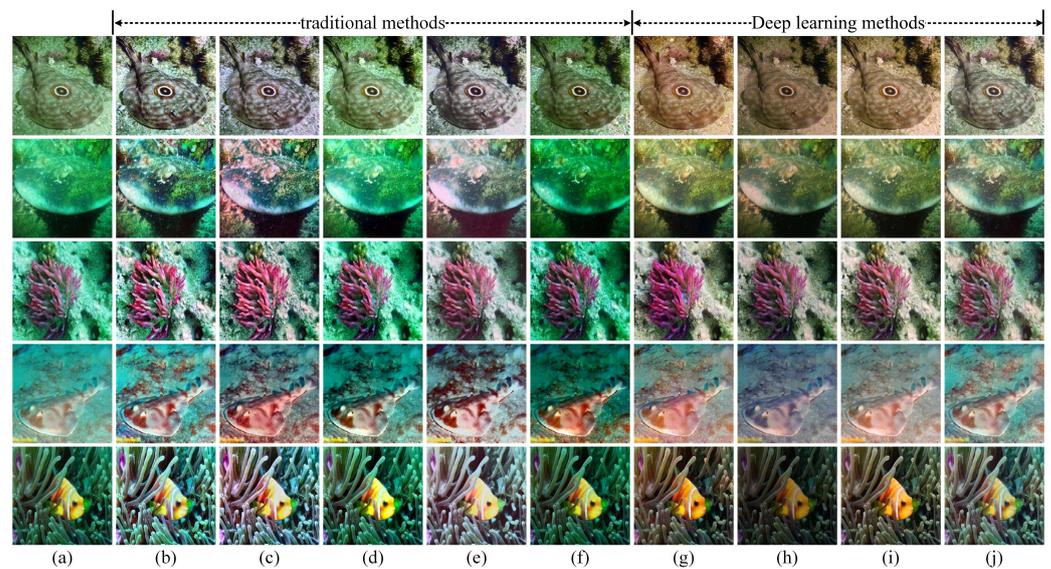
## 1. Introduction

Underwater images play an important role in information acquisition in the ocean. With the development of marine science, more and more underwater operations, such as underwater vehicle navigation and resource exploration [1,2], are greatly dependent on images. However, due to the special environment of the ocean, light attenuates in the water, which makes the contrast ratio of underwater images low. Different light has different attenuation in the water; the light with higher frequency penetrates the particles underwater with more ease, which produces blue–green images. Moreover, dissolved organic matter and suspended particles also introduce huge noise. All those elements make the underwater images seriously degraded [3]. Therefore, underwater image enhancement is particularly important in the ocean.

In the past decade, underwater image enhancement has made great progress, and a variety of methods have been proposed. According to the utilized technologies, algorithms can be divided into traditional physical model-based and non-physical model-based methods, as well as deep learning methods. The physical model-based methods consider the degradation principle of underwater images, then achieve image enhancement by establishing a compensation model. The non-physical model-based methods directly operate the pixel value to upgrade the visual quality. Deep learning is a data-driven method,

which extracts valuable features from a large amount of data. In the task of underwater image enhancement, it is assumed that there is a mapping between distorted image and high-quality image, and the mapping can be obtained by training neural networks. In this way, degraded images are translated into high-quality images.

Our pre-experiments found that traditional methods based on physical models or non-physical models usually have excellent performance in certain underwater images or scenes. However, it is difficult to deal with most other common scenes. These methods can enhance some features of an image, but often seriously distort the others, which does not improve the image quality in the end. For example, image contrast can be improved, but the color might become worse. However, deep learning methods usually offer a more balanced performance than traditional methods when dealing with complex underwater scenes, as shown in Figure 1.



**Figure 1.** Comparison between traditional methods and deep learning methods. (a) Raw underwater images. (b) CLAHE [4]. (c) Two-step [5]. (d) RGHS [6]. (e) GCF [7]. (f) UDCP2016 [8]. (g) FUnIEGAN [9]. (h) WaterNet [10]. (i) Shallow-Uwnet [11]. (j) MDNet.

Fusion-based idea [12] is a processing strategy that fuses images processed by multiple methods to achieve better results. The combination of multiple methods can provide multi-scale features, overcome the limitations of a single method to a certain extent, and achieve better results [13]. Therefore, this paper attempts to combine these two methods above to employ their respective advantages. The traditional method is used to preprocess underwater images, while deep learning is used for further enhancement. The main contributions of this paper are listed below:

1. The most advanced underwater image enhancement methods based on traditional methods and deep learning are compared and analyzed using three representative datasets (UIEB [10], EUVP [9] and U45 [14]). Our comparison and analysis provide comprehensive and valuable conclusion about the current underwater image enhancement.

2. A multi-input dense connection generator network, MDNet, is proposed. Raw images and processed images are input into the network together, and the shallow information is fully utilized by dense connection. MDNet has achieved state-of-the-art performance in some public datasets. Several examples are shown in Figure 1.

## 2. Related Work

As an important way to improve the visual quality of unclear pictures, image enhancement has attracted more and more researchers' attention. According to the used principles, underwater image enhancement technologies can be divided into traditional methods and

deep learning methods. The traditional methods mainly include non-physical model-based methods and physical model-based methods.

### 2.1. Physical Model-Based Method

A physical model-based method takes into account the mathematical modeling of underwater image degradation process. By estimating the model parameters, the degradation process is reversed to obtain high-quality images. Therefore, this approach is also called image restoration.

Dark channel prior (DCP) [15] is a commonly used underwater image enhancement method which was first utilized for image defogging by establishing a fog degradation model. There are similarities in modeling between haze images and underwater images, so DCP can also be applied to underwater image enhancement [16]. In the work of Ref. [17], DCP was combined with wavelength compensation to enhance underwater images. Considering that the red channel decays rapidly underwater and becomes unreliable, in [8], an underwater dark channel prior (UDCP2016) was proposed only for blue and green channels. In addition, there are others that perform enhancement processing according to optical characteristics. Ref. [18] predicted the depth of underwater scene according to the different attenuation of the RGB color channels, then completed enhancement modeling through the Markov random field. As the light attenuation varies in different water environments, Wang et al. [19] proposed a new adaptive attenuation prior curve to simulate the light attenuation process in different water environments. Ref. [20] used dark pixels and distance information (depth map) to estimate backscattering. The estimation of spatially varying light sources was then used to obtain the distance attenuation coefficient. According to the characteristics of underwater images, Ref. [21] first corrected the illumination and color of underwater images, and then realized dehazing based on the estimation of artificial background light and transmission map depth.

### 2.2. Non-Physical Model-Based Method

Non-physical model-based method neither relies on the mathematical model of underwater optical imaging nor considers the degradation process of the image. It improves the visual quality by adjusting pixel values directly. Image information is extracted without any prior knowledge [22], which makes it more universal.

