Cycle-SNSPGAN: Towards Real-World Image Dehazing via Cycle Spectral Normalized Soft Likelihood Estimation Patch GAN

Yongzhen Wang, Xuefeng Yan, Donghai Guan, Mingqiang Wei, *Senior Member, IEEE*, Yiping Chen, *Senior Member, IEEE*, Xiao-Ping Zhang, *Fellow, IEEE*, and Jonathan Li, *Senior Member, IEEE*

***Abstract*— Image dehazing is a common operation in autonomous driving, traffic monitoring and surveillance. Learning-based image dehazing has achieved excellent performance recently. However, it is nearly impossible to capture pairs of hazy/clean images from the real world to train an image dehazing network. Most of existing dehazing models that are learnt from synthetically generated hazy images generalize poorly on real-world hazy scenarios due to the obvious domain shift. To deal with this unpaired problem arisen by real-world hazy images, we present Cycle Spectral Normalized Soft likelihood estimation Patch Generative Adversarial Network (Cycle-SNSPGAN) for image dehazing. Cycle-SNSPGAN is an unsupervised dehazing framework to boost the generalization ability on real-world hazy images. To leverage unpaired samples of real-world hazy images without relying on their clean counterparts, we design an SN-Soft-Patch GAN and exploit a new cyclic self-perceptual loss which avoids using the ground-truth image to compute the perceptual similarity. Moreover, a significant color loss is adopted to brighten the dehazed images as human expects. Both visual and numerical results show clear improvements of the proposed Cycle-SNSPGAN over state-of-the-arts in terms of hazy-robustness and image detail recovery, with even only a small dataset training our Cycle-SNSPGAN. Code has been available at https://github.com/yz-wang/Cycle-SNSPGAN.**

Index Terms— **Cycle-SNSPGAN, image dehazing, unsupervised learning, generative adversarial network, unpaired data, object detection.**

I. Introduction

HAZE, as a common atmospheric phenomenon, unpredictably damages the visual quality of images captured by outdoor vision systems such as autonomous driving [1], [2], unmanned aerial vehicles (UAVs) [3], traffic monitoring [4], and surveillance [5]. Even top-performing vision approaches of object detection, classification and tracking suffer from significant accuracy degradations under hazy conditions. Image dehazing is an essential prerequisite for bridging the gap between hazy images and high-level vision tasks by enhancing image quality.

Image dehazing aims at recovering the underlying clean image from a hazy input. Current image dehazing techniques fall into two categories, i.e., the prior-based and the learning-based methods. Although both categories improve the overall visibility, there are still many challenging yet unsolved problems. First, learning-based methods commonly develop CNNs or other neural networks to restore degenerated hazy images from hazy inputs via paired data. Such paired data in the real world are difficult or even impossible to capture. Instead, pairs of hazy/clean images can be generated by computing a conventional atmospheric scattering equation with a known transmission map and a global light map. However, the wisdom of training on fake synthesized data will degrade the network's ability to deal with real-world scenarios. Second, these networks often introduce additional degradations such as the loss of image details, low contrasts, and halos, since they tend to overfit both synthesized and real hazy images, especially when only a small amount of training samples is available. Third, for the conventional prior-based methods, users have to tweak parameters multiple times to determine their optimal formulas. That means, users cannot apply filters without the laborious parameter tuning process and automatically achieve the optimal dehazing results.

Motivated by CycleGAN [10], we leverage unpaired data to develop a unsupervised image dehazing network, such that real-world hazy images can contribute to network training, thereby better recovering the underlying background details for the real-world hazy inputs. In view of this, we propose a Cycle Spectral Normalized Soft likelihood estimation Patch

Manuscript received March 23, 2021; revised July 28, 2021, November 7, 2021, and March 4, 2022; accepted April 22, 2022. This work was supported in part by the National Natural Science Foundation of China under Grant 62172218, in part by the Joint Fund of National Natural Science Foundation of China and Civil Aviation Administration of China under Grant U2033202, in part by the Free Exploration of Basic Research Project, Local Science and Technology Development Fund Guided by the Central Government of China under Grant 2021Szvup060 and in part by the 14th Five-Year Planning Equipment Pre-Research Program under Grant JCKY2020605C003. The Associate Editor for this article was Z. Duric. *(Corresponding author: Xuefeng Yan.)*

Yongzhen Wang, Xuefeng Yan and Donghai Guan are with the School of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China (e-mail: wangyz@nuaa.edu.cn; yxf@nuaa.edu.cn; dhguan@nuaa.edu.cn).

Mingqiang Wei is with the Shenzhen Research Institute, Nanjing University of Aeronautics and Astronautics, Shenzhen 518038, China (e-mail: mingqiang.wei@gmail.com).

Yiping Chen is with the Fujian Key Laboratory of Sensing and Computing for Smart Cities, School of Informatics, Xiamen University, Xiamen 361005, China (e-mail: chenyiping@xmu.edu.cn).

Xiao-Ping Zhang is with the Department of Electrical, Computer and Biomedical Engineering, Ryerson University, Toronto, ON M5B 2K3, Canada (e-mail: xzhang@ee.ryerson.ca).

Jonathan Li is with the Department of Geography and Environmental Management and the Department of Systems Design Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada (e-mail: junli@uwaterloo.ca).

Digital Object Identifier 10.1109/TITS.2022.3170328

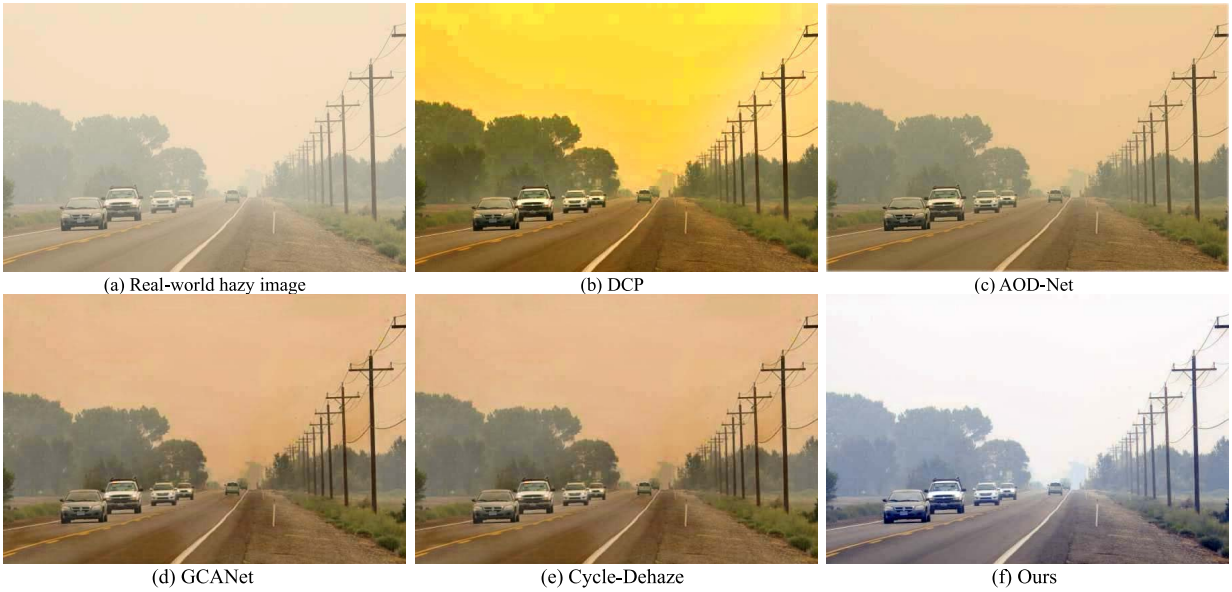


Fig. 1. Image dehazed results on a real-world hazy image. From (a) to (f): (a) the input hazy image, and the dehazing results of (b) DCP [6], (c) AOD-Net [7], (d) GCANet [8], (e) Cycle-Dehaze [9], and (f) our Cycle-SNSPGAN, respectively. Our Cycle-SNSPGAN can generate both haze-free and perceptually more pleasing results.

Generative Adversarial Network (termed a Cycle-SNSPGAN) for real-world image dehazing. Cycle-SNSPGAN is a unsupervised dehazing framework which can boost the generalization ability on real-world hazy scenarios. The critical component of our framework is the proposed SN-Soft-Patch GAN, which is an effective GAN that can leverage unpaired samples of real hazy images for training. In detail, we employ a spectral normalized GAN [11] and a patch discriminator [12] to improve the training speed and stability of the network. Also, we design a new soft likelihood estimation solution to calculate the output of the discriminator. Moreover, during the training process, we exploit the prior knowledge of color information as a loss function to brighten the dehazed images as humans expect, and develop a novel cyclic self-perceptual loss to make the dehazed images more realistic. Experiments on both synthetic datasets and real-world hazy images demonstrate that our model outperforms the state-of-the-art dehazing algorithms, even only a small dataset trains our Cycle-SNSPGAN.

Cycle-SNSPGAN leverages the previous expertise, as well as takes benefits from unpaired real-world training data. Fig. 1 exhibits image dehazing results of a real-world hazy image by different approaches, where the proposed Cycle-SNSPGAN produces much clearer and perceptually pleasing images, while existing approaches introduce additional artifacts such as low contrasts, halos, and haze residuals. Experiments on both synthetic datasets and real-world hazy images validate that our method performs favorably against the state-of-the-art dehazing approaches.

In summary, the contributions of our method are mainly four-fold.

- We propose an unsupervised real-world image dehazing framework via unpaired data, called Cycle Spectral Normalized Soft likelihood estimation Patch Generative Adversarial Network (Cycle-SNSPGAN). Collecting practical unpaired samples of real-world hazy images for training becomes possible, which boosts the network's generalization ability on unpredictable real-world hazy images.
- To improve the clearness of dehazed images, we propose an effective GAN, namely, SN-Soft-Patch GAN, which consists of a spectral normalized GAN, a patch discriminator, and the proposed solution of soft likelihood estimation.
- To preserve both large-scale structures and small-scale details from the hazy images, we develop a new cyclic self-perceptual loss to calculate the perceptual similarity without knowing the ground-truth clean image. Moreover, we regard the color features of the hazy images as prior knowledge, and introduce a color loss to make the dehazed images more realistic.

II. Related Work

In this section, we will introduce prior-based methods, and the learning-based methods, followed by the application of GANs in image dehazing.

A. Prior-Based Methods

Traditional prior-based methods resort to exploring several hand-crafted priors to restore the hazy images [6], [13]–[15]. The most representative work is the Dark Channel Prior (DCP) proposed by He *et al.* [6], which is based on the statistics of the clear non-sky images and can estimate the transmission map effectively. However, this prior is found to be unreliable in some cases, such as hazy images with sky regions. After that, many approaches have been designed to improve the performance of image dehazing based on different assumptions [15], [16]. Besides, Fattal [17] develop a color-line prior

for image dehazing based on the observation that the pixels of clear images patches exhibit a 1-dimensional distribution. Berman *et al.* [18] find that a haze-free image can be approximated by several hundred distinct colors, and then propose the non-local color prior (NCP) for image dehazing. Although these methods have achieved success in image dehazing, these priors often incur color distortions and users have to tweak parameters multiple times to determine their optimal formulas.

