

Article

Multi-Domain Rapid Enhancement Networks for Underwater Images

Longgang Zhao ¹ and Seok-Won Lee ^{1,2,*} 

¹ The Knowledge-Intensive Software Engineering (NiSE) Research Group, Department of Artificial Intelligence, Ajou University, Suwon City 16499, Republic of Korea; zhaolonggang@ajou.ac.kr

² Department of Software and Computer Engineering, Ajou University, Suwon City 16499, Republic of Korea

* Correspondence: leesw@ajou.ac.kr

Abstract: Images captured during marine engineering operations suffer from color distortion and low contrast. Underwater image enhancement helps to alleviate these problems. Many deep learning models can infer multi-source data, where images with different perspectives exist from multiple sources. To this end, we propose a multichannel deep convolutional neural network (MDCNN) linked to a VGG that can target multi-source (multi-domain) underwater image enhancement. The designed MDCNN feeds data from different domains into separate channels and implements parameters by linking VGGS, which improves the domain adaptation of the model. In addition, to optimize performance, multi-domain image perception loss functions, multilabel soft edge loss for specific image enhancement tasks, pixel-level loss, and external monitoring loss for edge sharpness preprocessing are proposed. These loss functions are set to effectively enhance the structural and textural similarity of underwater images. A series of qualitative and quantitative experiments demonstrate that our model is superior to the state-of-the-art Shallow UWnet in terms of UIQM, and the performance evaluation conducted on different datasets increased by 0.11 on average.

Keywords: underwater image enhancement; multi-domain machine learning; DCNN; domain adaptability; perceptual loss



Citation: Zhao, L.; Lee, S.-W. Multi-Domain Rapid Enhancement Networks for Underwater Images. *Sensors* **2023**, *23*, 8983. <https://doi.org/10.3390/s23218983>

Academic Editor: Inam Ullah

Received: 27 September 2023

Revised: 3 November 2023

Accepted: 3 November 2023

Published: 5 November 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

With the increasing demand for ocean cognition and situational awareness, exploring the mystery of ocean depths and obtaining high-quality underwater images are urgently needed by many marine enterprises [1,2]. Maritime enterprises rely on thousands of Internet of Things (IoT) sensors scattered underwater to collect data on the activity status of underwater creatures and moving images of seabed plates. This information is used to improve awareness of the continuous situation of the ocean. Owing to breakthroughs in key technologies such as big data, cloud computing, and the IoT, maritime enterprise information management can realize a clear perception of the entire underwater domain, providing important support for ocean exploration, national defense, and security [3].

In recent years, to solve engineering problems such as marine environmental monitoring, submarine surveying and mapping, submarine archaeological exploration, garbage collection, and underwater rescue, nations worldwide have vigorously developed underwater detection applications [4]. These applications require the real-time interpretation of images/videos so that underwater robots based on vision guidance can intelligently perceive the environment and set future execution tasks [5]. For example, Ref. [6] proposed an underwater image enhancement network named Ucolor, which uses medium transmission guidance, multicolor space embedding, and a combination of physical models and learning methods to solve the color deviation and low-contrast problems of underwater images. The authors of [7] proposed an underwater image enhancement method based on generative adversarial networks using a multiscale generator to generate clear underwater

images, effectively correcting color casts and contrast problems while protecting detailed information. In, [8] an L2UWE framework is proposed to efficiently enhance low-light underwater images by relying on local contrast and multiscale fusion technology to improve the clarity and brightness of the image. The authors of [9] constructed an underwater image enhancement benchmark (UIEB) and proposed the WaterNet underwater image enhancement network that can effectively correct color casts and restore image details. The authors of [10] proposed a generative adversarial network based on Pix2Pix and introduced technologies such as deep residual learning and multilayer perceptrons to remove the fog effect, correct color shift, and increase image details. Although these methods have made significant progress in improving the color cast, contrast, and brightness of underwater images, they fail to fully consider the relationship between the contrast, brightness, and color of underwater images and fail to adaptively balance these factors. Future research should continue to address this issue to improve the effectiveness of underwater image enhancement.

Contrast deterioration and color distortion in underwater images limit the application of underwater vision tasks [11]. In addition, the wavelength decays exponentially in deep water, resulting in the absorption of red wavelengths, leading to the dominance of green or blue hues in underwater images. These problems severely degrade the visual effects of underwater images. Consequently, the same scene in underwater images presents different background lights, forming a multi-domain problem [12]. For example, a dimly lit image is considered to be from the same domain and a well-lit image is considered to be from another domain.

Underwater images from different viewpoints or light backgrounds, as shown in Figure 1, can be regarded as data from different domains. Multi-domain machine learning shares model parameters through the model training of data in different domains to improve the model learning efficiency, which enables the enhancement of models for underwater images in different domains. However, most multi-domain machine learning methods assume the same distribution of data in different domains, ignoring non-IIDs, and such neglect leads to the inability of the model to make its image enhancement ability the strongest in each domain.



Figure 1. Multi-domain underwater image enhancement.

Scholars have proposed a variety of underwater image-enhancing methods, such as nonphysical and physical model-based methods [13]. The nonphysical model improves image quality by adjusting the image pixel value without relying on the underwater imaging model. However, nonphysical and physical models ignore the attenuation characteristics of underwater images in different domains and cannot achieve multi-domain underwater image enhancement. Data-driven methods use deep learning to learn nonlinear feature maps and enhance the underwater images. Convolutional neural networks (CNNs), which are widely used in computer vision tasks, are invariant to displacement and distortion, and have good recognition capabilities [14]. Many models based on CNNs (e.g., Shallow UWnet [15] and Uresnet [16]) and generative adversarial networks (GANs) (e.g., FUnIE GAN [17], Water Net [18], and Cycle-GAN [19]) have been used to enhance image quality by learning from a large amount of data. Significant progress has been made in image

super-resolution, denoising, deblurring, and dehazing techniques. However, deep learning models have a weak inference ability for solving multi-domain problems because they cannot effectively use the similarity of images between domains and ignore the local information of images in different domains, for example, the luminosity change in the same target [20].

This paper proposes a multi-domain underwater image fast enhancement model based on a multichannel deep convolutional neural network (DCNN) combined with a VGG network to improve the inference ability of multi-domain underwater image models called the MDCNN-VGG. The MDCNN-VGG has more layers and a more complex structure than the CNN. The DCNN can avoid the performance degradation of the network model caused by the data heterogeneity of underwater images in multiple domains. It uses DCNNs of different channels to mine the texture and color of underwater images in different domains, and this information is fed into the VGG. The VGG recognizes and classifies data elements in multiple domains to obtain specific feature representations in each domain, thereby achieving the rapid enhancement of underwater images.

The contribution of this study is as follows:

We designed a multi-domain underwater image enhancement model with a multi-channel DCNN linked to a VGG; specifically, the different network streams designed in the DCNN shared parameters through back-and-forth passing to enhance domain adaptation. The importance of different model parameters is also selected in the soft mask configuration model, such that important model parameters (e.g., texture structure and color) are input to the VGG, which in turn yields a specific feature representation in each domain to enhance underwater images.

To optimize the performance of the MDCNN-VGG, we designed a perceptual loss function for multi-domain underwater image enhancement. Multilabel soft-margin loss is used for specific tasks, and VGG perceptual loss is used for external supervision to achieve pixel-level loss and for preprocessing edge sharpness, thereby enhancing the structure and texture similarity of underwater images. In turn, we can optimally adjust the coefficients in the perceptual loss function to control the involvement of different functional loss terms in the model-training process to detect the focal region of the input image for target class enhancement.

Qualitative and quantitative experiments showed that the enhancement effect of this model on underwater image quality was better than that of the benchmark model.

