

Article

DFFA-Net: A Differential Convolutional Neural Network for Underwater Optical Image Dehazing

Xujia Hou, Feihu Zhang *, Zewen Wang, Guanglei Song, Zijun Huang and Jinpeng Wang

School of Marine Science and Technology, Northwestern Polytechnical University, Xi'an 710072, China; hxj1363947894@mail.nwpu.edu.cn (X.H.); wzv828@mail.nwpu.edu.cn (Z.W.); songguanglei@mail.nwpu.edu.cn (G.S.); zjhuang@mail.nwpu.edu.cn (Z.H.); qinlinglaosiji@mail.nwpu.edu.cn (J.W.)

* Correspondence: feihu.zhang@nwpu.edu.cn

Abstract: This paper proposes DFFA-Net, a novel differential convolutional neural network designed for underwater optical image dehazing. DFFA-Net is obtained by deeply analyzing the factors that affect the quality of underwater images and combining the underwater light propagation characteristics. DFFA-Net introduces a channel differential module that captures the mutual information between the green and blue channels with respect to the red channel. Additionally, a loss function sensitive to RGB color channels is introduced. Experimental results demonstrate that DFFA-Net achieves state-of-the-art performance in terms of quantitative metrics for single-image dehazing within convolutional neural network-based dehazing models. On the widely-used underwater Underwater Image Enhancement Benchmark (UIEB) image dehazing dataset, DFFA-Net achieves a peak signal-to-noise ratio (PSNR) of 24.2631 and a structural similarity index (SSIM) score of 0.9153. Further, we have deployed DFFA-Net on a self-developed Remotely Operated Vehicle (ROV). In a swimming pool environment, DFFA-Net can process hazy images in real time, providing better visual feedback to the operator. The source code has been open sourced.

Keywords: deep learning; image dehazing; underwater image processing; image analysis



Citation: Hou, X.; Zhang, F.; Wang, Z.; Song, G.; Huang, Z.; Wang, J. DFFA-Net: A Differential Convolutional Neural Network for Underwater Optical Image Dehazing. *Electronics* **2023**, *12*, 3876. <https://doi.org/10.3390/electronics12183876>

Academic Editor: Chunjie Zhang

Received: 8 August 2023

Revised: 31 August 2023

Accepted: 12 September 2023

Published: 14 September 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/>).

1. Introduction

The ocean is one of the largest ecosystems on Earth. Understanding the marine environment, including its biological and geological aspects, is of significant importance for environmental conservation, climate research, and resource development [1]. At present, underwater perception is mainly perceived by sonar at long distances and optical cameras at short distances [2,3]. However, underwater environments pose complex challenges around the propagation of light, including scattering, absorption, dispersion, and refraction [4,5]. These factors lead to issues such as blurriness, noise, and color distortion in underwater images, making underwater observation, photography, and visual perception difficult [6]. Therefore, underwater optical image dehazing is a research field that aims to improve the quality of underwater images and visual perception.

The field of image dehazing was initially extensively studied in the atmospheric environment. The atmospheric scattering model [7,8] can be described by Equation (1)

$$I(x) = t(x) \cdot J(x) + (1 - t(x)) \cdot A \quad (1)$$

where $I(x)$ represents the hazy image, $J(x)$ denotes the desired haze-free image, $t(x)$ represents a transmission factor that accounts for the light propagation, and A is the atmospheric light; $t(x)$ can be expressed as $t(x) = e^{-\beta d(x)}$, where $d(x)$ denotes the propagation distance and β represents the spectral attenuation coefficient. Building upon the foundation laid by Equation (1), we can derive the expression for the clear image $J(x)$, as described in Equation (2). Among the influential algorithms rooted in atmospheric environment models,

the Dark Channel Prior (DCP) [9] stands out. This algorithm's underlying observation is that in the majority of non-sky regions there exists at least one color channel with a low value. A and $t(x)$ are estimated according to this prior information, thereby obtaining the clear image $J(x)$.

$$J(x) = \frac{I(x) - A}{t(x)} + A \quad (2)$$

The core of traditional dehazing approaches lies in estimating the parameters A and $t(x)$. Thanks to the powerful fitting capability of neural networks, several end-to-end methods have been proposed, including AODNet [10], GCANet [11], FFANet [12], and more. These methods represent remarkable dehazing models for atmospheric environments, wherein a hazy input image is fed into the neural network to obtain a direct output of a clear image. The imaging principles of underwater environments share similarities with those of hazy image formation in atmospheric environments. Therefore, the aforementioned methods can be applied to underwater scenarios as well. By adapting and extending these techniques, they can be effectively utilized for underwater optical image dehazing.

One significant reason for the degradation of underwater images is the rapid attenuation of red light, resulting in low luminance values in the red channel of hazy underwater images. In this paper, we leverage this phenomenon as a prior information for underwater image dehazing. Building upon the FFA-Net architecture, we introduce two novel differential modules, one for B-R and one for G-R. The B-R and G-R modules capture the mutual information between the blue and red channels and between the green and red channels, respectively. Considering the statistical distribution of the hazy and clear images in the RGB channels, we designed a simple color channel-sensitive loss function. This loss function effectively guides the statistical distribution of the hazy image's color channels to better approximate those of the clear image, leading to an advanced underwater dehazing model. By incorporating these modifications, our proposed approach enhances the FFA-Net model, enabling it to better handle the specific challenges of underwater optical image dehazing and achieving superior performance in terms of both color fidelity and image quality. Furthermore, we have deployed the DFFA-Net on our self-developed ROV with the aim of enhancing the quality of the optical images available to the ROV operators.

In summary, in this paper we make the following key contributions:

- We propose a novel underwater image dehazing model, DFFA-Net, which utilizes color channel mutual information. Building on the FFA-Net architecture, we introduce the proposed color differential modules to facilitate learning of the domain mapping between hazy images and clear images.
- We introduce a color channel-sensitive loss function that guides the neural network to better align the color statistical distribution of hazy images with that of clear images.
- Compared to other convolutional neural network-based dehazing models, our proposed model demonstrates superior performance in terms of evaluation metrics. Additionally, we have made the code publicly available.
- We have successfully deployed the DFFA-Net on an ROV, allowing us to effectively obtain dehazed images. In a swimming pool environment, DFFA-Net provides real-time output images that ensure good camera image quality.

