Hindawi
Complexity
Volume 2021, Article ID 2352185, 13 pages
https://doi.org/10.1155/2021/2352185

Research Article

DefogNet: A Single-Image Dehazing Algorithm with Cyclic Structure and Cross-Layer Connections

Suting Chen,[1] **Wenhao Fan**,[1] **Shaw Peter**,[1] **Chuang Zhang**,[1] **Kui Chen**,[1] **and Yong Huang**[2]

[1]*Jiangsu Collaborative Innovation Center of Atmospheric Environment and Equipment Technology (CICAEET), Nanjing University of Information Science & Technology, Nanjing 210044, China*
[2]*The Anhui Province Meteorological Science Research Institute, Hefei 230061, China*

Correspondence should be addressed to Suting Chen; sutingchen@nuist.edu.cn

Received 31 July 2020; Revised 22 October 2020; Accepted 27 November 2020; Published 12 January 2021

Academic Editor: Zhijie Wang

Inspired by the application of CycleGAN networks to the image style conversion problem Zhu et al. (2017), this paper proposes an end-to-end network, DefogNet, for solving the single-image dehazing problem, treating the image dehazing problem as a style conversion problem from a fogged image to a nonfogged image, without the need to estimate a priori information from an atmospheric scattering model. DefogNet improves on CycleGAN by adding a cross-layer connection structure in the generator to enhance the network's multiscale feature extraction capability. The loss function was redesigned to add detail perception loss and color perception loss to improve the quality of texture information recovery and produce better fog-free images. In this paper, the novel Defog-SN algorithm is presented. This algorithm adds a spectral normalization layer to the discriminator's convolution layer to make the discriminant network conform to a 1-Lipschitz continuum and further improve the model's stability. In this study, the experimental process is completed based on the O-HAZE, I-HAZE, and RESIDE datasets. The dehazing results show that the method outperforms traditional methods in terms of PSNR and SSIM on synthetic datasets and Avegrad and Entropy on naturalistic images.

1. Introduction

Images collected under fog, haze, and other weather conditions often suffer from low contrast, unclear scenes, and large color errors, which are prone to adversely affect the application of computer vision algorithms such as target detection and semantic segmentation. Therefore, the method of dehazing a single-image directly without using any a priori information is of great significance to the field of computer vision. At present, the standard dehazing methods can be divided into three according to different principles: firstly, image enhancement techniques. These methods focus mainly on the contrast of the image itself and other information. Secondly, image restoration methods are based on the typical physics models; these methods are primarily through a priori knowledge and physics models to complete the dehazing operation. Thirdly, neural network-based dehazing methods mainly use neural networks to complete the haze feature extraction, thereby completing the dehazing process.

Image enhancement-based dehazing methods focus mainly on the haze image contrast, edge gradient, and other information. The most common dehazing methods include wavelet transform [1], Retinex method [2], and histogram equalization [3]. Jobson et al. [4] proposed a defogging method for the Retinex image enhancement algorithm. Tan [5] proposed a dehazing method that maximizes the local contrast of the image. Moreover, the result obtained using dehazing has excessive saturation and loses much detailed information. Kim [6] proposed a subblock partial overlap method based on the idea of local equilibrium; nevertheless, there is a blocking artifact in this method's results. Zuiderveld [7] proposed a contrast limited adaptive histogram equalization (CLAHE) method to address this problem by adaptively limiting the images' contrast. The information, such as the contrast of the haze image, reflects the severity of

the haze to a certain extent. Still, the method for this kind of intuitive information cannot investigate the haze image formation mechanism. Thus, it often loses detailed information in the process of dehazing. This limitation makes it difficult for this technique to achieve an acceptable defogging effect.

The physical model-based image recovery methods study the foggy images from the imaging mechanism. The dehazing operation is accomplished through a combination of a priori knowledge and the proposed physical model, albeit the priori information must be estimated. The most common methods include dehazing algorithms based on image depth information [8], dehazing algorithms based on atmospheric light polarization [9], and dehazing methods based on various types of a priori information [10–13]. Mccartney first proposed the atmospheric physics model in 1975 [14]. Oakley and Satherley [15] proposed a transmission valuation method for image pixels around this theoretical model; however, it requires information such as the depth of field of the image, which is not very practical. According to the polarization properties of atmospheric light, Schechner et al. [16] estimated the depth of field information for dehazing by setting the polaroid glass. Narasimhan and Nayar [17] proposed a multiscale dehazing algorithm and collected images of the same location at different times for comparison and dehazing. Tarel and Hautiere [18] estimated the atmospheric scattering function by using median filtering, but this method produces a Halo effect when the scene is transformed. Fattal [19] estimated the transmittance by using independent component analysis, but again, this method is not suitable for dense foggy weather. He et al. [10] introduced the dark-channel prior method, which is ideal for defogging and easy to implement; however, some areas with intense light, such as the sky, may cause significant interference to this method. The emergence of dark channels provides new ideas for image defogging. To solve the problems encounted in the dark-channel a priori method, many image defogging methods based on the dark-channel a priori theory have subsequently been produced [20–22].

In recent years, CNN-based image defogging methods have become the focus of research. These methods mainly use neural networks to learn haze image features. Cai et al. [23] proposed the end-to-end DehazeNet system, using a convolutional neural network to extract the haze features to optimize transmittance estimation. A multiscale depth haze-removal network (MSCNN) was proposed by Ren et al. [24] to directly study and estimate the relationship between transmittance and foggy image, solving the shortcomings of artificial features. Zhang et al. [25] improved the edge and detailed information of the image by constructing a pyramid densely connected dehazing network that jointly optimizes the transmittance and atmospheric light values. Goodfellow et al. [26] proposed a new-framework generative adversarial network (GAN), for estimating generative models according to the fixed-point theorem through the adversarial process. GANs [26] can make the distribution produced by the generator as close as possible to the distribution of real data. GANs [26] provided a theoretical basis for new methods of neural network dehazing. Many approaches to optimize the GANs algorithm have been proposed. The DCGAN model of Radford et al. [27] introduced CNNs into the structure of GANs [26]. This model completes the feature extraction operation through CNNs and obtains a higher stability when generating high-quality samples. Conditional GAN (cGAN) proposed by Mirza and Osindero [28] adds conditional information y to the GANs [26]; this improves the stability of the model and enhances the generator's expression capabilities. Isola et al. [29] proposed the pix2pix algorithm to use cGAN [28] to learn the mapping from input to output to complete various image conversion tasks. These methods have authenticity requirements for the foggy image and its corresponding nonfog image, the data requirements are high, and the acquisition is difficult.

