Article

Dilated Generative Adversarial Networks for Underwater Image Restoration

Jao-Chuan Lin [1], Chih-Bin Hsu [2], Jen-Chun Lee [3,*], Chung-Hsien Chen [4] and Te-Ming Tu [3]

1. Department of Marine Leisure Management, National Kaohsiung University of Science and Technology, Kaohsiung City 825, Taiwan; jcl@nkust.edu.tw
2. Department of Information Management, Chinese Naval Academy, Kaohsiung City 81345, Taiwan; bin7474@gmail.com
3. Department of Telecommunication Engineering, National Kaohsiung University of Science and Technology, Kaohsiung City 825, Taiwan; tr4707@gmail.com
4. Metal Industries Research & Development Centre (MIRDC), Kaohsiung City 811, Taiwan; jacky@mail.mirdc.org.tw
* Correspondence: i923002@nkust.edu.tw

Abstract: Underwater images often come with blurriness, lack of contrast, and low saturation due to the physics of light propagation, absorption, and scattering in seawater. To improve the visual quality of underwater images, many have proposed image processing methods that vary based on different approaches. We use a generative adversarial network (GAN)-based solution and generate high-quality underwater images equivalent to given raw underwater images by training our network to specify the differences between high-quality and raw underwater images. In our proposed method, which is called dilated GAN (DGAN), we add an additional loss function using structural similarity. Moreover, this method can not only determine the realness of the entire image but also functions with classification ability on each constituent pixel in the discriminator. Finally, using two different datasets, we compare the proposed model with other enhancement methods. We conduct several comparisons and demonstrate via full-reference and nonreference metrics that the proposed approach is able to simultaneously improve clarity and correct color and restores the visual quality of the images acquired in typical underwater scenarios.

Keywords: underwater image restoration; dilated generative adversarial networks; artificial intelligence; deep learning

1. Introduction

Marine resource exploration is very difficult due to image distortion caused by the underwater environment. Most images have complicated lighting conditions, color casts, color artifacts, and blurred details, which make image restoration more challenging. Therefore, underwater image enhancement-related innovations are essential to improve the visual quality and merit of images so as to accurately perceive the underwater world.

Researchers have addressed these problems with several image processing methods to improve the quality of images taken in underwater environments. One traditional approach is to use a degradation model to enhance underwater images, with a focus on adjusting image pixel values to produce appealing results without underwater physical parameters. However, important complex underwater physical and optical factors are required, which make such traditional methods difficult to implement. Because of the lack of important physical parameters, these methods exhibit poor visualization performance across a variety of underwater images. In addition, these models often are too computationally expensive for use in real-time applications.

With the increasing resources of data and computing power from time to time, traditional methods are being substituted by artificial intelligence (AI). Deep learning has gained

popularity due to its exceptional performance when trained with massive amounts of data. Furthermore, various convolutional neural network (CNN) architectures have been applied broadly to computer vision tasks, e.g., image deraining [1], image super-resolution [2], and image denoising [3], and deep learning has been applied to enhance underwater images [4–9]. Generative adversarial networks (GANs) have improved tremendously in the field of synthetic images, by which underwater image quality can be enhanced using a large collection of paired or unpaired data. The basic architecture of GANs are based on the structure of two neural networks, making one against another to create instances of data that are synthetic enough to pass the discriminator as real data. These networks are widely used in the field of generating images, videos, and voices.

In this paper, we propose a GAN-based deep model focusing on generating high-quality underwater images equivalent to given (input) real-world underwater images by removing color artifacts and casts. We term our solution dilated GAN (DGAN). DGAN uses the structural similarity index measure (SSIM) in its structure to increase the learning concentration of structural similarities between the enhanced image and ground-truth image domains. Furthermore, the pixel-based L1 norm and gradient operation constrain the architecture to produce high-quality underwater images. A general view of our proposed structure is shown in Figure 1. We go through an evaluation and comparison on the performance of the proposed architecture on two different datasets (containing both enhanced and real-world underwater image pairs). The experimental and comparative results show that the proposed method generates better high-quality underwater images when compared to state-of-the-art methods. Our contributions include the following:

1. We propose a GAN-based domain-transfer technique to produce a high-quality underwater image equivalent to a given (input) real-world underwater image, where an encoder–decoder architecture is used on both generator and discriminator.
2. In contrast to general generators, the encoder of the generator has a dilated convolutional to capture more contextual information.
3. We use an additional loss term based on the structural similarity index measure (SSIM), norm L1, and gradient operation for improved results.
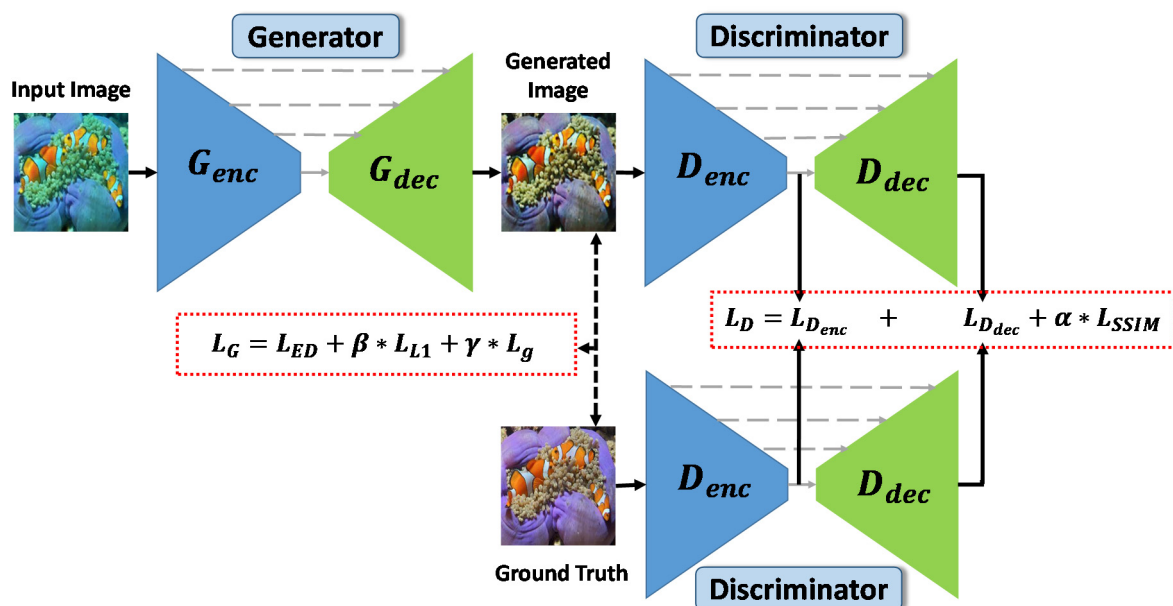4. We conduct experiments on two different large benchmark datasets.