B. Learning-Based Methods

With the rapid developments of CNNs, many learning-based methods have been proposed for image dehazing [7], [19]–[22]. Early methods mainly focused on estimating the transmission map from input hazy images. For example, Cai *et al.* [19] employ a coarse CNN to generate a holistic transmission map of the hazy input, and then recover the haze-free image via atmospheric scattering model. Ren *et al.* [21] develop the MSCNN framework that first generates a coarse transmission map and then refines the coarse result. However, inaccurate transmission map estimation affects the quality of dehazed results seriously. Recent learning-based methods learn to directly generate the haze-free image from a hazy input via supervised learning. AOD-Net proposed by Li *et al.* [7]. is the first model by using an end-to-end manner for image dehazing, which reconstructs the atmospheric scattering model and can output a haze-free image directly from a hazy image. Qin *et al.* [20] propose a novel end-to-end feature fusion network to directly restore the haze-free image, which is based on the attention mechanism and performed well on the SOTS (A benchmark dataset for testing dehazing results) [23]. Since these methods always require enough paired data to train the network, while such paired data are difficult to obtain. In light of this, the learning-based methods are often trained on synthesized hazy datasets. The gap between synthetic images and real-world images degrades their ability to remove haze.

C. GANs in Image Dehazing

Inspired by the success of GANs in other image processing tasks [24]–[26], Zhu *et al.* [27] first introduced the GAN in the field of image dehazing. They develop a network for image dehazing based on atmospheric scattering model and judge the output by a vanilla discriminator. Later, a series of GAN-based image dehazing works have sprung up [28]–[33]. Li *et al.* [30] propose a conditional GAN to directly generate haze-free results instead of solving the atmospheric scattering model. Qu *et al.* [34] develop a novel GAN-based dehazing network called EPDN, followed by a well-designed refiner without any physical model. Since these methods require paired data for training just like the learning-based methods, they have poor ability in dealing with real-world hazy images. Most recently, Zhu *et al.* [10] propose CycleGAN, which is a new image-to-image translation architecture based on unpaired data training. Motivated by this work, Engin *et al.* [9] develop the enhanced CycleGAN, i.e., Cycle-Dehaze for unpaired single image dehazing. Shao *et al.* [35] propose a novel domain adaptation network for image dehazing based on CycleGAN, which includes an image translation module and two image dehazing modules.

III. CYCLE-SNSPGAN

Due to the gap between synthetic images and real-world images, most learning-based dehazing methods learnt from synthetic images usually fail in generalizing well to real-world scenarios. To deal with this problem, we focus on learning and distilling global structures and small scale details rather than training the network on synthetic images with fake haze. And the haze-free images are represented by the CycleGAN under the guidance of unpredictable real-world haze conditions. Since our method can leverage real-world hazy images for network training, it will boost the generalization ability on real-world scenes. In the following, Cycle-SNSPGAN is introduced in detail including the overview, the architecture of the network, the new soft likelihood estimation solution, and the loss functions.

A. Overview

We propose Cycle-SNSPGAN to make full use of unpaired images for network training and enhance the dehazing ability on real-world scenarios. Cycle-SNSPGAN consists of two generators (G_A and G_B) and two discriminators (D_A and D_B), as demonstrated in Fig. 2. We adopt G_A to map the hazy images to haze-free images with D_A in a cycle-consistent manner. Similarly, G_B is employed to map the haze-free images to hazy images inversely. The discriminator D_A and D_B are designed to make the images generated by G_A and G_B have similar feature distributions with the target domain, thus making the generated images more realistic. In addition, our network in each domain shares the same network structure but with different inputs.

The whole pipeline can be described as follows (see Fig. 2): First, the unpaired hazy image x and haze-free image y are fed to the network simultaneously. Then, the generators G_A and G_B will map the input image to the fake haze-free image y' and hazy image x' respectively, i.e., $G_A(x) \rightarrow y'$ and $G_B(y) \rightarrow x'$. After that, the discriminators D_A and D_B will judge whether y' (or x') is a real image or a fake image outputted by the generator, which will improve the quality of the generated images. Next, we feed y' and x' to G_B and G_A to generate the reconstructed hazy image x'' and haze-free image y'', respectively. In this way, the cycle-consistency loss can be calculated by the reconstructed images and the input images, and the network can be trained in a cycle manner.

Remark: The purpose of adversarial training is to promote G_A and G_B to produce more realistic images. Thus we employ D_A and D_B to judge y' and x' for adversarial training with the generators. While in the training process, the parameters of G_A, G_B, D_A, and D_B are usually updated separately in one iteration. If the discriminators are employed to discriminate the reconstructed images (i.e., x'' and y'') simultaneously, the parameters of D_A and D_B will be updated twice in one iteration, and the update of the generators' parameters will also be affected. It is well known that the training of GAN itself is very difficult, and updating the parameters twice may

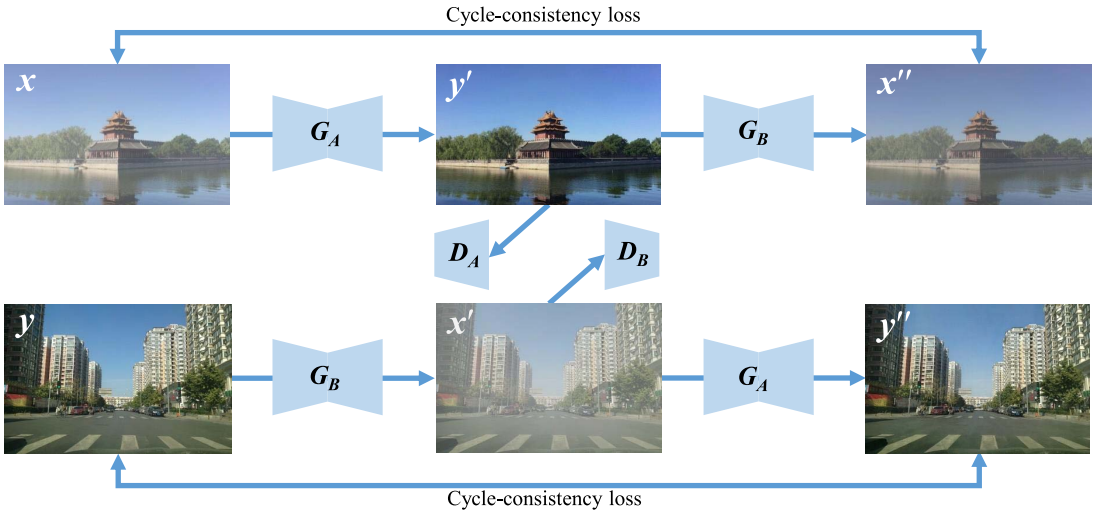
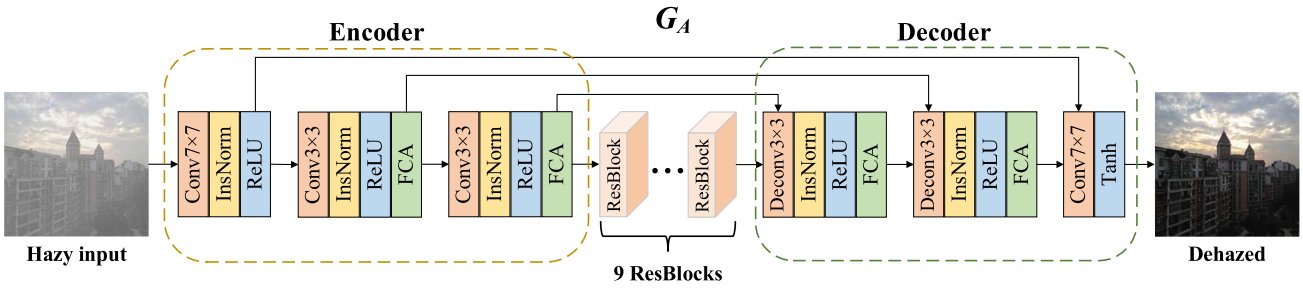
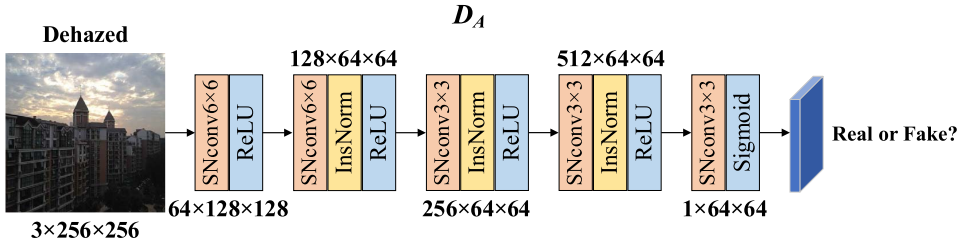


Fig. 2. Overview of Cycle-SNSPGAN. G_A & G_B are the generators, and D_A & D_B are the discriminators. G_A is used to map the hazy input to the haze-free image, while G_B is used to map the clear input to the hazy image.



(a) Generator. The generator outputs the dehazed image in an end-to-end manner. FCA refers to the frequency channel attention network.



(b) Discriminator. The discriminator can guide the generator to produce higher quality dehazed images with the help of spectral normalized GAN and patch GAN. The SNconv module refers to the spectral normalized convolutional network.

Fig. 3. The architecture of Cycle-SNSPGAN.

cause mode collapse, which further degrades the quality of the final generated images. In addition, although it is technically feasible to judge the reconstructed images (x'' and y''), this will result in a significant increase in computational complexity. Therefore, it only needs to judge the output images once in the training process for the purpose of adversarial training.

B. Network Architecture

Although CycleGAN is very effective, there are still some problems in the training process, such as mode collapse and convergence difficulties, which will degrade the quality of the final dehazed images. To address these problems, we develop a simple yet effective GAN, i.e., SN-Soft-Patch GAN, which consists of three parts: a spectral normalized GAN, a patch discriminator, and a novel solution of soft likelihood estimation. We first introduce the spectral normalized GAN [11] in the design of discriminator, which is an effective normalization approach that can improve the stability of training. Next, the patch GAN is employed to reduce the parameters of the model and boost the convergence speed of network training. Then, a novel soft likelihood estimation solution is proposed to enhance the overall quality of the final dehazed images, which will be described in the next subsection. Moreover, to further improve the dehazing ability of the model, an up-to-date attention mechanism (Frequency Channel Attention Network, i.e., FCANet) proposed by Qin *et al.* [36] is introduced in our network. Since G_B and D_B have the same architecture as G_A and D_A, we only present the architecture of G_A and D_A, which is illustrated in Fig. 3.

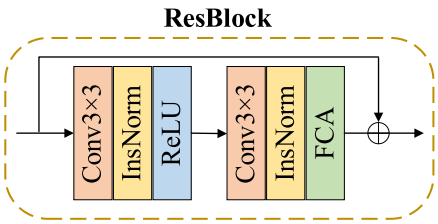


Fig. 4. Residual blocks. The residual blocks assist the generator to extract more complex features and remove haze.