2. Literature Review

2.1. Deep Learning

Image enhancement is a topic of significant interest within the fields of computer vision, signal processing, and others. Earlier works used artificially created filters to enhance local colors for contrast/brightness improvement and global enhancement based on scene assumptions (e.g., fog lines, dark channels, etc.). With the development of deep learning and driven by large-scale datasets, image enhancement has been remarkably successful. For example, deep CNN-based models have achieved superior performance in solving image coloring, color/contrast adjustment, and de-cluttering [21]. The recently proposed Shallow UWnet is a gated fusion CNN trained on the UIEB dataset for underwater image enhancement [22]. UResnet, a CNN-based residual network, is proposed in [23] as a more comprehensive supervised learning method for underwater image enhancement. In addition, GAN-based image style transformation or generation has achieved great success. For example, the recently proposed FUnIE GAN is assumed to perform nonlinear mapping between distorted and enhanced images, and it removes image blurring by constructing an image enhancement problem [24]. The conditional GAN proposed in [25] learns image enhancement through generalized training on large-scale datasets. In contrast, bidirectional GANs (e.g., CycleGAN, DualGAN [26], etc.) address the practical application of image enhancement by using a cyclic consistency loss. However, the above studies do not take

full advantage of the similarity of multi-domain images taken from different angles of the same target.

2.2. Physical-Based Methods

Traditional physics-based methods use atmospheric scattering models to estimate light transmission and ambient temperature scenes to recover the true pixel intensity [27,28]. The underwater image-defogging algorithm proposed in [29] reduces information loss in the output image [30]. In addition, multiband fusion-based enhancement [31] and blurred line-based color recovery (Uw-HL [32]) have been used to recover underwater image quality.

2.3. Nonphysical-Based Methods

Nonphysical-based methods directly modify image pixel values to produce satisfactory results without using a physical degradation model. In citeancuti2012 enhancing, a fusion-based method was proposed that can improve contrast and visual effects; however, an over-enhancement phenomenon exists. Recently, Ref. [33] modified [34] by introducing a novel white-balancing method to reduce over- and under-enhancement. Another study is based on the retinex model, [35] and uses color correction, post-enhancement, and layer decomposition to improve the underwater image quality.

In summary, deep learning techniques have achieved state-of-the-art performances in image enhancement tasks and can automatically learn relevant features from large datasets, thereby reducing the need for handcrafted features. However, deep learning models often require substantial computational resources and large training datasets. Overfitting can be a concern if the training data are not representative of the target domain [36].

Physics-based methods provide a solid theoretical foundation for image enhancement, allowing for accurate modeling of physical degradation. Physics-based methods are sensitive to the accuracy of assumed physical models, and deviations from these models can lead to errors.

Nonphysical-based methods are versatile and do not rely on explicit physical models, making them more flexible in a wider range of scenarios. An overreliance on nonphysical methods may lead to image artifacts or unrealistic enhancements.

3. MDCNN-VGG Hybrid Model Architecture

3.1. Overall

In this study, the MDCNN-VGG is proposed to make full use of the differential distribution of information in different domains while combining it with its parameter-sharing mechanism to enhance domain adaptability [29]. Extensive qualitative and quantitative experimental results demonstrate that the MDCNN-VGG has better underwater image quality enhancement than the benchmark model.

As shown in Figure 2, the proposed MDCNN-VGG consists of a multichannel DCNN with a VGG-16 model that combines the advantages of neural network classifiers and VGG [33] perceived loss.

The MDCNN consists of multiple DCNNs in parallel, and the specific structure of each DCNN, which consists of multiple fully connected CNN layers, is shown in Figure 2. DCNNs are applied using the same principles as traditional CNNs, which employ alternating convolutional layers and pooling in their network structure with fully connected network ends. The most distinguishable features are extracted from the original input images using supervised learning. The effective subregions are computed from the original underwater images of different domains using the perceptual field features of the DCNN [37]. To enhance the model domain adaptation capability, two DCNN network streams are used to share parameters between them, and the importance of different parameters of the model is configured using a soft mask to enhance the information of the network stream. Information such as the texture, structure, and color of the underwater images from different domains are mined and fed into the subsequent VGG. We set different CNN channels so

that in order to better extract underwater images in different domains, we map different channels to underwater images in different domains, and then perform feature fusion after different channels, so that the model can better obtain specially issued useful images, as well as better differentiation of different areas of underwater images.

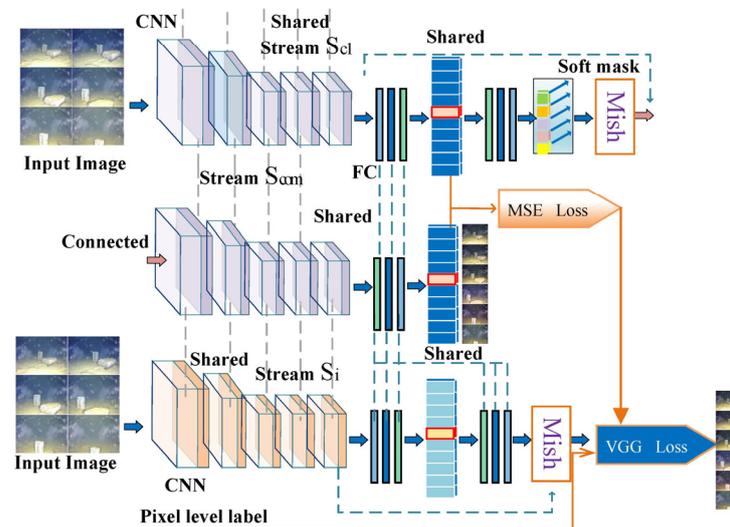


Figure 2. MDCNN-VGG architecture.

The VGG-16 identifies and classifies data elements belonging to different underwater background categories to obtain a feature representation of an underwater image for each domain. Specifically, the MDCNN-VGG identifies and classifies data elements belonging to different underwater background categories (e.g., water bodies appearing blue–green or dark blue, rocks underwater appearing silver) based on VGG perceptual loss for each domain.

3.2. Single-Channel DCNN

The architecture of the single-channel DCNN model is illustrated in Figure 3. The network model contains two network streams: S_{cl} and S_{com} . S_{cl} was developed with the goal of searching for regions that contribute to the identification of target objects in underwater images, and S_{com} ensures that all regions favorable for identification are found.

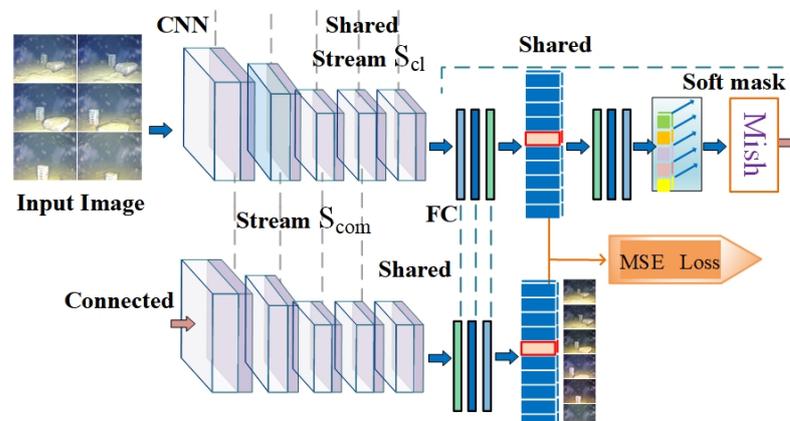


Figure 3. Single-channel DCNN framework diagram.

The network stream S_{cl} contains classical optimization techniques such as pooling, dropout, and other settings. The CNN is immediately followed by three fully connected layers. The parameters of this fully connected layer are shared with those of the fully

connected tail layer. Immediately thereafter, the soft mask feature (see attention mechanism [35]) is used to configure the importance of the different model parameters, allowing attention operations to be performed on different domain underwater images in the DCNN for the task of interest. In other words, the soft mask of the network is trained in an end-to-end manner to achieve precise enhancement of the attention content (e.g., the specific color enhancement of the corresponding target), and the soft mask has an enhancing effect on the information of the previous network stream, whose activation function uses Mish [38]. The results are fed into the subsequent network stream based on the idea of Resnet [39] that links the two network streams before and after, thereby avoiding gradient disappearance and sharing the model parameters of the CNN of the preceding network stream S_{cl} into the CNN of the following network stream. The specific implementation of parameter sharing in the model structure is shown by linking S_{cl} and S_{com} in Figure 3.