Figure 1. Overview of proposed DGAN architecture. The elements of the generator are shown in Figure 2 and the elements of the discriminator are shown in Figure 3.
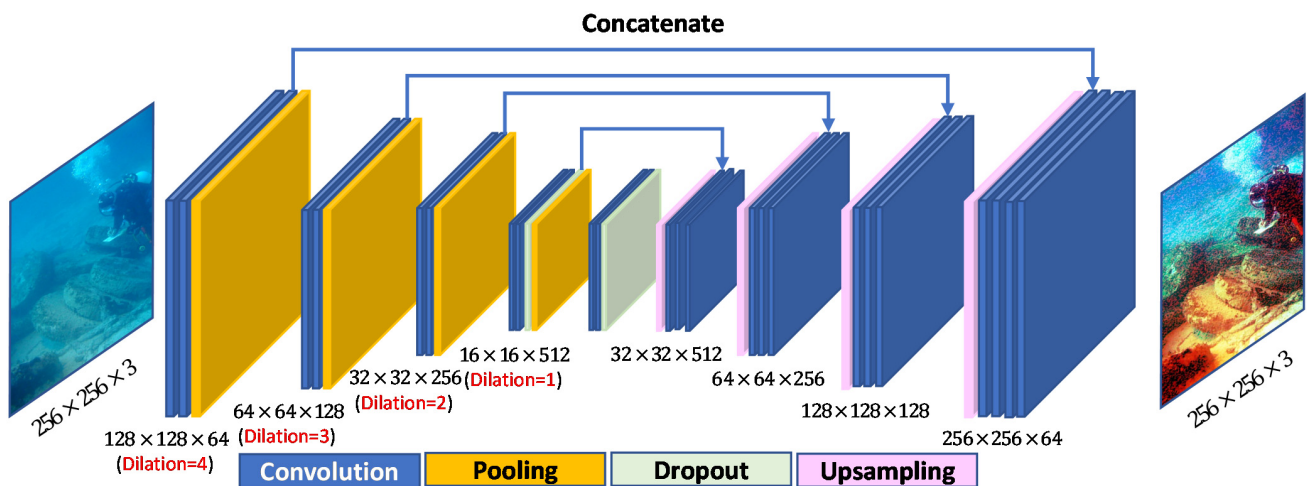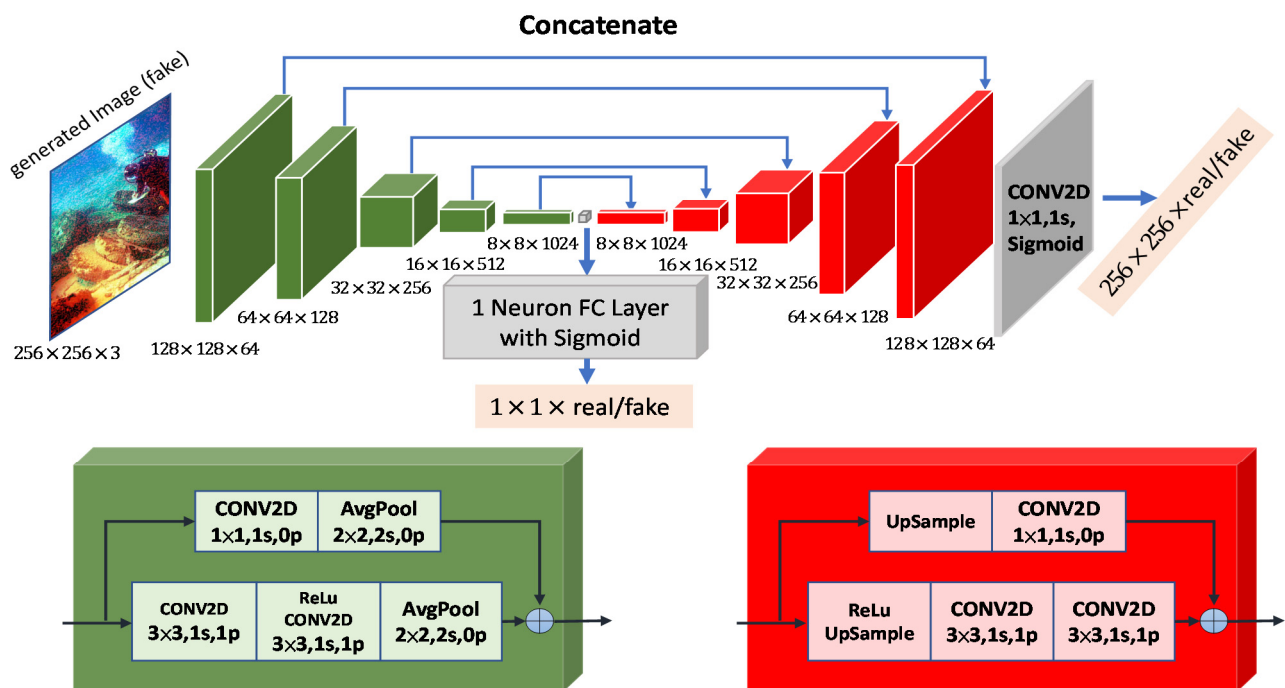
Figure 2. Generator architecture of proposed DGAN.