Based on GAN [26], Zhu et al. [30] proposed CycleGAN consisting of two unidirectional GANs. CycleGAN [30] suggests using cycle-consistent loss, which is mainly used to transform image styles and effectively overcomes the lack of constraint in the generated images. Albeit the CycleGAN [30] model suffers from noise and loss of texture details when completing the generation task for images with different texture complexity. The DefogNet network proposed in this paper mainly adds detail perception loss and color perception loss to CycleGAN [30]. It fully optimizes the CycleGAN [30] structure using cross-layer connections and the Defog-SN algorithm. Finally, it builds an end-to-end network without considering a priori information and without pairwise datasets.

The main contributions of this paper include the following:

Incomplete feature extraction, inefficient demist processing, and the need for a complex physical model are common issues of using an a priori physical model. To address this, we add the design of cross-layer connections in the generator, enhance the multiscale feature extraction capability of the model with feature pyramids, and obtain higher quality texture details of the generated images. By doing these modifications, we omit the manual design of the a priori model. Consequently, our approach requires neither fuzzy and real image samples nor any atmospheric scattering model parameters in the training and testing phases.

This study designed unique loss functions: detail perception loss and color perception loss to optimize DefogNet for single-image dehazing and to address the color shifts and high contrast resulting from the defogging operation.

We introduce spectral normalization in the discriminative network and propose the Defog-SN algorithm, which has strong generalization ability and effectively solves the problem of insufficient diversity of generated samples and improves the quality of defogged images enhancing the overall stability of the network as well as the convergence speed.

The rest of this paper is organized as follows: Section 2 mainly describes the methods proposed in this paper.

Section 3 draws relevant experimental conclusions and analyzes them in depth. Section 4 concludes the article.

2. Proposed Method

2.1. Structure of DefogNet. This paper presents a DefogNet-based single-image dehazing method that improves on the CycleGAN [30] network architecture. CycleGAN [30] adds cycle-consistent loss; the central role of cycle consistency loss is to extract the combination of high-level and low-level features in the VGG16 architecture [31], which creates constraints on the generator and preserves the original image structure. CycleGAN [30] can accurately calculate the $L1$ paradigm of the original image and the loop-through images to perform the task of unpaired image-to-image conversion. Still, the loss between the original image and the loop-through image does not allow the complete texture information to be recovered.

DefogNet is an improved version of CycleGAN [30], the structure of which is shown in Figure 1. In this paper, we propose the Defog-SN algorithm and add cross-layer connections to the generator, which optimizes the overall performance of the network and improves the quality of dehazed images. DefogNet redesigns the loss function and introduces detail perception loss and color perception loss biased on cycle-consistent loss. It then uses this function to calculate the total loss of the model. In this way, it preserves the initial information more completely in image reconstruction, effectively optimizing the dehazing function.

DefogNet consists of two generators, G and F, and two discriminators, D_X and D_Y. The two generative adversarial networks compete with each other as well as constraints. By training simultaneously, they learn the foggy features while keeping the image background structure almost unchanged and use unpaired datasets without manual labeling. This unsupervised approach dramatically simplifies the work in the data preparation stage.

In this paper, the foggy image dataset and the clear image dataset are used to train the model. The foggy image data and clear image data are defined as domain X and domain Y. There is no correspondence between the images in domain X and domain Y. After inputting the image in X into the generating model G, G will generate a new image Y_G. Hence, Y_G and the image in Y serve as the input to the discriminating model D_X, which determines when the input image is from the real clear data Y or the pseudoimage generated by the generator G. The result is fed back to G and used to strengthen G. Under the guidance of D_X, Y_G continuously reduces the gap between Y_G and the image in the clear data set Y.

Similarly, generator F acts to reduce Y_G and Y to foggy images and Y_G and Y are fed into the 2nd generating model F to generate a new image X_F, which is designed to make X_F as similar as possible to the image in X through a loss function to ensure that the learned mapping is meaningful. This model's cyclic structure allows the two GANs to generate increasingly realistic images, i.e., images in the fogged dataset X to complete the dehazing.

2.2. Cross-Layer Connection Structure on Generator. To properly train the sample distribution and effectively optimize the quality of the final generated image, the network structure was chosen was an encoder-transition-decoder. The role of the first part of the encoder layer is to complete the feature extraction process. The second part of the transition layer is to combine the image's different features through the residual network and reorganize the features. The third part of the decoder layer is to reconstruct the image.

When comparing the fogged image and the defogged image, it is found that the fogged image and the defogged image should have the same background, similar spatial structure, and details of the object. So, we need to preserve some of the input image structure, share some of the information between the input and output, and send this information directly to the decoding layer without going through the transformation layer. Based on this, our approach improves the generator structure and designs a codec network with a cross-layer connection structure to break the bottleneck of information loss during the codec process and discards the method of merely connecting all channels of the symmetric layer, as shown in Figure 2. The output of each convolution layer in the encoder will be directly inputted to the corresponding decoder simultaneously with the inverse convolution result of the output of the next convolution layer. This structure keeps the feature map's size consistent, effectively reducing the difference between the output and the original input; otherwise, the features of the original image will not be retained in the output, and the output will deviate from the background contour of the original image.