Specifically, in network stream S_{cl} , for a given domain of underwater images, I , $f_{l,k}$ is represented as the activation function of unit k in the l th layer of the soft mask. It can be seen that $f_{l,k}$ obtains the classification probability corresponding to domain category c for each underwater image to be enhanced, and the gradient obtained based on the activation function is used to update the weight, $\omega_{l,k}^c$, of the neurons through global average pooling, as shown in Equation (1).

$$\omega_{l,k}^c = GAP\left(\frac{\partial \omega^c}{\partial f_{l,k}}\right) \quad (1)$$

where $GAP(\cdot)$ denotes the global average pooling operation; at this time, there is no need to use the backward pass method to obtain $\omega_{l,k}^c$. $\omega_{l,k}^c$ denotes the importance of $f_{l,k}$ support c -class underwater image enhancement in the soft mask. To enhance the generalization ability of the overall network model, the weight is used to represent the importance of c -class image probability, the 2D convolution operation is performed on all f_l s to integrate the soft mask with the activation output of all layers, and then the Mish [38] operation is performed to obtain the soft mask (A^C)

$$A^C = \text{Mish}(\text{conv}(f_l, \omega^c)) \quad (2)$$

The soft mask applied to the original input underwater image is obtained through A^C using Equation (3) to obtain I^{*c} , which represents the semantic information in the c -class underwater image on which the network model is currently focused.

$$I^{*c} = I - (T(A^c) \odot I) \quad (3)$$

where $T(A^c)$ is a masking factor based on a threshold, and the sigmoid function is used as an approximation, as defined in Equation (4).

$$T(A^c) = \frac{1}{1 + \exp(-\omega(A^c - \sigma))} \quad (4)$$

where ω is the scaling parameter. It is then used as an input to the network stream to obtain enhanced information regarding the different domains of the underwater image. The attention mechanism in the designed model guides the network to focus on all regions of interest; that is, the high-response region in the soft mask contains an image quality that can be enhanced. The loss function uses the pixel MSE to calculate the difference between images I and I^{*c} .

$$L_{MSE} = \frac{1}{n} \sum_c s^c (I_i - I_i^{*c})^2 \quad (5)$$

To minimize the prediction probability error, we redesigned the MSE by adding $s^c(\bullet)$ to the constraint, which denotes the prediction probability of class c , where n is the number of images, I .

Considering the need to share model parameters for multi-domain underwater images, we set L_{self} for the multi-domain underwater image enhancement objective function using

multilabel soft marginal loss [40]. Alternative loss functions can be used for specific tasks to better separate individual domain categories. One of the simplest methods is to add margins to each domain, shown in Equation (6).

$$L_{cl} = -\frac{1}{|Q|} \sum_{(x,y) \in Q} \log \frac{e^{D(s^c(I_i - I_i^{*c}), r_c)}}{e^{D(s^c(I_i - I_i^{*c}), r_c)} + \sum_{k \in C} e^{D(s^c(I_i - I_i^{*c}), r_k) + m}} \quad (6)$$

where Q is the test set, and the representative points of each domain are represented by r_1, \dots, r_c (e.g., the centroids of all samples in the support set of each domain are used as representative points). $D(\cdot)$ is the metric module used to measure the cosine similarity of two feature vectors. The same margin, m , was added between two different domains, forcing a certain distance between the samples of different domains.

Subsequently, the overall loss function L_{self} of the model is the sum of the objective functions L_{self} and L_{MSE} of the multi-domain underwater image deep learning, as defined in Equation (7).

$$L_{self} = L_{cl} + \alpha L_{MSE} \quad (7)$$

where α is a weighting parameter, set empirically, and $\alpha = 1$ is used for all experiments in this study. Guided by L_{self} updates to the model weights, the network model learns to extend as much as possible the focal region of the input image that contributes to the target class enhancement, thus allowing the soft mask to be tailored to the task of interest (i.e., underwater image enhancement). The model was trained using a multinomial loss function that considered the resulting pixel-level loss to preprocess the sharpness of the edges and enhance the structural and textural similarity of underwater images.

3.3. MDCNN-VGG

In addition to allowing the network model to explore its own model weights, the network model itself can also employ additional supervised learning similar to the soft mask to make it suitable for the task of interest. We introduce the VGG [41] to integrate additional supervision into a supervised learning framework seamlessly.

There is a multi-domain phenomenon in underwater images in the application scenario studied in this paper; that is, there are multi-domain underwater images caused by factors such as different viewing angles or background light intensities. In our study, the multiple DCNNs described in Section 2.2 are used to integrate into multiple channels, and the advantages of the designed CNN model are described in Section 3.1. Based on the concept of ensemble learning [42], we set up multiple channels to learn the background types of underwater images in different domains. The model mines the effective information of underwater images of different domains from a local perspective with the help of DCNNs of different channels, while delivering this effective information to the VGG for accurate underwater image enhancement. The MDCNN-VGG effectively exploits the differential distribution of different domain information and applies a parameter-sharing-based mechanism to improve domain adaptation. In this study, the model learns the distribution bias of underwater image data from different domains to make it highly robust and to improve its generalization capability. The architecture of the designed MDCNN-VGG model is illustrated in Figure 3.

The design of the MDCNN-VGG fully considers the different background environments of different underwater images, and considers the situation of different domains (e.g., different shooting perspectives, different light intensities, etc.). Different channels input different domains of underwater image data, which enhances the model domain adaptation capability through the parameter-sharing mechanism [32], and designs the DCNN to effectively extract the underwater image features.

The objective function of our model is L_e for a newly designed externally supervised VGG-based loss function, in addition to L_{cl} and L_{MSE} , and is defined as

$$L_e = \frac{1}{n} \sum_c (A^c - H^c)^2 \quad (8)$$

where H^c denotes additional multi-domain supervision, for example, the multi-domain segmentation mask for Figure 1. Owing to the high time complexity of generating pixel-level segmentation maps, the model designed in this study was expected to use a small amount of data under external supervision to meet the requirements of practical scenarios.

The VGG was introduced into the model, and all parameters were shared between the two network streams and the VGG. The error values of the enhanced and real images were passed to the VGG to obtain the corresponding feature representations; that is, the distance between the two types of images, I and I^* , was calculated based on perceptual loss.

Naturally, the final MDCNN-VGG loss function, L , can be obtained using Equation (9).

$$L = L_{cl} + \alpha L_{MSE} + \beta L_e \quad (9)$$

where L_{cl} and L_{MSE} are shown in Section 3.2; α, β are coefficients that control the degree of involvement of the DCNN and additional supervision in the MDCNN-VGG training. The MDCNN-VGG can be easily improved for other vision tasks [33,34]. When the final network output $f_{l,k}$ is obtained, L_e is used to direct the network model to the key regions of the task of interest.

4. Experimental Design and Result Analysis

The main contents of the experiments in this section include the dataset and experimental setup, qualitative evaluation, quantitative evaluation, ablation experiment, multi-domain scenario, visual perception effect experiment, and inference time complexity.

4.1. Dataset and Experimental Setup

4.1.1. Dataset

We tested the MDCNN-VGG on real image datasets to demonstrate its ability to enhance underwater images from different datasets. The datasets used can be considered multi-domain underwater images, and the specific datasets are described as follows:

UFO-120 [43]: Clear images were collected from ocean soundings for different water types. The corresponding underwater images were generated using style transformation, where a subset of 120 images was used as the test set.

EUVP Dark [13]: A large collection of 10 K paired and 25 K unpaired images collected by data producers during ocean soundings under various visibility conditions, with both poor and good perceptual quality. It contains 5500 pairs of images with dark underwater backgrounds. In this study, 1000 images were used to test the model.

UIEBD [18]: Comprises 890 pairs of underwater images captured under different lighting conditions with different color ranges and contrasts.

4.1.2. Experimental Configuration

The model in this study was trained using the Adam optimizer, with the learning rate set to 0.0001, dropout set to 0.5, and batch size set to 1000. Approximately 10 h was required to optimize the model with more than 50 cycles. The experiment was run on an Intel(R) Core (TM) i7-10700k CPU with a 16 GB RAM and an NVIDIA GTX 1080 GPU.