Early approaches including white balance algorithm [23] and contrast limited histogram equalization (CLAHE) [4] all belong to non-physical model-based methods. In Ref. [24], the authors improved the color deviation and contrast of underwater images by directly stretching the range of dynamic pixels in the GRB and HIS color space. However, this operation often leads to the occurrence of over- or under-enhancement. In general, over- or under-enhancement are prevented by limiting the numerical range of pixels [25,26]. Underwater images are often accompanied by color deviation and low contrast. Therefore, researchers often use different methods to deal with them separately. Based on the gray world hypothesis [27], Fu et al. [5] proposed a two-step approach which includes a color correction algorithm and a contrast enhancement algorithm. In Ref. [28], further improvements were made by using adaptive gray world for color correction and employing Differential Gray-Levels Histogram Equalization to improve the contrast of the image. Huang et al. [6] corrected the image contrast by redistributing the histogram of each RGB channel, while color was corrected by stretching the "L" component and modifying the "a" and "b" components in the CIE-Lab color space.

Another approach attempts to enhance underwater images based on the Retinex model. In Ref. [29], the authors proposed a novel variational framework for Retinex to decompose reflectance and illumination. Different strategies were adopted to improve reflectance and illumination so as to improve the detail and brightness of the image. An improvement was made by Xiong et al. [7], who imposed Gaussian curvature priors on reflectance and illumination to prevent over-enhancement.

### 2.3. Deep Learning Method

Deep learning is a data-driven approach which automatically learn features from massive data. During the past years, underwater image enhancement has made remarkable progress due the development of deep learning and the availability of large-scale datasets. Deep learning assumes that there is a mapping between raw underwater images and high-quality images. Underwater image enhancement is realized by neural network models which are trained through paired or unpaired images.

Deep learning heavily relies on large-scale datasets, and the quality of the datasets determines the upper limit of the effectiveness that deep learning can achieve. Due to the unique underwater environment, it has always been a challenge to obtain the ground truth of an underwater image. The usual approach is to use different methods to enhance underwater images and select the best one as the ground truth. However, such generated ground truth still does not perform well in some challenging situations. Due to the lack of ground truth, Li et al. [30] used generative adversarial network (GAN) to simulate underwater images from the in-air images. With the synthetic dataset, they proposed a two-stage convolutional neural network (CNN) to enhance underwater images. With the good performance of GAN in the field of image translation [31], GAN has also been successfully used for underwater image enhancement [32]. In Ref. [33], the authors proposed a multiscale dense GAN in which a large number of residual connections were added. Considering the lack of physical model constraints in end-to-end learning, Ref. [34] embedded the physical model into the network design, and GAN was adopted for coefficient estimation based on Akkaynak and Treibitz's work [20]. Based on the fusion-based idea, Li et al. [10] inputted the preprocessed images and the raw images together to train an enhancement neural network and proposed a CNN with multiple inputs and single output named WaterNet. Usually, neural networks require a lot of computing resources, which makes it difficult to deploy on mobile devices. To solve this problem, Refs. [9,11] proposed lightweight neural networks, respectively, which can achieve a good running speed with limited computing resources while ensuring the performance and improved the deployable performance of the model.

Previous deep learning methods usually use a single input to enhance the underwater image. However, one of the advantages of deep learning method is that it allows flexible multiple inputs. Inspired by the fusion-based idea, this paper proposes a multi-input dense connection generator network, MDNet, which inputs the preprocessed images and the raw images into the generator together. It ingeniously combines the traditional method with the deep learning method and continues to use the deep learning idea for further enhancement. In addition, the traditional methods and deep learning methods are compared and analyzed, which provides a comprehensive perspective for the current underwater image enhancement.

## 3. The Proposed Method

### 3.1. Model Structure

This paper mixes the preprocessed images with raw images, then inputs the mixed images into the neural network for processing. At the same time, GAN is used for training in the generative and adversarial manner. High-quality underwater enhanced images are generated by a well-trained generator. The schematic diagram of the proposed method is shown in Figure 2.

GAN is used to enhance the underwater images. Firstly, raw images are pretreated by traditional methods. The pretreated images and the raw images are both imported to the generator that produces the enhanced images. Then, the enhanced images together with the ground truth are input into the discriminator that determines whether the input images are "true" (ground truth). In this way, the generator and the discriminator are trained circularly to continuously improve the generator's ability of producing "false images" (enhanced images) and the discriminator's ability of identifying "true or false". In this way,

the generated images become more and more close to the ground truth. Finally, we obtain a generator network that can improve the visual quality of underwater images.

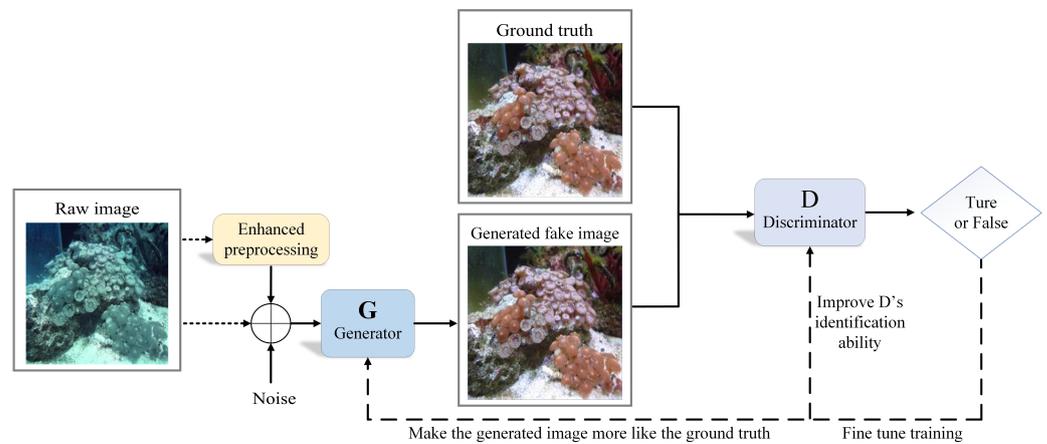


Figure 2. Diagram of the proposed method.

### 3.1.1. The Generator

Different from semantic segmentation and object detection [35,36] which emphasize pixel level understanding, image enhancement focuses on mapping degraded images with brighter and clearer images to improve visual quality. This puts forward higher requirements for the retention and restoration of image details. In the process of neural network processing, high-level semantic features can be extracted from the image, but shallow details are also lost. Therefore, we designed a multi-input dense connection generator network to improve the utilization of shallow information, as shown in Figure 3. The proposed generator is composed of a full convolution network and three densely connected convolution blocks in series. The inputs are raw images and preprocessed images; CLAHE is selected for image preprocessing. The two inputs are first fused together, and the feature map after one convolution layer is then connected to the output of each convolution block by skipping. The final output is processed by a convolution layer again.