1) Generator: The function of the generator is to generate the dehazed images from the corresponding hazy images (G_A), or generate the hazy images from the corresponding haze-free images (G_B). Motivated by "U-Net" [37] and "ResNet" [38], we develop an encoder-decoder network as the generator and introduce the skip connections to avoid the problem of gradient vanishing. As demonstrated in Fig. 3(a), the generator can directly output the dehazed (hazy) images in an end-to-end manner, which contains an encoding module, a feature extraction module and a decoding module.

The encoding module is composed of an initial layer and two down-sampling layers. Thus, the feature information of the input hazy images can be encoded into the feature maps. The feature extraction module consists of 9 residual blocks (see Fig. 4), which can further extract more complex and deeper features from the input features, and remove the haze simultaneously. After the feature extraction, two up-sampling operations and a $Tanh$ activation function are adopted in the decoding module to output the final dehazed images. Furthermore, as mentioned above, we introduce FCANet in the design of the generator to enhance the performance of our model.

Remark: Using more complex network architectures, such as GNNs and Transformers [39], [40], possibly enhances the dehazing performance of existing models. However, a light-weight network is more desirable to deploy for image dehazing tasks, due to the fact that the hardware and computing ability of intelligent transportation systems (e.g., unmanned aerial vehicles, intelligent vehicles) are usually limited. Therefore, we employ a simple yet effective ResNet-based generator to achieve a better parameter-performance trade-off.

2) Attention Mechanism: In view of the fact that in a hazy image with uneven haze distribution, we always want to dehaze the hazy regions rather than the haze-free regions, thereby making the output image have good visibility. Since the attention mechanism has been widely used to improve the performance of the neural networks [41]–[43]. Inspired by a recent work [36], we adopt FCANet in the design of the generator and the residual blocks to improve the dehazing ability of our model, as shown in Fig. 3(a) and Fig. 4. FCANet combines the channel attention mechanism with the Discrete Cosine Transform (DCT) cleverly, and expands on the basis of SENet [43] to obtain a new multi spectral channel attention mechanism.

FCANet enables our model to learn the weights from different feature maps adaptively, thus boosting the dehazing ability of the network. Moreover, after the introduction of FCANet,

(a) Hazy image (b) Dehazing result

Fig. 5. Image dehazing results on the NH-HAZE dataset [44]. Cycle-SNSPGAN can effectively remove haze in the uneven haze distribution scenarios.

the results of ablation study indicate that the performance of the proposed model can be improved significantly. We test the proposed Cycle-SNSPGAN on the NH-HAZE dataset [44], to better understand its performance of solving the uneven haze distribution scenarios. As illustrated in Fig. 5, our Cycle-SNSPGAN can remove the unevenly-distributed haze effectively.

3) Discriminator: For adversarial training, the discriminator is constructed in a fully convolution fashion and a novel SN-Soft-Patch discriminator is proposed to improve the performance of the vanilla patch discriminator. As demonstrated in Fig. 3(b), we first design a spectral normalized convolutional layer instead of the traditional convolutional layer to make the training process more stable. Next, we employed 4 non-linear ReLU layers, 3 instance normalization layers, and a sigmoid activation function to output a 64×64 patch. Finally, the values of these patches will be employed to judge whether the input image is a real image or a fake image generated by G_A. Additionally, to further enhance the quality of the final dehazed images, we propose a novel soft likelihood estimation solution to calculate the output of these patches. The proposed method can enhance the discriminating ability of the discriminator, thus promoting the generator to output higher quality images.

C. Soft Likelihood Estimation Solution

Inspired by Versteegen *et al.* [45], we develop a new solution to compute the final output of the patch discriminator rather than directly calculating the average of all patches. The proposed soft likelihood estimation solution can encourage the discriminator to focus on the detailed features of the input images, thus improving the overall quality of the final dehazed images. Specifically, we first obtain the weight matrix by calculating the proximity of each patch to the lowest patch likelihood. Here, the SoftMax/SoftMin functions are used to smooth the max/min objective functions while solving the weight matrix. After that, we multiply the weight matrix on

each patch with the corresponding likelihoods and sum them to produce the final output. Finally, the output will be judged for the input image, whether it is a real or fake image. In this way, the discriminator will pay more attention to the variations in the low probability regions, and the generator will produce images with rich details. The specific calculation process can be expressed as formula (1):

$$P = \sum_i (W_i \times \ln l_i), \tag{1}$$

where

$$W_i = \frac{e^{\alpha \times \ln l_i}}{\sum_i e^{\alpha \times \ln l_i}}. \tag{2}$$

In the above formulas, P is the final output of the discriminator, i denotes each patch in the discriminator, l represents the patch-region likelihoods, W_i denotes the weight-matrix and α represents the SoftMax or SoftMin operations during the weight-matrix calculation. We set α to 1 to calculate the SoftMax function and -1 to calculate the SoftMin function.

D. Loss Functions

Different with most existing dehazing solutions, we comprehensively consider all the positive factors that can enhance the quality of the final dehazed images and make them much clearer and more realistic. Therefore, we exploit a comprehensive loss that contains adversarial loss, cycle-consistency loss, identity loss, color loss, the Total Variation (TV) loss, as well as the new cyclic self-perceptual loss, which is formulated as

$$L_{Total} = \lambda_1 L_{adv}(G) + \lambda_2 L_{cyc} + \lambda_3 L_{ide} + \lambda_4 L_{color} + \lambda_5 L_{tv} + \lambda_6 L_{self-per}. \tag{3}$$

1) Adversarial Loss: The adversarial loss is employed to match the distribution of the generated images with the data distribution in the target domain. We adopt the Least-Squares GAN loss [46] to improve the stability of the training process and enhance the quality of the generated images. The definition of adversarial loss is shown in the following formulas:

$$L_{adv}(G) = E_{G_A(x)\sim P_{fake}}\left[(D_A(G_A(x)) - 1)^2\right] + E_{G_B(y)\sim P_{fake}}\left[(D_B(G_B(y)) - 1)^2\right], \tag{4}$$

$$L_{adv}(D_A) = E_{y\sim P_{real}}\left[(D_A(y) - 1)^2\right] + E_{G_A(x)\sim P_{fake}}\left[(D_A(G_A(x)))^2\right], \tag{5}$$

$$L_{adv}(D_B) = E_{x\sim P_{\text{real}}}\left[(D_B(x) - 1)^2\right] + E_{G_B(y)\sim P_{fake}}\left[(D_B(G_B(y)))^2\right], \tag{6}$$

where $G_A(x)$ and $G_B(y)$ are the fake haze-free and hazy images generated by G_A and G_B, respectively.

2) Cycle-Consistency Loss: Thanks to the cycle-consistency loss, it is possible to collect practical unpaired samples of real hazy images for training. The cycle consistency loss can be expressed by formula (7):

$$L_{cyc} = E_{x\sim P_{data(X)}}\left[\|(G_B(G_A(x)) - x)\|_1\right] + E_{y\sim P_{data(Y)}}\left[\|(G_A(G_B(y)) - y)\|_1\right], \tag{7}$$

where $G_B(G_A(x))$ and $G_A(G_B(x))$ are the reconstructed hazy and haze-free images respectively. The purpose of cycle-consistency loss is to make the reconstructed images match closely to the original input images, i.e., $G_B(G_A(x))\approx x$, and $G_B(G_A(x))\approx y$.

3) Identity Loss: Besides the adversarial loss and cycle-consistency loss, we adopt the identity loss to make the images generated by the generator consistent with the tone of the input images, which is shown as the following formula (8):

$$L_{ide} = E_{y\sim P_{data(Y)}}\left[\|(G_A(y) - y)\|_1\right] + E_{x\sim P_{data(X)}}\left[\|(G_B(x) - x)\|_1\right]. \tag{8}$$

The identity loss can make the output image have the same structure as the input image, thus improving the quality of the final generated image. We surprised find that the identity loss is very helpful to enhance the details of the dehazed images, while many image-to-image translation works pay less attention to employing this loss function in the training process [9], [47], [48].

4) Color Loss: In light of the fact that the brightness and contrast of hazy images are usually lower than haze-free images, we regard this common phenomenon as prior knowledge, and introduce the color loss [49] to measure the color difference between the dehazed images and haze-free images. This loss function can brighten the dehazed images as humans expect, thus making the dehazed images more realistic. To this end, we first employ a Gaussian filter over the dehazed image and haze-free image to obtain the blur representations, and then the mean squared error between them is computed by the following formula (9):

$$L_{color} = MSE(G_A(x)_{blur}, y_{blur}), \tag{9}$$

where $G_A(x)_{blur}$ and y_{blur} are the blurred images of $G_A(x)$ and y, respectively. Since we mainly focus on image dehazing, color loss is only utilized on G_A. Fig. 6 exhibits three real-world hazy samples and the corresponding dehazing results obtained by the proposed Cycle-SNSPGAN. As observed, the introduction of color loss enables Cycle-SNSPGAN to produce both brighter and more realistic images. Also, Cycle-SNSPGAN has a certain capability of color correction.

5) Total Variation Loss: In order to remove the noise of the generated images and make them much clearer, we introduce the TV loss [50] to enforce spatial smoothness of the generated images, which is shown as follows:

$$L_{tv} = \|\nabla_x G_A(x) + \nabla_y G_A(x)\|_1 + \|\nabla_x G_B(y) + \nabla_y G_B(y)\|_1, \tag{10}$$

where ∇_x and ∇_y denote the horizontal and vertical differential operation matrices, respectively. This loss function can improve the quality of the generated images and make them much clearer.

6) Cyclic Self-Perceptual Loss: The above loss functions cannot recover all the texture information from the hazy image, especially in the case of thick haze. Inspired by [51] and [52], we propose a novel cyclic self-perceptual loss to preserve both large-scale structures and small-scale details from hazy input.

(a) Real-world hazy image (b) Dehazing result

Fig. 6. Image dehazing results on the real-world hazy images by the proposed Cycle-SNSPGAN. Our approach can produce both brighter and more realistic images.

(a) Hazy image (b) Dehazing result

Fig. 7. Image dehazing results on the Dense-Haze dataset [53]. The proposed Cycle-SNSPGAN can effectively remove thick haze.

Different from the vanilla perceptual loss, our training process is based on a unsupervised learning framework without ground-truth images to calculate perceptual similarity. Therefore, we propose to directly measure the perceptual similarity between the dehazed result and the corresponding hazy input, rather than the ground-truth image. This perceptual similarity can make the dehazed image have similar structure and texture features to the input image, so that more detailed features can be preserved. We calculate the perceptual loss in a cyclic manner, i.e., the perceptual similarities of $G_A(x)$ and $G_B(y)$ are calculated simultaneously. According to the deep features extracted by a pre-trained VGG network, the perceptual similarities are calculated by the following formula (11):

$$L_{self-per} = \|\psi(G_A(x)) - \psi(x)\|_1 + \|\psi(G_B(y)) - \psi(y)\|_1, \quad (11)$$

where $\psi(.)$ denotes the feature maps extracted from the 2^{nd} and 5^{th} pooling layers within the VGG-16 network pre-trained on ImageNet. To verify the use of cyclic self-perceptual loss that can effectively remove thick haze, the Dense-Haze dataset [53] is employed to qualitatively evaluate the proposed Cycle-SNSPGAN, as shown in Fig. 7. Although there is still a little haze left in the scene, most of the thick haze has been removed and the visibility of the image has been considerably improved.