4.1.3. Baseline

Shallow UWnet [4]: A shallow CNN for underwater image enhancement. Three methods, white balance, gamma correction, and histogram equalization, were used to pre-process the WaterNet enhancement input for the characteristics of the blurred background environment of underwater images.

UResnet [9]: A CNN-based residual network is a more comprehensive supervised learning method for underwater image enhancement.

FUnIE GAN [10]: Assumes nonlinear mapping between distorted and enhanced images, and the blurring of images is removed by constructing image enhancement.

CycleGAN [12]: A technique that automatically performs image-to-image transformation without pairwise examples, using a batch of images from the source and target domains that do not need to be correlated and trained in an unsupervised manner.

UGAN-P [34]: Underwater GAN with gradient penalty.

Uw HL [44]: Color recovery based on fuzzy lines. This method was based on a physical model design scheme.

4.1.4. Evaluation Metrics

The standard metrics used in this study, namely the peak signal-to-noise ratio (PSNR) and structural similarity index metric (SSIM), were quantitatively evaluated for the output images of the proposed model. The PSNR and SSIM quantify the reconstruction quality and structural similarity of the output images with regard to the corresponding reference images [45]. In addition, the output image quality was analyzed in this study using the non-reference underwater image quality metric (UIQM). The UIQM is composed of three attribute metrics: image color (UICM), sharpness (UISM), and contrast (UIConM), where each attribute assesses the quality of the underwater image from a single dimension.

4.2. Multi-Domain Scenarios

In this study, the image enhancement effect of the model was verified through multi-domain underwater image enhancement experiments. The enhancement effects of the different models and algorithms for different domain images are shown in Figures 4 and 5, respectively.

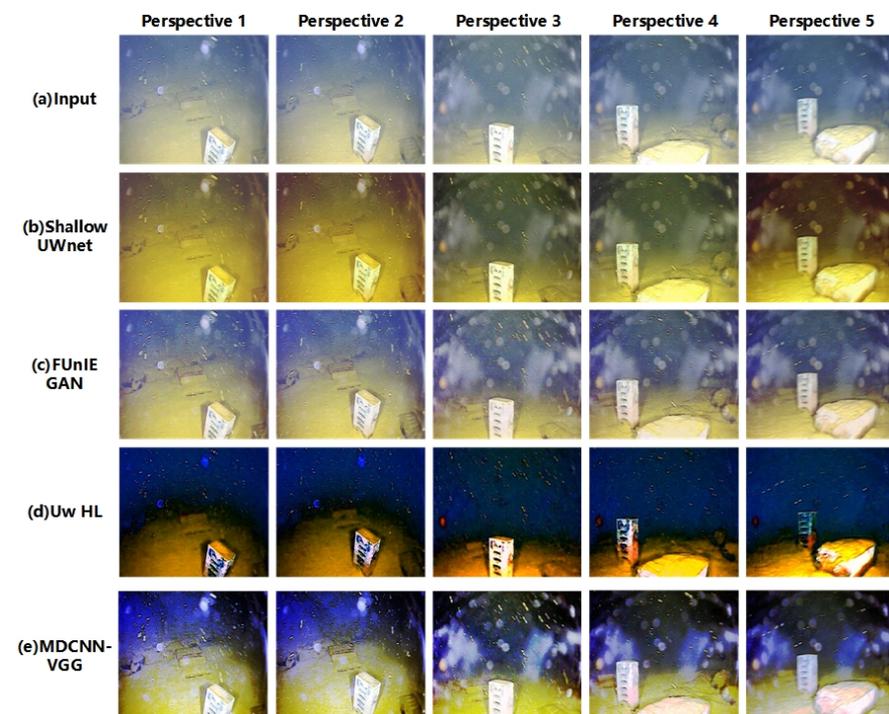


Figure 4. Submarine target scene.

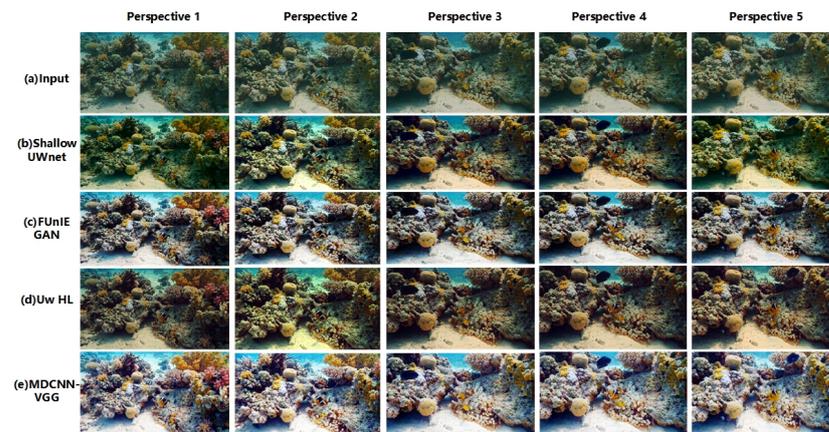


Figure 5. Submarine coral group scene.

From Figure 4, it is clear that the different methods have some enhancement effect, but poor performance for different domain images. As shown in Figure 4, the Shallow UWnet caused overexposure, whereas the FUnIE GAN deepened the background color of the water body. However, the physics-based scheme, Uw-HL, exhibits image oversaturation, which makes the contrast of underwater images in different domains too high. The reason for the failure of these schemes is that they ignore the correlation between underwater images in different domains. The proposed model has the best enhancement effect for images in different domains, where the targets are still clearly visible, even for images in different viewpoints. The success of the model in this study can be attributed to the designed multichannel and fusion of different loss terms that are used to achieve a better enhancement effect for underwater images from different viewpoints. In other scenarios, such as those shown in Figure 5, the enhancement performance of the model in this study remained the best.

Figure 6 shows qualitative comparisons of the different underwater image enhancement schemes.

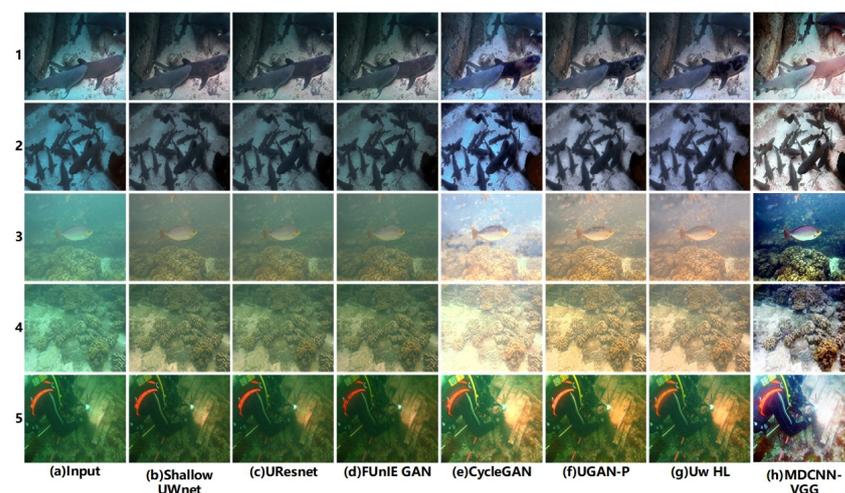


Figure 6. Qualitative comparisons of different underwater image enhancement schemes.

As shown in Figure 6, FUnIE GAN and CycleGAN often appear oversaturated, whereas UResnet and Shallow UWnet usually fail to correct the green tones in the images, owing to the greater depth of the above network models and the tendency of the model to overfit. UGAN-P and Uw-HL performed better and their enhancement performances were similar to those of the MDCNN-VGG. However, the UGAN-P and Uw-HL models are susceptible to the influence of bright objects in the scene, and the oversaturation phenomenon, particularly Uw-HL, fails to enhance the global brightness in some cases,

which shows that the above two schemes cannot explicitly pre-estimate the targets in underwater images and fail to improve the visual perception of the images. However, it can be observed from Figure 6 that the multichannel setup of the MDCNN-VGG achieves color consistency and hue correction, and enhances the reference color or texture information in the loss function using the multichannel fusion processing of underwater images from different domains. Overall, the multi-domain underwater image enhancement scheme achieves the same performance as the physical-based model without using scene depth or unknown water body information and outperforms other baselines [46].