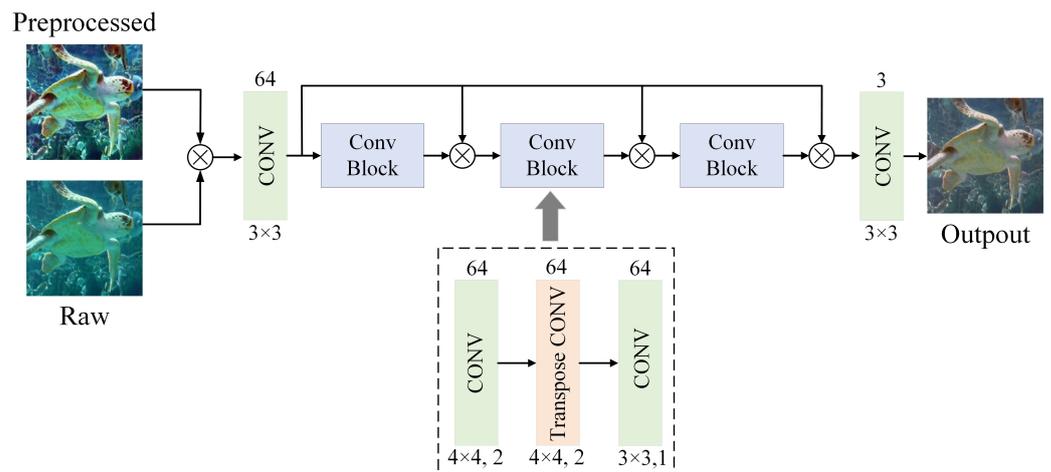


Figure 3. Diagram of the generator.

Convolution block: The convolution block includes three parts: down-sampling, up-sampling and convolution. Down-sampling is realized by a convolution with a kernel size of  $4 \times 4$  and a stride size of 2. Compared with pool operation, this can extract more image features, which is conducive to information fusion. Up-sampling is realized by a transpose convolution with a kernel size of  $4 \times 4$  and a stride size of 2. The feature map is

extracted by down-sampling, then restored to the original size by up-sampling, and further processed by a convolution layer. It can be expressed by the following formula:

$$I_{out} = Conv(TranConv(Pool(I_{in}))) \tag{1}$$

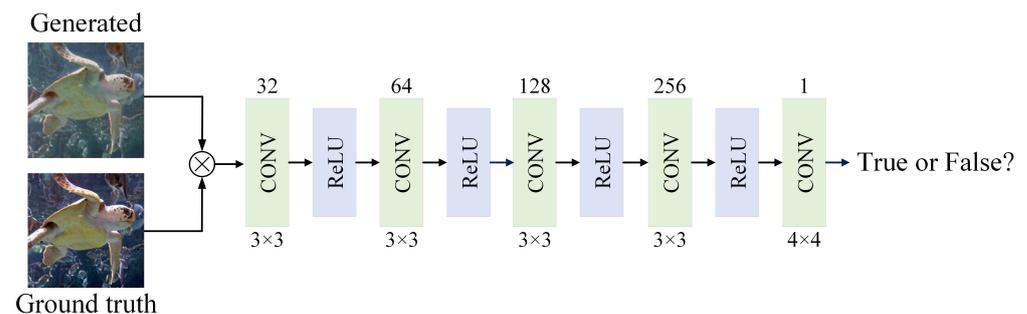
where  $I_{in}$  and  $I_{out}$  refer to input and output, respectively,  $Pool$  refers to pooling operation,  $TranConv$  refers to transpose convolution,  $Conv$  refers to convolution operation.

**Skipping connection:** The input is connected (by channel accumulation) with the output of each convolution block by skipping. When the gradient disappears, this skipping connection imposes a greater weight on the channel related to the original input than the output channel with a separate convolution block. This ensures that the features can be learned from each convolution block while retaining the basic features of the input. In addition, the generator model is fully convoluted and does not use a full connection layer, so it has a more flexible size on the input image.

### 3.1.2. The Discriminator

For the discriminator, we refer to [9] and use the structure of Markovian PatchGANs [31]. This kind of structure processes the image layer by layer, and a characteristic matrix of  $n \times n \times 1$  size is obtained. Each element in the matrix corresponds to a patch in the image. The whole image is discriminated by discriminating all elements. This kind of structure assumes the independence of pixels beyond the patch-size (that pixels other than patch-size are independent), that is, images are discriminated only based on patch-level information. This assumption is very important for effectively capturing high-frequency features such as local textures and styles [37]. In addition, this structure has higher efficiency in calculation as it requires fewer parameters due to the comparison based on patch-level.

As shown in Figure 4, the generated image and ground truth are superimposed together for identification. After 5 convolution layers and 4 nonlinear activation functions, the image is judged with a result of “true or false”. The first four layers in the discriminator adopt a  $3 \times 3$  convolution kernel with a step size of 2. The last layer uses a  $4 \times 4$  convolution kernel with a step size of 1. ReLU is used as the nonlinear activation function.



**Figure 4.** Diagram of discriminator structure.

### 3.2. Loss Function

A standard conditional GAN-based model learns a mapping  $G : (X, Z) \rightarrow Y$ , where  $X$  represents the input,  $Z$  represents random noise,  $Y$  is obtained by generator  $G$ , and the conditional adversarial loss function is as follows:

$$L_{cGAN} = E_{X,Y}[\log D(Y)] + E_{X,Y}[\log(1 - D(X, G(X, Z)))] \tag{2}$$

Generator  $G$  tries to minimize  $L_{cGAN}$  while  $D$  tries to maximize  $L_{cGAN}$ . Based on this, we add three additional aspects, global similarity, image content, local texture and style to quantify the perceived image quality.

**Global similarity:** The existing methods have shown that adding an  $L_1(L_2)$  loss to the objective function enables  $G$  to learn to sample from a globally similar space in an

$L_1(L_2)$  sense [31,38]. Since the  $L_1$  loss function does not easily cause ambiguity, we add the following loss term in the objective function:

$$L_1(G) = E_{X,Y,Z}[||Y - G(X, Z)||_1] \tag{3}$$

**Image content:** The loss of content is added to the objective function to encourage  $G$  to generate enhanced images (i.e., representation) similar to the target image (ground truth). Inspired by [39,40], the advanced features extracted from the block5-conv2 layer of the pre-trained VGG-19 network are used as the image content function  $\Phi(\cdot)$ . The content loss function is defined as follows:

$$L_{con}(G) = E_{X,Y,Z}[||\Phi(Y) - \Phi(G(X, Z))||_2] \tag{4}$$

**Local texture and style:** Markovian PatchGANs can effectively capture high-frequency information about local textures and styles. Hence, discriminator  $D$  is used to emphasize the consistency of local texture and style in an antagonistic way [31].

**Structural similarity:** At the same time, we add structural similarity ( $SSIM$ ) loss to the loss function to standardize the structure and texture of the generated image:

$$L_{SSIM}(G) = 1 - \frac{1}{N} \sum_{i=1}^N SSIM(x_i, y_i) \tag{5}$$

$SSIM$  compares image blocks based on brightness, contrast and structure:

$$SSIM(x, y) = \left( \frac{2\mu_x\mu_y + c_1}{\mu_x^2 + \mu_y^2 + c_1} \right) \left( \frac{2\sigma_{xy} + c_2}{\sigma_x^2 + \sigma_y^2 + c_2} \right) \tag{6}$$

where  $\mu_x(\mu_y)$  donates the average of  $x(y)$ ,  $\sigma_x^2(\sigma_y^2)$  indicates the variance of  $x(y)$ ,  $\sigma_{xy}$  is the covariance between  $x$  and  $y$ . In addition,  $c_1 = (255 \times 0.01)^2$  and  $c_2 = (255 \times 0.03)^2$  are constants that ensure numeric stability.