IV. Experiment

A. Implementation Details

1) Dataset: Since Cycle-SNSPGAN is trained in an unsupervised learning manner, the real-world hazy images can be chosen for training. We randomly choose unpaired hazy and haze-free images from OTS (Outdoor Training Set), RTTS (Real-world Task-driven Testing Set) and URHI (Unannotated Real Hazy Images) from RESIDE dataset [23] as our training set. In detail, the training set is composed of 3000 hazy images chosen from RTTS and URHI respectively. For haze-free images, since there are no haze-free images in RTTS and URHI, we train G_B by randomly choosing 6000 images from OTS. Although the dataset is very small, the final dehazing results of the model are favorably compared to the results by the learning-based methods training on the entire ITS dataset (Indoor Training Set, containing 100000 indoor hazy/haze-free image pairs).

2) Training Details: We implement our network using Pytorch 1.5 on a system with Intel Xeon E5-2698 v4 CPU and NVIDIA Tesla V100 GPU. For accelerating the training procedure, Cycle-SNSPGAN is trained by the Adam optimizer with a batch size of 2, where the momentum parameters β_1 and β_2 take the values of 0.5 and 0.999, respectively. The initial learning rate l is set to 2×10^{-4} for both generators and discriminators. Since GANs are difficult to train, we fix l during the first 100 epochs of training, and decay l to 0 linearly in the next 100 epochs. In the following experiments, the loss weights are set with $\lambda_1 = 1$, $\lambda_2 = 10$, $\lambda_3 = 5$, $\lambda_4 = 0.5$, $\lambda_5 = 0.5$ and $\lambda_6 = 5$. For all these loss weights, a large number of experiments are performed to ensure their optimum values.

In addition, we also find that the quality of the generator's output will be degraded due to the strong ability of SN-Soft-Patch discriminator. Therefore, the network training is divided into the following two stages: we first use the traditional patch discriminator to train the generator, and then add the soft likelihood estimation solution after the generator becomes stable. It is worth noting that, by experiments, the model has the best performance when the soft likelihood estimation solution is added after the 28th epoch.

3) Evaluation Settings: The proposed method is compared quantitatively and qualitatively with several previous state-of-the-art approaches. These methods can fall into three

TABLE I

QUANTITATIVE COMPARISON (PSNR/SSIM) OF VARIOUS STATE-OF-THE-ART IMAGE DEHAZING APPROACHES ON SYNTHETIC DATASETS. OUR CYCLE-SNSPGAN PERFORMS FAVORABLY AGAINST OTHER DEHAZING ALGORITHMS

Method	Publication	Type	SOTS outdoor		HSTS	
			PSNR↑	SSIM↑	PSNR↑	SSIM↑
DCP [6]	TPAMI'11	Prior	18.38	0.819	17.01	0.803
BCCR [15]	ICCV'13	Prior	15.71	0.769	15.21	0.747
NCP [18]	CVPR'16	Prior	18.07	0.802	17.62	0.798
MSCNN [21]	ECCV'16	Supervised	19.56	0.863	18.27	0.841
AOD-Net [7]	ICCV'17	Supervised	20.08	0.861	19.68	0.835
GFN [54]	CVPR'18	Supervised	21.49	0.838	22.94	0.874
DCPDN [29]	CVPR'18	Supervised	22.49	0.857	20.07	0.847
GCANet [8]	WACV'19	Supervised	21.66	0.867	21.37	0.874
EPDN [34]	CVPR'19	Supervised	22.57	0.863	20.37	0.877
MSCNN-HE [55]	IJCV'20	Supervised	22.72	0.871	21.23	0.851
GFN-IJCV [56]	IJCV'20	Supervised	24.21	0.849	23.17	0.829
Interleaved CSF [57]	TIP'20	Supervised	24.17	0.923	22.94	0.907
CycleGAN [10]	ICCV'17	Unsupervised	17.32	0.706	16.05	0.703
Cycle-Dehaze [9]	CVPRW'18	Unsupervised	18.60	0.797	17.96	0.777
Dehaze-GLCGAN [47]	arXiv'20	Unsupervised	23.03	0.917	-	-
LIGHT-Net [58]	TETCI'20	Unsupervised	23.11	0.917	22.27	0.906
Deep DCP [59]	TIP'20	Unsupervised	20.99	0.893	21.20	0.871
YOLY [60]	IJCV'21	Unsupervised	20.39	0.889	21.02	0.905
Cycle-SNSPGAN	Ours	Unsupervised	**24.28**	**0.925**	**25.30**	**0.935**

TABLE II

QUANTITATIVE COMPARISONS (NIQE/BRISQUE/SSEQ/FADE) WITH SOTAS ON THE HSTS DATASET. RED AND BLUE COLORS ARE USED TO INDICATE THE 1^{st} AND 2^{nd} RANKS, RESPECTIVELY

Method	Type	NIQE↓	BRISQUE↓	SSEQ↓	FADE↓
DCP [6]	Prior	3.213	14.388	13.899	0.436
AOD-Net [7]	Supervised	3.139	18.155	25.017	0.738
GCANet [8]	Supervised	3.125	15.177	34.156	0.620
EPDN [34]	Supervised	3.214	14.323	21.334	0.568
GFN-IJCV [56]	Supervised	3.191	17.002	35.135	0.618
CycleGAN [10]	Unsupervised	4.239	16.135	21.510	0.493
Cycle-Dehaze [9]	Unsupervised	3.659	20.883	26.817	0.422
Cycle-SNSPGAN	Unsupervised	3.120	14.275	20.545	0.431

categories: prior-based, supervised-based and unsupervised-based. For prior-based methods, we compare with DCP [6], BCCR [15] and NCP [18]. For supervised-based approaches, we compare with MSCNN [21], AOD-Net [7], GFN [56], DCPDN [29], GCANet [8], EPDN [34], MSCNN-HE [55], GFN-IJCV [56], and Interleaved CSF [57]. For unsupervised-based methods, we compare with CycleGAN [10], Cycle-Dehaze [9], Dehaze-GLCGAN [47], LIGHT-Net [58], Deep DCP [59], and a more recent unsupervised and untrained image dehazing framework YOLY [60]. Both the SOTS (Synthetic Object Testing Set) and HSTS (Hybrid Subjective Testing Set) from RESIDE are employed as the test set. We adopt average Peak Signal to Noise Ratio (PSNR) and Structural Similarity index (SSIM) for quantitative evaluation of the recovered images, which are the most widely used image objective evaluation indexes. PSNR is employed to evaluate the quality of an image compared to the original image after processing, while SSIM is employed to measure image similarity from three aspects: brightness, contrast, and structure. Larger values of PSNR or SSIM usually indicate better results.

Additionally, we utilize four well-known no-reference image assessment metrics to further measure the quality of dehazed images from multiple perspectives, including NIQE [61], BRISQUE [62], SSEQ [63], and FADE [64]. Both NIQE and BRISQUE are used to assess the overall quality of the given images, and lower values indicate better results. SSEQ assesses image quality by computing the entropy in the spatial and frequency domains of image patches, while FADE is a criterion for evaluating the haze density of the given images.

B. Comparison Results

1) Comparison on Synthetic Dataset: Fig. 8 and Fig. 9 show qualitative comparisons on synthetic outdoor images from SOTS and HSTS. Since the real-world haze always exists outdoors, we mainly focus on the dehazing of outdoor datasets. DCP tends to cause color distortions and introduce artifacts in the sky regions, which affects the visibility of dehazing images. For AOD-Net, GCANet and Deep DCP, although they have largely overcome the problem of color distortion, none of them can remove the haze completely. Although Cycle-Dehaze succeeds in removing the haze to a certain extent, it also causes color distortion and introduces halo artifacts into the whole image. Compared with these state-of-the-arts, the Cycle-SNSPGAN can generate much clearer dehazed images with perceptually pleasing.

Moreover, the averaged PSNR and SSIM of all networks are reported for quantitative evaluation, as shown in Tab. I. For all the compared methods, we either retrain their models on the entire ITS dataset or directly adopt the pre-trained models provided by the authors for evaluation. It is obvious that our Cycle-SNSPGAN outperforms other state-of-the-arts in terms of PSNR. We also achieved impressive performance in terms of SSIM. Although the Cycle-SNSPGAN is only trained on 6000 unpaired image pairs, it exhibits better on SOTS and is generalized smoothly on HSTS.

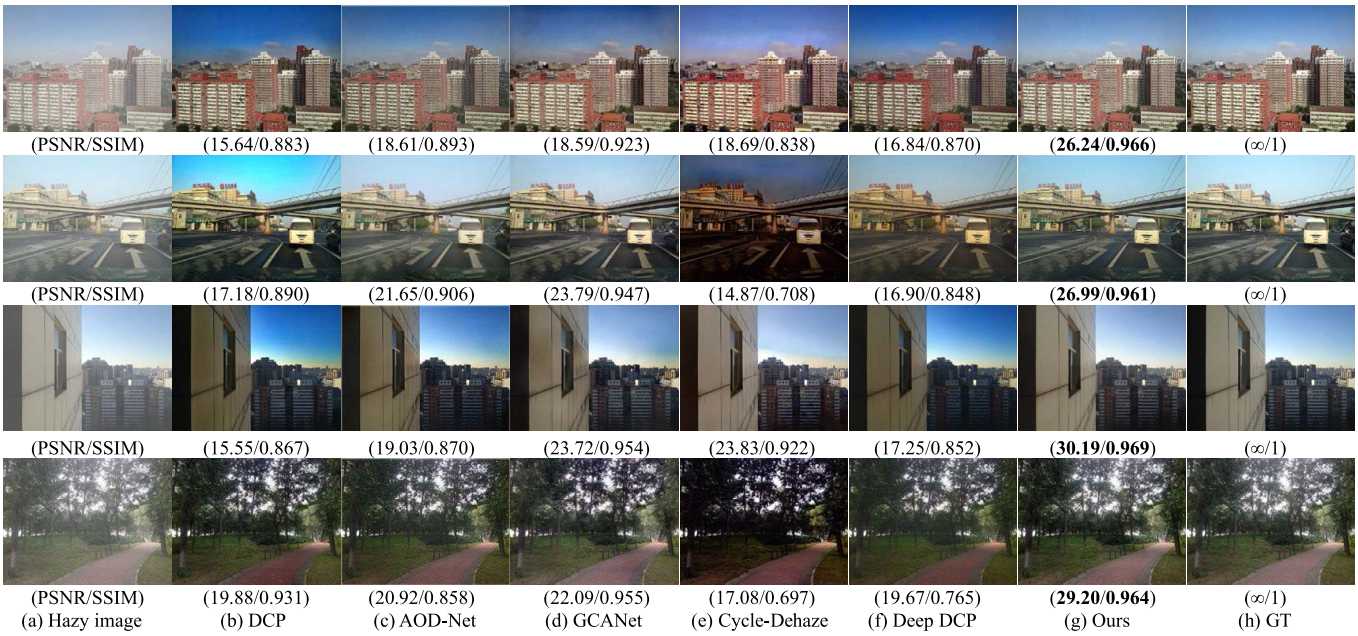


Fig. 8. Image dehazing results on the SOTS outdoor dataset. From (a) to (h): (a) the hazy image, and the dehazing results of (b) DCP [6], (c) AOD-Net [7], (d) GCANet [8], (e) Cycle-Dehaze [9], (f) Deep DCP [59], (g) our Cycle-SNSPGAN, respectively, and (h) the ground-truth image. Our Cycle-SNSPGAN can produce much clearer results with less color distortion and fewer artifacts.