2. Related Work

Presently, image dehazing methods fall into two main categories: atmospheric scattering model-based solutions and data-driven domain style learning. Among the atmospheric scattering model-based methods, exemplified by the DCP [9], parameter estimation occurs through a comparison of the disparities between hazy and clear images within the dark channel. These methods leverage the dark channel prior in conjunction with Equation (1) to derive a clear image. While this approach generally proves effective, it may exhibit subpar performance in intricate scenarios, primarily because of its dependence on scene-specific assumptions and priors. Numerous dehazing techniques rely on prior knowledge, including methods that utilize the non-correlation between object surface chromaticity and

the transfer function to estimate transmittance [13]. However, the accuracy of estimation can be affected when the amount of color in the image is small. The Color Attenuation Prior (CAP) [14] is a dehazing method that relies on prior knowledge related to color attenuation. It makes use of the correlations between brightness, saturation, and depth of field to extract relevant information for the dehazing process. Nevertheless, it is important to note that this method faces challenges in terms of sample collection and possesses an incomplete theoretical foundation. The Non-Local Denoising (NLD) [15] method is a global transmittance estimation technique that relies on non-local color prior information. This approach can effectively derive depth of field information and dehaze images. However, it may struggle to accurately detect hazy lines in scenarios with intense illumination. On the other hand, the Boundary Constraint and Contextual Regularization (BCCR) [16] stands out as an efficient image dehazing approach. It incorporates boundary constraints and contextual regularization, resulting in the preservation of finer image details. Radiance–Reflectance Optimization (RRO) [17] combines both radiation and reflection components. It adopts a structure-oriented approach, and refines additional norm filters for enhanced image dehazing. In summary, atmospheric scattering model-based methods are generally effective in achieving significant results in image dehazing. However, their robustness is limited because they rely heavily on expert knowledge and specific assumptions about the scene.

Data-driven domain-style learning approaches offer an alternative method for image dehazing. They operate through supervised learning, circumventing the need to estimate parameters such as A and $t(x)$. Instead, these approaches directly acquire the haze-free image through neural networks. DehazeNet [18] is an end-to-end network for estimating transmittance in which the maxout network is added to the feature extraction layer of the Convolutional Neural Network (CNN) deep architecture. DehazeNet represents an amalgamation of pertinent dehazing theories, including the color attenuation prior, dark channel prior, hue difference, and maximum contrast. It additionally introduces a novel nonlinear activation function into its architecture, contributing to the enhancement of image quality in the dehazing process. AOD-Net [10] is a pioneering model in the field of dehazing, introducing a groundbreaking approach by substituting a novel parameter for the transmission matrix and atmospheric light parameters. This innovation allows for end-to-end computation using a simple neural network. Water-Net [19] introduces a gated fusion-based CNN designed to adapt to the characteristics of degraded underwater images. Water-Net consists of feature transformation units composed of white balancing, gamma correction, and histogram equalization methods, ultimately achieving underwater image enhancement. Multi-Scale CNN (MSCNN) [20] employs a multi-scale neural network approach to achieve precise estimation of the transmission map. GCANet [11] introduces an end-to-end contextual aggregation network that utilizes dilated convolutions to prevent grid distortion and incorporates gated subnetworks for feature fusion at various levels. There approaches utilizing adversarial neural networks have been proposed for underwater dehazing as well [21]; however, these require depth information as input.

3. Method

The task of underwater image dehazing requires taking a hazy underwater image as input and generating a clear image as output. In this paper, we propose DFFA-Net to address this task. The network architecture of DFFA-Net is illustrated in Figure 1.

DFFA-Net has three branches: the main branch, the G–R branch, and the B–R branch. Each branch is composed of a convolution layer and several group structure modules. The input of the main branch is an RGB three-channel image; the number of output channels is increased to 64 after the first convolution layer, while the width and height remain unchanged. The features are further extracted through three group structures, each of which maintains the same topology and is composed of N block structures, a convolution layer, and local residual connections. The input of the G–R branch and the B–R branch consists of the information shared between the green and red channel and the blue and red

channel, respectively. The G–R branch and B–R branch first increase the number of channels to three through a convolution layer, then further increase the number of channels to 64 through another convolution layer. The two branches then pass through a group structure to obtain the output. After the output of the three branches is spliced in the channel dimension, it is sent to the Feature Attention (FA) module. The FA module is composed of the channel attention and pixel attention. Finally, the dehazed image is obtained by postprocessing the two convolution layers and the global residual.

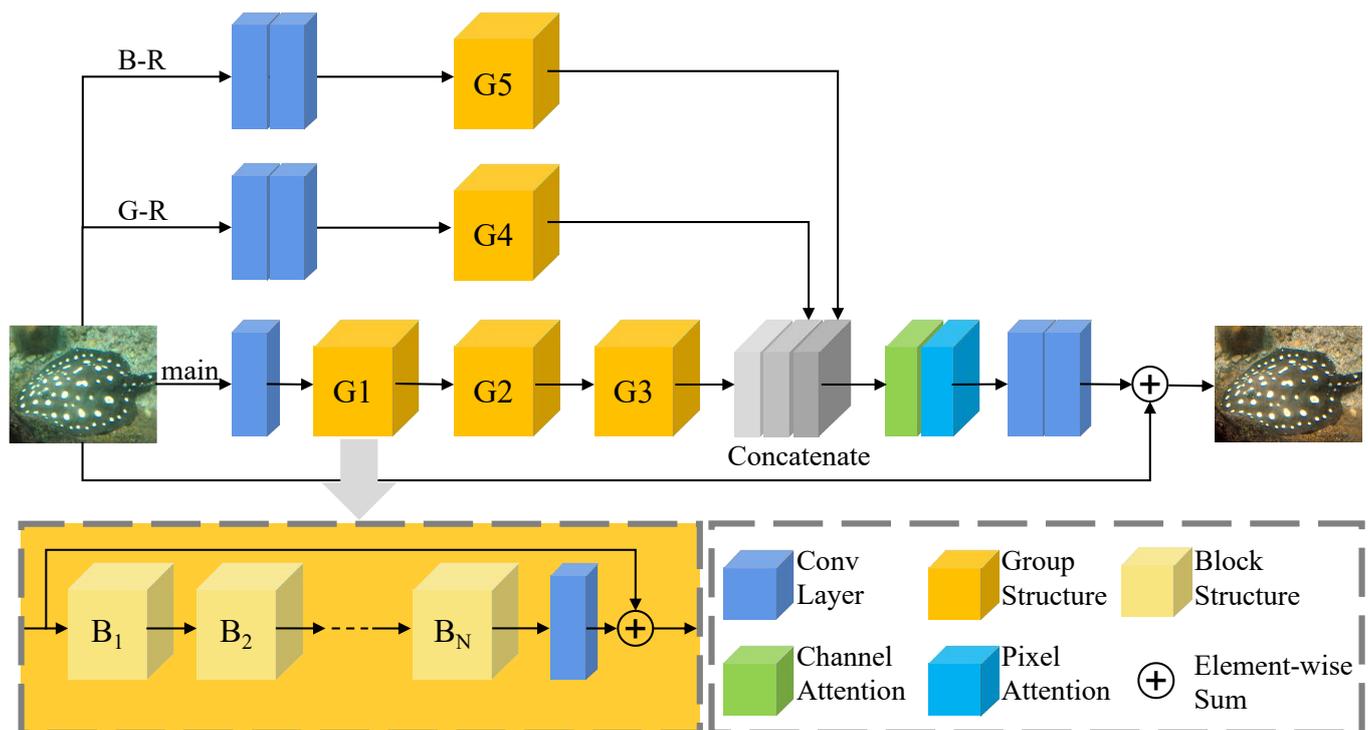


Figure 1. Network structure diagram of DFFA-Net. After obtaining the input image, DFFA-Net first extracts the corresponding features through three branches (main, G–R, and B–R), then these features are spliced and the degree of hazy of pixels at different positions is further estimated by the FA module. Finally, the clear image is obtained by two convolution layers and the global residual.