The final objective loss function is as follows:

$$G^* = \underset{G}{agr \min} \underset{D}{\max} L_{cGAN}(G, D) + L_1(G) + L_{con}(G) + L_{SSIM}(G) \tag{7}$$

## 4. Results and Discussion

### 4.1. Datasets

We collected some datasets of ocean scenes and made full use of them to train and evaluate our proposed methods.

(1) UIEB dataset: This dataset contained a large number of images collected from real underwater environment, Google, YouTube and the related papers. The reference images in the dataset were generated by 12 different enhancement methods. The ground truth was selected from the images by 50 persons with respect to their visual quality. The dataset contains a total of 890 paired images and 60 challenging images (without ground truth). Referring to the practice of [10], 800 pairs of images were randomly selected from 890 pairs of images as the training set, and the remaining 90 pairs of images were used as the test set. In order to facilitate the training and testing process, images in the dataset were uniformly scaled to the size of  $256 \times 256$ .

(2) EUVP dataset: This dataset contained more than 12 thousand paired images and 8 thousand unpaired images. During the production of the dataset, seven different cameras were used to shoot underwater, and the best image was selected as the reference image. It also contains some carefully selected images from YouTube. EUVP contains three subsets. We chose the EUVP-Imagenet subset with rich scenarios as our training and test set. EUVP-Imagenet contains 4260 paired images. A total of 4000 images were selected randomly for training, and the remaining 260 images were used as test set. The size of the images is  $256 \times 256$ .

(3) U45 Dataset: This dataset collected 45 real underwater images (without ground truth), including three types: green, blue and fog, corresponding, respectively, to the color casts, low contrast, and haze-like effects of underwater degradation. The size of the images is  $256 \times 256$ .

We used the EUVP and UIEB datasets to train and evaluate the proposed network, respectively. Because there is no ground truth in the U45 dataset, we used the network trained by the UIEB dataset to enhance U45, and we used non-reference evaluation indicators for quantitative evaluation.

#### 4.2. Implementation Details

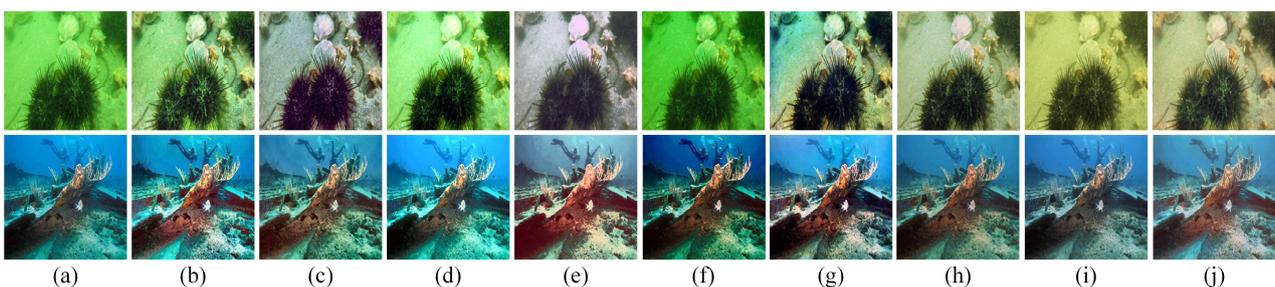
(1) Experimental details: We implemented the proposed method on a laptop equipped with 11800H CPU, RTX 3070 GPU and 16 G memory. A total of 200 iterations of training were carried out on the PyTorch framework. The batch-size was set to 8. The Adam optimizer was used to optimize the training, and the learning rate was set to  $3 \times 10^{-4}$ , the decay of 1st-order momentum was 0.5, and the decay of 2nd-order momentum was 0.99.

(2) Comparison Methods: The methods for comparison included CLAHE [4], the state-of-the-art deep learning methods included FUnIE-GAN [9], WaterNet [10], Shallow-Uwnet [11], and the traditional methods included two-step [5], RGHS [6], GCF [7], UDCP2016 [8]. To ensure a fair comparison, we adopted the same setting in all competing methods as provided by the original papers.

#### 4.3. Qualitative Evaluation

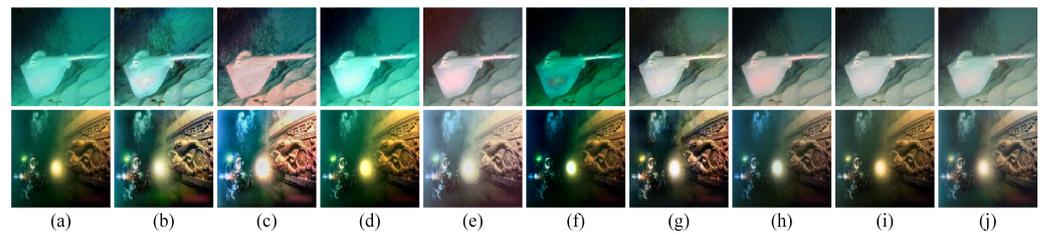
We divided the underwater scene into four categories: color deviation scenes, light spot scenes, scattering scenes and extreme degeneration scenes. The results of different methods and the corresponding reference images are shown in Figures 5–8.

The absorption level of light in water depends on its wavelength. Red light attenuates the quickest while blue light does slowest [41]. The attenuation brings color deviation to the image, resulting in a greenish or bluish color phenomenon. Examples are shown in Figure 5. For the green scenes, Two-step and GCF can effectively eliminate color deviation. However, WaterNet and MDNet (proposed) make the image slightly yellow. In case of a blue scene, WaterNet, Shallow-Uwnet and MDNet (proposed) can effectively remove color distortion, while CLAHE, GCF and FUNIE-GAN aggravate the red component in the image and cause distortion.



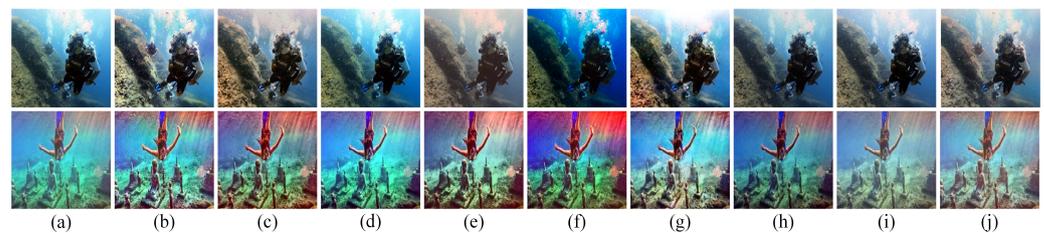
**Figure 5.** Comparison of the color deviation scenes. (a) Raw images. (b) CLAHE. (c) Two-step. (d) RGHS. (e) GCF. (f) UDCP2016. (g) FUnIE-GAN. (h) WaterNet. (i) Shallow-Uwnet. (j) MDNet.