In this paper, the batch normalization layer is removed for each unit in the transformation layer, and the SeLU activation function [32] is used to automatically normalize the sample distribution to a mean of zero and a standard deviation of one. The relevant formula is

$$\operatorname{selu}(x) = \lambda \begin{cases} x, & x > 0, \\ \alpha e^{x} - \alpha, & x \leq 0. \end{cases} \tag{1}$$

The structure can ensure the high consistency between the defogged image and the original banded fog image except for haze and improve the multiscale fusion feature extraction capability of the network for haze.

2.3. Defog-SN Algorithm on Discriminator. CycleGAN [30] consists of two unidirectional GANs, which have the disadvantages of unstable training and collapse-prone models due to the discriminant network's poor control performance. In this paper, the Defog-SN algorithm is to solve this problem. This algorithm incorporates spectral normalization [33] into the discriminant network by inserting a spectral normalization layer into each convolutional layer to optimize the quality of dehazed images. The discriminator mainly consists of 6 layers; the first four layers are used to complete the input image's feature extraction process and then the last layer. The last fully-connected layer outputs Wasserstein distance. The discriminator structure is shown in Figure 3.

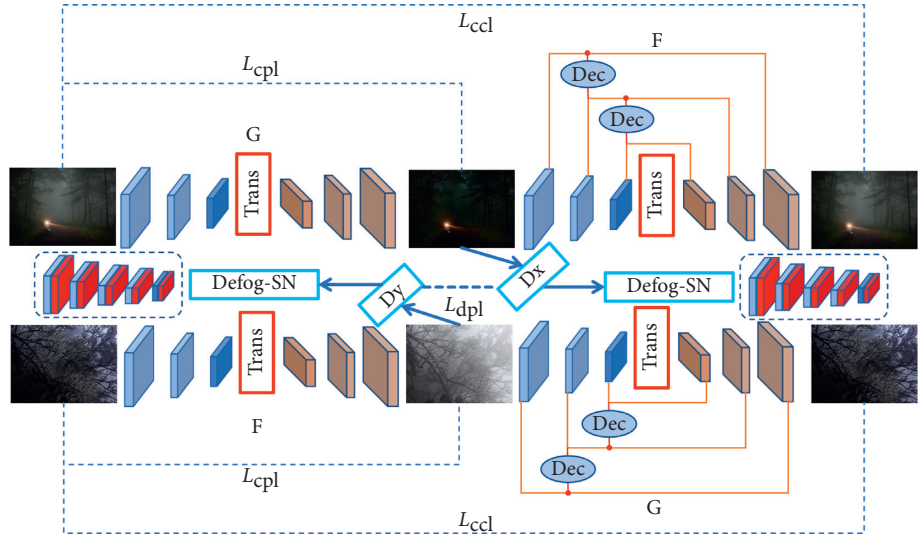


Figure 1: DefogNet network structure.

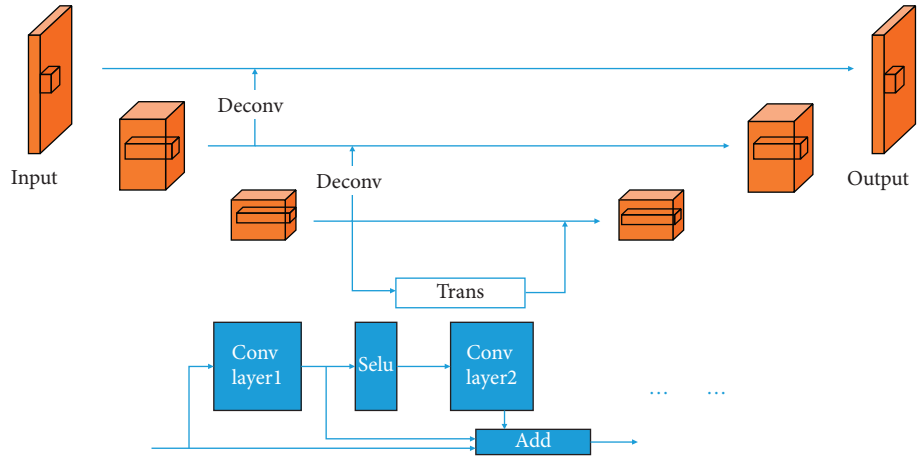


Figure 2: Structure of the generator.

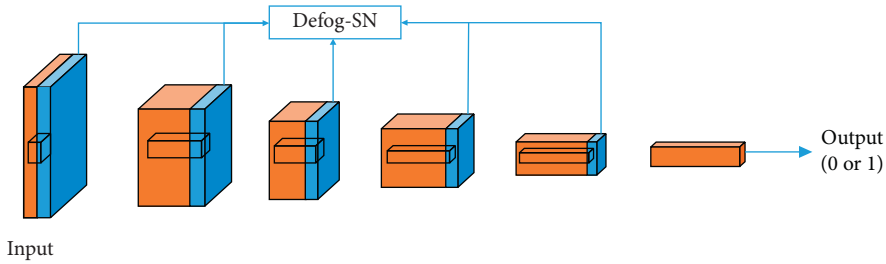


Figure 3: Discriminator structure.

The GAN stability theorem states that if the discriminant network can conform to a 1-Lipschitz continuum in terms of the input, the output, and the control of the discriminant network can be significantly optimized. The stability of the GAN training process can be further enhanced [34]. According to the Lipschitz theory of complex functions, typical neural networks consist of multilayer structures, which can be viewed as a more complicated complex function. If the individual functions can satisfy the 1-Lipschitz continuum, then the composite functions that combine these functions also satisfy the continuum. DefogNet's discriminant network's activation functions are all Leaky

ReLU [35] functions, which satisfy 1-Lipschitz continuity, so, as long as each convolutional layer of the discriminant network satisfies 1-Lipschitz continuity, the whole discriminant network satisfies 1-Lipschitz continuity. The formula for the Lipschitz constraint is

$$\frac{\|f(x)-f(x')\|_2}{\|x-x'\|_2}\le\beta, \tag{2}$$

where β is a constant and the gradient of a function satisfying the formula is always limited to a range less than or equal to β. The Defog-SN algorithm uses spectral normalization to provide a global normal for the discriminant network. The maximum singularity value can have a decisive effect on the Lipschitz continuity constant of the linear operator, thus enabling the parameter matrix to use more features when generating images, thus effectively enhancing the diversity of samples and optimizing the quality of image defogging.

