

Overwater Image Dehazing via Cycle-Consistent Generative Adversarial Network

Shunyuan Zheng, Jiamin Sun, Qinglin Liu, Yuankai Qi, and Shengping Zhang

School of Computer Science and Technology, Harbin Institute of Technology, Weihai
264209, China
{sawyer0503,sunjiamin17}@gmail.com

Abstract. In contrast to images taken on land scenes, images taken over water are more prone to degradation due to the influence of the haze. However, existing image dehazing methods are mainly developed for land scenes and perform poorly when applied to overwater images. To address this problem, we collect the first overwater image dehazing dataset and propose an OverWater Image Dehazing GAN (OWI-DehazeGAN). Due to the difficulties of collecting paired hazy and clean images, the dataset is composed of unpaired hazy and clean images taken over water. The proposed OWI-DehazeGAN learns the underlying style mapping between hazy and clean images in an encoder-decoder framework, which is supervised by a forward-backward translation consistency loss for self-supervision and a perceptual loss for content preservation. In addition to qualitative evaluation, we design an image quality assessment network to rank the dehazed images. Experimental results on both real and synthetic test data demonstrate that the proposed method performs superiorly against several state-of-the-art land dehazing methods.

Keywords: Image dehazing · Overwater image · Unpaired data · Generative adversarial networks

1 Introduction

Images of overwater scenes play an important role in human image galleries. However, these images are prone to degradation due to thick mist that are often appearing over lakes, rivers, and seas. Although numerous image dehazing methods have been developed [1–5], our experiments show that these methods perform far from satisfying since they are originally designed for land scene images, of which the data distribution differs significantly.

Hazy images are usually modeled as $I(x) = J(x)t(x) + A(1 - t(x))$, where $I(x)$ and $J(x)$ are the observed hazy image and the scene, respectively [6, 7]. The symbol x denotes a pixel index, and A is the global atmospheric light. $t(\cdot)$ denotes the transmission map, which describes the portion of light that is not scattered and reaches the camera sensors. When the haze is homogeneous, $t(\cdot)$ can be defined as: $t(x) = e^{-\beta d(x)}$, where β is the scattering coefficient and $d(x)$ is the distance between objects and the camera.

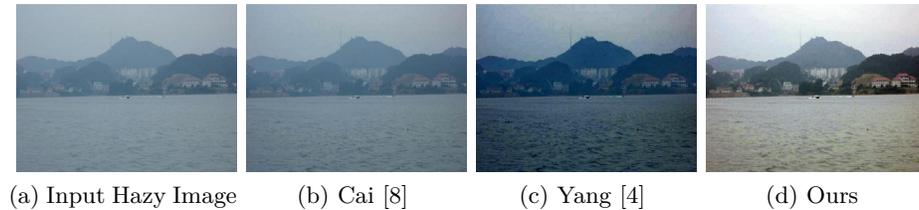


Fig. 1. Overwater image dehazing example. The proposed method generates more clear images compared to state-of-the-art methods.

Existing methods fall into two categories according to the type of features they used: methods based on hand-crafted features [1, 2, 9–12] or methods based on CNN features [3–5, 8, 13–15]. The former generally focuses on estimating the global atmospheric light intensity and the transmission map, and hence their performance are susceptible to estimation errors of $A(\cdot)$ or $t(\cdot)$. To alleviate these limitations, the latter, which is based on CNNs [16] or Generative Adversial Networks (GANs) [17], aims to directly estimate clean images in a data-driven scheme. Although promising dehazing results have been achieved, existing CNN- or GAN- based methods perform not well on overwater images as shown in Figure 1b and 1c. Another issue is that existing image dehazing datasets [18–20] are dominated by land scenes.

In this paper, we address both the above mentioned issues. *First*, we construct a new dataset, named OverwaterHaze, specially for dehazing overwater images. Since collecting paired hazy and clear images is difficult and expensive, the OverwaterHaze dataset is composed of unpaired hazy and clean overwater images. *Second*, we propose an OverWater Image Dehazing GAN (OWI-DehazeGAN) inspired by CycleGAN [21] to directly recover clean images. Although the unpaired character challenges most of existing methods, we demonstrate that satisfying dehazing performance could be achieved by the proposed method.

Our contributions are summarized as follows:

- We create the first overwater image dehazing dataset, and we hope this dataset is able to facilitate the research in this field.
- We propose an OWI-DehazeGAN to dehaze overwater images, which is based on but performs superior to Cycle-GAN. The proposed network is able to utilize unpaired training data and preserve image details simultaneously.
- We propose an image quality assessment network to rank the generated dehazed images, which facilitates the comparison of different algorithms.

2 Related Work

2.1 Methods Based on Hand-crafted Features

Many efforts have been devoted to image dehazing in the past decades based on hand-crafted features [1, 2, 9–12]. Tan *et al.* [9] propose a contrast maximizing

approach using markov random fields (MRF) based on the observation that clean images have higher contrast than hazy ones. In [10] Tarel *et al.* propose a fast dehazing method by combining atmospheric veil inference, image restoration and smoothing tone mapping. Later, He *et al.* [11] estimate the transmission map by utilizing dark-channel prior (DCP). Meng *et al.* [12] explore the inherent boundary constraint on the transmission function. In order to recover depth information, Zhu *et al.* [1] propose a color attenuation prior (CAP) by creating a linear model on local priors. Different from previous methods that use various patch-based priors, Berman *et al.* [2] present a new image dehazing algorithm based on non-local prior so that a haze-free image is able to be well approximated by a few distinct colors.