Artificial light sources are often used for light compensation in ocean exploration. Generally, artificial light sources radiate from a fixed point to the surrounding area. In this way, there is an obvious light spot on the center of the target and the light source, which results in insufficient light conditions in other areas. Comparisons of light spot scenes are shown in Figure 6. GCF, WaterNet, Shallow-Uwnet and MDNet can well restore the details of the periphery of the spot, but GCF and WaterNet also unnecessarily increase the brightness of the spot area.



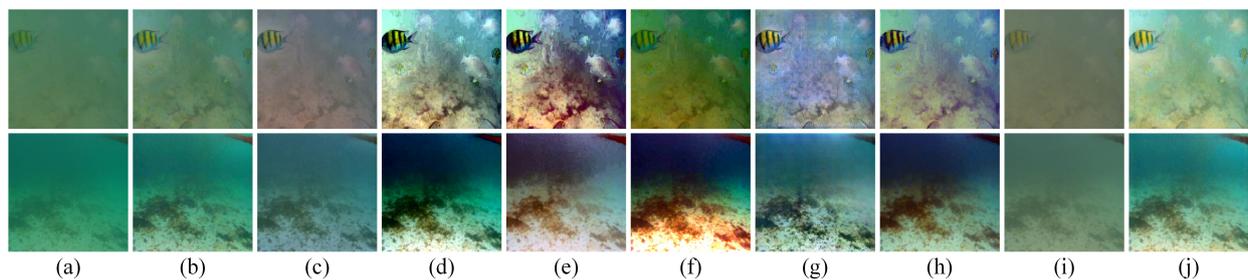
**Figure 6.** Comparison of the light spot scenes. (a) Raw images. (b) CLAHE. (c) Two-step. (d) RGHS. (e) GCF. (f) UDCP2016. (g) FUnIE-GAN. (h) WaterNet. (i) Shallow-UwNet. (j) MDNet.

Scattering causes complex changes in the propagation path of light, thus changing the distribution of light in space. In underwater images, the scattering effect depends on the distance  $d$  between targets and the sensor. In Figure 7 (the top line), the blurring effect and color degradation are easily revised in low-scatter scenes ( $d$  is small) except UDCP2016. In high-scatter scenes ( $d$  is big) (the bottom line in Figure 7), the color distortion and blurring effect of the image are difficult to revise. CLAHE, two-step, GCF and UDCP2016 aggravate the red distortion, while WaterNet, Shallow-UwNet and MDNet achieve relatively better overall results.



**Figure 7.** Comparison of the scattering scenes. (a) Raw images. (b) CLAHE. (c) Two-step. (d) RGHS. (e) GCF. (f) UDCP2016. (g) FUnIE-GAN. (h) WaterNet. (i) Shallow-UwNet. (j) MDNet.

In an extremely degraded environment with weak light and serious scattering of suspended particles, underwater images present quite low quality and include color distortion, low contrast and blurring. This brings huge challenges to enhancement algorithms. Comparisons are shown in Figure 8. Only RGHS, WaterNet and MDNet performed well in this task. On the contrary, GCF and UDCP2016 caused serious distortion.



**Figure 8.** Comparison of the extreme degeneration scenes. (a) Raw images. (b) CLAHE. (c) Two-step. (d) RGHS. (e) GCF. (f) UDCP2016. (g) FUnIE-GAN. (h) WaterNet. (i) Shallow-UwNet. (j) MDNet.

#### 4.4. Quantitative Evaluation

Two kinds of indicators were used to evaluate the enhancement performance: the full reference evaluation indicators and the non-reference evaluation indicator. Full reference evaluation indicators include peak signal to noise ratio ( $PSNR$ ) and  $SSIM$ . The formula of  $PSNR$  is as follows:

$$PSNR(x, y) = 10 \times \log_{10}\left(\frac{255}{MSE(x, y)}\right) \quad (8)$$

*PSNR* compares the generated image  $x$  with the ground truth  $y$  based on the mean-square error (*MSE*). *SSIM* is shown in Formula 6.

Underwater image quality measure (*UIQM*) [42] was selected as a non-reference evaluation indicator. It is a specified indicator that can comprehensively evaluate underwater images from three perspectives: color, sharpness and contrast. The expression is as follows:

$$UIQM = c_1 \times UICM + c_2 \times UISM + c_3 \times UIConM \quad (9)$$

where *UICM*, *UISM* and *UIConM* represent the color index, sharpness index and contrast index of the image, respectively [42]. The coefficient settings are  $c_1 = 0.0282$ ,  $c_2 = 0.2953$ ,  $c_3 = 3.5753$ . The underwater image quality can be thoroughly evaluated through the linear combination of the three indexes.

We used the original image as the only input and conducted a single-input ablation study (called MDNet-). The experimental results of different methods on UIEB dataset are shown in Table 1. In traditional methods, RGHS achieved the best performance on *PSNR*, CLAHE achieved the best performance on *SSIM*, while GCF obtained the highest score on *UIQM*. What needs to be noted is that the original UIEB dataset used the two-step method to generate ground truth. Therefore, the two-step method definitely acquires the best results on this dataset. Consequently, we did not use the two-step approach for comparison to make it fair. In deep learning methods, the proposed MDNet performed the best on all three indicators. Although MDNet and WaterNet have the same mean value of a *SSIM* indicator, the variance of MDNet is much smaller. It can be concluded that MDNet has a competent performance on the UIEB dataset on various indicators. The results of the ablation study on UIEB (including the other two datasets) also demonstrate the improved effect brought by multiple inputs.

**Table 1.** Results of different methods on UIEB dataset.

Methods	PSNR↑	SSIM↑	UIQM↑
CLAHE [4]	20.84 ± 3.59	<b>0.85 ± 0.08</b>	2.94 ± 0.57
RGHS [6]	<b>23.24 ± 5.51</b>	0.83 ± 0.10	2.55 ± 0.70
GCF [7]	20.43 ± 4.22	0.82 ± 0.09	<b>3.16 ± 0.43</b>
UDCP2016 [8]	15.65 ± 3.93	0.68 ± 0.12	2.15 ± 0.64
FUnIE-GAN [9]	20.29 ± 3.12	0.79 ± 0.06	3.08 ± 0.29
WaterNet [10]	19.52 ± 4.00	0.82 ± 0.10	2.98 ± 0.39
Shallow-Uwnet [11]	20.78 ± 5.13	0.81 ± 0.10	2.84 ± 0.52
MDNet-	20.67 ± 4.39	0.78 ± 0.09	2.91 ± 0.43
<b>MDNet</b>	<b>22.27 ± 3.51</b>	<b>0.82 ± 0.06</b>	<b>3.09 ± 0.34</b>

The results of different methods on EUVP-Imagenet dataset are shown in Table 2. Among the traditional methods, RGHS achieved the best performance on *PSNR* and *SSIM*, while GCF achieved the best performance on *UIQM*. In deep learning methods, Shallow-Uwnet was the best on *PSNR* and *SSIM*, while MDNet followed closely. The proposed MDNet showed the best results on *UIQM*. On this dataset, deep learning methods performed generally better than traditional methods on *PSNR* and *SSIM*, while traditional methods were better than learning methods on *UIQM*. Thus, the proposed approach not only kept the good results on the *PSNR* and *SSIM* indicators, but also largely improved the *UIQM* value that other learning methods did not have.