Figure 3. Proposed U-Net-based discriminator architecture. The symbol ⊕ denotes elementwise summation.

Following the introduction in Section 1, Section 2 briefly introduces the related work. Section 3 presents a more detailed representation of our proposed strategy, including the visual enhancement of underwater imagery. We discuss experimental results and comparisons in Section 4, followed by a conclusion in Section 5.

2. Related Work

Given the importance of underwater vision, many techniques have been proposed to enhance underwater images. We summarize these approaches based on traditional methods and deep-learning-based approaches.

2.1. Traditional Methods

The main goal of techniques of underwater image enhancement is to improve the quality of underwater images by manufacturing a degradation model, which is later run through estimation of model parameters. The most popular method is perhaps the dark

channel prior (DCP) technique [10], which is applied primarily to dehaze fogged images. DCP does well in dehazing fogged images, but fails to restore underwater images. DCP-based variants have been developed for underwater imaging [11,12]. Drews et al. [11] propose a method that specifies and applies the DCP only on the blue and green channels apart from the red channel to predict the transmission map. Galdran et al. [12] propose a novel method based on DCP theory to improve clarity and correct color in the underwater environment. Peng et al. [13] propose a method, generalized dark channel prior (GDCP), to perform an image restoration process. This method reduces the variables to several DCP variants for special cases of conditions of ambient light and turbid medium. Instead of the DCP, Li et al. [14] employ a hybrid method to improve the visual quality of degraded underwater images, which includes color correction and underwater image dehazing. Another traditional method modifies image pixel values to improve visual quality. Iqbal et al. [15] propose an unsupervised color correction method (UCM) based on color balancing, contrast correction of the RGB color model, and an HSI color model for underwater image enhancement. Ancuti et al. [16] propose a method that combines two images processed by white balance adjustment and contrasts limited adaptive histogram equalization. Fu et al. [17] employ a retinex-based approach to enhance a single underwater image that includes color correction, lightening of dark regions, naturalness preservation, and edge and detail enhancement. In 2017, the same authors [18] proposed a two-step approach, which includes two algorithms, color correction and contrast enhancement, for underwater image enhancement. Despite the excellent results obtained from these models, the complexity of the algorithm and lack of important physical parameters prevent their application for general users.

2.2. Deep-Learning-Based Methods

Another approach to underwater image enhancement employs learning-based methods. Deep learning is popular as a powerful tool for solving low-level vision problems, and has boosted this already rapidly developing field. Most of the literature on generating high-quality underwater images introduces generative adversarial network (GAN)-based architectures [5–8]. These methods heavily rely on synthetic pairs of high-quality images and degraded counterparts due to the training method. Fabbri et al. [7] demonstrate that CycleGAN can be used to perform dataset generation, which contains paired images for the restoration model of the underwater image. Li et al. [6] employ an unsupervised generative network and propose a real-time color correction method (WaterGAN) for monocular underwater images. Wang et al. [5] propose an underwater GAN (UWGAN) for generating realistic underwater images, and use U-net with combined loss functions for degraded underwater image enhancement. In 2020, Guo et al. [8] proposed a multiscale dense generative adversarial network (GAN) for enhancing underwater images. This method can perform color correction and detail restoration simultaneously.

Our work differs from the abovementioned works in its introduction of an unconventional GAN-based solution, which uses an encoder–decoder-based architecture as the discriminator structure, and loss function based on SSIM, L1, and gradient loss. Furthermore, for comparison of the proposed solution and three recently proposed GAN-based architectures (UGAN [7], UWGAN [5], and WaterGAN [6]), we utilize two datasets in our experiments. There are several differences between our network and the most relevant literature, which include: (i) the usage of the pixel-based SSIM loss in the discriminator loss function, (ii) an increase in the dilated convolution in the encoder part of the generator architecture to adapt the receptive fields for images of different scales, and (iii) in the discriminator architecture, we include a decoder part to generate underwater images with high quality.

3. Proposed Architecture

Briefly, a GAN is a type of network that consists of a generator network and a discriminator network. Each network attempts to beat one another in an adversarial

relationship [19]. The goal of the discriminator is to determine whether the input image is real or fake by discriminating the output of the generator from real images, and the goal of the generator is to produce fake images that imitate real images so well that it fools the discriminator.

Our network architectures are based on two distinct U-Net structures. We use the first in our generator (see Figure 2), and the second in the discriminator (see Figure 3), which is similarly structured as U-Net [20]. In both Figures 2 and 3, the block structures are used in the encoder and decoder networks; each of those blocks is coded with color based on their type. Conversely, the upsampling layer of the discriminator is composed of bilinear interpolation with a scale rate of 2, which is shown in Figure 3. Moreover, an average pooling with stride 2 to downscale the input to one-half resolution forms the downsampling layer. In the figures, stride value uses *s*, padding value uses *p*, and kernel sizes are 1×1, 2×2, and 3×3.