The Defog-SN algorithm adds spectral normalization after each convolutional layer to make each layer conform to the 1-Lipschitz continuity. The details of the algorithm are as follows: for the convolutional layer t: $h_{\text{in}} \longrightarrow h_{\text{out}}$, A refers to the matrix of convolutional layer parameters, and its spectral parameter $\sigma(A)$ is calculated by the following equation [36]:

$$\sigma(A)=\max_{h\neq 0}\frac{\|Ah\|_2}{\|h\|_2}=\max_{\|h\|_2\le 1}\|Ah\|, \tag{3}$$

where $\sigma(A)$ is equal to the maximum singularity of matrix A and the Lipschitz continuity constant of the convolutional layer t is equal to the spectral paradigm of its convolutional layer parameter matrix. The spectral paradigm $\sigma(W)$ of the convolutional layer parameter matrix, W, is calculated so that the maximum singularity of the convolutional layer parameter matrix W_{SN} after completion of the normalization process is equal to 1. In this way, the input and output satisfy the 1-Lipschitz condition. The formula for W_{SN} is as follows:

$$W_{\text{SN}}=\frac{W}{\sigma(W)}. \tag{4}$$

The vector u is randomly initialized as the right singularity eigenvector of the parameter matrix W. The following formula calculates the left singularity eigenvector v:

$$\begin{aligned} v&=\frac{\|W^T u\|}{\|W^T u\|_2},\\ u&=\frac{\|Wv\|}{\|Wv\|_2}. \end{aligned} \tag{5}$$

After iterations, it is possible to calculate the opening $\sqrt{\lambda_1}$ of the maximum eigenvalue of W^TW, i.e., the maximum singular value of the matrix W:

$$\sqrt{\lambda_1}=u^TWv. \tag{6}$$

The update of the convolutional layer parameter matrix W^i is accomplished as follows:

$$W^i\leftarrow W^i\leftarrow\alpha\nabla_{W^i}\left(W^i_{\text{SN}}\left(W^i\right),X,Y\right), \tag{7}$$

where α is the learning rate.

We use deconvolution to simplify and speed up the calculation of the spectral norm of convolution. Hanie et al. [37] developed a method to calculate all singular values, including the spectral norm; however, this method is only suitable for convolution filters with step size 1 and 0 padding. In the training process, the standardization factor depends on the step size and filling scheme of the control convolution operation.

This paper proposed an efficient method to calculate the maximum singular value (i.e., the spectral norm) of a 6-layer convolution layer with an arbitrary step size and filling scheme. The output feature map ψ of layer i in the neural network can be expressed as a linear operation of input X:

$$\psi_i(X)=\sum_{j=1}^{M}F_{i,j}*X_j, \tag{8}$$

where M is the feature map of the input and $F_{i,j}$ is a filter. Here, we ignore the additional bias term. We vectorize X and let $T_{i,j}$ express the overall linear operation related to $F_{i,j}$:

$$\psi_i(X)=\begin{bmatrix}T_{1,1} & \cdots & T_{1,M}\end{bmatrix}X. \tag{9}$$

Then, the convolution operation can be expressed as

$$\psi(X)=\begin{bmatrix}T_{1,1} & \cdots & T_{1,M}\\ \vdots & \ddots & \vdots\\ T_{N,1} & \cdots & T_{N,M}\end{bmatrix}X=WX. \tag{10}$$

By convolution transposition, we can obtain the spectral norm associated with W. By using W^T effectively, we can implement this matrix multiplication more efficiently, without explicitly constructing W. The correlation spectrum norm $\sigma(W)$ can be obtained by the power iteration method, and appropriate step size and filling parameters are added in convolution and convolution transposition operation. We use the same value u repeatedly, and only update W once per step. We use a wider range to restrict $(W)\le\beta$:

$$W_{SN}=\frac{W}{\max\left(1,\left(\sigma(W)/\beta\right)\right)}, \tag{11}$$

which results in a faster training speed.

We now use Wasserstein distance as a criterion to measure the generated distribution p_g and the real distribution p_{data}. Due to the introduction of 1-Lipschitz continuity, we need to restrict the variation range of network parameters within a certain range; the change range of parameters should not exceed a certain constant in each update. Therefore, the Wasserstein distance between the real data distribution p_{data} and the generated data distribution p_g can be expressed as follows:

$$D_w=E_{x\sim p_{\text{data}}}\left[f_w(x)\right]-E_{x\sim p_g}\left[f_w(x)\right]. \tag{12}$$

The smaller the D_w, the more likely the generated distribution p_g is to be close to the true distribution p_{data}. Due

to the introduction of spectral normalization, the function is differentiable in all cases, allowing us to solve the problem of gradient vanishing in the GAN model's training process [38]. Therefore, the objective function of the DefogNet discriminator is as follows:

$$\text{obj}^{D} = \min\Big(E_{x\sim p_g}[f_w(x)] - E_{x\sim p_{\text{data}}}[f_w(x)]\Big). \tag{13}$$

Next, we enhance the Lipschitz constraint by gradient penalty. Firstly, we use the random sampling method to obtain the True sample X_{data}, False sample X_g, and a random number ϑ in the range of [0,1]. Then, we randomly interpolate the sample between X_{data} and X_g:

$$\widehat{X} = \vartheta X_{\text{data}} + (1-\vartheta)X_g. \tag{14}$$

The distribution satisfied by $\widehat{X}$ is denoted as $p_{\widehat{X}}$, and the improved objective function of the DefogNet is

$$\text{obj}^{(G,D)} = \min\left(\begin{array}{c} E_{x\sim p_g}[D_w(x)] - E_{x\sim p_{\text{data}}}[D_w(x)] + \\ E_{\widehat{X}\sim p_{\widehat{X}}}\left[\left\|\nabla_{\widehat{X}} D_w(\widehat{X})\right\|_2 - \left(D_w(x) - D_w(\widehat{X})\right)^2\right]^2 \end{array}\right). \tag{15}$$