The full reference evaluation indicators require ground truth; thus, only *UIQM* was used to evaluate the results of different methods on the U45 dataset (without ground truth), as shown in Table 3. Among the traditional methods, GCF achieved the best performance on *UIQM*. In deep learning methods, FUnIE-GAN was the best on *UIQM*, while MDNet followed closely. MDNet has reached the top three position on this dataset.

**Table 2.** Results of different methods on EUVP-Imagenet dataset.

Methods	PSNR↑	SSIM↑	UIQM↑
CLAHE [4]	18.72 ± 1.97	0.66 ± 0.08	2.92 ± 0.35
Two-step [5]	18.50 ± 2.77	0.65 ± 0.10	3.04 ± 0.25
RGHS [6]	<b>22.56 ± 2.89</b>	<b>0.69 ± 0.07</b>	2.30 ± 0.41
GCF [7]	18.18 ± 4.00	0.66 ± 0.18	<b>3.42 ± 0.27</b>
UDCP2016 [8]	19.32 ± 4.42	0.64 ± 0.09	2.20 ± 0.45
FUnIE-GAN [9]	23.00 ± 2.20	0.75 ± 0.07	3.14 ± 0.31
WaterNet [10]	19.89 ± 4.59	0.74 ± 0.08	2.89 ± 0.42
Shallow-Uwnet [11]	<b>24.86 ± 2.68</b>	<b>0.79 ± 0.07</b>	3.08 ± 0.34
MDNet-	23.04 ± 2.54	0.75 ± 0.08	3.19 ± 0.30
<b>MDNet</b>	23.47 ± 2.46	0.76 ± 0.08	<b>3.21 ± 0.30</b>

**Table 3.** Results of different methods on U45dataset.

Methods	UIQM↑
CLAHE [4]	3.04 ± 0.62
Two-step [5]	3.31 ± 0.28
RGHS [6]	2.57 ± 0.74
GCF [7]	<b>3.36 ± 0.27</b>
UDCP2016 [8]	2.41 ± 0.70
FUnIE-GAN [9]	<b>3.40 ± 0.25</b>
WaterNet [10]	3.27 ± 0.27
Shallow-Uwnet [11]	3.03 ± 0.42
MDNet-	3.32 ± 0.25
<b>MDNet</b>	3.34 ± 0.28

#### 4.5. Analysis of Results

(1) Performance analysis: The proposed MDNet combines traditional methods with deep learning methods and further improves the quality of underwater images by using deep learning methods on the basis of preprocessing with traditional methods. In qualitative evaluation, MDNet can effectively enhance underwater images and improve visual quality for different scenes. In a quantitative evaluation, compared with other deep learning methods, MDNet has achieved excellent performance on all three datasets. Moreover, the image preprocessing method of MDNet can be easily replaced by others, thus making the approach quite flexible.

(2) Comparison with ground truth: In this work, we found that the performance of the proposed MDNet is very close to ground truth on many images intuitively. In some cases, it is even better than the ground truth of poor quality (as shown in Figure 9). This indicates the strong capability of the proposed method in enhancing underwater images.

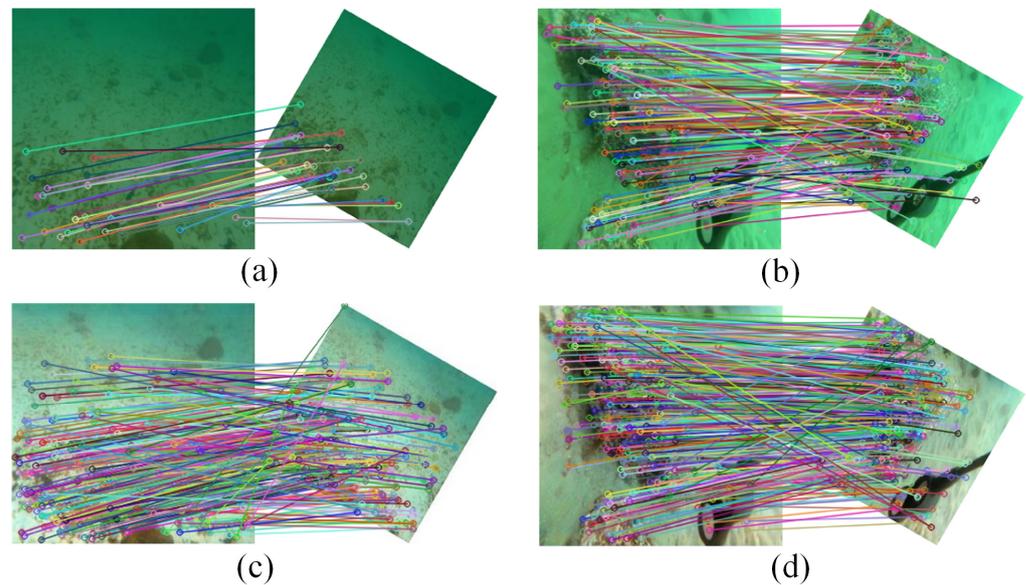
(3) Comparison between traditional methods and deep learning methods: It is worth noting that traditional methods tend to perform better on certain underwater images or scenes (as shown in Figure 5, the two-step and GCF results in the top line), which is also consistent with the fact that images processed by traditional methods are often selected as ground truth in dataset generation [10]. However, traditional methods have difficulty dealing with complex and diverse underwater scenes, and often produce over- or under-enhancement (such as two-step and GCF in Figure 8). In this situation, deep learning approaches display stronger generalization performance, and their overall performance is better than that of the traditional approaches (it can also be seen from the test indicators). Therefore, for a single underwater scene, an appropriate traditional method would be usually qualified; however, in terms of more complex and changeable scenes, a deep learning method can be a better choice.