While the afore-mentioned methods have achieved promising results, they perform far from satisfying when applied to overwater images. MRF [9] tends to produce over-saturated images. The enhanced images of FVR [10] often contain distorted colors and severe halos. DCP [11] does not work well when it comes to the sky regions, as the scene objects are similar to the atmospheric light.

2.2 Methods Based on CNN Features

Deep convolutional neural networks have shown promising success in various computer vision tasks. Cai *et al.* [8] propose an end-to-end DehazeNet with non-linear regression layers to estimate medium transmission. Instead of estimating the transmission map or the atmospheric light firstly, AOD-Net [15] predicts the haze-free images directly using a light-weight CNN. Proximal Dehaze-Net [4] attaches the advantages of traditional prior-based dehazing methods to deep learning technologies by incorporating the haze-related prior learning.

Since Goodfellow [17] proposed the GAN method in 2014, there have been many effective variants tailored to different computer vision tasks [21–23]. Motivated by the success of GANs in those regions, several GAN-based methods have been proposed for image dehazing. In [5], a Densely Connected Pyramid Dehazing Network (DCPDN) is proposed to jointly learn the transmission map, atmospheric light and dehazing result all together. Yang *et al.* [24] propose to loose the paired training constraint by introducing a disentanglement and reconstruction mechanism. Li *et al.* [13] design a solution based on a cGAN network [22] to directly estimate the clean image. Ren *et al.* [3] adopt an ensemble strategy to take advantage of information in white balance, contrast enhancing, and gamma correction images. Overall, these methods are trained on paired data, which is unsuitable for the proposed overwater image dehazing task, where only unpaired training data is available.

2.3 Image Dehazing Dataset

Image dehazing tasks profit from the continuous efforts for large-scale data. Several datasets [18–20] have been introduced for image dehazing.

MSCNN [14] and AOD-Net [15] utilize the indoor NYU2 Depth Database [25] and the Middlebury Stereo database [26] to synthesize hazy images using the

known depth information. O-HAZE [19] is an outdoor scenes dataset, which is composed of pairs of real hazy and corresponding clean data. I-HAZE [19] is a dataset that contains 35 image pairs of hazy and corresponding ground-truth indoor images. Li *et al.* [20] have launched a new large-scale benchmark which is made up of synthetic and real-world hazy images, called Realistic Single Image Dehazing (RESIDE). However, most datasets are synthetic and not tailored to handling overwater image dehazing. Different from the above datasets, we collect a dataset containing real data which is specially for dehazing overwater images.

3 Proposed Method

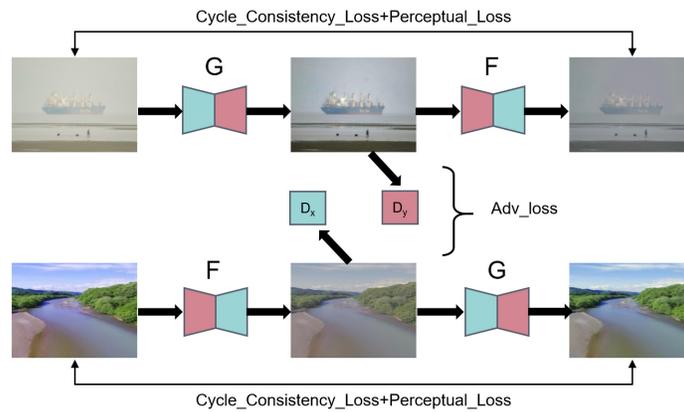


Fig. 2. The main architecture of the proposed OWI-DehazeGAN network. G and F denote generators, where $G : X \rightarrow Y$ generates clean images from hazy images and $F : Y \rightarrow X$ vice versa. D_x and D_y denote discriminators. Adversarial loss, cycle consistency loss and perceptual loss are employed to train the network.

Figure 2 shows the main architecture of the proposed OWI-DehazeGAN. Unlike traditional GANs, OWI-DehazeGAN consists of two generators (G and F) and two discriminators (D_x and D_y) in order to be trainable with unpaired training data. Specifically, generator G predicts clean images Y from hazy images X , and F vice versa. D_x and D_y distinguish hazy images and clean images, respectively. Below we provide more details about each component.

3.1 Generator

We adopt the same structure for the two generators G and F . Both generators are divided into three parts: encoding, transformation, and decoding. The architecture of the generator is shown in Figure 3a.

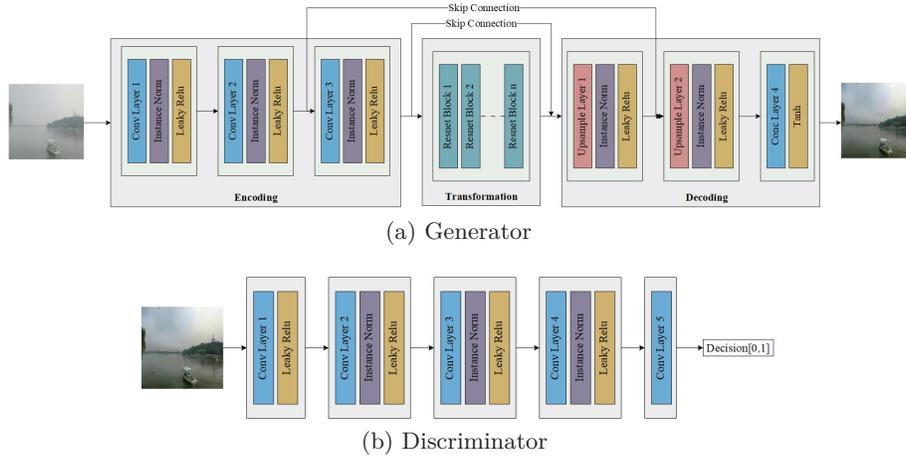


Fig. 3. Architecture of our generator and discriminator. The Generator consists of encoding, transformation, and decoding three parts.