2.4. Loss Function of DefogNet. CycleGAN's [30] loss function consists of the generator and discriminator's adversarial loss function and a cycle-consistent loss function. The generated adversarial loss function consists of the discriminator's probability estimate of the true sample and the discriminator's probability estimate of the generated sample:

$$\begin{aligned}\min_G \max_D V(D,G) = {} & E_{x\sim P_{\text{data}(x)}}[\log D(x)] \\ & + E_{z\sim P_{z(x)}}[\log(1 - D(G(z)))].\end{aligned} \tag{16}$$

The first step, for this stage, is to use the generator to transform the real image in domain X into a false image in domain Y. Then, use the generator to complete the reconstruction process to obtain the reconstructed image Y_G, preserving all the image's original information. Then, this and the real image Y are passed to the discriminator D_X to determine the authenticity, thereby obtaining a complete one-way GAN. The correlation loss function is

$$\begin{aligned} L_{\text{GAN}}(G, D_Y, X, Y) = {} & E_{y\sim P_{\text{data}(y)}}[\log D_w(y)] \\ & + E_{x\sim P_{\text{data}(x)}}[\log(1 - D_w(f_w(x)))], \\ L_{\text{GAN}}(F, D_X, X, Y) = {} & E_{x\sim P_{\text{data}(x)}}[\log D_w(x)] \\ & + E_{y\sim P_{\text{data}(y)}}[\log(1 - D_w(f_w(y)))]. \end{aligned} \tag{17}$$

The cycle-consistent loss function is introduced into the network to learn the mapping of $G_{X\longrightarrow Y}$ and $F_{Y\longrightarrow X}$ and can convert X to Y and back again successfully, thus avoiding that all images are mapped to the same image in Y:

$$L_{\text{CCL}} = \|\phi(x) - \phi(F(G(x)))\|_2^2 + \|\phi(y) - \phi(G(F(y)))\|_2^2. \tag{18}$$

The CycleGAN [30] loss function is

$$L_{\text{CYC}} = L_{\text{GAN}}(G, D_Y, X, Y) + L_{\text{GAN}}(F, D_X, X, Y) + \gamma L_{\text{CCL}}, \tag{19}$$

where X and Y refer to the two data domains, and x and y are the above data domains within the sample data, G refers to the X to Y mapping function, F refers to the Y to X mapping function, D_X and D_Y refer to the discriminator, and γ is the weight of the cycle-consistent loss.

The least-squares loss method used by CycleGAN [30] penalizes the outlier samples too much, which reduces the diversity of the generated samples. A single loss function does not guarantee that the model will successfully map a single input X_i to the expected output Y_i. This study adds color perception loss function and detail perception loss function based on the original loss function to train the network more optimally on unpaired images. The loss is mainly used to estimate the image's differences after dehazing and minimize the change to the original image. Discriminator G:

$$L_{\text{dpl}}(G_{X\longrightarrow Y}) = E_{y\sim P_{\text{data}(y)}}[D_X(F(y)) - 1]^2 + \frac{1}{2}E_{y\sim P_{\text{data}(y)}}\sqrt{[F(y)-1]^2 - y^2}. \tag{20}$$

Discriminator F:

$$L_{\text{dpl}}(F_{Y\longrightarrow X}) = E_{x\sim P_{\text{data}(x)}}[D_Y(G(x)) - 1]^2 + \frac{1}{2}E_{x\sim P_{\text{data}(x)}}\sqrt{[G(x)-1]^2 - x^2}. \tag{21}$$

We combine the equations to form the detail perception loss.

$$L_{\text{dpl}} = L_{\text{dpl}}\left(F_{X\longrightarrow Y}\right) + L_{\text{dpl}}\left(G_{Y\longrightarrow X}\right). \tag{22}$$

Because the dehazing process must be completed for r, g, b three types of channels to complete the operation, but need to maintain the image after the completion of defogging does not produce large color differences. Therefore, it is necessary to add the color perception L_{cpl} when generating the image:

$$L_{\text{cpl}}(I) = \sum_{w=1}^{W}\sum_{h=1}^{H}\begin{cases} \cos^{-1}\left\{\dfrac{0.5[(r-g)+(r-b)]}{\left[(r-g)^2+(r-b)(g-b)\right]^{1/2}}\right\}, & b \le g, \\ 2\pi - \cos^{-1}\left\{\dfrac{0.5[(r-g)+(r-b)]}{\left[(r-g)^2+(r-b)(g-b)\right]^{1/2}}\right\}, & b > g. \end{cases} \tag{23}$$

W and H are the width and height of the image. The final loss function is

$$L_{\text{defog}} = L_{\text{dpl}} + L_{\text{cpl}}(I) + L_{\text{CYC}}. \tag{24}$$

3. Experiences and Results

In the experimental section, the method will be compared with the results of several advanced methods for hazing, including CycleGAN [30], providing qualitative and quantitative analysis of the experimental results on both synthetic and naturalistic datasets. This experiment completes all training as well as testing procedures on TensorFlow [39]. The NVIDIA Tesla V100 GPU is used for model training, optimized using the Adadelta algorithm [40] with a good adaptive learning rate, a batch size of 1, and a learning rate of $2 * 10^{-4}$.