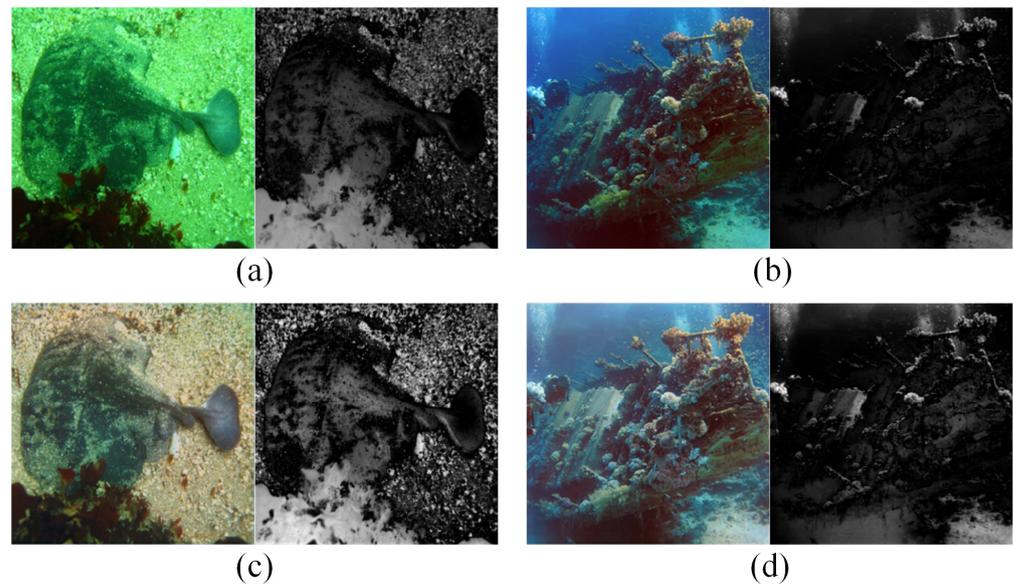
**Figure 9.** Some examples of enhanced images that are better than the ground truth in UIEB dataset.

4.6. Application Tests

To verify the validity of the proposed method in this paper, we conducted the SIFT-based key point matching test [43] and the global contrast-based salient region detection test [44]. The numbers of key points matched to the original underwater image in Figure 10a,b are 45 and 222, respectively. After processing with the method, the numbers of key points that can be matched are displayed in Figure 10c,d, which are 335 and 342, respectively. The results show that the images processed with our method can match more key points. The salient region detection tests are presented in Figure 11. Comparing our results with the original underwater images, it can be concluded that the proposed method leads to a clearer saliency map, which provides more image details.



**Figure 10.** Image key point matching test. (a,b): Key point matching result of original images; (c,d): Key point matching result of enhanced images.



**Figure 11.** Image salient region detection test. (a,b): Original images salient region detection result map; (c,d): Enhanced images salient region detection result map.

## 5. Conclusions and Future Work

This paper proposes an underwater image enhancement method, MDNet, which combines the traditional method with the deep learning method. Qualitative and quantitative evaluations have demonstrated the excellent performance of MDNet, and application tests have also validated its effectiveness. In addition, the method allows the use of arbitrary enhancement methods for preprocessing which offers strong flexibility to the approach. The quality of the images enhanced by MDNet can even be better than some ground truth in many cases. In the meantime, we found that traditional methods usually work better in the case of individual images or certain simple scenes. In complex environments, deep learning methods offer better performance in general cases.

In this paper, only CLAHE was used and discussed in image preprocessing. More algorithms can be introduced to our approach. Deep learning part of the algorithm can also be further improved to obtain better performance. The quality of datasets will have great influence on deep learning-based approaches. However, there are still some low-quality ground truths in many datasets, such as UIEB and EUVP. Therefore, it merits to build a high-quality underwater image enhancement dataset in the future.

**Author Contributions:** Proposed the basic route of the research, wrote part of the paper, and completed the finalization of the paper, R.Z.; conceived the idea, guided and revised the paper, D.A. and D.L.; guided and revised the paper, S.Z. (Song Zhang); and S.Z. (Shili Zhao) conducted experiments and analyzed the experimental results. All authors have read and agreed to the published version of the manuscript.

**Funding:** This paper was supported by The National Natural Science Foundation of China (NO.32273188) and Key Research and Development Plan of the Ministry of Science and Technology (NO.2022YFD2001700).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Conflicts of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## References

1. Bingham, B.; Foley, B.; Singh, H.; Camilli, R.; Delaporta, K.; Eustice, R.; Mallios, A.; Mindell, D.; Roman, C.; Sakellariou, D. Robotic tools for deep water archaeology: Surveying an ancient shipwreck with an autonomous underwater vehicle. *J. Field Robot.* **2010**, *27*, 702–717. [\[CrossRef\]](#)
2. Wu, Y.; Ta, X.; Xiao, R.; Wei, Y.; An, D.; Li, D. Survey of underwater robot positioning navigation. *Appl. Ocean Res.* **2019**, *90*, 101845. [\[CrossRef\]](#)
3. Lu, H.; Li, Y.; Serikawa, S. Underwater image enhancement using guided trigonometric bilateral filter and fast automatic color correction. In Proceedings of the 2013 IEEE International Conference on Image Processing, Melbourne, Australia, 15–18 September 2013; pp. 3412–3416.
4. Reza, A.M. Realization of the contrast limited adaptive histogram equalization (CLAHE) for real-time image enhancement. *J. VLSI Signal Process. Syst. Signal Image Video Technol.* **2004**, *38*, 35–44. [\[CrossRef\]](#)
5. Fu, X.; Fan, Z.; Ling, M.; Huang, Y.; Ding, X. Two-step approach for single underwater image enhancement. In Proceedings of the 2017 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS), Xiamen, China, 6–9 November 2017; pp. 789–794.
6. Huang, D.; Wang, Y.; Song, W.; Sequeira, J.; Mavromatis, S. Shallow-water image enhancement using relative global histogram stretching based on adaptive parameter acquisition. In Proceedings of the International Conference on Multimedia Modeling, Bangkok, Thailand, 5–7 February 2018; pp. 453–465.
7. Xiong, J.; Dai, Y.; Zhuang, P. Underwater Image Enhancement by Gaussian Curvature Filter. In Proceedings of the 2019 IEEE 4th International Conference on Signal and Image Processing (ICSIP), Wuxi, China, 19–21 July 2019; pp. 1026–1030.
8. Drews, P.L.; Nascimento, E.R.; Botelho, S.S.; Campos, M.F.M. Underwater depth estimation and image restoration based on single images. *IEEE Comput. Graph. Appl.* **2016**, *36*, 24–35. [\[CrossRef\]](#) [\[PubMed\]](#)
9. Islam, M.J.; Xia, Y.; Sattar, J. Fast underwater image enhancement for improved visual perception. *IEEE Robot. Autom. Lett.* **2020**, *5*, 3227–3234. [\[CrossRef\]](#)
10. Li, C.; Guo, C.; Ren, W.; Cong, R.; Hou, J.; Kwong, S.; Tao, D. An underwater image enhancement benchmark dataset and beyond. *IEEE Trans. Image Process.* **2019**, *29*, 4376–4389. [\[CrossRef\]](#)
11. Naik, A.; Swarnakar, A.; Mittal, K. Shallow-uwnet: Compressed model for underwater image enhancement (student abstract). In Proceedings of the AAAI Conference on Artificial Intelligence, Virtually, 22 February–1 March 2021; pp. 15853–15854.
12. Ancuti, C.; Ancuti, C.O.; Haber, T.; Bekaert, P. Enhancing underwater images and videos by fusion. In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012; pp. 81–88.
13. Ancuti, C.O.; Ancuti, C.; De Vleeschouwer, C.; Bekaert, P. Color balance and fusion for underwater image enhancement. *IEEE Trans. Image Process.* **2017**, *27*, 379–393. [\[CrossRef\]](#)
14. Li, H.; Li, J.; Wang, W. A fusion adversarial underwater image enhancement network with a public test dataset. *arXiv* **2019**, arXiv:1906.06819.
15. He, K.; Sun, J.; Tang, X. Single image haze removal using dark channel prior. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *33*, 2341–2353.
16. Chao, L.; Wang, M. Removal of water scattering. In Proceedings of the 2010 2nd International Conference on Computer Engineering and Technology, Chengdu, China, 16–18 April 2010; pp. V2-35–V32-39.
17. Chiang, J.Y.; Chen, Y.-C. Underwater image enhancement by wavelength compensation and dehazing. *IEEE Trans. Image Process.* **2011**, *21*, 1756–1769. [\[CrossRef\]](#)
18. Carlevaris-Bianco, N.; Mohan, A.; Eustice, R.M. Initial results in underwater single image dehazing. In Proceedings of the Oceans 2010 Mts/IEEE Seattle, Seattle, WA, USA, 20–23 September 2010; pp. 1–8.
19. Wang, Y.; Liu, H.; Chau, L.-P. Single underwater image restoration using adaptive attenuation-curve prior. *IEEE Trans. Circuits Syst. I Regul. Pap.* **2017**, *65*, 992–1002. [\[CrossRef\]](#)
20. Akkaynak, D.; Treibitz, T. Sea-thru: A method for removing water from underwater images. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 1682–1691.
21. Muniraj, M.; Dhandapani, V. Underwater image enhancement by combining color constancy and dehazing based on depth estimation. *Neurocomputing* **2021**, *460*, 211–230. [\[CrossRef\]](#)
22. Raveendran, S.; Patil, M.D.; Birajdar, G.K. Underwater image enhancement: A comprehensive review, recent trends, challenges and applications. *Artif. Intell. Rev.* **2021**, *54*, 5413–5467. [\[CrossRef\]](#)
23. Liu, Y.-C.; Chan, W.-H.; Chen, Y.-Q. Automatic white balance for digital still camera. *IEEE Trans. Consum. Electron.* **1995**, *41*, 460–466.
24. Iqbal, K.; Odetayo, M.; James, A.; Salam, R.A.; Talib, A.Z.H. Enhancing the low quality images using unsupervised colour correction method. In Proceedings of the 2010 IEEE International Conference on Systems, Man and Cybernetics, Istanbul, Turkey, 10–13 October 2010; pp. 1703–1709.
25. Ghani, A.S.A.; Isa, N.A.M. Underwater image quality enhancement through integrated color model with Rayleigh distribution. *Appl. Soft Comput.* **2015**, *27*, 219–230. [\[CrossRef\]](#)
26. Ghani, A.S.A.; Isa, N.A.M. Enhancement of low quality underwater image through integrated global and local contrast correction. *Appl. Soft Comput.* **2015**, *37*, 332–344. [\[CrossRef\]](#)
27. Buchsbaum, G. A spatial processor model for object colour perception. *J. Frankl. Inst.* **1980**, *310*, 1–26. [\[CrossRef\]](#)