3.1. Feature Attention (FA)

In previous dehazing methods, the influence of different channels and pixels was assumed to be consistent throughout the image. However, in reality the distribution of haze across an image is uneven due to variations in the depth of field. Therefore, FFA-Net introduces the FA module to address this issue. The FA module consists of Channel Attention (CA) and Pixel Attention (PA) components.

CA is responsible for determining the weights of different channels in the overall performance of the network. The implementation process of CA is as follows:

- 1 A global adaptive average pooling layer is applied to obtain a $B \times C \times 1 \times 1$ output.
- 2 Two convolutional layers are used to downsample the number of channels from C to $C//8$ and then upsample it back to C . The downsampled output is then further passed through a ReLU activation function and the upsampled output is passed through a sigmoid activation function.
- 3 The original input is multiplied element-wise by the output from Step 2.

PA, on the other hand, determines the weights for different pixel positions corresponding to the dehazing effect. The implementation process of PA is as follows:

- 1 Two convolutional layers are used to downsample the number of channels to $C//8$ and 1, respectively. The output of the first convolutional layer is passed through a

ReLU activation function and the output of the second convolutional layer is passed through a sigmoid activation function.

- 2 The original input is multiplied element-wise by the output from Step 1.

By incorporating CA and PA the network is able to selectively treat different channels and pixel regions, allowing for more flexible and adaptive dehazing performance.

3.2. Group Structure

The group structure in DFFA-Net consists of N block structures, a Conv layer, and a local residual connection. The block structure is the fundamental module of DFFA-Net, and is composed of a local residual connection and FA module, as shown in Figure 2. This block structure has been proven to enhance network performance and training stability. By employing multiple local residual connections, the network can focus on extracting more relevant information. The cascaded combination of N basic structures increases the network's depth, providing stronger learning capabilities. Additionally, the use of multiple residuals helps to overcome training difficulties. In the main branch of DFFA-Net, N is set to 10.

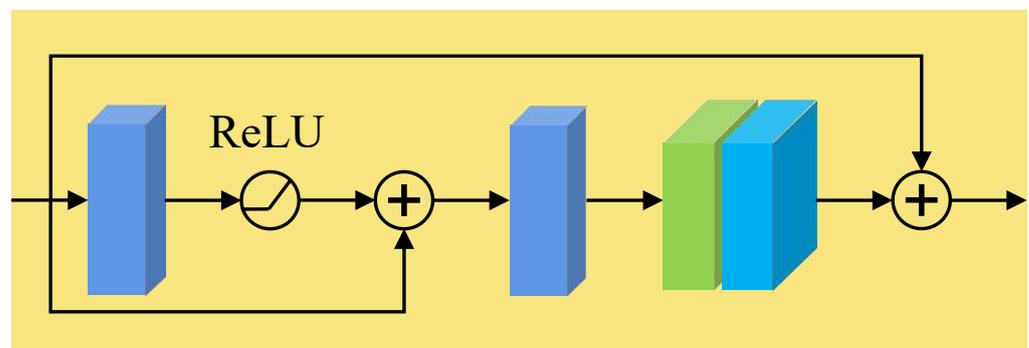


Figure 2. Basic network structure diagram.

3.3. B–R and G–R

The design of the differential modules (B–R and G–R) is a key aspect of DFFA-Net. As shown in Figure 3, the color distribution of hazy underwater images is distorted due to the rapid attenuation of red light in water. In clear underwater images, this phenomenon is resolved. Therefore, underwater image dehazing can be understood as a transformation performed on the statistical distribution of the RGB channels.

Considering that the green and blue channels exhibit relatively consistent distributions in both clear and hazy images, we designed two different branches outside the main branch of the network. The respective inputs to these branches consist of the differences between the blue channel and the red channel and the differences between the green channel and the red channel. We believe that these two branches are able to capture the mutual information between the blue/green channels and the red channel, thereby helping the neural network to transform the statistical distribution of the RGB channels.

Similar to the main branch, the outputs of the B–R and G–R modules pass through a Conv layer and a group structure before being merged into the main branch. In this case, $N = 6$ for the group structure in the B–R and G–R modules.

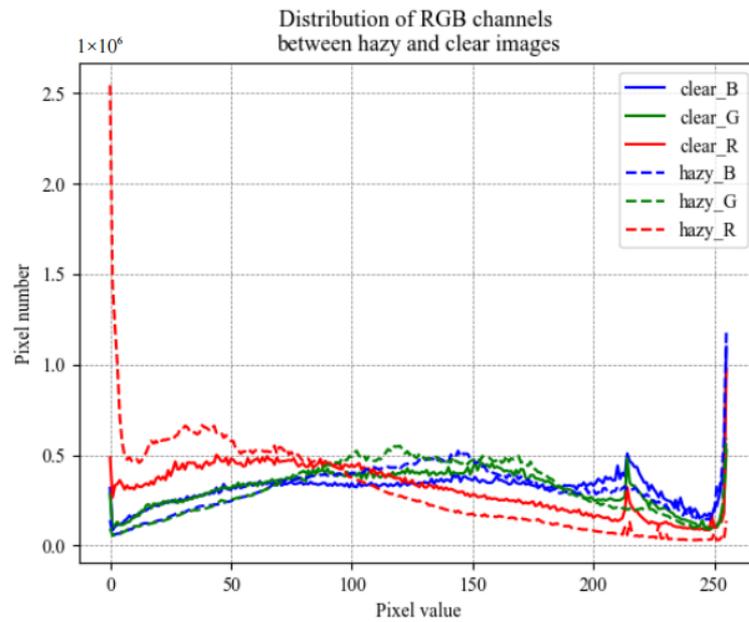


Figure 3. Statistical distribution comparison of underwater hazy images and clear images in RGB channels.