3.1. Datasets. Experiments were conducted on the I-HAZE [41], O-HAZE [42], and RESIDE [43] datasets. The outdoor dataset of RESIDE [43] contains 8970 clear images and 31,950 foggy images synthesized from clear images. O-HAZE [42] is an outdoor scene database containing 45 pairs of realistic foggy images of outdoor scenes and the corresponding nonfoggy images. I-HAZE [41] contains 35 pairs of real indoor scenes with fog and the corresponding nonfog images. I-HAZE [41] and O-HAZE [42] are common experimental data sets of dehazing scenes taken from controllable scenes made by professional fogging machines, which have similar lighting conditions. This experiment randomly selects 4900 foggy and nonfog images from I-HAZE [41], O-HAZE [42], and RESIDE [43] for training and validation. To facilitate the comparison with existing advanced methods, PSNR and SSIM are used as evaluation metrics in this paper to compare the synthetic target test set containing 400 indoor images and 400 outdoor images. AveGrad and Entropy are selected as evaluation metrics for evaluating the method's performance on natural and realistic images, and experiments are conducted on the RTTS dataset in RESIDE [43]. The photos needed to be resized to 256×256 in size when imported to the network.

3.2. Validation of DefogNet Method. To verify the performance of the generator's improved cross-layer connection structure, we compare the DefogNet model using different generator structures under the same conditions and compare the network using the DefogNet without a cross-layer connection structure. The experiment is set to $300^{*}10^{3}$ iterations. The network model is saved and output after every $5^{*}10^{3}$ training iterations, and the PSNR index evaluation of the model in the RESIDE data set is performed. The results show that the model using the cross-layer connection structure has the highest PSNR evaluation index in the experimental iterative-training process compared with the original model. The comparison results are shown in Figures 4 and 5.

For the neural network based on Gan, the network's convergence speed and stability are important indexes to evaluate its performance. To verify the performance of DefogNet, we train the model processed by Defog-SN to compare with those without this processing and compare the convergence speed of DefogNet's discriminator and CycleGAN's discriminator. In the process of $300 * 10^{3}$ iterations, our method's convergence speed is higher than that of other methods, which also indicates the effectiveness of Defog-SN in improving the network performance. The comparison results are shown in Figure 6.

Several models with different loss functions are trained under the same conditions to verify the multi-loss function's performance. One-hundred images were randomly selected from RESIDE [43] for the experiment. Defognet (NET-4) is compared with different combination loss models in Table 1. The design of the performance verification model is shown in Table 1. The results show that the model using multiple loss fusion has the highest PSNR evaluation index in the $300 * 10^{3}$ iterations training process compared with the model using other loss functions. The comparison results are shown in Figures 7 and 8.

3.3. Results on Synthetic Datasets. For RESIDE [43], I-HAZE [41], and O-HAZE [42] datasets, we choose PSNR and SSIM as the quantitative evaluation indexes for the experimental results. The posttraining evaluation of this paper's algorithm

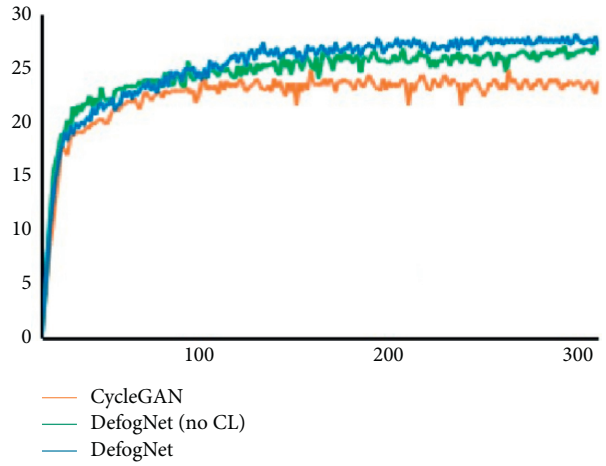


Figure 4: The convergence of PSNR values.

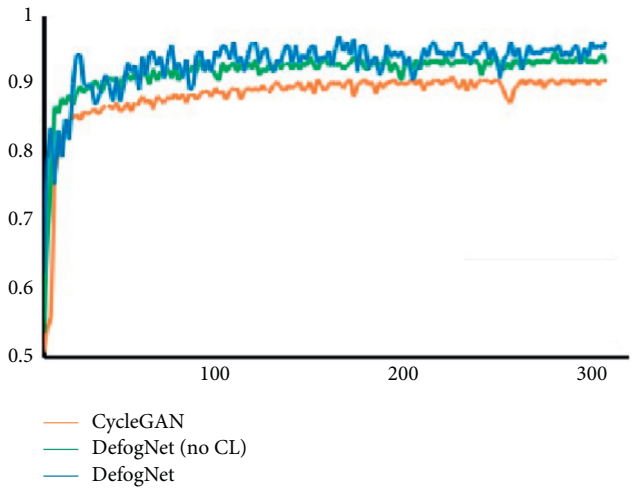


Figure 5: The convergence of SSIM values.

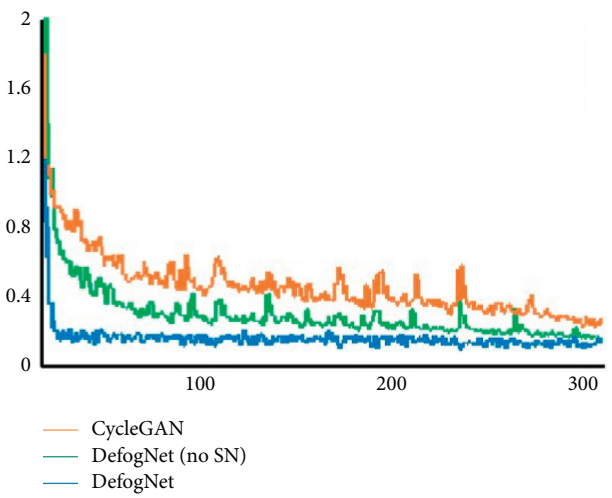


Figure 6: The gradient convergence of discriminator.

Table 1: Neural networks with different loss functions.