3.1. Generator Network

This generator network uses the U-Net framework [21] as initialization to resolve the process of image-to-image translation tasks. Low-level features are hierarchically extracted and recombined into higher-level features into the encoder by the U-Net architecture, followed by a multiple-feature, elementwise classification in the decoder. The encoder–decoder architecture of our generator is shown in Figure 2. In the encoder stage, there are four downsampling blocks. Each downsampling block consists of three network layers, namely the convolution layer, ReLU layer, and MaxPooling layer, to extract features. Similarly, each upsampling block of the decoder stage also consists of three network layers (upsampling layer, convolution layer, and ReLU layer) to reconstruct the image at the same resolution. For each block in the encoder, the convolutional result is transferred to the decoder symmetrically before max-pooling. On the other hand, each block in the decoder receives the feature representation that learned from the encoder and simultaneously concatenates it with the output from the deconvolutional layer. In addition, for the necessity of generating high-quality results, the increasing of feature resolution is inevitable, recent state-of-the-art methods rely heavily on the use of dilated convolution [22], in which the spacing between each pixel in the convolutional filter represents the dilation rate. By increasing the field of view, the receptive field arises as an additional advantage. Thus, this helps the filter capture more contextual information. Therefore, for each block in the encoder, to extract more information in multiscale feature maps, we adopt a dilated convolution with rates of 1, 2, 3, and 4, as shown in Figure 2.

3.2. Discriminator Network

In general, based on their output format, discriminator architectures are classified as either discriminators for real-or-fake classification of the entire input image or discriminators that judge both the entire image and its pixels. Figure 4 compares the two types. A recent study [20] reports that the improvement of utilizing U-Net architecture for the discriminator increases the performance of the overall GAN architecture by looking into each individual pixel. To classify each pixel individually in our discriminator for generating high-quality underwater images, we use an architecture inspired by this work. We use bilinear interpolation in the discriminator for the upsampling stages and an average pooling layer for the downsampling stages to limit the number of learnable parameters, as using upsampling and average-pooling functions reduces memory consumption when compared to convolutional and deconvolutional layers.
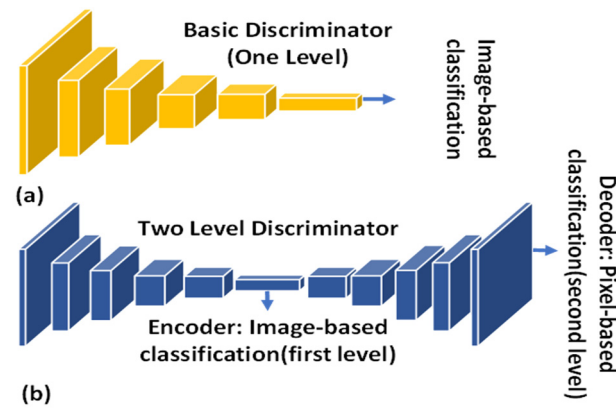
Figure 4. Comparison between the one-level discriminator and two-level discriminator: (**a**) the entire image is classified by the one-level discriminator as either fake or real, whereas (**b**) the two-level discriminator judges both constituent pixels and the whole image.

A package of multiple layers with skip connection(s) is called a residual block. Two types of residual blocks are utilized in our discriminator. We downsample its input as the first type of residual block (green boxes in Figure 4) and upsample its input (red boxes) using the second type. We include 2D convolutional layers with 3×3 and 1×1 kernel sizes and the same padding in both residual blocks. Before applying convolutional layers, the red block uses interpolation to upsample its input, the green block uses average pooling to downsample its input after applying convolutional layers. In the end, both blocks perform channel-wise summation on two parallel branches and pass the output to the next block.

3.3. Discriminator Loss

Using encoder–decoder architecture, our discriminator contains three loss terms that are given in Equation (1). The first two discriminator terms focus on the whole image (see Equation (2)) and each constituent pixel (see Equation (3)) as fake or real. The last term is SSIM-based loss (L_{SSIM}).

$$L_D = L_{D_{enc}} + L_{D_{dec}} + \alpha \times L_{SSIM} \qquad (1)$$

where $L_{D_{enc}}$ represents the loss, which is obtained at the end of the encoder for image-based loss, and $L_{D_{dec}}$ represents the loss, which is the decoder for pixel-based loss at the end. The definition is shown as follows:

$$L_{D_{enc}} = -\mathbb{E}_y[\log D_{enc}(Y)] - \mathbb{E}_x[\log(1 - D_{enc}(G(X)))] \qquad (2)$$

$$L_{D_{dec}} = -\mathbb{E}_y\left[\sum_{i,j} \log [D_{dec}(Y)]_{i,j}\right] - \mathbb{E}_x\left[\sum_{i,j} \log (1 - [D_{dec}(G(X))]_{i,j})\right] \qquad (3)$$

where Y is the ground-truth, high-quality underwater image, $G(X)$ is the generator's high-quality underwater image output, $\mathbb{E}(.)$ refers to the expected value, $D(Y)$ and $D(G(X))$ are the discriminator outputs, and (i, j) refer to both images' coordination of pixels (where both input and output have the same dimensions). The decoder output of the discriminator yields the probability of each pixel being real or fake in a separate way.