3.4. Loss

For the purposes of this paper, we designed a combined loss function consisting of two parts. The first part is designed to be sensitive to the color channel for the color distribution phenomenon in Figure 3, and can be expressed as follows:

$$L_{channel} = \sum \omega_i \sum |C_{gt}^i - pred(C_{haze}^i)|, i = R, G, B \tag{3}$$

where C_{haze}^i represents the matrix of the hazy image in the corresponding color channel, $pred$ represents the output obtained by DFFA-Net, C_{gt}^i represents the matrix of the label image in the corresponding color channel, and ω is a hyperparameter, with $\omega_R = 0.6$ and $\omega_G = \omega_B = 0.2$ in the experiments.

The second part is the Learned Perceptual Image Patch Similarity (LPIPS) [22], which has been proven to be a reliable loss function in the field of image reconstruction. LPIPS uses a trained VGG19 network [23] to pass two images through the network and calculate the output value difference in each layer of the network, as shown Equation (4):

$$d(X_1, X_2) = \sum_l \frac{1}{H_l W_l} \sum_{h,w} \left\| w_l \odot \left(\hat{Y}_{1hw}^l - \hat{Y}_{2hw}^l \right) \right\|_2^2 \tag{4}$$

where X_1 and X_2 represent the two images of the input VGG, d is the distance between X_1 and X_2 , H and W are the feature height and width in the corresponding l layer, \odot stands for the inner product, \hat{Y}_{1hw}^l and \hat{Y}_{2hw}^l are the outputs obtained by the l layer of the VGG location in (h, w) , and w_l is the scaling factor.

Therefore, the total loss function can be expressed as follows:

$$L_{loss} = L_{channel} + W \cdot d(X_1, X_2) \tag{5}$$

where $W = 0.04$.

4. Experiment and Analysis

4.1. Dataset

We utilized the popular UIEB as our dataset [19]. UIEB consists of hazy images and corresponding clear images. The hazy images consist of 890 real-world underwater images

captured under natural light, artificial light, or a combination of both. The clear images were obtained by applying twelve image enhancement methods to each hazy image, and the ground truth (GT) was selected based on subjective judgments from volunteers among all the enhanced results. The twelve methods specifically refer to: (1) fusion-based [24]; (2) two-step-based [25]; (3) retinex-based [26]; (4) DCP; (5) UDCP [27]; (6) regression-based [28]; (7) GDCP [29]; (8) red channel-based [30]; (9) histogram prior [31]; (10) blurriness-based [32]; (11) MSCNN [20]; and (12) dive+ [33]. We split the UIEB dataset into training, validation, and testing sets in a ratio of 700:90:100, respectively. Selected images are shown in Figure 4.



Figure 4. Selected images from the UIEB dataset, showing examples taken from rivers, oceans, and other environments. The photographed objects include marine life and underwater structures.

4.2. Training

Our experiments were conducted on the Ubuntu 20.04 LTS operating system using CUDA 11.6, Python 3.9, and PyTorch 1.12.1. The hardware setup included four RTX 3090 GPUs, 256 GB of RAM, and an Intel Xeon Silver 4210R processor. We set the initial learning rate to 0.0001 and used a cosine annealing strategy to slowly decay the learning rate to zero as a function of the cosine.

4.3. Evaluation Index of the Model

We used PSNR and SSIM to measure our dehazing performance. PSNR is a measure of image enhancement performance. With a given clean image X_1 of size $W \times H \times C$ and a noisy image X_2 , the mean squared error (MSE) is defined by Equation (6):

$$MSE = \frac{1}{W \cdot H \cdot C} \sum_{i=0}^{W-1} \sum_{j=0}^{H-1} \sum_{k=0}^{C-1} [X_1(i, j, k) - X_2(i, j, k)]^2 \quad (6)$$

while PSNR (dB) is defined by Equation (7):

$$PSNR = 10 \cdot \log_{10} \left(\frac{MAX_{X_1}^2}{MSE} \right) \quad (7)$$

where MAX_{X_1} is the maximum pixel value of the image. SSIM [34] measures the similarity between two images, mainly judged by focusing on the similarity of edges and textures. Its calculation formula is shown in Equation (8):

$$SSIM(X_1, X_2) = L(X_1, X_2) \times C(X_1, X_2) \times S(X_1, X_2) \quad (8)$$

where L represents the brightness similarity, C represents the contrast similarity, S represents the structure score, and L , C , and S are respectively calculated as in Equation (9).

$$\begin{aligned} L(X_1, X_2) &= \frac{2u_{X_1}u_{X_2} + C_1}{u_{X_1}^2 + u_{X_2}^2 + C_1} \\ C(X_1, X_2) &= \frac{2\sigma_{X_1}\sigma_{X_2} + C_2}{\sigma_{X_1}^2 + \sigma_{X_2}^2 + C_2} \\ S(X_1, X_2) &= \frac{\sigma_{X_1X_2} + C_3}{\sigma_{X_1}\sigma_{X_2} + C_3} \end{aligned} \quad (9)$$

In the above equations, u_{X_1} and u_{X_2} represent the mean of images X_1 and X_2 , σ_{X_1} and σ_{X_2} represent the standard deviation, $\sigma_{X_1X_2}$ represents the covariance, and C_1 , C_2 , and C_3 are constants used to avoid divisibility by 0; in these experiments, $C_1 = 0.01$, $C_2 = 0.03$, and $C_3 = C_2/2$.

4.4. Results

We conducted an investigation into the selection of the N parameter within the group structure. Specifically, we designed experiments for the main branch encompassing six different parameter selections along with four separate groups of experiments for the G-R and B-R branches. The results of these parameter choices were meticulously arranged and analyzed. The evaluation scores based on SSIM are depicted in Figure 5a, while those based on PSNR are illustrated in Figure 5b. Notably, it is evident that under constant conditions a larger N value within the main branch leads to higher scores on the evaluation metrics. However, a critical turning point is observed at $N = 10$, beyond which the rate of the score increase diminishes significantly. Consequently, we determined that $N = 10$ is the optimal choice for the main branch in DFFA-Net. Moreover, when we kept the N value fixed in the main branch we observed that both the B-R and G-R branches exhibited varying degrees of score improvement with increasing N values. Notably, $N = 6$ and $N = 10$ yielded nearly identical improvements in model performance, indicating that in these cases $N = 6$ in can satisfy the model's requirements. Thus, we selected $N = 6$ for both the B-R and G-R branches.