Encoding. The encoding module extracts image features by three convolution layers, which serve as down-sampling layers to decrease the resolution of the original input. Each convolution layer is followed by an instance normalization and a Leaky ReLU. Since image dehazing can be treated as a domain adaptation problem, instance normalization is more suitable than batch normalization.

Transformation. The transformation module translates information from one domain to another via nine ResNet blocks [27]. The ResNet block in our network contains two 3×3 convolution layers with the same number of filters. Due to the results of image dehazing need to retain the characteristics of the original image, the ResNet block is well suited to accomplish these transformations.

Decoding. The decoding module includes up-sampling operations and non-linear mappings. There are several choices for upsampling, such as deconvolution [28], sub-pixel convolution [29] and resize convolution [30]. In order to reduce checkerboard artifacts [30] caused by deconvolution or sub-pixel convolution, we use the resize convolution for decoding. Inspired by the success of U-Net [31], we introduce two symmetric skip connections to deliver information between encoding and decoding modules. Finally, images are recovered through convolution and tanh activation.

3.2 Discriminator

We use two discriminators D_x and D_y to distinguish the input hazy images and clean images, respectively. The discriminator is implemented in a fully convolution fashion, as shown in Figure 3b. We use four convolution blocks in discriminator. The first block consists of a convolution layer and a Leaky ReLU, the last block only contains a convolution layer and the remaining blocks are composed of a convolution layer, an instance normalization and a Leaky ReLU.

3.3 Loss Function

We utilize three kinds of losses to enable the proposed network trainable with unpaired data and preserve image details simultaneously.

Adversarial loss. As done in CycleGAN, we use the adversarial loss and the cycle consistency loss for unpaired training data. $x \in X$, $y \in Y$ are a hazy image and an unpaired clean image, respectively. For the generator G and discriminator D_y , the adversarial loss is formulated as:

$$L_{GAN}(G, D_y, x, y) = \log(D_y(y)) + \log(1 - D_y(G(x))) \quad (1)$$

Correspondingly, the constraint on generator F and its discriminator D_x is

$$L_{GAN}(F, D_x, x, y) = \log(D_x(x)) + \log(1 - D_x(F(y))) \quad (2)$$

However, the above losses are prone to unstable training and generating low quality images. To make the training more robust and achieve high quality images, we use a least squares loss [23] instead of the negative log likelihood objective [17]. Therefore, Eq (1) and (2) are modified as:

$$L_{adv}(G, D_y, x, y) = \frac{1}{2} * [(D_y(G(x)) - 1)^2 + (D_y(y) - 1)^2 + D_y(G(x))^2] \quad (3)$$

$$L_{adv}(F, D_x, x, y) = \frac{1}{2} * [(D_x(F(y)) - 1)^2 + (D_x(x) - 1)^2 + D_x(F(y))^2] \quad (4)$$

And the final adversarial loss is denoted as:

$$L_{adv}(G, F, D_x, D_y) = L_{adv}(G, D_y, x, y) + L_{adv}(F, D_x, x, y) \quad (5)$$

Cycle consistency loss. CycleGAN introduces a cycle consistency loss to solve the problem that an adversarial loss alone cannot ensure the matching between the output distribution and the target distribution. For each image x , $F(G(x))$ is able to bring $G(x)$ back to the original image. Similarly, $G(F(y))$ is able to bring $F(y)$ back to the original image y . $F(G(x))$ is the cyclic image of input x , and $G(F(y))$ is the cyclic image of the original image y . To train generators G and F at the same time, the consistency loss includes two constraints: $F(G(x)) \approx x$, $G(F(y)) \approx y$. The cycle consistency loss is defined to calculate L1-norm between the input and the cyclic image for unpaired image dehazing:

$$L_{cyc}(G, F) = \|F(G(x)) - x\|_1 + \|G(F(y)) - y\|_1 \quad (6)$$

Perceptual loss. We introduce perceptual loss to restrict the reconstruction of image details. Instead of measuring per-pixel difference between the images, perceptual loss is concerned with the distinction between feature maps, which comprises various aspects of content and perceptual quality. The perceptual loss is defined as:

$$L_{per}(G, F) = \|\theta(x) - \theta(F(G(x)))\|_2^2 + \|\theta(y) - \theta(G(F(y)))\|_2^2 \quad (7)$$

Here, θ represents the feature maps which generated from the relu4_2 on pretrained VGG-16 [32] network.

Objective function. We define the loss as a weighted sum of previous losses:

$$L(G, F, D_x, D_y) = L_{adv}(G, F, D_x, D_y) + \lambda L_{cyc}(G, F) + \mu L_{per}(G, F) \quad (8)$$

Where coefficients λ and μ represent the weights of cycle consistency loss and perceptual loss, respectively. We found that giving an over-weight to perceptual loss may cause the instability of training process thus the weight of perceptual loss should be much less than the weight of cyclic consistency loss. We minimize the generators G, F and maximize the discriminators D_x, D_y in training process.