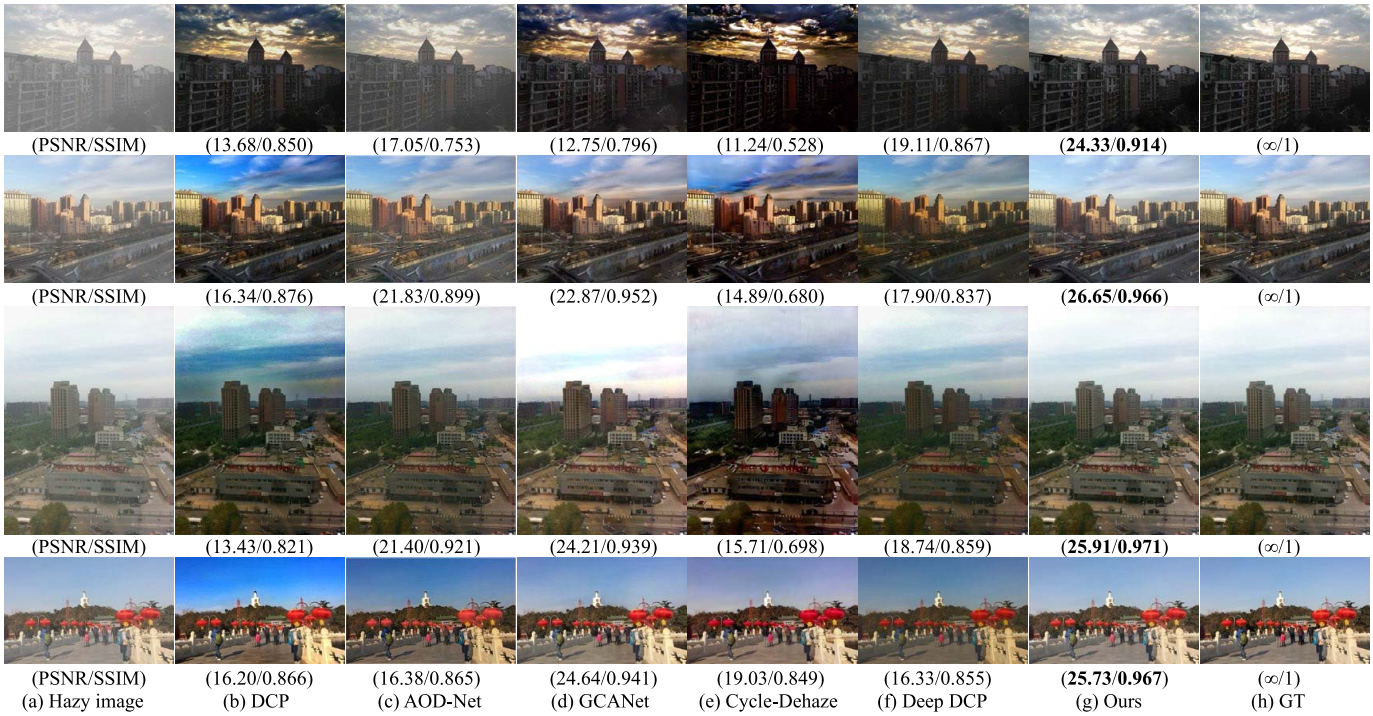


Fig. 9. Image dehazing results on the HSTS dataset. From (a) to (h): (a) the hazy image, and the dehazing results of (b) DCP [6], (c) AOD-Net [7], (d) GCANet [8], (e) Cycle-Dehaze [9], (f) Deep DCP [59], (g) our Cycle-SNSPGAN, respectively, and (h) the ground-truth image. Our Cycle-SNSPGAN can generate much clearer dehazed images with perceptually pleasing.

In addition to PSNR and SSIM, we also report the averaged NIQE, BRISQUE, SSEQ, and FADE for a comprehensive evaluation of different dehazing methods, as exhibited in Tab. II. Obviously, Cycle-SNSPGAN outperforms other dehazing algorithms in terms of NIQE and BRISQUE, which indicates that the overall quality of the images restored by our method is better. Although compared with DCP and Cycle-Dehaze, our Cycle-SNSPGAN is not the best in terms of SSEQ and FADE, it still achieves impressive performance and ranks second among the eight dehazing approaches.

2) Comparison on Real-World Hazy Images: We also compare our approach with several state-of-the-art methods on

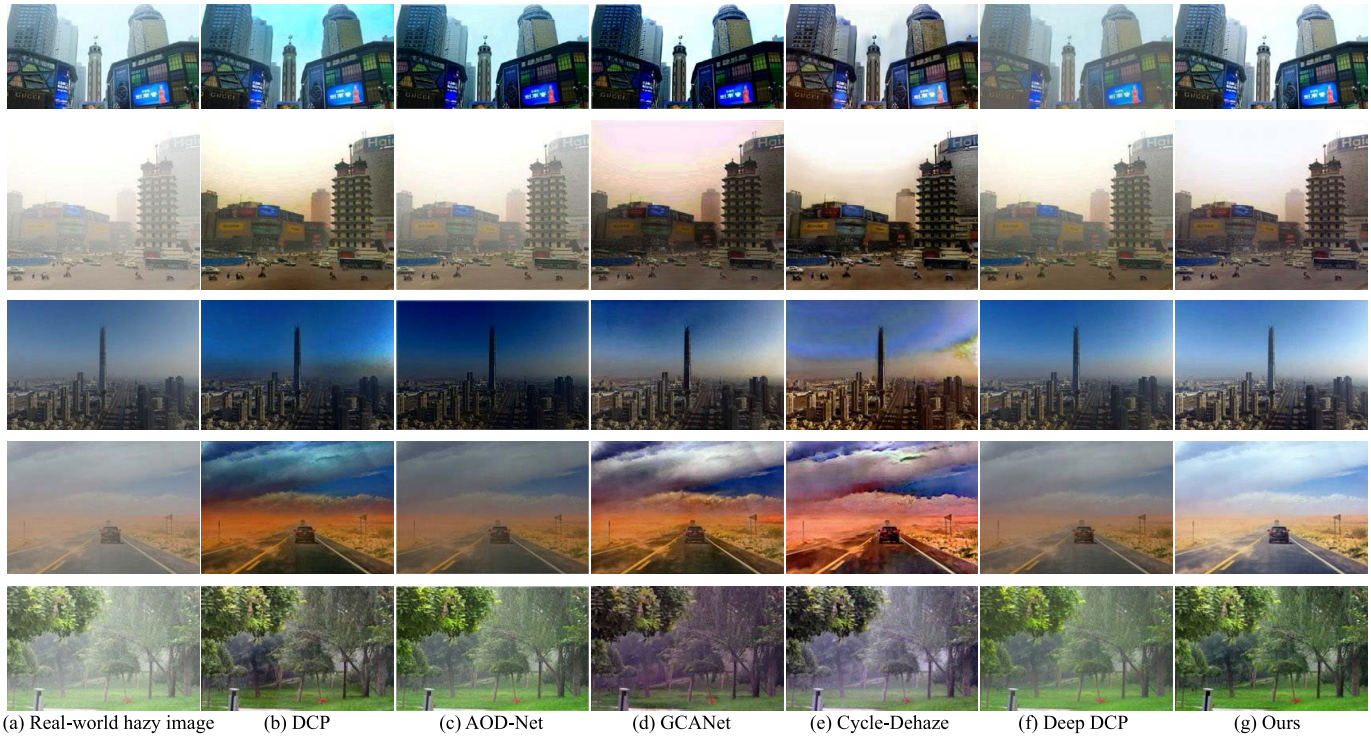
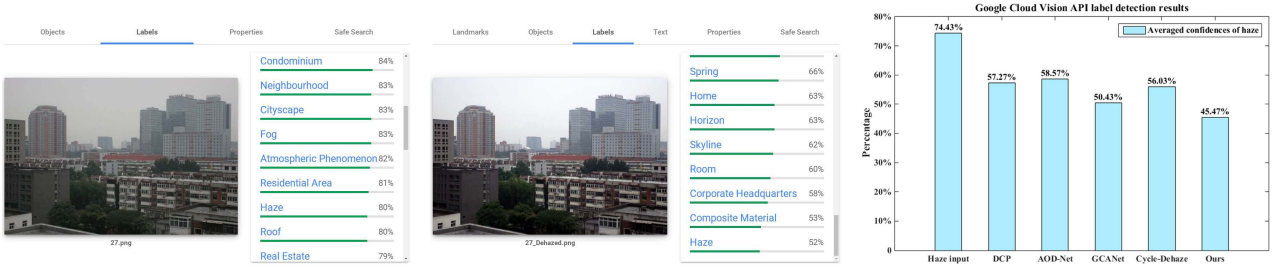


Fig. 10. Image dehazing results on the real-world hazy images. From (a) to (g): (a) the input hazy image, and the dehazing results of (b) DCP [6], (c) AOD-Net [7], (d) GCANet [8], (e) Cycle-Dehaze [9], (f) Deep DCP [59], and (g) our Cycle-SNSPGAN, respectively. Our Cycle-SNSPGAN can produce both clearer and more realistic dehazing results.



(a) (b) (c)

Fig. 11. The dehazing results tested on the Google Cloud Vision API. From (a)-(c): (a) label detection result in the real-world hazy image, (b) label detection result after dehazing by our Cycle-SNSPGAN, and (c) the averaged confidences in recognizing haze from 30 sets of the real-world hazy images and dehazed images of DCP [6], AOD-Net [7], GCANet [8], Cycle-Dehaze [9], and our Cycle-SNSPGAN, respectively.

real-world hazy images (the ground-truth image is not available). Fig. 10 illustrates five real hazy samples and the corresponding dehazing results obtained by different algorithms. Similar to the results in Fig. 8 and Fig. 9, the dehazing results in Fig. 10 generated by DCP lead to color distortion, which makes the dehazed images look darker. The results of AOD-Net, GCANet, and Deep DCP still contain the haze residuals. For Cycle-Dehaze, the dehazed images look darker and have significant artifacts. In contrast, our Cycle-SNSPGAN can generate both haze-free and perceptually pleasing images.