4.3. Qualitative Evaluation

Table 1 presents a comparison of the average PSNR, structural similarity (SSIM), and UIQM of the test images for each model. The results demonstrate that the MDCNN-VGG achieves the best UIQM values in UFO-120; however, the PSNR and SSIM values are 0.30 relatively.

Table 1. Quantitative metrics of underwater image enhancement.

	PSNR			SSIM			UIQM		
	EUVP Dark	UFO-120	UIEB	EUVP Dark	UFO-120	UIEB	EUVP Dark	UFO-120	UIEB
Shallow UWnet	20.83	18.45	21.24	0.90	0.73	0.90	2.71	2.56	2.50
UResnet	27.61	21.24	24.98	0.97	0.78	0.95	2.40	2.27	2.38
FUnIE GAN	28.68	30.38	38.75	0.96	0.81	1.00	2.95	2.89	3.08
CycleGAN	8.79	16.23	17.24	0.84	0.68	0.79	2.95	2.89	2.77
UGAN-P	27.61	15.23	24.96	0.97	0.67	0.95	2.40	2.73	2.38
Uw HL	39.91	30.38	38.75	1.00	0.81	0.99	2.71	2.56	2.50
MDCNN-VGG	27.49	25.27	19.09	0.82	0.74	0.75	3.00	2.88	2.80

Weak in the paired EUVP dataset. The UGAN-P and UResNet produced better results for the paired data. Similar analyses were performed in [38,39], which quantified the sharpness, clarity, and contrast of underwater images. The UIQM results presented in Table 1 show that the MDCNN-VGG outperforms state-of-the-art methods, and the best UIQM values show that the resulting image has balanced color, clarity, and contrast. In this paper, we hypothesize that the global similarity loss in the MDCNN-VGG and FUnIE GAN, or the gradient penalty term in UGAN-P contributes to such enhancement tasks, owing to the fact that all the above methods add L1 terms to the adversarial objective. It is evident from Table 1 that the MDCNN-VGG contributes to an average improvement over the state-of-the-art Shallow Uwnet in the UIQM metric for different datasets by 0.30, where comparable statistics of performance improvement are observed for the PSNR and SSIM [47,48].

CNNs have a wide range of applications in computer vision, and owing to their advantages, they have been promoted for applications in underwater imaging. The MDCNN-VGG maintained excellent quantitative performance. The enhancement capability of the MDCNN-VGG for underwater images in different datasets is shown in Figures 7–9. Notably, the MDCNN-VGG can test its generalization capability on different datasets, making it more widely applicable to various types of underwater scenes with different degradation levels.

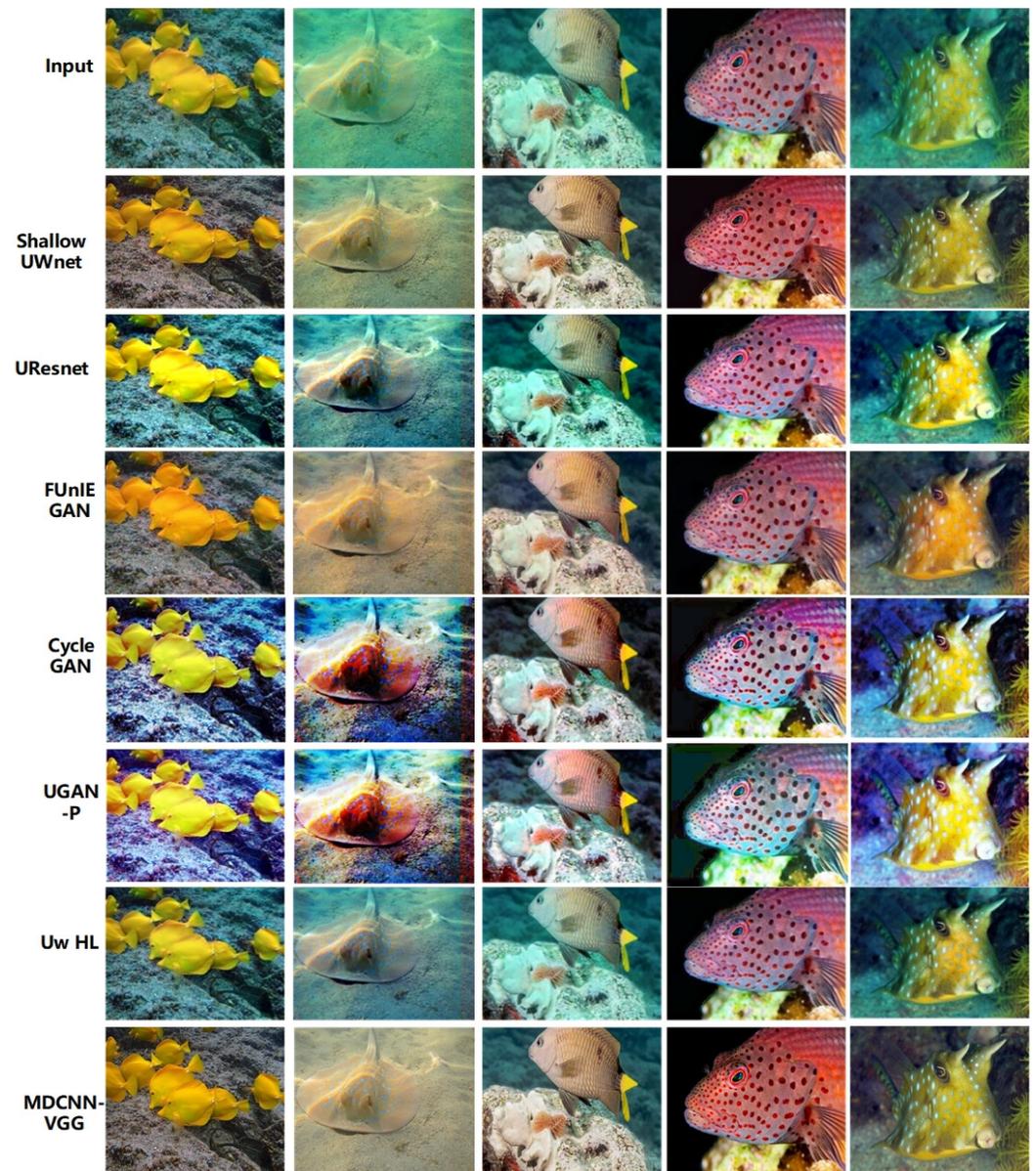


Figure 7. Underwater image enhancement results of UFO-120.