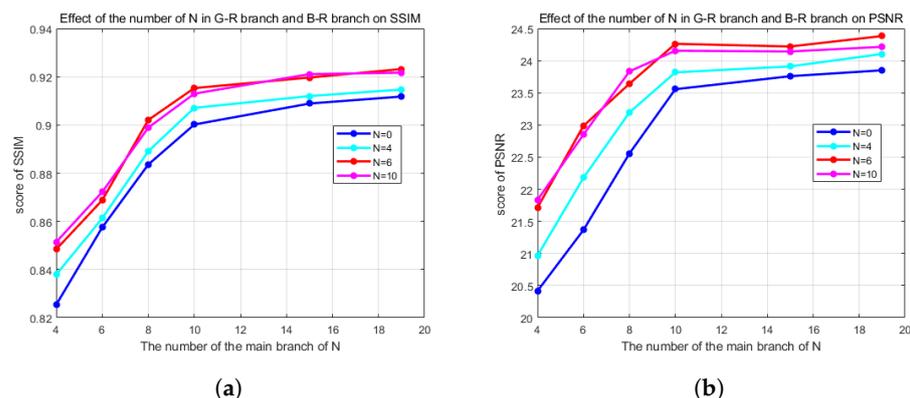


Figure 5. Effect of different values of of N on the group structure in terms of SSIM and PSNR. The horizontal axis shows the value of N in the main branch, while the value of N in the B-R and G-R branches is shown in the legend. The vertical axis shows the respective SSIM or PSNR score. (a) Effect of different values of N on the main branch, B-R branch, and G-R branch in terms of SSIM. (b) Effect of different values of N on the main branch, B-R branch, and G-R branch in terms of PSNR.

We conducted qualitative and quantitative experiments to compare our approach with five mainstream dehazing algorithms, including GCANet, Water-Net, and others. As shown in Figure 6, the DCP algorithm suffers from color distortion due to the failure of its underlying assumptions. AOD-Net tends to produce darker images in comparison. When compared to the aforementioned methods, Water-Net, GCANet, and FFA-Net exhibit

relatively superior dehazing performance. However, failure to correct the color channels results in persistent discrepancies with the GT color patterns. From a visual standpoint, our proposed method demonstrates the closest resemblance to the GT.

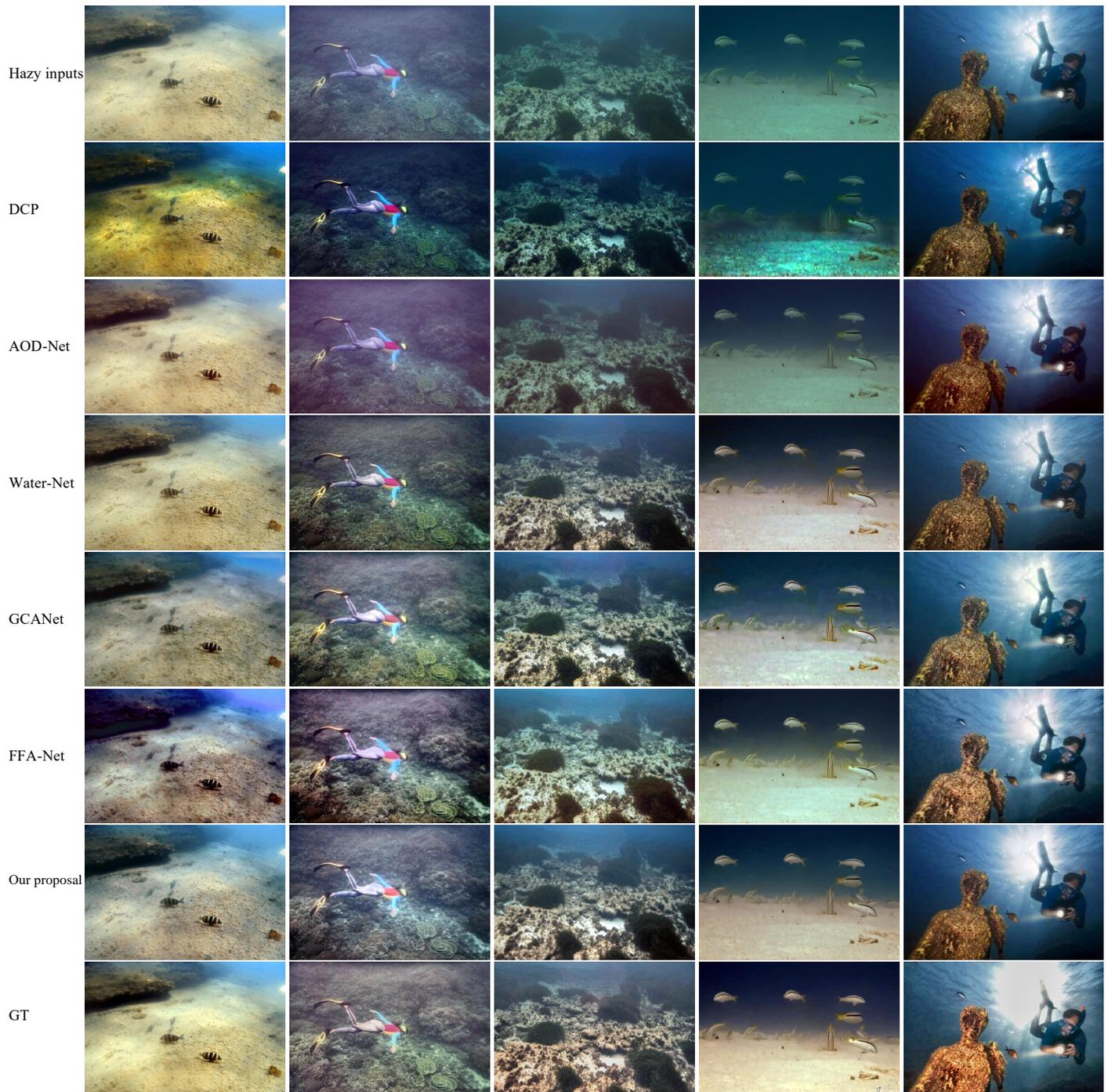


Figure 6. Comparison of results between our proposed approach and other dehazing algorithms based on convolutional neural networks.

The quantitative evaluation scores for the experimental results are presented in Table 1, with our method achieving the highest scores in terms of both SSIM and PSNR. In terms of model size, because DCP is implemented by conventional methods, its model size cannot be evaluated. AODNet takes the least time to process a single image. At the same time, it should be noted that our proposal shows improvements in all indicators compared with

FFA-Net. The accuracy of the model is improved while ensuring a high inference speed. These results indicate the effectiveness of our approach in underwater image dehazing, as it outperforms other methods.

Table 1. Comparison of different dehazing algorithms. \uparrow means that higher scores indicate better performance.

Method	SSIM \uparrow	PSNR \uparrow	Model Size (MB)	Inference Time (ms)
DCP	0.7468	15.2346	-	166.43
AODNet	0.8244	19.2519	0.14	29.13
WaterNet	0.9115	23.4502	4.17	68.86
GCANet	0.9025	23.1840	4.42	76.97
FFANet	0.9118	23.5567	17.74	110.81
Our proposal	0.9153	24.2631	13.20	90.21

5. Application

In this section, in order to further verify the reliability and practicability of our proposed DFFA-Net, we design a specific application experiment involving the application of our proposed method to a self-developed ROV.