To better understand the performance of the proposed method on real-world images, the Google Cloud Vision API is applied to evaluate the dehazing results. Specifically, we employ Vision API to test 30 sets of real-world hazy images and dehazed images of our method and several dehazing approaches [6]–[9]. As illustrated in Fig. 11, when the hazy images are dehazed by the Cycle-SNSPGAN, the confidences of Vision API in recognizing haze are reduced significantly. Moreover, our method can effectively remove the haze in the real world and performs well against the other approaches.

Additionally, we further conduct the user study to quantitatively evaluate the performance of real-world haze removal. Specifically, we first prepare for 30 real-world hazy images randomly selected from the RTTS dataset (different from our training set). Then, we employ several state-of-the-art dehazing approaches to remove the haze from these 30 images. After that, 10 participants with 5 females and 5 males are asked to score the dehazing results on a scale from 1 (worst quality) to 5 (best quality). For each participant, we present him/her

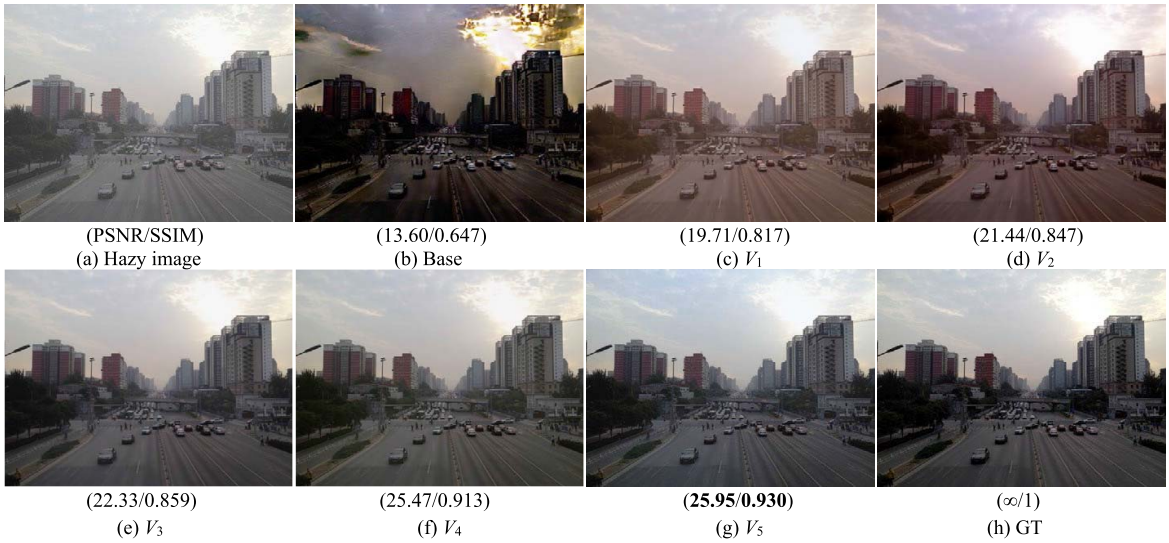


Fig. 12. Visual results of ablation studies. From (a) to (h): (a) the input hazy image, and the results of (b) Base, (c) V_1, (d) V_2, (e) V_3, (f) V_4, (g) V_5, respectively, and (h) the ground-truth image.

TABLE III

USER STUDY RESULTS. MEAN RATINGS GIVEN BY THE PARTICIPANTS ON THE VARIOUS DEHAZING METHODS

Method	Rating (mean & standard dev.)
DCP [6]	2.76 ± 0.71
AOD-Net [7]	2.72 ± 0.33
GCANet [8]	3.09 ± 0.44
Cycle-Dehaze [9]	2.51 ± 0.64
Deep DCP [59]	2.84 ± 0.41
Cycle-SNSPGAN (ours)	**3.41** ± **0.47**

TABLE IV

QUANTITATIVE COMPARISONS (NIQE/BRISQUE/SSEQ/FADE) WITH STATE-OF-THE-ART DEHAZING ALGORITHMS ON 30 REAL-WORLD IMAGES. RED AND BLUE INDICATE THE 1^{st} AND 2^{nd} RANKS, RESPECTIVELY

Method	Type	NIQE↓	BRISQUE↓	SSEQ↓	FADE↓
Hazy	-	4.442	30.560	30.333	2.361
DCP [6]	Prior	4.028	28.625	32.047	0.632
AOD-Net [7]	Supervised	4.252	29.327	32.435	1.224
GCANet [8]	Supervised	3.922	24.985	30.559	0.682
Cycle-Dehaze [9]	Unsupervised	4.061	24.772	29.507	0.686
Deep DCP [59]	Unsupervised	4.070	28.615	33.474	1.144
Cycle-SNSPGAN	Unsupervised	3.915	23.839	29.105	0.677

the 180 dehazing results in a random order without knowing the corresponding approaches. The statistic result is reported in Tab. III, showing that our Cycle-SNSPGAN achieves the best performance for real-world haze removal.

For the quantitative comparison, we report the averaged NIQE, BRISQUE, SSEQ, and FADE values of different dehazing algorithms in Tab. IV. All these metrics are evaluated on the 30 images prepared for the user study. As a display, our Cycle-SNSPGAN wins the first place in terms of NIQE, BRISQUE, and SSEQ, indicating that the dehazed images restored by our method are much more satisfactory. Moreover, the proposed Cycle-SNSPGAN also achieves impressive performance in terms of FADE. In general, Cycle-SNSPGAN wins three of the four metrics, which further verifies the superiority of our method on real-world dehazing tasks.

C. Ablation Study

In this subsection, ablation studies are performed to explore the contribution of the components of Cycle-SNSPGAN towards effective dehazing. Our nine configurations to train the proposed Cycle-SNSPGAN are:

1) base network (original CycleGAN) + SN-Soft-Patch GAN architecture → V_1,
2) V_1 + cyclic self-perceptual loss → V_2,
3) V_2 + color loss → V_3,
4) V_3 + attention module → V_4,
5) V_4 + total variation loss → V_5 (full model),
6) V_5 − SN-Soft-Patch GAN architecture → V_6,
7) V_5 − cyclic self-perceptual loss → V_7,
8) V_5 − color loss → V_8,
9) V_5 − attention module → V_9.

All these variants are retrained in the same way with 100 epochs and are tested on the SOTS outdoor dataset. The experimental results are shown in Tab. V and Fig. 12.

As demonstrated in Fig. 12 and Tab. V, each component contributes to the image dehazing quality, especially the SN-Soft-Patch architecture and the proposed self-perceptual loss. The introduction of the attention module and the color loss have greatly improved the performance of the model in terms of PSNR, while the use of total variation loss improves the network performance in terms of SSIM. In short, if we make full use of the implementation details in this paper, the results will outperform other competitive dehazing approaches.

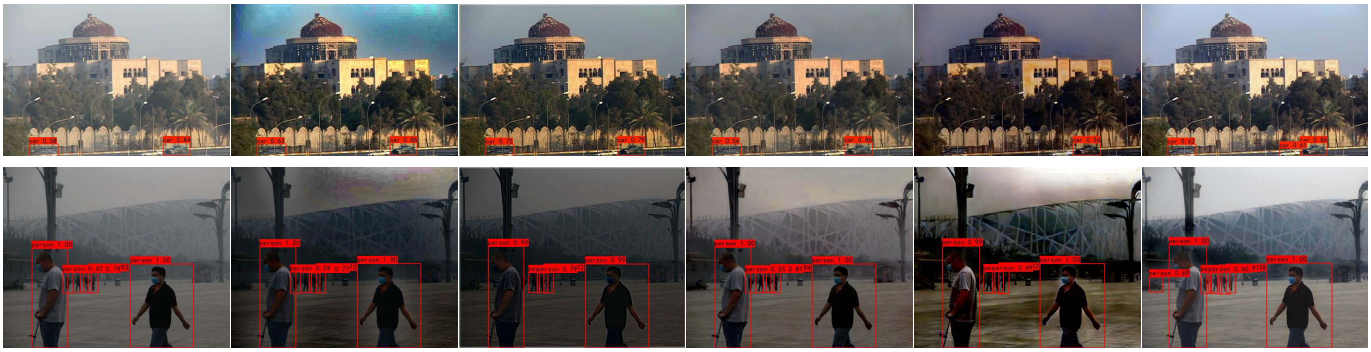
D. Application

The accuracy of object detection can be significantly degraded under hazy conditions (see Fig. 13 (a)). To demon-

TABLE V

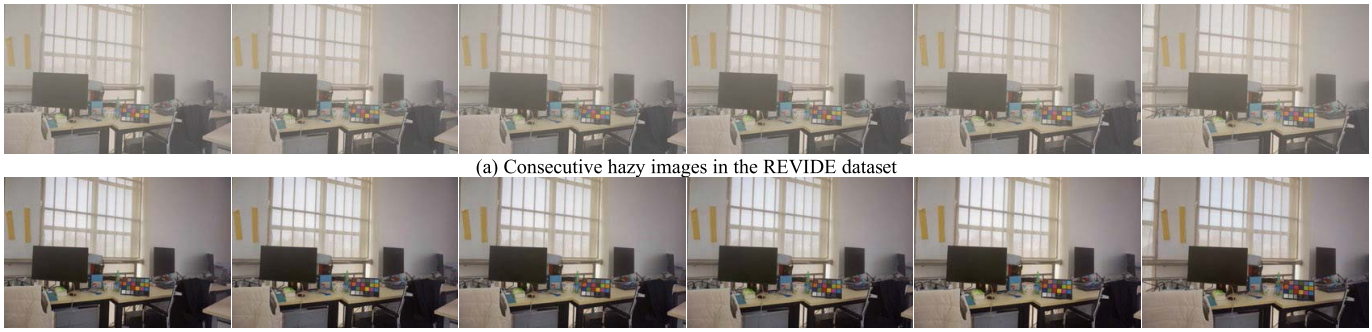
ABLATION ANALYSIS OF DIFFERENT TRAINING STRATEGIES ON SOTS OUTDOOR DATASET. OUR FULL MODEL OUTPERFORMS OTHER ALTERNATIVES

Variants	Base	V_1	V_2	V_3	V_4	V_5	V_6	V_7	V_8	V_9
SN-Soft-Patch architecture	w/o	✓	✓	✓	✓	✓	w/o	✓	✓	✓
Cyclic self-perceptual loss	w/o	w/o	✓	✓	✓	✓	✓	w/o	✓	✓
Color loss	w/o	w/o	w/o	✓	✓	✓	✓	✓	w/o	✓
Attention module	w/o	w/o	w/o	w/o	✓	✓	✓	✓	✓	w/o
Total variation loss	w/o	w/o	w/o	w/o	w/o	✓	✓	✓	✓	✓
PSNR	16.74	20.41	21.69	22.77	23.04	**23.34**	22.07	22.23	22.64	22.71
SSIM	0.684	0.810	0.857	0.866	0.881	**0.911**	0.853	0.874	0.879	0.871



(a) Real-world hazy image (b) DCP (c) AOD-Net (d) GCANet (e) Cycle-Dehaze (f) Ours

Fig. 13. Object detection results on the real-world hazy images and images after dehazing by different methods. From (a) to (f): (a) object detection results on the hazy images, and the object detection results after dehazing by (b) DCP [6], (c) AOD-Net [7], (d) GCANet [8], (e) Cycle-Dehaze [9], and (f) our Cycle-SNSPGAN, respectively.