The final objective function is:

$$\langle G^*, F^* \rangle = \arg \min_{G, F} \max_{D_x, D_y} L(G, F, D_x, D_y) \quad (9)$$

3.4 Dehazed Image Quality Assessment

In order to verify the effectiveness of the proposed OWI-DehazeGAN, we design a dehazed image quality assessment model based on natural image statistics and VGG network. Natural images are directly captured from natural scenes, so they have some natural properties. By making statistics on these properties, natural scene statistics (NSS[33]) of images can be obtained. NSS has been widely used in image quality assessment, especially no-reference image quality assessment.

In this paper, the NSS we use is mean subtracted contrast normalization (MSCN [34]) coefficients, which is used to normalize a hazy image. After normalization pixel intensities of haze-free by MSCN coefficient follow a Gaussian Distribution while pixel intensities of hazy images do not. The deviation of the distribution from an ideal bell curve is therefore a measure of the amount of distortion in the image. To calculate the MSCN Coefficients, the image intensity $I(i, j)$ at pixel (i, j) is transformed to the luminance $\hat{I}(i, j)$. $\hat{I}(i, j)$ is defined as:

$$\hat{I}(i, j) = \frac{I(i, j) - \mu(i, j)}{\sigma(i, j) + C} \quad (10)$$

where $\mu(i, j)$ and $\sigma(i, j)$ represent the local mean field and local variance field obtained by calculating the image using a gaussian window with a specific size. Local Mean Field μ is the Gaussian Blur of the input image. Local Variance Field σ is the Gaussian Blur of the square of the difference between original image and μ . In case the denominator is zero, a constant C is added. When a dehazed image is normalized by MSCN coefficient, only the uniform appearance and the edge information are retained. Human eyes are very sensitive to edge information, so the normalized image is consistent with human vision.

The proposed IQA model for dehazed images is consist of luminance normalization, feature extraction and regression of evaluation score. The dehazed images are firstly normalized by MSCN coefficient, which provides a good normalization of image luminance and does not have a strong dependence on the intensity of texture. Then a VGG-16 model is used to extract features for finally predicting an image quality score between 0 and 9 through two fully connected

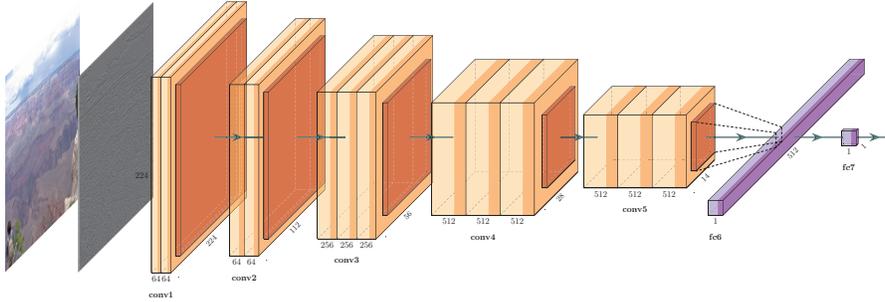


Fig. 4. Architecture of the proposed IQA model for dehazed images.

layers, whose units are 512 and 1 respectively. The architecture of the IQA model is shown in Figure 4. The loss function of this IQA model is MAE. The loss is defined as:

$$loss_{IQA} = \frac{1}{N} \sum_{i=1}^n |y_i - y_i^*| \quad (11)$$

where N represents the number of images in the training set, y_i and y_i^* denote target data and output data, respectively. The optimization goal of the IQA model in the training phase is to minimize the average absolute error loss. Learning the mapping between dehazed images and corresponding Mean Opinion Scores (MOS [35]) is achieved by minimizing the loss between the predicted score y_i^* and the corresponding ground truth y_i .

4 Experiment

4.1 Dataset

We collect a real unpaired image dataset called OverwaterHaze for image dehazing in overwater scenes. The training set consists of 4531 unpaired images, which are 2090 hazy images and 2441 clean images, all images are crawled from Google. These training images are resized to 640×480 . Figure 5 illustrates some examples of our dataset. There are three differences between the proposed dataset and the existing datasets: (1) The OverwaterHaze dataset is a large-scale natural dataset with hazy images and unpaired haze-free images, as the previous datasets are only composed of synthetic data; (2) The OverwaterHaze dataset is tailored to the task of overwater image dehazing, rather than focusing on indoor or outdoor scenes; (3) The proposed dataset is much more challenging because the regions of sky and water surface make up a large part of the image.

In order to evaluate different image dehazing methods in overwater scene subjectively, we introduce a natural overwater testing set, which contains 127



Fig. 5. Examples of the training set in OverwaterHaze dataset (best viewed in color). (a)–(f) Hazy images. (g)–(l) Clean images.

challenging hazy images collected from overwater scenes. To quantitatively compare different image dehazing methods, we select 90 (30 images \times 3 medium extinction coefficients β) overwater hazy images with corresponding ground-truth from the RESIDE OTS dataset [20]. We apply SSIM, PSNR and CIEDE2000 to the dehazed results on synthetic images. Based on the OverwaterHaze dataset, we compare our proposed method against several state-of-the-art dehazing methods in real and synthetic data, including: DCP [11], FVR [10], BCCR [12], CAP [1], DehazeNet [8], MSCNN [14], AOD-Net [15], dehaze-cGAN [13], Proximal Dehaze-Net [4], CycleGAN [21].