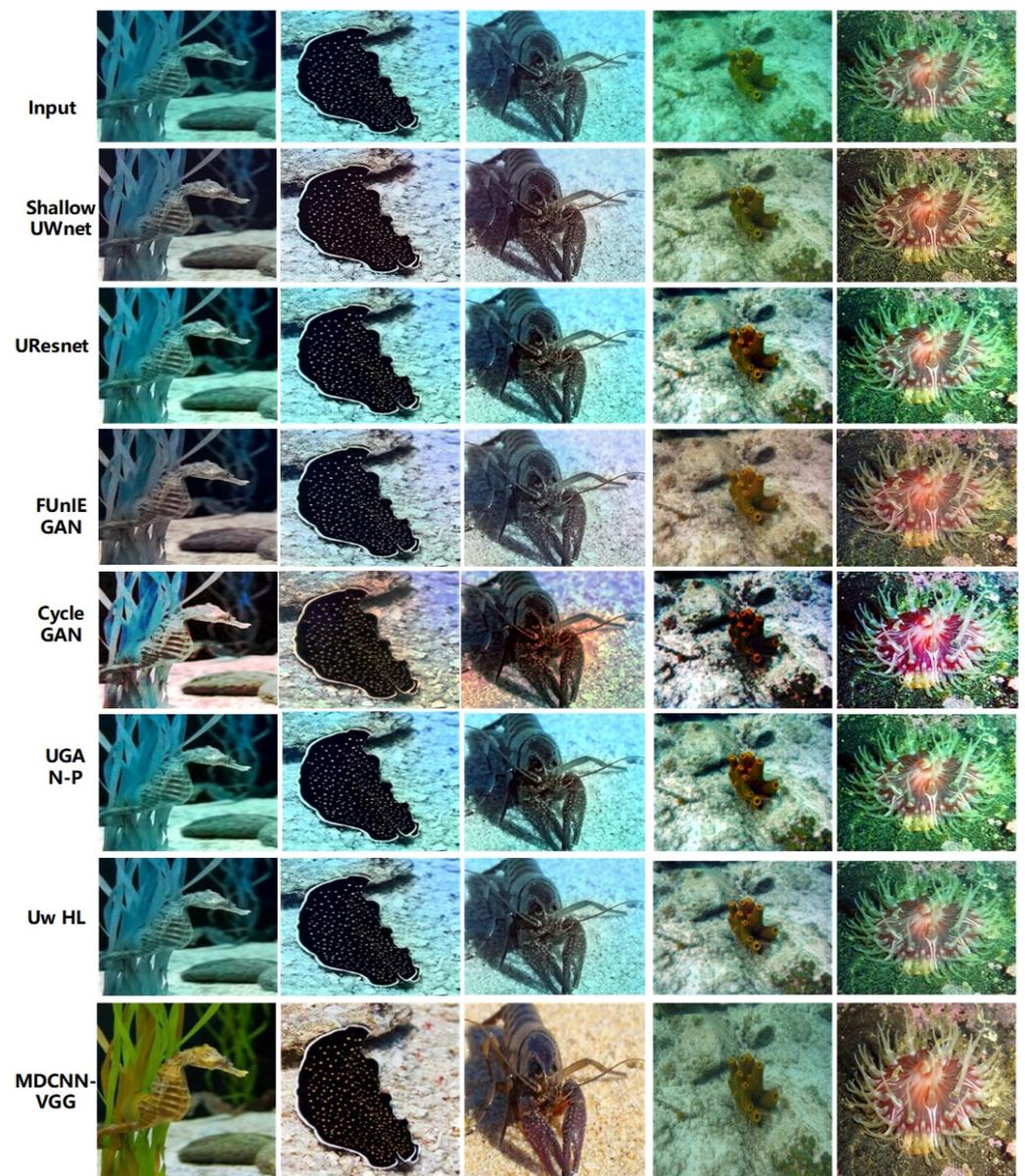


Figure 8. Underwater image enhancement results of EUVP.

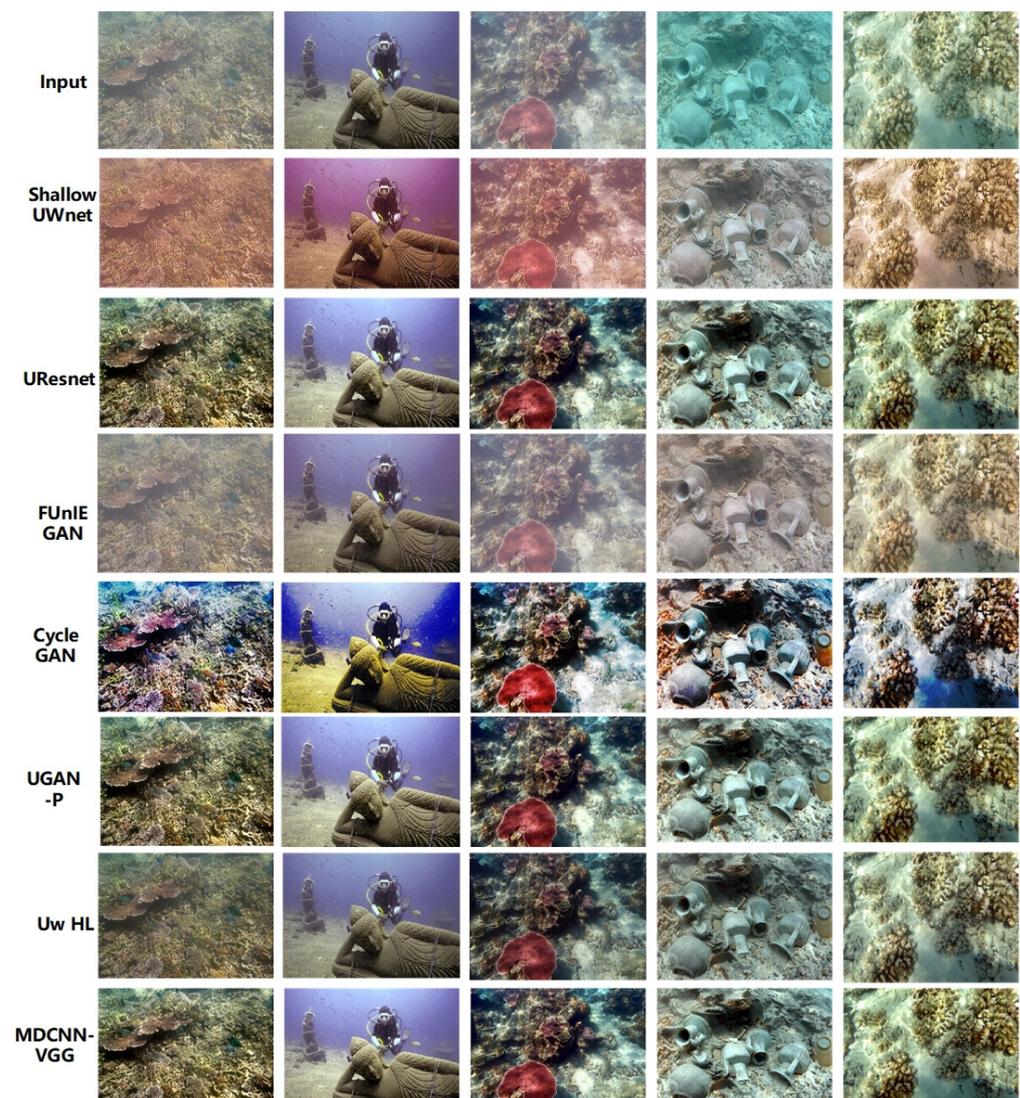


Figure 9. Underwater image enhancement results of UIEB.

4.4. Ablation Experiments

In this study, we first qualitatively analyzed the enhanced the color and sharpness of images generated with the MDCNN-VGG and compared them with their respective corresponding baselines. As shown in Figure 10, the enhanced underwater images mainly recover their true colors and sharpness. In addition, the color correction and global contrast enhancement results plotted for underwater images of different hues, as shown in Figure 11, clearly show the distinct texture and vivid colors of the local images after underwater image enhancement with the proposed model. This is due to the additional multi-domain supervised VGG setup that allows for the pixel-level segmentation masking of underwater images, using Le to guide the network model to focus on the critical tasks of interest regions [49].



Figure 10. Underwater image enhancement results.

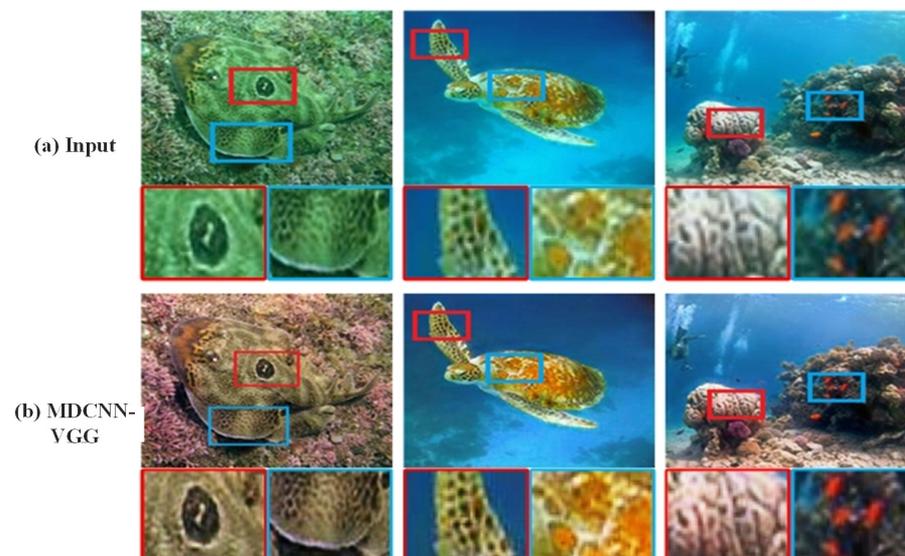


Figure 11. Underwater image detail enhancement results.