(a) Consecutive hazy images in the REVIDE dataset

(b) Dehazing results of our Cycle-SNSPGAN

Fig. 14. Consecutive hazy images in the REVIDE dataset [65] and the corresponding dehazing results of the proposed Cycle-SNSPGAN.

strate that Cycle-SNSPGAN can benefit vision-based applications like traffic monitoring and surveillance, we apply a pre-trained YOLOv4 [66] to detect objects of interests on the real-world hazy images and images after dehazing by different methods. As shown in Fig. 13, after dehazing, the confidences in detecting objects of interests are significantly improved. clearly, the proposed model performs favorably against the other methods for object detection on the real-world hazy images. Additionally, we also report the mean Average Precision (mAP) before and after dehazing on 30 real-world images selected from the RTTS dataset, as tabulated in Tab. VI. Obviously, the proposed Cycle-SNSPGAN achieves the best performance, which indicates that our method is effective in real-world dehazing tasks and can benefit downstream vision systems.

Furthermore, dehazing is often a pre-processing step in many automated systems, where the input of such systems may be a sequence of images (i.e., video). Therefore, we test our Cycle-SNSPGAN on a recently released hazy video dataset, i.e., REal-world VIdeo DEhazing dataset (REVIDE) [65]. Fig. 14 exhibits six consecutive hazy images in the REVIDE dataset and the corresponding dehazing results of Cycle-SNSPGAN. As observed, our method can effectively address the consecutive hazy images to produce the realistic dehazing results. However, Cycle-SNSPGAN currently focuses on dehzing single images, thus overlooking the temporal redundancy from neighborhood hazy frames. Although it is challenging to exploit the temporal redundancy among the hazy frames, we will attempt to explore a fusion strategy in the future work, making Cycle-SNSPGAN enjoy such the temporal

TABLE VI

COMPARISON OF OBJECT DETECTION PRECISION MEASURED ON 30 REAL-WORLD IMAGES SELECTED FROM THE RTTS DATASET. THE OBJECT DETECTOR YOLOV4 IS TRAINED ON THE VOC2007 DATASET

Method	mAP(%)
Hazy	34.91
DCP [6]	35.73
AOD-Net [7]	34.82
GCANet [8]	38.47
Cycle-Dehaze [9]	37.21
Cycle-SNSPGAN (ours)	**41.77**

TABLE VII

RUNTIME (IN SECONDS) COMPARISON OF DIFFERENT DEHAZING METHODS TESTED ON THE HSTS DATASET

Method	Platform	Average time
DCP [6]	Python (CPU)	1.37
AOD-Net [7]	PyTorch (GPU)	**0.10**
GFN [54]	PyTorch (GPU)	0.41
CycleGAN [10]	TensorFlow (GPU)	2.11
Cycle-Dehaze [9]	TensorFlow (GPU)	1.92
GCANet [8]	PyTorch (GPU)	0.21
Deep DCP [59]	TensorFlow (GPU)	0.63
Cycle-SNSPGAN (ours)	PyTorch (GPU)	0.19

redundancy information to further enhance the dehazing quality.

E. Runtime Analysis

Efficiency should be considered as an important component for intelligent transportation systems like autonomous driving, traffic monitoring, and surveillance. We report the execution speed of some typical dehazing approaches by listing their averaged running time in Tab. VII. Note that all the methods are implemented on a system with an Intel(R) Xeon E5-2698 v4 CPU, 64 GB RAM, and an NVIDIA Tesla V100 GPU. Cycle-SNSPGAN takes about 0.19*s* to handle one hazy image from the HSTS dataset on average. The speed of Cycle-SNSPGAN is acceptable, since it ranks second among the 8 dehazing algorithms.

V. CONCLUSION

In this work, we have proposed a unsupervised dehazing framework, called Cycle-SNSPGAN, to enhance the generalization ability on real-world hazy scenarios. In order to leverage unpaired samples of real-world hazy images for training, we design an SN-Soft-Patch GAN and exploit a novel cyclic self-perceptual loss which avoids using the ground-truth image to compute the perceptual similarity. This loss function can help our approach preserve both large-scale structures and small-scale details from the hazy images. In addition, the color features of the hazy images are taken as prior knowledge, and a color loss is introduced to make the dehazed images more realistic. Experiments on both synthetic dataset and real-world hazy images demonstrate that our method performs favorably against the state-of-the-art dehazing approaches, even only a small dataset trains our Cycle-SNSPGAN.

ACKNOWLEDGMENT

The authors thank the anonymous reviewers for their careful reading and valuable comments.

REFERENCES

[1] M. Negru, S. Nedevschi, and R. I. Peter, "Exponential contrast restoration in fog conditions for driving assistance," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 4, pp. 2257–2268, Aug. 2015.

[2] X. Min, G. Zhai, K. Gu, X. Yang, and X. Guan, "Objective quality evaluation of dehazed images," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 8, pp. 2879–2892, Aug. 2019.

[3] J. R. Peters, A. Surana, G. S. Taylor, T. S. Turpin, and F. Bullo, "UAV surveillance under visibility and dwell-time constraints: A sampling-based approach," 2019, *arXiv:1908.05347*.

[4] D.-Y. Choi, J.-H. Choi, J. Choi, and B. C. Song, "Sharpness enhancement and super-resolution of around-view monitor images," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 8, pp. 2650–2662, Aug. 2018.

[5] M. Nasir, K. Muhammad, J. Lloret, A. K. Sangaiah, and M. Sajjad, "Fog computing enabled cost-effective distributed summarization of surveillance videos for smart cities," *J. Parallel Distrib. Comput.*, vol. 126, pp. 161–170, Apr. 2019.

[6] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 12, pp. 2341–2353, Dec. 2011.

[7] B. Li, X. Peng, Z. Wang, J. Xu, and D. Feng, "AOD-Net: All-in-one dehazing network," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 4780–4788.

[8] D. Chen *et al.*, "Gated context aggregation network for image dehazing and deraining," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2019, pp. 1375–1383.

[9] D. Engin, A. Genc, and E. H. Kemal, "Cycle-dehaze: Enhanced cyclegan for single image dehazing," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2018, pp. 825–833.

[10] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2242–2251.

[11] T. Miyato, T. Kataoka, M. Koyama, and Y. Yoshida, "Spectral normalization for generative adversarial networks," in *Proc. Int. Conf. Learn. Represent.*, Apr. 2018, pp. 1–26.

[12] P. Isola *et al.*, "Image-to-image translation with conditional adversarial networks," in *Proc. CVPR*, Jul. 2017, pp. 5967–5976.

[13] R. T. Tan, "Visibility in bad weather from a single image," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2008, pp. 1–8.

[14] Q. Zhu, J. Mai, and L. Shao, "A fast single image haze removal algorithm using color attenuation prior," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3522–3533, Nov. 2015.

[15] G. Meng, Y. Wang, J. Duan, S. Xiang, and C. Pan, "Efficient image dehazing with boundary constraint and contextual regularization," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 617–624.

[16] J.-P. Tarel and N. Hautiere, "Fast visibility restoration from a single color or gray level image," in *Proc. IEEE 12th Int. Conf. Comput. Vis. (ICCV)*, Sep. 2009, pp. 2201–2208.

[17] R. Fattal, "Dehazing using color-lines," *ACM Trans. Graph.*, vol. 34, no. 1, pp. 1–14, Nov. 2014.

[18] D. Berman, T. Treibitz, and S. Avidan, "Non-local image dehazing," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1674–1682.

[19] B. Cai, X. Xu, K. Jia, C. Qing, and D. Tao, "DehazeNet: An end-to-end system for single image haze removal," *IEEE Trans. Image Process.*, vol. 25, no. 11, pp. 5187–5198, Nov. 2016.

[20] X. Qin, Z. Wang, Y. Bai, X. Xie, and H. Jia, "FFA-Net: Feature fusion attention network for single image dehazing," in *Proc. Conf. Artif. Intell. (AAAI)*, New York, NY, USA, Feb. 2020, pp. 11908–11915.

[21] W. Ren, S. Liu, H. Zhang, J. Pan, X. Cao, and M.-H. Yang, "Single image dehazing via multi-scale convolutional neural networks," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Sep. 2016, pp. 154–169.

[22] A. Dudhane, K. M. Biradar, P. W. Patil, P. Hambarde, and S. Murala, "Varicolored image de-hazing," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 4564–4573.

[23] B. Li *et al.*, "Benchmarking single-image dehazing and beyond," *IEEE Trans. Image Process.*, vol. 28, no. 1, pp. 492–505, Jan. 2019.

[24] C. Ledig *et al.*, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 105–114.

[25] W. Zhang, Y. Liu, C. Dong, and Y. Qiao, "RankSRGAN: Generative adversarial networks with ranker for image super-resolution," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 3096–3105.

[26] H. Zhang, V. Sindagi, and V. M. Patel, "Image de-raining using a conditional generative adversarial network," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 11, pp. 3943–3956, Nov. 2020.

[27] H. Zhu, X. Peng, V. Chandrasekhar, L. Li, and J.-H. Lim, "DehazeGAN: When image dehazing meets differential programming," in *Proc. 27th Int. Joint Conf. Artif. Intell.*, Jul. 2018, pp. 1234–1240.

[28] P. L. Suarez, A. D. Sappa, B. X. Vintimilla, and R. I. Hammoud, "Deep learning based single image dehazing," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2018, pp. 1169–1176.

[29] H. Zhang and V. M. Patel, "Densely connected pyramid dehazing network," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 3194–3203.

[30] R. Li, J. Pan, Z. Li, and J. Tang, "Single image dehazing via conditional generative adversarial network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2018, pp. 8202–8211.

[31] Y. Dong, Y. Liu, H. Zhang, S. Chen, and Y. Qiao, "FD-GAN: Generative adversarial networks with fusion-discriminator for single image dehazing," in *Proc. Conf. Artif. Intell. (AAAI)*, 2020, vol. 34, no. 7, pp. 10729–10736.

[32] X. Liu, Z. Shi, Z. Wu, and J. Chen, "GridDehazeNet+: An enhanced multi-scale network with intra-task knowledge transfer for single image dehazing," 2021, *arXiv:2103.13998*.

[33] J. Zhang and D. Tao, "FAMED-Net: A fast and accurate multi-scale end-to-end dehazing network," *IEEE Trans. Image Process.*, vol. 29, pp. 72–84, 2019.

[34] Y. Qu, Y. Chen, J. Huang, and Y. Xie, "Enhanced pix2pix dehazing network," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 8160–8168.

[35] Y. Shao, L. Li, W. Ren, C. Gao, and N. Sang, "Domain adaptation for image dehazing," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 2805–2814.

[36] Z. Qin, P. Zhang, F. Wu, and X. Li, "FcaNet: Frequency channel attention networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 763–772.

[37] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, Oct. 2015, pp. 234–241.

[38] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.