5.1. ROV Design

The exterior of the ROV is shown in Figure 7. It has a streamlined shape and aesthetic design, taking into account body weight and structural strength. The ROV is equipped with inertial navigation and depth sensors as well as self-developed motion control algorithms for leading motion stability and accuracy. It has three motion modes (manual mode, stable mode, and fixed depth mode), allowing it to adapt to different types of job tasks. In terms of power, the ROV is equipped with four vertical thrusters and four horizontal thrusters, with six degrees of freedom characteristics. For observation and operation, the ROV is equipped with an underwater camera with a head, underwater lights, and a gripper located within the camera's field of view. The ROV performance characteristics are shown in Table 2.



Figure 7. Three-dimensional schematic model of the ROV.

Table 2. Functions and features of the ROV.

Maneuvering ability	Forward, backward, transverse, snorkeling, pitch, roll, yaw
Moving speed	Stepless speed regulation
Whether to support multiple maneuvering operations at the same time	Support
Attitude control	Pitch Angle holding function ($\pm 20^\circ$) Roll Angle holding function ($\pm 20^\circ$)
Input hold function	One key lock the current control execution command, so that the robot maintains the current speed, depth, heading Angle and attitude of continuous navigation
Control mode	Manual mode, stable mode, fixed depth mode
Video recording capability	Support
Hand bite force	10 Kg
Design depth	50 m
Endurance time	2 to 3 h

The ROV adopts the open rack layout. All control system components and other electronic circuit components are enclosed in watertight control compartments. A watertight connector is installed on the end cover of the watertight cabin to meet the need to route cables through the cabin. The lithium battery is packaged in a watertight battery compartment, and the power transmission cable transmits power to the control compartment through a watertight connector on the end cover. The design is equipped with eight thrusters (four vertical thrusters and four horizontal thrusters) to meet the needs of the robot's six degrees of freedom for movement (forward/backward, transverse, snorkeling, pitching, rolling and yawing). The robot claw and underwater light are installed in the front and lower part of the control pod, while the camera and camera head are installed in the head of the control pod and are observed through an optical acrylic sphere.

5.2. Control System Design

The control system has good hardware and software architecture, as shown in Figure 8. At the hardware level, the mode of an autonomous pilot and auxiliary computer is adopted. The autonomous pilot is responsible for all motion-related operations, including converting control instructions into Pulse Width Modulation (PWM) signals of each propeller, calculating attitude and depth and feedback, Proportion Integration Differentiation (PID) closed-loop control, etc. The autonomous pilot itself has built-in inertial navigation, which provides attitude angle and speed information in combination with depth sensor readings to provide the required feedback for the entire system. The autopilot is responsible for controlling the camera head, underwater lights and robot claws. The auxiliary computer is responsible for encoding, decoding and diverting the control information, ROV feedback information, video streams, etc. Remote real-time communication and image transmission between the ROV and the shore system are both realized through the power carrier module.

A modular design is implemented at the software level, as shown in Figure 9. The whole system includes the underlying drive module of the airborne application layer, its motion solution, communication, and video streaming modules, the operator client of the ground station, and the driver module of the handle of the ground station. All of the self-developed motion control algorithms are encapsulated in the motion solution module of the airborne application layer. Each module independently develops a shared interface, which can be called when calculating, ensuring the logical clarity and stability of the whole system.

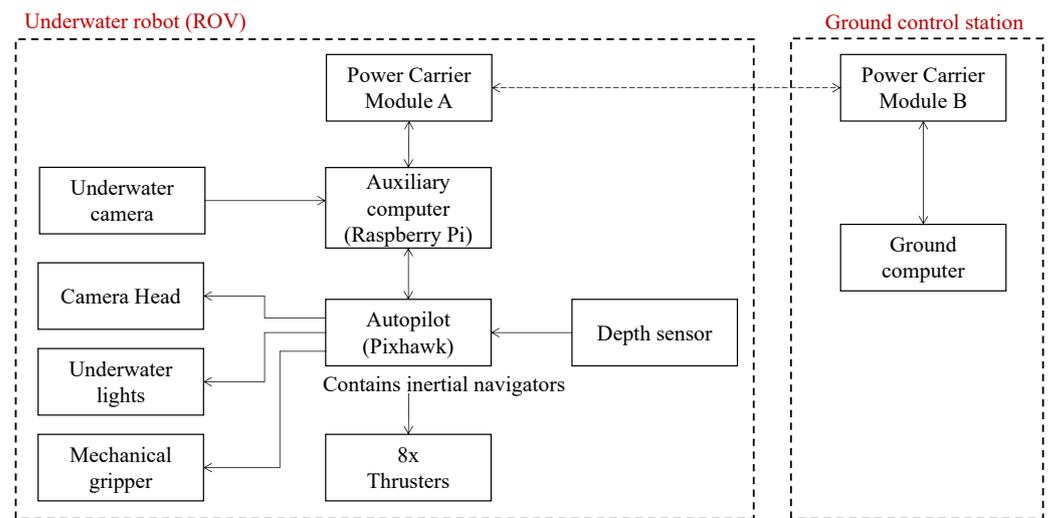


Figure 8. Control system hardware architecture.