To verify the contribution of each loss term to the underwater image enhancement (see Figure 12), we have looked at their effect on the image enhancement with and without these loss terms. Figure 12 shows that the different loss terms clearly contribute to the image.

Enhancement can be analyzed more intuitively via a direct comparison using the UIQM values. The calculation shows that in the top row of Figure 12, the UIQM value in Figure 12c is higher than the UIQM result of Figure 12b by 0.038, whereas the UIQM result in Figure 12d is higher than that of Figure 12b by 0.321. The UIQM result of the model utilizing all loss terms is higher than that of Figure 12b by 0.546. This clearly shows that the different loss terms designed in this study can result in better image enhancement and compensate for each other's deficiencies. The image in the bottom row of Figure 12 shows the effect of the above analysis.

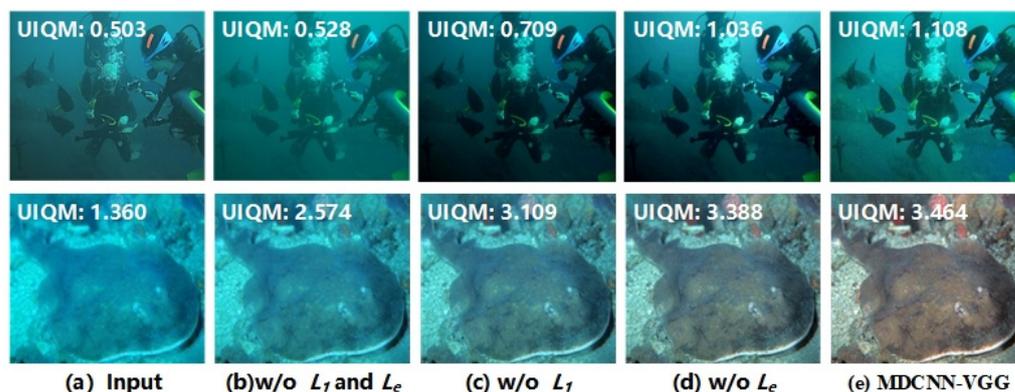


Figure 12. Ablation experiment results of different loss items in the MDCNN-VGG.

5. Conclusions

In this study, we propose a deep learning model, the MDCNN-VGG, which is an underwater image enhancement technique based on a hybrid model, to achieve the fast enhancement of multi-domain underwater images. In the MDCNN-VGG, a DCNN using different channels can effectively mine the local information of underwater images in different domains and pass the above local information to the VGG to enhance underwater images accurately. In this study, the model was based on the global color and structural content of the image, the local texture, and style information, and the perceived loss function was established by evaluating the image quality. We performed extensive qualitative and quantitative evaluations as well as multi-domain image enhancement studies. The results show that the MDCNN-VGG contributes a 0.11 average improvement over the state-of-the-art Shallow UWnet for different datasets in terms of UIQM values, and other performance metrics are similarly improved.

In the future, we believe that (1) we can improve the design structure of the model to further enhance its application inference capability, such as considering image pair applications for underwater images in a small sample drive [47]; (2) even though the model in this study enhances image texture and color enhancement capability through different domains of underwater images collaboratively, there are still problems such as blurred details, color bias, and overexposure. For the different loss terms in this study, robust optimization through adversarial learning was used to enhance the model's ability to deal with these detailed problems and further suppress the perturbation of samples to the model through robust optimization.

Author Contributions: Conceptualization, L.Z. and S.-W.L.; methodology, L.Z.; software, L.Z.; validation, L.Z.; formal analysis, L.Z.; investigation, L.Z.; resources, L.Z.; data curation, L.Z.; writing—original draft preparation, L.Z.; writing—review and editing, L.Z.; visualization, L.Z.; supervision, S.-W.L.; project administration, S.-W.L.; funding acquisition, S.-W.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Institute of Information & communications Technology Planning & Evaluation (IITP) under the Artificial Intelligence Convergence Innovation Human Resources Development (IITP-2023-No.RS-2023-00255968) grant funded by the Korean government (MSIT). The BK21 FOUR program of the National Research Foundation of Korea was funded by the Ministry of Education (NRF5199991014091).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The authors confirm that the data supporting the findings of this study are available from the corresponding author, upon reasonable request.

Acknowledgments: We thank Ashray Adhikari for providing the multi-domain underwater dataset for this study. We thank the anonymous reviewers for their critical comments on this manuscript.