When humans observe an image, people extract the exact information of the image structurally, without the deviation between the corresponding pixels of images. However, as an evaluation criterion based on structural information to measure the degree of similarity between images, the structural similarity can not only overcome the influence of the changing texture caused by light changes, but it is also more suitable for human subjective

visual effects. We add the additional L_{SSIM} to ensure that the generated output resembles the given input image structurally.

$$L_{SSIM} = \frac{1}{m} \sum_{i=0}^{m-1} (1 - SSIM(G(X_i), Y_i)) , \tag{4}$$

where m is the batch size. As a similarity metric measuring function between images Y and $G(X)$, the definition of $SSIM$ [23] is shown as follows:

$$SSIM(Y', Y) = \frac{2\mu_{y'}\mu_y + C_1}{\mu_{y'}^2 + \mu_y^2 + C_1} \cdot \frac{2\sigma_{y'}\sigma_y + C_2}{\sigma_{y'}^2 + \sigma_y^2 + C_2} \cdot \frac{\sigma_{y'y} + C_3}{\sigma_{y'}\sigma_y + C_3} \tag{5}$$

where C_1, C_2, and C_3 are parameters to ensure division stability [24]. We set $C_3 = C_2/2$ while $C_1 = (0.03 * L)^2$ and $C_2 = (0.05 * L)^2$, where we use L as the representative of the image pixels' range. As described above, Y' is the generated output high-quality underwater image $G(X)$, Y is the ground truth (high-quality underwater image), $\sigma_{y'}$ and σ_y are modified standard deviations for Y' and Y, and $\mu_{y'}$ and μ_y are the modified mean values of their respective images. Finally, the modified covariance is referred as $\sigma_{y'y}$. The computation of $SSIM$'s detail can be found in [24].

3.4. Generator Loss

Three losses in different terms are also used in our generator (L_G). The first term is the ED (encode/decode) loss (L_{ED}), which focuses on image global structures and local details. We additionally use L1-distance loss (L_{L1}) and gradient loss (L_g) as

$$L_G = L_{ED} + \beta * L_{L1} + \gamma * L_g \tag{6}$$

where the terms β and γ define the hyperparameters for the generator network loss function. While the luminance information of the underwater image is pixel-intensity characterized, a gradient can partially be used as an indicator of texture detail for underwater image enhancement. Therefore, to obtain a high-quality underwater image with an intensity similar to the ground-truth underwater image, we use L_{L1} and L_g to describe the content loss of the image during the generation process. For a single image, L_{L1} and L_g are defined as

$$L_{L1} = \mathbb{E}[||Y - G(X)||_1] \tag{7}$$

$$L_g = \mathbb{E}[|||\nabla Y - \nabla G(X)|||] \tag{8}$$

where Y and $G(X)$ represent the ground-truth (high-quality) and generated underwater image, respectively, $|| \ ||_1$ denotes the L_1 norm of the matrix, and ∇ denotes the gradient operator.

$$L_{ED} = -\mathbb{E}_x \left[\log D_{enc}(G(X)) + \sum_{i,j} \log[D_{dec}(G(X))]_{i,j} \right] \tag{9}$$

In Equation (9), the objective function of ED loss (L_{ED}) forces $D(G(X))$ through maximization to deceive the discriminator. That is, when the generator generates an image, the higher the score is for realness, the more realistic it is. A generator's loss is calculated as if the discriminator classifies the realness of the generated images.

4. Experimental Results and Analysis

This section describes the datasets and metrics which are used to evaluate the proposed method. On both underwater images, either synthetic or real-world images, we compared the proposed method with state-of-the-art methods of underwater image enhancement. All experiments were implemented based on the CUDA C++ API on an i7 Intel CPU with 16 GB RAM and GTX 1080Ti GPU with 11 GB on-chip memory.

4.1. Underwater Dataset

In these experiments, we used two publically available underwater datasets, EUVP [25] and UIEB [26], to illustrate the robustness of the proposed method. By containing separate sets of 11,435 paired image samples of poor and good perceptual quality in the Enhancing Underwater Visual Perception (EUVP) dataset, supervised training of underwater image enhancement models can be facilitated. Each image is 256×256. The EUVP test set consists of 515 image pairs of the same size. There are 890 raw underwater images with corresponding reference images of high quality contained by the Underwater Image Enhancement Benchmark (UIEB) datum. In the UIEB dataset, we used a random subset of 800 images for training and the remaining 90 images for testing, and used five-fold cross-validation. In our training process, we used Adam optimization with a learning rate of 2×10^{-4} for the generator and a learning rate of 2×10^{-6} for the discriminator. We set the batch size to 32. With random initialization, networks were trained from scratch. We set $\alpha = 1.0$, $\beta = 2.0$, and $\gamma = 1.5$ in our experiments.

4.2. Quantitative Evaluation

We employed two full-reference metrics to secure the performance evaluation for each network: the structural similarity index measure (Equation (5)) and the peak signal-to-noise ratio (PSNR). A reflection of the proximity to the reference can be understood from these two metrics, where a high level of PSNR value represents closer image content and a reflection of more similarity between structures and textures represented by a higher SSIM value. To make our results more convincing, when evaluating the underwater images, we employed a nonreference metric: the underwater image quality measure (UIQM) [27]. A higher UIQM score suggests a more consistent result with human visual perception. The Frechet inception distance (FID) [28] has been used for quality measurement of the generated images. The FID is a method for comparing the statistics of two distributions by computing the distance between them. Mathematically, it can be written as

$$FID(r,g) = ||\mu_r - \mu_g||_2^2 + T_r\left(\Sigma_r + \Sigma_g - 2(\Sigma_r \Sigma_g)^{\frac{1}{2}}\right) \tag{10}$$

where μ_r, μ_g represent the means of real-world underwater images and generated underwater images, respectively; Σ_r, Σ_g represent the covariances of real-world and generated underwater images; and T_r denotes the matrix trace. A lower FID score means the generated image is more realistic.