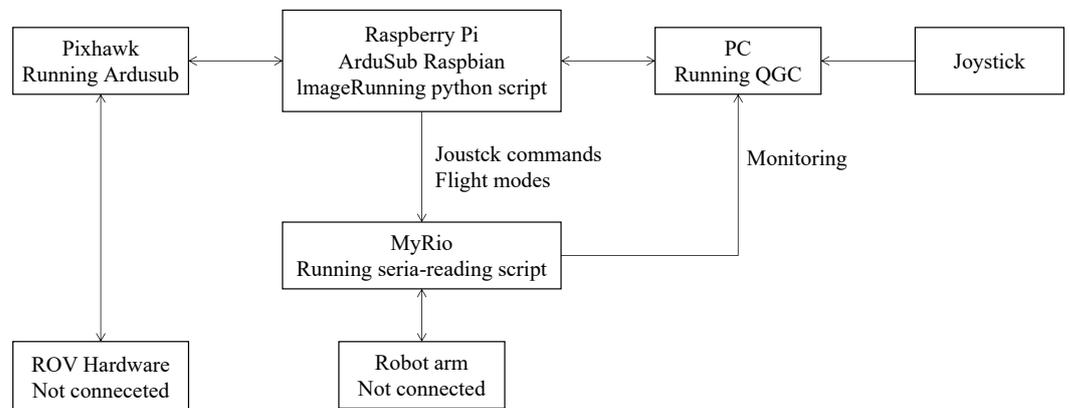
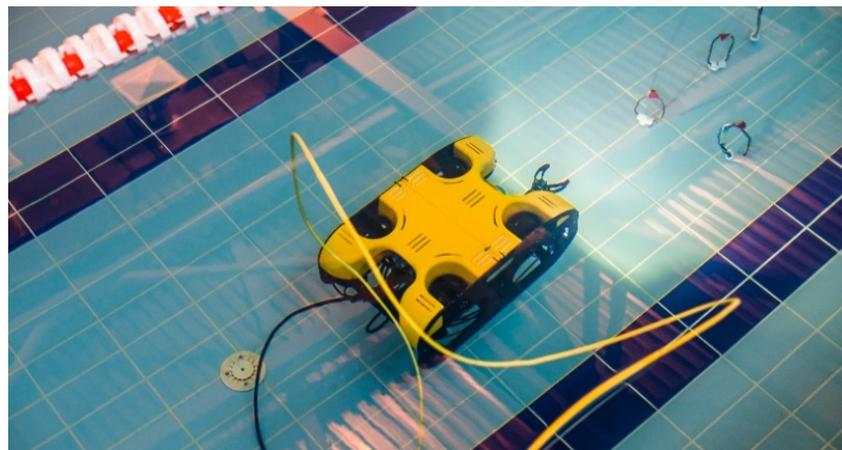


Figure 9. Control system software architecture.

5.3. Real-Time Underwater Image Dehazing

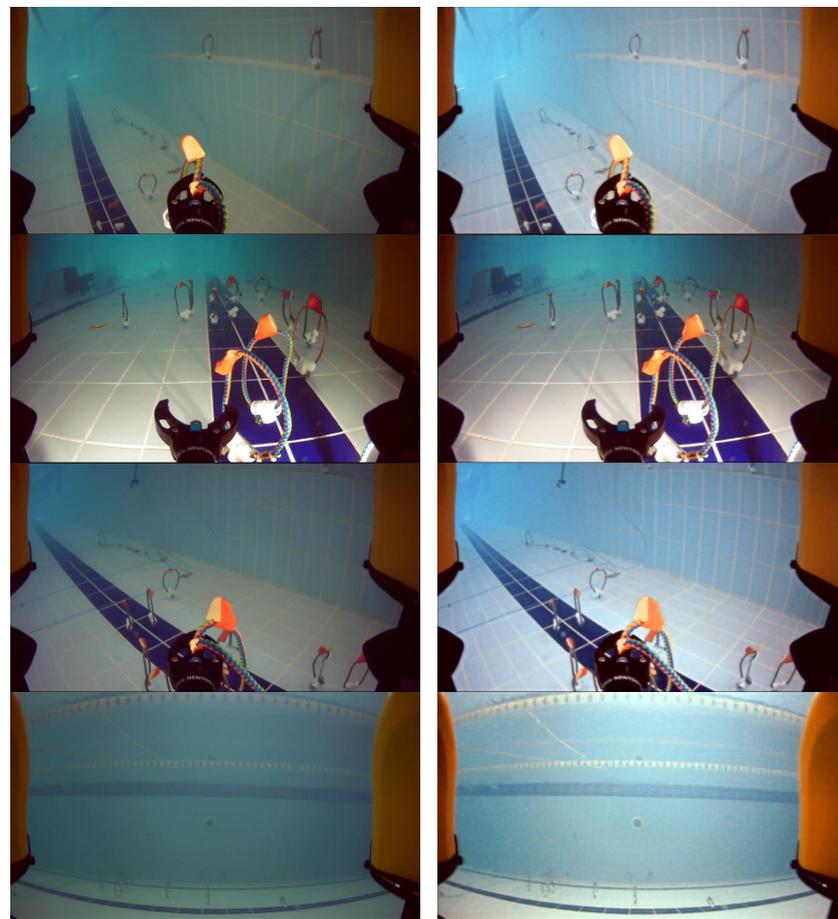
For a scenario involving the need to ensure good air tightness, we positioned the ROV within an indoor swimming pool. Various items were strategically placed within the pool for the ROV’s robotic arm to manipulate. The ROV’s live video feed was continuously transmitted to the ground computer via cable. However, it is worth noting that the image quality of the optical camera on the ROV did not meet our expectations due to the presence of chemicals such as copper sulfate and disinfectants within the pool water. To mitigate this issue, we employ DFFA-Net to process the video stream received from the ground computer in real time. The result of this processing, depicting the video stream after the removal of haze, is illustrated in Figure 10b. The image processed by DFFA-Net shows a significant improvement in color, contrast, and saturation. Compared with the hazy image directly received by ROV, the processed image has a longer perception range and better visual effect. With the help of DFFA-Net, the ROV operator can engage in remote control operation with greater efficiency.



(a)

The ROV camera reception

DFFA-Net output



(b)

Figure 10. Views obtained during the ROV experiment in the pool. The operator controlled the ROV by observing images from the optical camera and using the feed to perform tasks such as grabbing items. (a) View of the ROV. (b) Left: the original view obtained by the ROV camera; right: the view after DFFA-Net treatment.

6. Conclusions

In this paper, we introduce an end-to-end solution for underwater optical image de-hazing named DFFA-Net. To address the issue of imbalanced color channel distribution that is commonly encountered in underwater images, DFFA-Net innovatively incorporates

a differential module setup into its architecture. This modular setup plays a crucial role in effectively mitigating the aforementioned problem, ensuring improved image quality. Additionally, a novel loss function is devised to guide the network when learning the respective distributions of the transformations between the color channels. Our experimental results showcase the superiority of DFFA-Net, as it attains the highest scores in terms of both the PSNR and SSIM metrics when compared against other convolutional neural network-based dehazing methods. This success opens up possibilities for broader applications of our approach in related domains, including image reconstruction and super-resolution. Furthermore, we have successfully deployed DFFA-Net on our ROV. Through practical experiments, DFFA-Net demonstrates real-time processing capabilities for underwater images, producing clear and vivid optimized images. This can extend the operational vision of ROV operators, enhancing their capabilities in challenging underwater environments. Looking ahead, our future work will focus on optimizing the network architecture of DFFA-Net to achieve a more lightweight model in order to increase its efficiency and applicability in various underwater imaging scenarios.

Author Contributions: Methodology, F.Z. and X.H.; data curation, Z.W.; WaterNet, J.W.; AODNet, Z.H.; GCANet, G.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (52171322) and the Graduate Innovation Seed Fund of Northwestern Polytechnical University (PF2023066, PF2023067).

Data Availability Statement: The source code has been made available on: <https://gitee.com/hhsupremehh/dffa-net.git> (accessed on 25 June 2023).