4.2 Experimental Settings

The input images of generators and discriminators are set to 256×256 during training. We use an Adam solver to optimize gradient with a learning rate of $2e-4$. The batch size is 1. The weights of cyclic consistency loss λ and perceptual loss μ are 10 and 0.0001, respectively. The coefficient a of Leaky ReLU is 0.2. The update proportion is 1 for generators G , F and discriminators D_x , D_y .

4.3 Qualitative Results on Real Images

Figure 6 shows an example of dehazing results of the proposed algorithm against the state-of-the-art methods. DCP [11] tends to overestimate the thickness of the haze and produce dark results (Figure 6b). The dehazed images by FVR [10] and BCCR [12] have significant color distortions and miss most details as shown in Figure 6c~6d. The best performer in the hand-crafted prior based methods is CAP [1], which generally reconstructs details of haze-free images. The deep learning based approach achieve comparable results, such as DehazeNet [8], MSCNN [14] and dehaze-cGAN [13]. But these results indicate that existing methods cannot handle overwater hazy images well. For example, the dehazed results by MSCNN and DehazeNet (Figure 6f~6g) have a similar problem that tends to magnify the phenomenon of color cast and have some remaining haze.

The illumination appears dark in the results of Proximal Dehaze-Net [4] and AOD-Net [15], as shown in Figure 6h~6j. From Figure 6k, CycleGAN [21] generates some pseudo-colors to a certain degree, which makes it quite different from the original colors. Meanwhile, its result generates extensive checkerboard artifacts in the sky regions. In contrast, the dehazed result by our method shown in Figure 6l is visually pleasing in the mist condition.

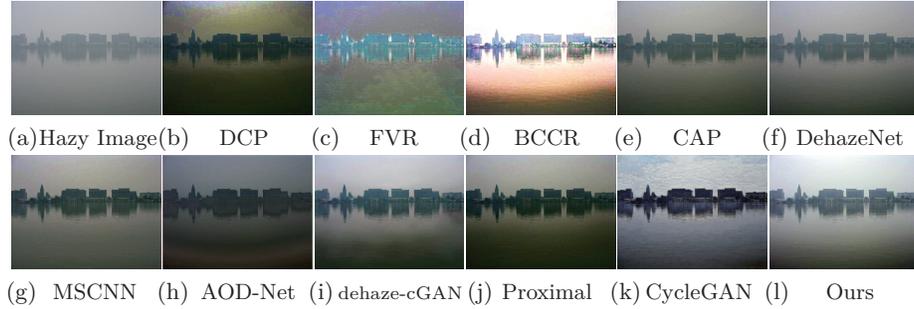


Fig. 6. Real hazy images and corresponding dehazing results from several state-of-the-art methods (best viewed in color).

4.4 Qualitative and Quantitative Results on Synthetic Images

We further conduct some experiments based on synthetic hazy images. Although the proposed method is trained on real unpaired data, we note that it can be applied for synthetic images as well. Figure 7 shows some dehazed images generated by various methods. Figure 7a shows the groundtruth as reference. As shown in Figure 7b~7d, the results of DCP[11], FVR [10], and BCCR[12] have some distortions in colors or details. The dehazed results processed by CAP [1] (Figure 7e), DehazeNet [8] (Figure 7f), MSCNN [14] (Figure 7g), AOD-Net [15] (Figure 7h), dehaze-cGAN [13] (Figure 7i), and Proximal Dehaze-Net [4] (Figure 7j) are closer to groundtruth 7a than the results based on priors. However, there still exists some remaining haze as shown in Figure 7e~7h. The result generated by CycleGAN [21] in Figure 7k exists serious color cast and losses some color information. The dehazed result generated by our approach in Figure 7l, by contrast, is visually close to the groundtruth image.

An advantage of testing on synthetic data is able to objectively evaluate experimental results via SSIM, PSNR and CIEDE2000. Higher SSIM score indicates that the generated results are more consistent with human perception. PSNR forecasts the effectiveness of image dehazing, and CIEDE2000 presents that smaller scores indicate better color preservation. In Figure 7, the SSIM and PSNR values also indicate that our method surpass other methods. From Table 1, our method get higher PSNR and SSIM. Remarkably, the SSIM and PSNR of our model are significantly better than CycleGAN.

Table 1. Average PSNR, SSIM and, CIEDE2000 values of the dehazed results on the new synthetic dataset. The best result, the second result, and the third place result are represented by red, blue, and green, respectively.