4.3. Discussion

First, we evaluated the proposed method on the test sets of the two datasets with traditional and deep-learning methods. The quantitative evaluation is summarized in Table 1. The best results of different metrics are marked in bold. Table 1 shows that no method is best in terms of all metrics. However, the results demonstrate that DGAN outperforms other methods in terms of both PSNR and SSIM. To demonstrate the advantages of the proposed DGAN, visual comparisons are presented in Figure 5a,b. Although the two images may give the impression that the UDCP results are sharper, a careful inspection reveals that UDCP leads to excessive enhancement and oversaturation. Similarly, the UWGAN images are unnatural and excessively enhanced. As opposed to UWGAN, the proposed method presents promising results without introducing artificial colors and casting colors on real-world imagery. The main reason is that the generator network of the proposed model uses dilated convolution layers to control the receptive fields of a filter without changing the dimension of the data via polling or striding, and without increasing the number of parameters. This property is especially useful in tasks where the precise pixel position is important. Finally, the restored images by DGAN obtained better image enhancement results on the two datasets, with improvements in both visual quality and quantitative metrics compared to existing methods. Overall, deep-learning-based methods produce better restoration results than traditional methods, which produce blurry

and colorcast results. This is exhibited for DCP [10], UDCP [11], and the retinex-based method [17], as the ineffective result shows the unsuccessful removal of the green-bluish tone in the underwater images. Deep-learning-based methods have fewer color artifacts and more high-fidelity object areas.

Table 1. Comparison of different algorithms on two different datasets.

| EUVP Dataset	SSIM↑	PSNR↑	UIQM↑	FID↓
DCP	0.825	27.44	3.17	27.54
UDCP	0.745	27.83	3.81	25.87
Retinex based	0.821	27.94	3.64	39.42
UGAN	0.775	28.56	**4.82**	21.49
UWGAN	0.793	27.69	4.12	25.74
WaterGAN	0.834	**28.73**	4.37	**18.76**
DGAN	**0.857**	28.20	4.39	19.55
UIEB Dataset	**SSIM↑**	**PSNR↑**	**UIQM↑**	**FID↓**
DCP	0.863	27.76	2.17	29.54
UDCP	0.667	28.14	3.81	22.87
Retinex based	0.779	27.94	3.64	32.42
UGAN	0.763	28.28	4.11	**16.49**
UWGAN	0.841	28.35	3.77	17.74
WaterGAN	0.784	27.33	**4.27**	18.76
DGAN	**0.894**	**28.57**	3.90	16.55

(a)