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

References

1. Zhou, J.-c.; Zhang, D.-h.; Zhang, W.-s. Classical and state-of-the-art approaches for underwater image defogging: A comprehensive survey. *Front. Inf. Technol. Electron. Eng.* **2020**, *21*, 1745–1769. [[CrossRef](#)]
2. Zhou, J.; Sun, J.; Zhang, W.; Lin, Z. Multi-view underwater image enhancement method via embedded fusion mechanism. *Eng. Appl. Artif. Intell.* **2023**, *121*, 105946. [[CrossRef](#)]
3. Zhou, J.; Pang, L.; Zhang, D.; Zhang, W. Underwater image enhancement method via multi-interval subhistogram perspective equalization. *IEEE J. Ocean. Eng.* **2023**, *48*, 474–488. [[CrossRef](#)]
4. Li, C.; Guo, J.; Guo, C. Emerging from water: Underwater image color correction based on weakly supervised color transfer. *IEEE Signal Process. Lett.* **2018**, *25*, 323–327. [[CrossRef](#)]
5. Li, C.; Anwar, S.; Hou, J.; Cong, R.; Guo, C.; Ren, W. Underwater image enhancement via medium transmission-guided multi-color space embedding. *IEEE Trans. Image Process.* **2021**, *30*, 4985–5000. [[CrossRef](#)]
6. Peng, L.; Zhu, C.; Bian, L. U-shape transformer for underwater image enhancement. *IEEE Trans. Image Process.* **2023**, *32*, 3066–3079. [[CrossRef](#)]
7. Cui, B.; Zhang, B.; Zhao, Q. Underwater Image Enhancement Method Based on Improved Generative Adversarial Network. In Proceedings of the 2023 5th Asia Pacific Information Technology Conference, Ho Chi Minh City, Vietnam, 9–11 February 2023; pp. 62–68.
8. Marques, T.P.; Albu, A.B. L²UWE: A framework for the efficient enhancement of low-light underwater images using local contrast and multi-scale fusion. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Seattle, WA, USA, 14–19 June 2020; pp. 2286–2295.
9. Chen, Y.; Li, H.; Yuan, Q.; Wang, Z.; Hu, C.; Ke, W. Underwater Image Enhancement based on Improved Water-Net. In Proceedings of the 2022 IEEE International Conference on Cyborg and Bionic Systems (CBS), Wuhan, China, 24–26 March 2023; pp. 450–454.
10. Bhat, A.; Narang, Y.; Goyal, Y. Underwater Image Enhancement with Feature Preservation using Generative Adversarial Networks (UIEFP GAN). In Proceedings of the 2022 6th International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 25–27 May 2022; pp. 1023–1029.
11. Mathur, P.; Soni, B. Exploring Color Models for Enhancement of Underwater Image. In *Data Driven Approach Towards Disruptive Technologies: Proceedings of MIDAS 2020*; Springer: Singapore, 2021; pp. 325–336.
12. Zhang, C.; Li, M.; Wu, D. Federated Multidomain Learning With Graph Ensemble Autoencoder GMM for Emotion Recognition. *IEEE Trans. Intell. Transp. Syst.* **2023**, *24*, 7631–7641. [[CrossRef](#)]
13. Anwar, S.; Li, C. Diving deeper into underwater image enhancement: A survey. *Signal Process. Image Commun.* **2020**, *89*, 115978. [[CrossRef](#)]
14. Li, C.; Cong, R.; Kwong, S.; Hou, J.; Fu, H.; Zhu, G.; Zhang, D.; Huang, Q. ASIF-Net: Attention steered interweave fusion network for RGB-D salient object detection. *IEEE Trans. Cybern.* **2020**, *51*, 88–100. [[CrossRef](#)]
15. Naik, A.; Swarnakar, A.; Mittal, K. Shallow-uwnet: Compressed model for underwater image enhancement (student abstract). *Proc. AAAI Conf. Artif. Intell.* **2021**, *35*, 15853–15854. [[CrossRef](#)]
16. Liu, P.; Wang, G.; Qi, H.; Zhang, C.; Zheng, H.; Yu, Z. Underwater image enhancement with a deep residual framework. *IEEE Access* **2019**, *7*, 94614–94629. [[CrossRef](#)]
17. Islam, M.J.; Xia, Y.; Sattar, J. Fast underwater image enhancement for improved visual perception. *IEEE Robot. Autom. Lett.* **2020**, *5*, 3227–3234. [[CrossRef](#)]
18. Erdem, F.; Bayram, B.; Bakirman, T.; Bayrak, O.C.; Akpinar, B. An ensemble deep learning based shoreline segmentation approach (WaterNet) from Landsat 8 OLI images. *Adv. Space Res.* **2021**, *67*, 964–974. [[CrossRef](#)]
19. Zong, X.; Chen, Z.; Wang, D. Local-CycleGAN: A general end-to-end network for visual enhancement in complex deep-water environment. *Appl. Intell.* **2021**, *51*, 1947–1958. [[CrossRef](#)]
20. Smith, L.N.; Topin, N. Super-convergence: Very fast training of neural networks using large learning rates. In Proceedings of the Artificial Intelligence and Machine Learning for Multi-Domain Operations Applications, Baltimore, MD, USA, 14–18 April 2019; Volume 11006, p. 1100612.
21. Jia, G.; Lu, Y.; Lu, W.; Shi, Y.; Yang, J. A Novel Verification Method for Chinese Aviation Radiotelephony Readbacks Based on LSTM-RNN. *Electron. Lett.* **2017**, *53*, 401–403. [[CrossRef](#)]
22. Kumar, A.; Ghosh, A.K. Decision Tree- and Random Forest- Based Novel Unsteady Aerodynamics Modeling Using Flight Data. *J. Aircr.* **2019**, *56*, 403–409. [[CrossRef](#)]
23. Goodfellow, I.J.; Warde-Farley, D.; Mirza, M.; Courville, A.; Bengio, Y. Maxout Networks. In Proceedings of the 30th International Conference on Machine Learning, Atlanta, GA, USA, 16–21 June 2013.
24. Xu, J.; Zhang, Y.; Miao, D. Three-way confusion matrix for classification: A measure driven view. *Inf. Sci.* **2020**, *507*, 772–794. [[CrossRef](#)]
25. Zhang, H.; Sun, L.; Wu, L.; Gu, K. DuGAN: An effective framework for underwater image enhancement. *IET Image Process.* **2021**, *15*, 2010–2019. [[CrossRef](#)]

26. Lin, Q.; Fouchez, D.; Pasquet, J. Galaxy image translation with semi-supervised noise-reconstructed generative adversarial networks. In Proceedings of the 2020 25th International Conference on Pattern Recognition (ICPR), Milan, Italy, 10–15 January 2021; pp. 5634–5641.
27. Ignatov, A.; Kobyshev, N.; Timofte, R.; Vanhoey, K. DSLR-quality photos on mobile devices with deep convolutional networks. In Proceedings of the 2017 IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 3297–3305.
28. Regan, D.; Padmavathi, C. A Study on Performance of Bilateral and Trilateral Filters in the Mixed Noise Environment. *J. Comput. Theor. Nanosci.* **2018**, *15*, 2089–2091. [[CrossRef](#)]
29. 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](#)]
30. 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](#)] [[PubMed](#)]
31. Dai, C.; Lin, M.; Wu, X.; Wang, Z.; Guan, Z. Single underwater image restoration by decomposing curves of attenuating color. *Opt. Laser Technol.* **2020**, *123*, 105947. [[CrossRef](#)]
32. Berman, D.; Levy, D.; Avidan, S.; Treibitz, T. Underwater single image color restoration using haze-lines and a new quantitative dataset. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *43*, 2822–2837. [[CrossRef](#)] [[PubMed](#)]
33. Zhou, J.; Zhang, D.; Zhang, W. A multifeature fusion method for the color distortion and low contrast of underwater images. *Multimed. Tools Appl.* **2021**, *80*, 17515–17541. [[CrossRef](#)]
34. 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.
35. 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.
36. Zhu, L.; Fan, H.; Luo, Y.; Xu, M.; Yang, Y. Temporal cross-layer correlation mining for action recognition. *IEEE Trans. Multimed.* **2021**, *24*, 668–676. [[CrossRef](#)]
37. Zoumpourlis, G.; Doumanoglou, A.; Vretos, N.; Daras, P. Non-linear Convolution Filters for CNN-Based Learning. In Proceedings of the 2017 IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017.
38. Zhuang, P.; Li, C.; Wu, J. Bayesian retinex underwater image enhancement. *Eng. Appl. Artif. Intell.* **2021**, *101*, 104171. [[CrossRef](#)]
39. Sengupta, A.; Ye, Y.; Wang, R.; Liu, C.; Roy, K. Going deeper in spiking neural networks: VGG and residual architectures. *Front. Neurosci.* **2019**, *13*, 95. [[CrossRef](#)]
40. Islam, M.J.; Fulton, M.; Sattar, J. Toward a generic diver-following algorithm: Balancing robustness and efficiency in deep visual detection. *IEEE Robot. Autom. Lett.* **2018**, *4*, 113–120. [[CrossRef](#)]
41. Zhang, Y.; Li, F.; Zhang, S.; Zhu, T. Leads Detection Using Mixture Statistical Distribution Based CRF Algorithm from Sentinel-1 Dual Polarization SAR Imagery. In Proceedings of the EGU General Assembly 2017 Conference, Vienna, Austria, 23–28 April 2017.
42. Felix, G.; Claudia, N.; Sebastian, H.; Mariette, V.; Christoph, R.; Emanuele, S.; Simonetta, P.; Wolfgang, W. The Added Value of the VH/VV Polarization-Ratio for Global Soil Moisture Estimations from Scatterometer Data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 3668–3679.
43. Huang, F.; Ash, J.; Langford, J.; Schapire, R. Learning deep resnet blocks sequentially using boosting theory. In Proceedings of the 35th International Conference on Machine Learning, Stockholm, Sweden, 10–15 July 2018; pp. 2058–2067.
44. Cao, Z.; Simon, T.; Wei, S.E.; Sheikh, Y. Realtime multi-person 2D pose estimation using part affinity fields. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1302–1310.
45. 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](#)]
46. Ward, R.; Wu, X.; Bottou, L. AdaGrad stepsizes: Sharp convergence over nonconvex landscapes. In Proceedings of the 36th International Conference on Machine Learning, Long Beach, CA, USA, 9–15 June 2019; pp. 6677–6686.
47. Li, L.; Lv, M.; Jia, Z.; Ma, H. Sparse representation-based multi-focus image fusion method via local energy in shearlet domain. *Sensors* **2023**, *23*, 2888. [[CrossRef](#)] [[PubMed](#)]
48. Li, L.; Ma, H. Pulse coupled neural network-based multimodal medical image fusion via guided filtering and WSEML in NSCT domain. *Entropy* **2021**, *23*, 591. [[CrossRef](#)] [[PubMed](#)]
49. Li, L.; Si, Y.; Wang, L.; Jia, Z.; Ma, H. A novel approach for multi-focus image fusion based on SF-PAPCNN and ISML in NSST domain. *Multimed. Tools Appl.* **2020**, *79*, 24303–24328. [[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.