[39] X. Fu, Q. Qi, Z. Zha, Y. Zhu, and X. Ding, "Rain streak removal via dual graph convolutional network," in *Proc. AAAI Conf. Artif. Intell. (AAAI)*, Feb. 2021, pp. 1352–1360.

[40] D. Zhao, J. Li, H. Li, and L. Xu, "Complementary feature enhanced network with vision transformer for image dehazing," 2021, *arXiv:2109.07100*.

[41] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image super-resolution using very deep residual channel attention networks," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Sep. 2018, pp. 294–310.

[42] K. Xu *et al.*, "Show, attend and tell: Neural image caption generation with visual attention," in *Proc. Int. Conf. Mach. Learn. (ICML)*, Jun. 2015, pp. 2048–2057.

[43] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. Eur. Conf. Comput. Vis.*, Munich, Germany, Sep. 2018, pp. 7132–7141.

[44] C. O. Ancuti, C. Ancuti, and R. Timofte, "NH-HAZE: An image dehazing benchmark with non-homogeneous hazy and haze-free images," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2020, pp. 1798–1805.

[45] R. Versteegen, G. Gimel'farb, and P. Riddle, "Texture modelling with nested high-order Markov–Gibbs random fields," *Comput. Vis. Image Understand.*, vol. 143, pp. 120–134, Feb. 2016.

[46] X. Mao, Q. Li, H. Xie, R. Y. K. Lau, Z. Wang, and S. P. Smolley, "Least squares generative adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2813–2821.

[47] Z. Anvari and V. Athitsos, "Dehaze-GLCGAN: Unpaired single image de-hazing via adversarial training," 2020, *arXiv:2008.06632*.

[48] C. Liu, J. Fan, and G. Yin, "Efficient unpaired image dehazing with cyclic perceptual-depth supervision," 2020, *arXiv:2007.05220*.

[49] A. Ignatov, N. Kobyshev, K. Vanhoey, R. Timofte, and L. Van Gool, "DSLR-quality photos on mobile devices with deep convolutional networks," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2017, pp. 3297–3305.

[50] H. A. Aly and E. Dubois, "Image up-sampling using total-variation regularization with a new observation model," *IEEE Trans. Image Process.*, vol. 14, no. 10, pp. 1647–1659, Oct. 2005.

[51] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *Proc. 14th Eur. Conf. Comput. Vis. (ECCV)*, Amsterdam, The Netherlands, Oct. 2016, pp. 694–711.

[52] Y. Jiang, X. Gong, D. Liu, Y. Cheng, and C. Fang, "EnlightenGAN: Deep light enhancement without paired supervision," *IEEE Trans. Image Process.*, vol. 30, pp. 2340–2349, 2021, doi: 10.1109/TIP.2021.3051462.

[53] C. O. Ancuti, C. Ancuti, M. Sbert, and R. Timofte, "Dense-haze: A benchmark for image dehazing with dense-haze and haze-free images," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2019, pp. 1014–1018.

[54] W. Ren *et al.*, "Gated fusion network for single image dehazing," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 3253–3261.

[55] W. Ren, J. Pan, H. Zhang, X. Cao, and M.-H. Yang, "Single image dehazing via multi-scale convolutional neural networks with holistic edges," *Int. J. Comput. Vis.*, vol. 128, no. 1, pp. 240–259, Jan. 2020.

[56] X. Zhang, H. Dong, Z. Hu, W. Lai, F. Wang, and M. Yang, "Gated fusion network for degraded image super resolution," *Int. J. Comput. Vis.*, vol. 128, no. 6, pp. 1699–1721, Jan. 2020.

[57] Q. Wu, W. Ren, and X. Cao, "Learning interleaved cascade of shrinkage fields for joint image dehazing and denoising," *IEEE Trans. Image Process.*, vol. 29, pp. 1788–1801, 2020.

[58] A. Dudhane, P. W. Patil, and S. Murala, "An end-to-end network for image de-hazing and beyond," *IEEE Trans. Emerg. Topics Comput. Intell.*, vol. 6, no. 1, pp. 159–170, Feb. 2022.

[59] A. Golts, D. Freedman, and M. Elad, "Unsupervised single image dehazing using dark channel prior loss," *IEEE Trans. Image Process.*, vol. 29, pp. 2692–2701, 2020.

[60] B. Li, Y. Gou, S. Gu, J. Z. Liu, J. T. Zhou, and X. Peng, "You only look yourself: Unsupervised and untrained single image dehazing neural network," *Int. J. Comput. Vis.*, vol. 129, no. 5, pp. 1754–1767, Mar. 2021.

[61] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a 'completely blindr' image quality analyzer," *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 209–212, Mar. 2013.

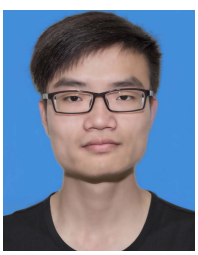
[62] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Trans. Image Process.*, vol. 21, no. 12, pp. 4695–4708, Dec. 2012.

[63] L. Liu, B. Liu, H. Huang, and A. C. Bovik, "No-reference image quality assessment based on spatial and spectral entropies," *Signal Process., Image Commun.*, vol. 29, no. 8, pp. 856–863, 2014.

[64] L. K. Choi, J. You, and A. C. Bovik, "Referenceless prediction of perceptual fog density and perceptual image defogging," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3888–3901, Nov. 2015.

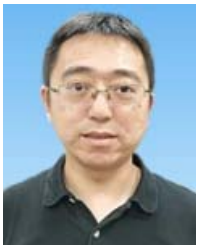
[65] X. Zhang *et al.*, "Learning to restore hazy video: A new real-world dataset and a new method," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 9239–9248.

[66] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal speed and accuracy of object detection," 2020, *arXiv:2004.10934*.



Yongzhen Wang received the M.S. degree in 2019. He is currently pursuing the Ph.D. degree with the School of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics (NUAA), Nanjing, China. His research interests include deep learning, image processing, and computer vision, particularly in the domains of image dehazing, image deraining, and object detection. He has served as a PC Member for AAAI 2022.

Xuefeng Yan received the Ph.D. degree from the Beijing Institute of Technology in 2005. He was a Visiting Scholar at Georgia State University in 2008 and 2012. He is currently a Professor with the School of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics (NUAA), China. His research interests include intelligent computing, MBSE/complex system modeling, simulation, and evaluation.



Donghai Guan received the Ph.D. degree from Kyung Hee University (KHU), South Korea, in 2009. He is currently an Associate Professor with the School of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, China. His research interests focus on data mining and machine learning.

Mingqiang Wei (Senior Member, IEEE) received the Ph.D. degree in computer science and engineering from The Chinese University of Hong Kong (CUHK) in 2014. He is currently a Professor with the School of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics (NUAA). Before joining NUAA, he served as an Assistant Professor at the Hefei University of Technology and a Post-Doctoral Fellow at CUHK. His research interests focus on 3D vision, computer graphics, and deep learning. He was a recipient of the CUHK Young Scholar Thesis Awards in 2014. He is currently an Associate Editor of the *Visual Computer* journal, *Journal of Electronic Imaging*, and *International Journal of Image and Graphics*; and a Guest Editor of IEEE TRANSACTIONS ON MULTIMEDIA.

Yiping Chen (Senior Member, IEEE) received the Ph.D. degree in information and communications engineering from the National University of Defense Technology, Changsha, China, in 2011. She is currently a Research Associate Professor with the Fujian Key Laboratory of Sensing and Computing for Smart Cities, School of Informatics, Xiamen University, China. From 2007 to 2011, she was an Assistant Researcher with The Chinese University of Hong Kong, China. She has published more than 70 articles in referred journals, including IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING; and conferences, including CVPR, IGARSS, and ISPRS. Her current research interests include remote sensing image processing, mobile laser scanning data analysis, 3D point cloud computer vision, and autonomous driving. She was a recipient of the 2020 Best Reviewer of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING.

Xiao-Ping Zhang (Fellow, IEEE) received the B.S. and Ph.D. degrees in electronic engineering from Tsinghua University in 1992 and 1996, respectively, and the M.B.A. degree (Hons.) in finance, economics and entrepreneurship from the University of Chicago Booth School of Business, Chicago, IL, USA.

Since Fall 2000, he has been with the Department of Electrical, Computer and Biomedical Engineering, Ryerson University, Toronto, ON, Canada, where he is currently a Professor and the Director of the Communication and Signal Processing Applications Laboratory. He has served as the Program Director of graduate studies. He is cross-appointed to the Finance Department, Ted Rogers School of Management, Ryerson University. He was a Visiting Scientist with the Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA, USA, in 2015 and 2017. He is a frequent consultant for biotech companies and investment firms. His research interests include image and multimedia content analysis, sensor networks and the IoT, machine learning, statistical signal processing, and applications in big data, finance, and marketing.

Dr. Zhang is a fellow of the Canadian Academy of Engineering and the Engineering Institute of Canada; a Registered Professional Engineer in Ontario, Canada; and a member of Beta Gamma Sigma Honor Society. He received the 2020 Sarwan Sahota Ryerson Distinguished Scholar Award—the Ryerson University highest honor for scholarly, research and creative achievements. He is selected as an IEEE Distinguished Lecturer by the IEEE Signal Processing Society from 2020 to 2021 and by the IEEE Circuits and Systems Society from 2021 to 2022. He is the General Co-Chair for the 2021 IEEE International Conference on Acoustics, Speech, and Signal Processing; the 2017 GlobalSIP Symposium on Signal and Information Processing for Finance and Business; and the 2019 GlobalSIP Symposium on Signal, Information Processing and AI for Finance and Business. He was an Elected Member of the ICME Steering Committee. He is the General Chair of the 2015 IEEE International Workshop on Multimedia Signal Processing. He is the Editor-in-Chief of the IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING. He is a Senior Area Editor of the IEEE TRANSACTIONS ON IMAGE PROCESSING. He served as a Senior Area Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING; and an Associate Editor for the IEEE TRANSACTIONS ON IMAGE PROCESSING, the IEEE TRANSACTIONS ON MULTIMEDIA, the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, the IEEE TRANSACTIONS ON SIGNAL PROCESSING, and the IEEE SIGNAL PROCESSING LETTERS.

Jonathan Li (Senior Member, IEEE) received the Ph.D. degree in geomatics engineering from the University of Cape Town, South Africa, in 2000. He is currently a Professor with the Department of Geography and Environmental Management and cross-appointed with the Department of Systems Design Engineering, University of Waterloo, Canada. He has coauthored over 500 publications, including more than 300 in refereed journals and more than 200 in conference proceedings. His main research interests include image and point cloud analytics, mobile mapping, and AI-powered information extraction from LiDAR point clouds and earth observation images. He is a fellow of the Engineering Institute of Canada. He was a recipient of the 2021 Geomatica Award, the 2020 Samuel Gamble Award, and the 2019 Outstanding Achievement Award in mobile mapping technology. He is currently serving as the Editor-in-Chief for *International Journal of Applied Earth Observation and Geoinformation*; and an Associate Editor for IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, and *Canadian Journal of Remote Sensing*.