Figure 5. *Cont.*

(**b**)
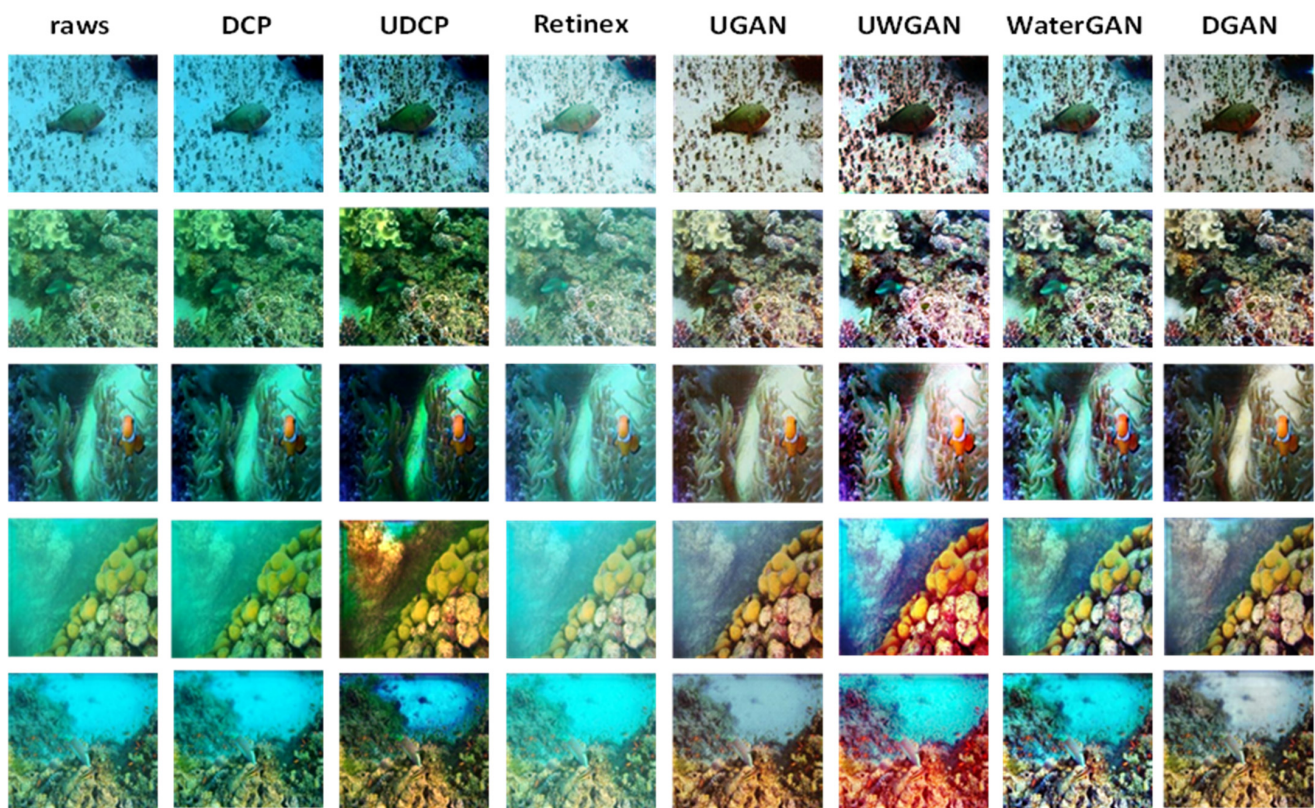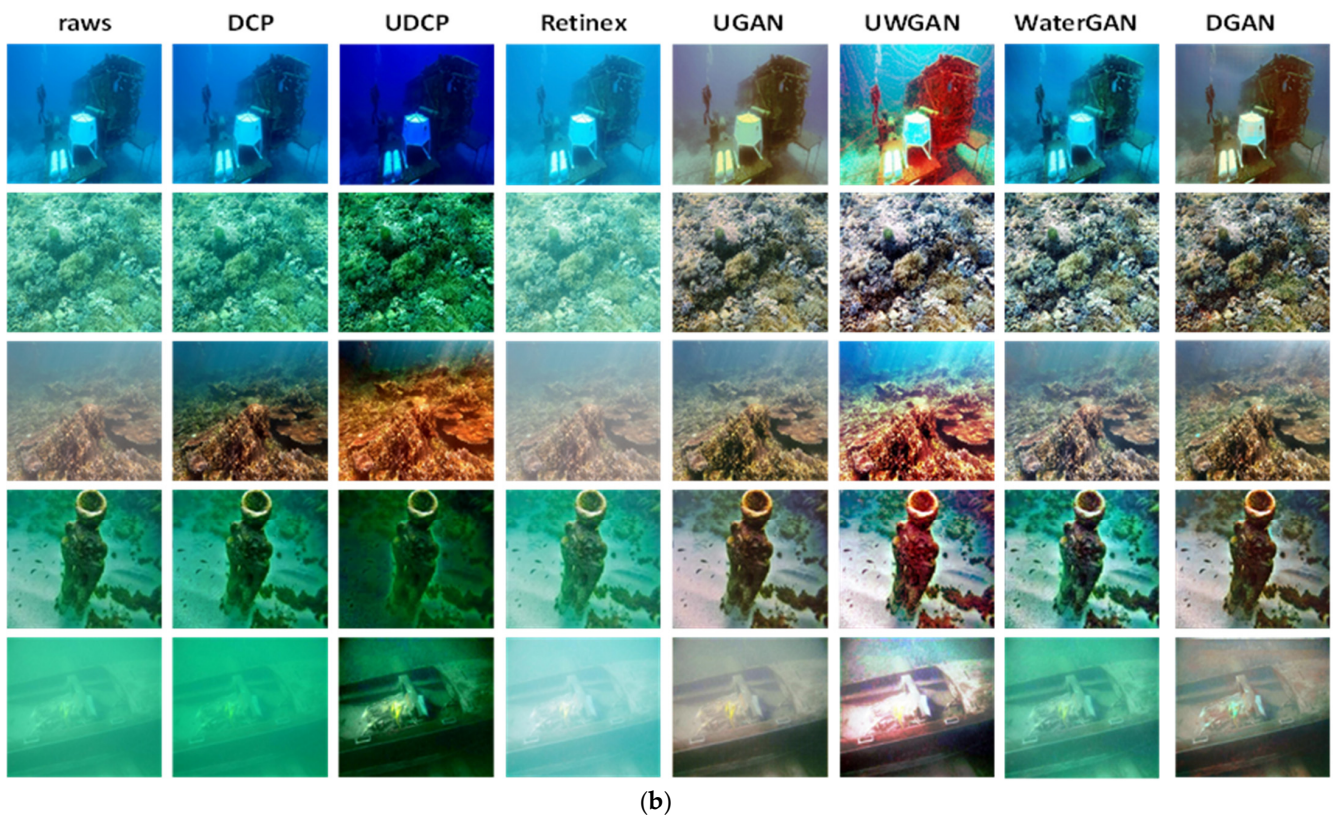
Figure 5. Qualitative underwater image enhancement performance on (**a**) EUVP dataset and (**b**) UIEB dataset.

Considering underwater scenes, we also employed the UIQM metric to evaluate underwater image enhancement. Note that UIQM is a valuable reference but cannot give absolute justifications, for they are nonsensitive to color artifacts and casts and biased to some features. In Table 1, the UIQM results were obtained by applying different methods in different underwater scenes. Although the proposed DGAN does not achieve the best UIQM results, it does exhibit good performance in underwater images in various scenes. In addition, Figure 6 shows the qualitative results of three examples with zoomed patches; we explored the block effect of enhancement underwater images. In this figure, we contrast the featured image patches by local pattern. The results without dilated convolution (UGAN) exhibit an obvious block effect. However, in an effort to achieve better subjective perception, DGAN with dilated concatenation removes this block effect at the cost of decreased UIQM performance.

Figure 6. Visual comparison with UGAN on EUVP dataset. The first row is UGAN, and the second row is the proposed method.

5. Conclusions

As the use of marine resources increases, the demand to improve the visual quality of underwater images has also increased. After enhancing underwater images, we promote underwater object detection and classification performance. In this paper, we present DGAN, a method for enhancing underwater imagery. The U-Net architecture of the discriminator and the dilation convolution of the generator are used to improve visual perception when we construct the DGAN neural network. In addition, to establish a robust connection between local patches and global contents, we change both the loss function and basic architecture. Our results have shown high performance, which is in a dominant position compared to state-of-the-art underwater image enhancement methods. In future work, more effective underwater image enhancement frameworks via unsupervised learning to reduce the burden of paired training data needs to be exploited.