Models	Loss function
Net-1	$L_{CYC} + L_{cpl}(I)$
Net-2	$L_{dpl} + L_{CYC}$
Net-3	$L_{dpl} + L_{cpl}(I)$
Net-4	$L_{dpl} + L_{CYC} + L_{cpl}(I)$

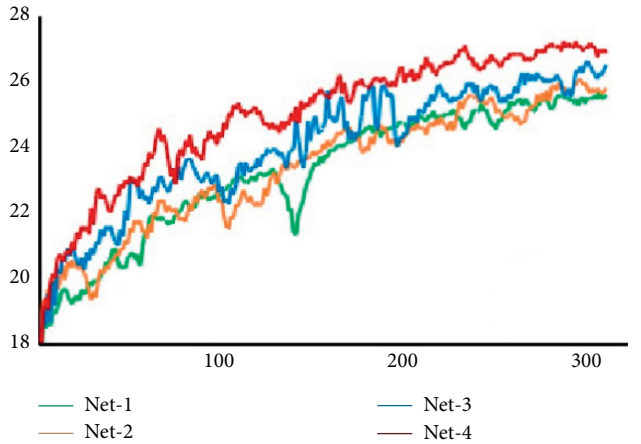


Figure 7: Comparison of PSNR values using different loss function models.

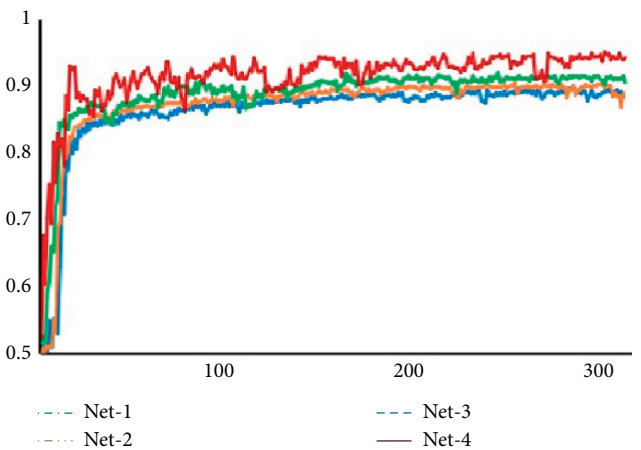
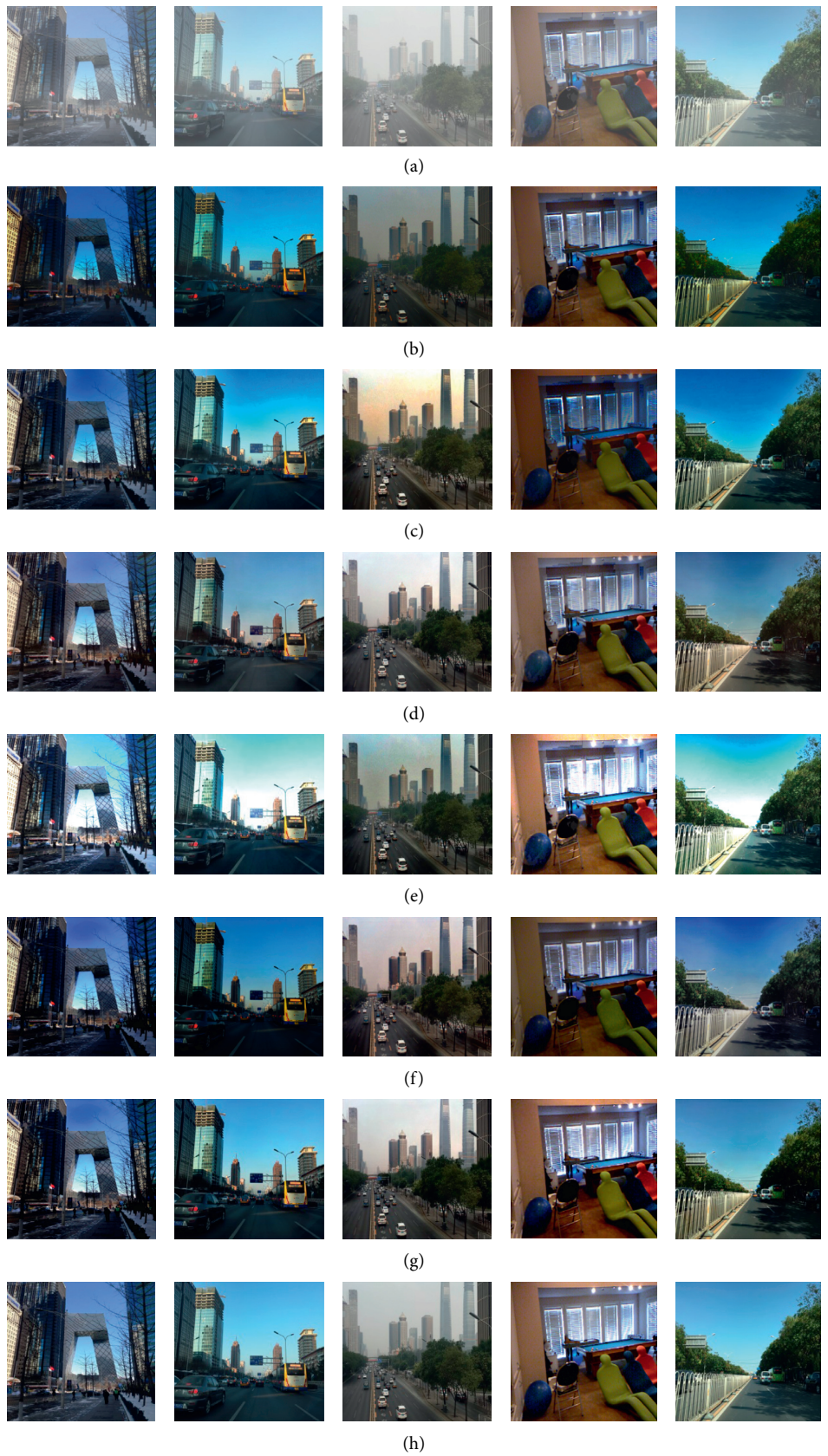


Figure 8: Comparison of SSIM values using different loss function models.

is compared with several advanced single-image dehazing methods, as well as the original CycleGAN effect under the same parameter setting environment, and the results are shown in Table 2 and Figure 9. It can be seen that the dehazing results produced by this algorithm have higher PSNR and SSIM values (Figures 9).