	DCP	FVR	BCCR	CAP	DehazeNet	MSCNN	AOD-Net	dehaze-cGAN	Proximal	CycleGAN	Ours
SSIM	0.717	0.817	0.680	0.825	0.867	0.850	0.835	0.876	0.820	0.584	0.893
PSNR	14.57	16.56	13.92	19.50	23.27	20.55	19.33	22.77	17.79	18.31	24.93
CIEDE2000	14.92	11.71	15.35	9.37	6.23	7.81	9.11	6.51	10.14	11.5	5.98

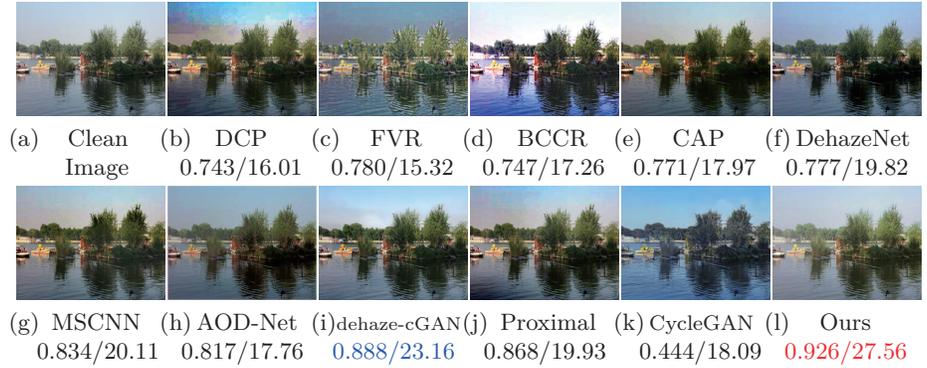


Fig. 7. Comparison in terms of SSIM/PSNR for different dehazing methods.

4.5 Dehazed Images Ranking

The proposed IQA model for dehazed images is pre-trained on TID2013 [35] and then fine-tuned using the IVC Dehazing Dataset [18]. The TID2013 includes different types of image distortion, while IVC Dehazing Dataset is designed to evaluate the quality of dehazed images. Predicted scores are used to qualitatively rank photos as shown in Figure 8. Ranking scores and the ranking are presented below each image, where ‘1’ denotes the best visual perception and ‘10’ for the worst image quality. Figure 8 shows that the quality of overwater dehazed images generated by OWI-DehazeGAN is better than other methods. For a comprehensive comparison, we also report the dehazed image quality measured by four typical image quality assessment methods in Table 2. The best results are shown in red font. Table 2 shows that the proposed method achieves the best performance in terms of almost all metrics.

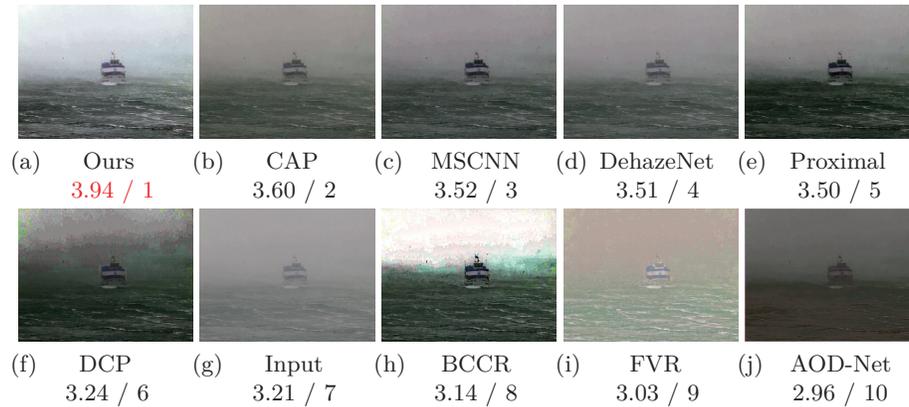
5 Analysis and Discussion

5.1 Effect of Resize Convolution

In the decoding process of the generator, we use the resize convolution to increase the resolution of feature maps, rather than deconvolution or sub-pixel convolution. To better understand how the resize convolution contributes to our

Table 2. Comparison of dehazed image quality using four image quality assessment methods. The top three results are in red, blue, and green font, respectively.

	FADE ↓	SSEQ ↓	BLINDS-2 ↓	NIMA ↑
Ours	1.95	36.24	31.50	6.47
CAP	3.03	40.97	51.00	5.95
MSCNN	2.38	45.56	50.50	6.23
DehazeNet	3.11	42.49	49.50	6.12
Proximal-Dehaze	1.69	43.69	54.50	6.06
DCP	1.45	43.32	51.50	5.36
BCCR	1.17	45.76	48.50	6.08
FVR	2.67	41.71	48.00	4.86
AOD-Net	2.30	38.62	54.50	6.15

**Fig. 8.** Comparison via the proposed IQA model. Ranking scores and the ranking are shown below each image.

proposed method, we train three end-to-end networks with different upsample mode: (i) deconvolution, (ii) sub-pixel convolution, and (iii) resize convolution.

Figure 9 shows the results of three upsampling mode in our network. In comparison, the result of resize convolution is best viewed from the perspective of the human perception and retain more detailed information. From Figure 9c, we observe plenty checkerboard pattern of artifacts caused by deconvolution. Although the sub-pixel convolution (Figure 9d) can alleviate the ‘checkerboard artifacts’ to some extent, the result of sub-pixel convolution is rough and unsatisfying. Compared with the first two approaches, resize convolution recover most scene details and maintain the original colors. From Table 3, the introduced resize convolution gains higher PSNR, SSIM scores and a lower CIEDE2000 score than deconvolution and sub-pixel convolution, which indicate resize convolution can generate visually perceptible images.

Table 3. Average scores in terms of PSNR, SSIM, and CIEDE2000 for three upsampling convolutions on the synthetic test set of the OverwaterHaze dataset.