References

1. Li, R.; Cheong, L.-F.; Tan R., T. Heavy rain image restoration: Integrating physics model and conditional adversarial learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–20 June 2019; pp. 1633–1642.
2. Yang, W.; Zhang, X.; Tian, Y.; Wang, W.; Xue, J.H.; Liao, Q. Deep Learning for Single Image Super-Resolution: A Brief Review. *IEEE Trans. Multimed.* **2019**, *21*, 3106–3121. [CrossRef]
3. Wang, Z.; Wang, L.; Duan, S.; Li, Y. An Image Denoising Method Based on Deep Residual GAN. *J. Phys. Conf. Ser.* **2020**, *1550*, 032127. [CrossRef]
4. Zhu, J.-Y.; Park, T.; Isola, P.; Efros, A.A. Unpaired image-to-image translation using cycle-consistent adversarial networks 2017. In Proceedings of the IEEE international conference on computer vision, Venice, Italy, 30 March 2017; pp. 2223–2232.

5. Wang, N.; Zhou, Y.; Han, F.; Zhu, H.; Yao, J. UWGAN: Underwater GAN for Real-world Underwater Color Restoration and Dehazing. *arXiv* **2019**, arXiv:1912.10269.
6. Li, J.; Skinner, K.; Eustice, R.; Johnson-Roberson, M. WaterGAN: Unsupervised generative network to enable real-time color correction of monocular underwater images. *IEEE Robot. Autom. Lett.* **2018**, *3*, 387–394. [CrossRef]
7. Fabbri, C.; Islam, M.; Sattar, J. Enhancing underwater imagery using generative adversarial networks. In Proceedings of the 2018 IEEE International Conference on Robotics and Automation, Brisbane, QLD, Australia, 21–25 May 2018; pp. 7159–7165.
8. Guo, Y.; Li, H.; Zhuang, P. Underwater Image Enhancement Using a Multiscale Dense Generative Adversarial Network. *IEEE J. Ocean. Eng.* **2020**, *45*, 862–870. [CrossRef]
9. Islam, M.J.; Xia, Y.; Sattar, J. Fast Underwater Image Enhancement for Improved Visual Perception. *IEEE Robot. Autom. Lett.* **2020**, *5*, 3227–3234. [CrossRef]
10. He, K.; Sun, J.; Tang, X. Single image haze removal using dark channel prior. *IEEE Trans. Pattern Anal. Mach. Intell.* **2011**, *33*, 2341–2353. [PubMed]
11. Drews, P.; Nascimento, E.; Moraes, F.; Botelho, S.; Campos, M. Transmission estimation in underwater single images. In Proceedings of the IEEE international conference on computer vision workshops, Sydney, Australia, 2–8 December 2013; pp. 825–830.
12. Galdran, A.; Pardo, D.; Picn, A. Automatic Red-Channel underwater image restoration. *J. Vis. Commu. Image Repre.* **2015**, *26*, 132–145. [CrossRef]
13. Peng, Y.; Cao, T.; Cosman, P. Generalization of the dark channel prior for single image restoration. *IEEE Trans. Image Processing* **2018**, *27*, 2856–2868. [CrossRef] [PubMed]
14. Fu, X.; Fan, Z.; Ling, M.; Huang, Y.; Ding, X. A hybrid method for underwater image correction. *Pattern Rec. Lett.* **2017**, *94*, 62–67.
15. Iqbal, K.; Odetayo, M.; James, A. Enhancing the low quality images using unsupervised colour correction method. In Proceedings of the 2010 IEEE International Conference on Systems, Man and Cybernetics, Istanbul, Turkey, 10–13 October 2010; pp. 1703–1709.
16. 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.
17. 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.
18. Fu, X.; Fan, Z.; Ling, M. Two-step approach for single underwater image enhancement. In Proceedings of the 2017 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS), Xiamen, China, 6–9 November 2017; pp. 789–794.
19. Goodfellow, I.J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.C.; Bengio, Y. Generative adversarial nets. In Proceedings of the NIPS, Montreal, QC, Canada, 8–13 December 2014.
20. Schonfeld, E.; Schiele, B.; Khoreva, A. A u-net based discriminator for generative adversarial networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 8207–8216.
21. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*; Springer: Berlin/Heidelberg, Germany, 2015; pp. 234–241.
22. Yu, F.; Koltun, V. Multi-scale context aggregation by dilated convolutions. In Proceedings of the ICLR, San Juan, Puerto Rico, 2–4 May 2016.
23. Sheikh, H.R.; Sabir, M.F.; Bovik, A.C. A statistical evaluation of recent full reference image quality assessment algorithms. *IEEE Trans. Image Processing* **2006**, *15*, 3440–3451. [CrossRef] [PubMed]
24. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Processing* **2004**, *13*, 600–612. [CrossRef] [PubMed]
25. EUVP Dataset. Available online: http://irvlab.cs.umn.edu/resources/euvp-dataset (accessed on 22 February 2022).
26. 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 Processing* **2020**, *29*, 4376–4389. [CrossRef] [PubMed]
27. Panetta, K.; Gao, C.; Agaian, S. Human-Visual-System-Inspired Underwater Image Quality Measures. *IEEE J. Ocean. Eng.* **2016**, *41*, 541–551. [CrossRef]
28. Heusel, M.; Ramsauer, H.; Unterthiner, T.; Nessler, B.; Hochreiter, S. Gans trained by a two time-scale update rule converge to a local Nash equilibrium. In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; pp. 6626–6637.