Acknowledgments: We would like to acknowledge the facilities and technical assistance provided by the Key Laboratory of Unmanned Underwater Transport Technology.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Liu, Y.; Anderlini, E.; Wang, S.; Ma, S.; Ding, Z. Ocean explorations using autonomy: Technologies, strategies and applications. In Proceedings of the Offshore Robotics, Xi'an, China, 30 May–5 June 2021; Springer: Berlin/Heidelberg, Germany, 2022; pp. 35–58.
2. Egg, L.; Pander, J.; Mueller, M.; Geist, J. Comparison of sonar-, camera- and net-based methods in detecting riverine fish-movement patterns. *Mar. Freshw. Res.* **2018**, *69*, 1905–1912. [[CrossRef](#)]
3. Terayama, K.; Shin, K.; Mizuno, K.; Tsuda, K. Integration of sonar and optical camera images using deep neural network for fish monitoring. *Aquac. Eng.* **2019**, *86*, 102000. [[CrossRef](#)]
4. Schettini, R.; Corchs, S. Underwater image processing: State of the art of restoration and image enhancement methods. *Eurasip J. Adv. Signal Process.* **2010**, *2010*, 1–14. [[CrossRef](#)]
5. Sahu, P.; Gupta, N.; Sharma, N. A survey on underwater image enhancement techniques. *Int. J. Comput. Appl.* **2014**, *87*, 333–338. [[CrossRef](#)]
6. Zhang, W.; Zhuang, P.; Sun, H.H.; Li, G.; Kwong, S.; Li, C. Underwater image enhancement via minimal color loss and locally adaptive contrast enhancement. *IEEE Trans. Image Process.* **2022**, *31*, 3997–4010. [[CrossRef](#)] [[PubMed](#)]
7. Narasimhan, S.G.; Nayar, S.K. Chromatic framework for vision in bad weather. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2000 (Cat. No. PR00662), Hilton Head Island, SC, USA, 15 June 2000; IEEE: New York, NY, USA, 2000; Volume 1, pp. 598–605.
8. Narasimhan, S.G.; Nayar, S.K. Vision and the atmosphere. *Int. J. Comput. Vis.* **2002**, *48*, 233. [[CrossRef](#)]
9. He, K.; Sun, J.; Tang, X. Single image haze removal using dark channel prior. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *33*, 2341–2353.
10. Li, B.; Peng, X.; Wang, Z.; Xu, J.; Feng, D. Aod-net: All-in-one dehazing network. In Proceedings of the IEEE International Conference on Computer Vision, Vancouver, BC, Canada, 7–14 July 2017; pp. 4770–4778.
11. Chen, D.; He, M.; Fan, Q.; Liao, J.; Zhang, L.; Hou, D.; Yuan, L.; Hua, G. Gated context aggregation network for image dehazing and deraining. In Proceedings of the 2019 IEEE Winter Conference on Applications of Computer Vision (WACV), Waikoloa Village, HI, USA, 7–11 January 2019; IEEE: New York, NY, USA, 2019; pp. 1375–1383.
12. Qin, X.; Wang, Z.; Bai, Y.; Xie, X.; Jia, H. FFA-Net: Feature fusion attention network for single image dehazing. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; Volume 34, pp. 11908–11915.
13. Fattal, R. Single image dehazing. *Acml Trans. Graph. Tog* **2008**, *27*, 1–9. [[CrossRef](#)]

14. Zhu, Q.; Mai, J.; Shao, L. A fast single image haze removal algorithm using color attenuation prior. *IEEE Trans. Image Process.* **2015**, *24*, 3522–3533.
15. Berman, D.; Avidan, S. Non-local image dehazing. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 1674–1682.
16. Meng, G.; Wang, Y.; Duan, J.; Xiang, S.; Pan, C. Efficient image dehazing with boundary constraint and contextual regularization. In Proceedings of the IEEE International Conference on Computer Vision, Washington, DC, USA, 1–8 December 2013; pp. 617–624.
17. Shin, J.; Kim, M.; Paik, J.; Lee, S. Radiance–reflectance combined optimization and structure-guided ℓ_0 -Norm for single image dehazing. *IEEE Trans. Multimed.* **2019**, *22*, 30–44. [[CrossRef](#)]
18. Cai, B.; Xu, X.; Jia, K.; Qing, C.; Tao, D. Dehazenet: An end-to-end system for single image haze removal. *IEEE Trans. Image Process.* **2016**, *25*, 5187–5198. [[CrossRef](#)] [[PubMed](#)]
19. 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.* **2020**, *29*, 4376–4389. [[CrossRef](#)] [[PubMed](#)]
20. Ren, W.; Liu, S.; Zhang, H.; Pan, J.; Cao, X.; Yang, M.H. Single image dehazing via multi-scale convolutional neural networks. In Proceedings of the Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, 11–14 October 2016; Part II 14; Springer: Berlin/Heidelberg, Germany, 2016; pp. 154–169.
21. 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](#)]
22. Zhang, R.; Isola, P.; Efros, A.A.; Shechtman, E.; Wang, O. The Unreasonable Effectiveness of Deep Features as a Perceptual Metric. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018.
23. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv* **2014**, arXiv:1409.1556.
24. 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; IEEE: New York, NY, USA, 2012; pp. 81–88.
25. 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; IEEE: New York, NY, USA, 2017; pp. 789–794.
26. 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; IEEE: New York, NY, USA, 2014; pp. 4572–4576.
27. 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](#)]
28. Li, C.; Guo, J.; Guo, C.; Cong, R.; Gong, J. A hybrid method for underwater image correction. *Pattern Recognit. Lett.* **2017**, *94*, 62–67. [[CrossRef](#)]
29. Peng, Y.T.; Cao, K.; Cosman, P.C. Generalization of the dark channel prior for single image restoration. *IEEE Trans. Image Process.* **2018**, *27*, 2856–2868. [[CrossRef](#)]
30. Galdran, A.; Pardo, D.; Picón, A.; Alvarez-Gila, A. Automatic red-channel underwater image restoration. *J. Vis. Commun. Image Represent.* **2015**, *26*, 132–145. [[CrossRef](#)]
31. Li, C.Y.; Guo, J.C.; Cong, R.M.; Pang, Y.W.; Wang, B. Underwater image enhancement by dehazing with minimum information loss and histogram distribution prior. *IEEE Trans. Image Process.* **2016**, *25*, 5664–5677. [[CrossRef](#)]
32. Peng, Y.T.; Cosman, P.C. Underwater image restoration based on image blurriness and light absorption. *IEEE Trans. Image Process.* **2017**, *26*, 1579–1594. [[CrossRef](#)]
33. 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](#)] [[PubMed](#)]
34. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [[CrossRef](#)] [[PubMed](#)]

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.