Average Metrics	SSIM	PSNR	CIEDE2000
Deconvolution	0.758	20.34	9.12
Sub-pixel Convolution	0.643	20.21	10.86
Resize Convolution	0.819	22.19	7.41

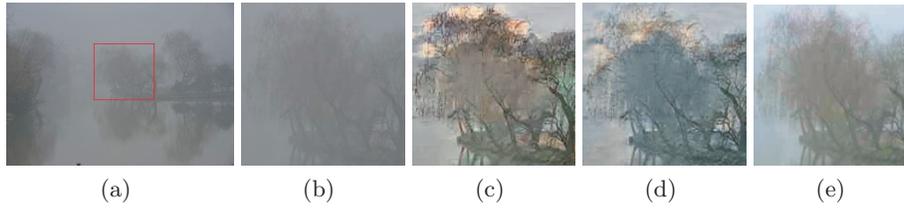


Fig. 9. Effectiveness of the proposed network with resize convolution. (a) and (b) are input hazy images. (b)~(e) are the zoom-in views. (c)~(e) are the dehazing results of deconvolution, sub-pixel convolution and resize convolution, respectively.

Table 4. Effect of perceptual loss in terms of SSIM, PSNR, and CIEDE2000.

Average Metrics	SSIM	PSNR	CIEDE2000
CycleGAN loss	0.819	22.19	7.41
CycleGAN loss+VGG loss	0.893	24.93	5.98

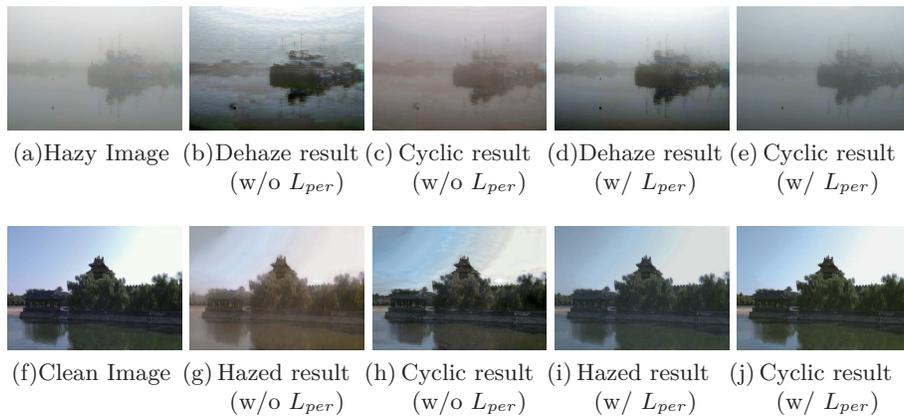


Fig. 10. Comparison of dehazing with and without perceptual loss. (a)~(e) represents the generation direction of $X \rightarrow Y \rightarrow X$. (f)~(j) says the direction of formation is $Y \rightarrow X \rightarrow Y$. ‘(w/o L_{per})’ denotes the network without perceptual loss, and ‘(w/ L_{per})’ denotes the network with perceptual loss.

5.2 Effect of Perceptual Loss

To show the effectiveness of our loss function, we train an overwater image dehazing network without perceptual loss additionally. The result of a comparative experiment of dehazing with and without perceptual loss is shown in Figure 10, the generated direction for images in the first row is $X \rightarrow G(x) \rightarrow F(G(x))$, the second row is the opposite of the first row $Y \rightarrow F(Y) \rightarrow G(F(Y))$. We can observe from the Figure 10c, 10d, 10g, 10h that the estimated haze-free images and cyclic images lack fine details and the regions of the sky do not match with the input hazy image, which leads to the dehazed results containing halo artifacts when the perceptual loss is not used. Through the comparison of Figure 10b and 10d we can also find that the perceptual loss is favorable for the reconstruction of the sky regions, which is very necessary for the overwater image dehazing.

From Table 4, we observe that our network with perceptual loss gains higher PSNR, SSIM scores and a lower CIEDE2000 score. Higher SSIM and PSNR scores suggest the proposed method with perceptual loss is consistent with human perception. Lower CIEDE2000 means the less color difference between dehazed image and groundtruth. The above experiments show that the proposed loss is effective for the overwater image dehazing task.

6 Conclusion

In this paper, we formulate an overwater image dehazing task, create the first overwater image dehazing dataset, and propose the OWI-DehazeGAN to dehaze overwater images. Compared to previous CNN-based methods which require paired training data, our OWI-DehazeGAN is able to be trained on unpaired images. Our method directly predicts clean images from hazy input bypassing to estimate transmission maps and global atmospheric lights. We utilize the perceptual loss and the resize convolution to preserve detailed textures and alleviate checkerboard artifacts. Extensive experiments demonstrate that our method produces superior results than most of the state-of-the-art dehazing methods.

Acknowledgement. This work was supported in part by National Natural Science Foundation of China (Nos. 61902092 and 61872112), Fundamental Research Funds for the Central Universities Grant No.HIT.NSRIF.20200005, and National Key Research and Development Program of China (Nos. 2018YFC0806802 and 2018YFC0832105).