28. Wong, S.-L.; Paramesran, R.; Taguchi, A. Underwater image enhancement by adaptive gray world and differential gray-levels histogram equalization. *Adv. Electr. Comput. Eng.* **2018**, *18*, 109–116. [[CrossRef](#)]
29. Fu, X.; Zhuang, P.; Huang, Y.; Liao, Y.; Zhang, X.-P.; Ding, X. A retinex-based enhancing approach for single underwater image. In Proceedings of the 2014 IEEE International Conference on Image Processing (ICIP), Paris, France, 27–30 October 2014; pp. 4572–4576.
30. Li, J.; Skinner, K.A.; Eustice, R.M.; Johnson-Roberson, M. WaterGAN: Unsupervised generative network to enable real-time color correction of monocular underwater images. *IEEE Robot. Autom. Lett.* **2017**, *3*, 387–394. [[CrossRef](#)]
31. Isola, P.; Zhu, J.-Y.; Zhou, T.; Efros, A.A. Image-to-image translation with conditional adversarial networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1125–1134.
32. Fabbri, C.; Islam, M.J.; Sattar, J. Enhancing underwater imagery using generative adversarial networks. In Proceedings of the 2018 IEEE International Conference on Robotics and Automation (ICRA), Brisbane, Australia, 21–25 May 2018; pp. 7159–7165.
33. Guo, Y.; Li, H.; Zhuang, P. Underwater image enhancement using a multiscale dense generative adversarial network. *IEEE J. Ocean. Eng.* **2019**, *45*, 862–870. [[CrossRef](#)]
34. Liu, X.; Gao, Z.; Chen, B.M. IPMGAN: Integrating physical model and generative adversarial network for underwater image enhancement. *Neurocomputing* **2021**, *453*, 538–551. [[CrossRef](#)]
35. Mo, Y.; Wu, Y.; Yang, X.; Liu, F.; Liao, Y. Review the state-of-the-art technologies of semantic segmentation based on deep learning. *Neurocomputing* **2022**, *493*, 626–646. [[CrossRef](#)]
36. Padilla, R.; Netto, S.L.; Da Silva, E.A. A survey on performance metrics for object-detection algorithms. In Proceedings of the 2020 International Conference on Systems, Signals and Image Processing (IWSSIP), Niterói, Brazil, 1–3 July 2020; pp. 237–242.
37. Yi, Z.; Zhang, H.; Tan, P.; Gong, M. Dualgan: Unsupervised dual learning for image-to-image translation. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2849–2857.
38. Yu, X.; Qu, Y.; Hong, M. Underwater-GAN: Underwater image restoration via conditional generative adversarial network. In Proceedings of the International Conference on Pattern Recognition, Beijing, China, 20–24 August 2018; pp. 66–75.
39. Ignatov, A.; Kobyshev, N.; Timofte, R.; Vanhoey, K.; Van Gool, L. Dslr-quality photos on mobile devices with deep convolutional networks. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 3277–3285.
40. Johnson, J.; Alahi, A.; Fei-Fei, L. Perceptual losses for real-time style transfer and super-resolution. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016; pp. 694–711.
41. Akkaynak, D.; Treibitz, T.; Shlesinger, T.; Loya, Y.; Tamir, R.; Iluz, D. What is the space of attenuation coefficients in underwater computer vision? In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4931–4940.
42. Panetta, K.; Gao, C.; Agaian, S. Human-visual-system-inspired underwater image quality measures. *IEEE J. Ocean. Eng.* **2015**, *41*, 541–551. [[CrossRef](#)]
43. Lowe, D.G. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110. [[CrossRef](#)]
44. Cheng, M.-M.; Mitra, N.J.; Huang, X.; Torr, P.H.; Hu, S.-M. Global contrast based salient region detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2014**, *37*, 569–582. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.