References

1. Zhu, Q., Mai, J., Shao, L.: A fast single image haze removal algorithm using color attenuation prior. *IEEE Transactions on Image Processing* **24** (2015) 3522–3533
2. Berman, D., treibitz, T., Avidan, S.: Non-local image dehazing. In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. (2016) 1674–1682

3. Ren, W., Ma, L., Zhang, J., Pan, J., Cao, X., Liu, W., Yang, M.H.: Gated fusion network for single image dehazing. In: IEEE Conference on Computer Vision and Pattern Recognition. (2018) 3253–3261
4. Yang, D., Sun, J.: Proximal dehaze-net: A prior learning-based deep network for single image dehazing. In: The European Conference on Computer Vision (ECCV). (2018) 729–746
5. Zhang, H., Patel, V.M.: Densely connected pyramid dehazing network. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (2018) 3194–3203
6. McCartney, E.J.: Optics of the atmosphere: scattering by molecules and particles. New York, John Wiley and Sons, Inc., 1976. 421 p. (1976)
7. Narasimhan, S.G., Nayar, S.K.: Chromatic framework for vision in bad weather. In: Proceedings IEEE Conference on Computer Vision and Pattern Recognition. CVPR 2000 (Cat. No. PR00662). Volume 1., IEEE (2000) 598–605
8. Cai, B., Xu, X., Jia, K., Qing, C., Tao, D.: Dehazenet: An end-to-end system for single image haze removal. IEEE Transactions on Image Processing **25** (2016) 5187–5198
9. Tan, R.T.: Visibility in bad weather from a single image. In: 2008 IEEE Conference on Computer Vision and Pattern Recognition, IEEE (2008) 1–8
10. Tarel, J.P., Hautiere, N.: Fast visibility restoration from a single color or gray level image. In: 2009 IEEE 12th International Conference on Computer Vision, IEEE (2009) 2201–2208
11. He, K., Sun, J., Tang, X.: Single image haze removal using dark channel prior. IEEE transactions on pattern analysis and machine intelligence **33** (2011) 2341–2353
12. Meng, G., Wang, Y., Duan, J., Xiang, S., Pan, C.: Efficient image dehazing with boundary constraint and contextual regularization. In: The IEEE International Conference on Computer Vision (ICCV). (2013) 617–624
13. Li, R., Pan, J., Li, Z., Tang, J.: Single image dehazing via conditional generative adversarial network. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (2018) 8202–8211
14. Ren, W., Liu, S., Zhang, H., Pan, J., Cao, X., Yang, M.H.: Single image dehazing via multi-scale convolutional neural networks. In: European Conference on Computer Vision. (2016) 154–169
15. 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. (2017) 4770–4778
16. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems. (2012) 1097–1105
17. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. In: Advances in neural information processing systems. (2014) 2672–2680
18. Ma, K., Liu, W., Wang, Z.: Perceptual evaluation of single image dehazing algorithms. In: 2015 IEEE International Conference on Image Processing (ICIP), IEEE (2015) 3600–3604
19. Ancuti, C., Ancuti, C.O., Timofte, R., De Vleeschouwer, C.: I-haze: a dehazing benchmark with real hazy and haze-free indoor images. In: International Conference on Advanced Concepts for Intelligent Vision Systems, Springer (2018) 620–631

20. Li, B., Ren, W., Fu, D., Tao, D., Feng, D., Zeng, W., Wang, Z.: Benchmarking single-image dehazing and beyond. *IEEE Transactions on Image Processing* **28** (2019) 492–505
21. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: *The IEEE International Conference on Computer Vision (ICCV)*. (2017) 2242–2251
22. Mirza, M., Osindero, S.: Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784* (2014)
23. Mao, X., Li, Q., Xie, H., Lau, R.Y., Wang, Z., Paul Smolley, S.: Least squares generative adversarial networks. In: *Proceedings of the IEEE International Conference on Computer Vision*. (2017) 2794–2802
24. Yang, X., Xu, Z., Luo, J.: Towards perceptual image dehazing by physics-based disentanglement and adversarial training. In: *Thirty-Second AAAI Conference on Artificial Intelligence*. (2018) 7485–7492
25. Silberman, N., Hoiem, D., Kohli, P., Fergus, R.: Indoor segmentation and support inference from rgbd images. In: *European Conference on Computer Vision*, Springer (2012) 746–760
26. Scharstein, D., Szeliski, R.: High-accuracy stereo depth maps using structured light. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings. Volume 1., IEEE* (2003) 195–202
27. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. (2016) 770–778
28. Noh, H., Hong, S., Han, B.: Learning deconvolution network for semantic segmentation. In: *Proceedings of the IEEE international conference on computer vision*. (2015) 1520–1528
29. Shi, W., Caballero, J., Huszar, F., Totz, J., Aitken, A.P., Bishop, R., Rueckert, D., Wang, Z.: Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. (2016) 1874–1883
30. Odena, A., Dumoulin, V., Olah, C.: Deconvolution and checkerboard artifacts. *Distill* (2016)
31. 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 (2015) 234–241
32. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014)
33. Ruderman, D.L.: The statistics of natural images. *Network Computation in Neural Systems* **5** (2009) 517–548
34. Lark Kwon, C., Jaehee, Y., Alan Conrad, B.: Referenceless prediction of perceptual fog density and perceptual image defogging. *IEEE Transactions on Image Processing* **24** (2015) 3888–3901
35. Ponomarenko, N., Ieremeiev, O., Lukin, V., Egiazarian, K., Jin, L., Astola, J., Vozel, B., Chehdi, K., Carli, M., Battisti, F.: Color image database tid2013: Peculiarities and preliminary results. In: *European workshop on visual information processing (EUVIP)*, IEEE (2013) 106–111