According to the method proposed in this paper, the PSNR and SSIM values of the results are compared with those of CycleGAN and other methods. It can be found that DefogNet's structural optimization method is useful in



(a) (b) (c) (d) (e) (f) (g) (h)

Figure 9: Comparison of results on synthetic datasets: (a) Haze, (b) Cai [23], (c) He [10], (d) Zhang [25], (e) Ren [24], (f) Zhu [30], (g) this study, and (h) GT.

(a)

(b)

(c)

(d)

(e)

(f)

Figure 10: Continued.

(g)

Figure 10: Comparison of results on natural, realistic images. (a) Haze, (b) Cai [23], (c) He [10], (d) Zhang [25], (e) Ren [24], (f) Zhu [30], and (g) Our study.

Table 2: Quantitative comparisons of image dehazing on the I-HAZE, O-HAZE, and RESIDE data sets.

		Cai [23]	He [10]	Zhang [25]	Ren [24]	Zhu [30]	Ours
First image	PSNR	19.13	15.66	24.57	18.22	23.12	29.45
	SSIM	0.79	0.84	0.88	0.85	0.89	0.95
Second image	PSNR	20.86	19.45	25.10	20.43	25.72	27.95
	SSIM	0.81	0.84	0.89	0.82	0.88	0.93
Third image	PSNR	21.96	18.66	23.18	19.77	22.28	28.31
	SSIM	0.85	0.79	0.82	0.86	0.85	0.96
Fourth image	PSNR	23.50	20.07	26.13	22.79	24.33	29.37
	SSIM	0.79	0.80	0.87	0.83	0.86	0.91
Fifth image	PSNR	22.48	21.09	23.64	22.17	23.94	26.95
	SSIM	0.83	0.83	0.86	0.85	0.85	0.94

substantially improving the performance of various aspects of the original CycleGAN. Although He's method [10] can remove some haze, it produces artifacts and color distortion, especially in the sky and light-colored areas. Cai's method [23] and Ren's method [24] need to estimate the transmittance, and due to inaccurate transmittance estimation, the defogging results still contain more artifacts and haze residue. Zhang's method can generate clear images; however, compared with our algorithm's dehazing images, there are more unclear object outlines. The overall image color is darker, and the sky's color in the images is slightly distorted. Zhu's method [30] effectively avoids artifacts but suffers from color shifts and distortions. These methods have more or less color distortion. Our method has achieved high PSNR and SSIM values and sound visual effects. It also indicates that this article's loss function has an improved impact on color constraints. This algorithm avoids estimating the transmittance and atmospheric light values and produces sharper dehazed images than other algorithms. Not only does the method obtain higher quality evaluation scores, but the defogging results also yield sharper image edge details and less noise while avoiding color distortion problems (Table 2).

3.4. Results on Natural Realistic Images. At present, the neural network used for the image dehazing problem is challenging to achieve good results on natural real image data sets. Because CNN tends to have overfitting problems, it is more likely to solve the task for a specific scene or a particular dataset. This paper chose a cross-data set method for experimental analysis in the training and testing stage and selected an everyday, natural, image data set to compare with other methods' dehazing results. Since there is no original image for comparison, the information entropy (Entropy) and average gradient (AveGrad), which reflect the image clarity, are used as the evaluation criteria for the defogging results in this paper. The dehazing results of this method and other methods, such as CycleGAN [30], are presented for comparison. Figure 10 shows the foggy images of three real scenes and the corresponding dehazing results generated by several algorithms.

From the data in Table 3, it follows that the gradient value of the original foggy images is low, most of the edge details of buildings or plants in the images were obscured by the haze, and the images are not clear (i.e., the image information entropy value is low). In comparison, the gradient and entropy values of He's algorithm [10] and Zhang's algorithm [25] are relatively large. Still, the color of the sky part of the image (i.e., the upper part of the image) is more distorted in He's algorithm. Due to the shortcomings of the dark-channel prior algorithm, there is still haze residue and dark color in the dehazed image. Although Cai's algorithm [23] and Ren's algorithm [24] have relatively large Entropy values, there is still haze residue in the image after dehazing, which affects the detailed display of objects in the image. Zhu's algorithm [30] achieves a high-quality score, but it is challenging to ensure precise object edges and details, and the overall color of the image is not natural enough. Compared with other algorithms, the gradient value and information entropy of the algorithm in this paper were larger than other algorithms. The details in the image are

Table 3: Quantitative comparisons of image dehazing on the RTTS dataset from RESIDE.

		Cai [23]	He [10]	Zhang [25]	Ren [24]	Zhu [30]	Ours
First image	AveGrad	5.8524	6.0103	6.5384	5.4105	6.3822	6.6396
	Entropy	7.2023	6.925	7.4511	7.1409	7.2946	7.3975
Second image	AveGrad	6.2126	6.4438	6.7374	5.8781	6.7822	6.9413
	Entropy	7.2384	7.7843	7.8394	7.4974	7.7908	7.8233

retained, and the clarity is higher. The overall color of the dehazing result is more natural, which indicates that this method can effectively solve the overfitting problem for similar data.

4. Conclusion

This paper proposes an effective new defogging method DefogNet. This method is optimized based on the CycleGAN method, completely omits artificially extracting features. and does not require scene prior information. It is a method with a wide range of adaptations. The optimized network architecture and loss function in this study solve the problem of difficult training sample collection encountered when using deep learning methods for dehazing research, greatly reducing the difficulty of obtaining training datasets, making this method more practical and accurate with other deep learning-based dehazing method. This paper adds cross-layer connections in the generator, which optimizes the high and low-level fusion feature extraction capability of the model, effectively avoiding overfitting and improving the generated images' quality. A unique loss function is designed to add detail perception loss and color perception loss to avoid the color differences and reconstruction loss in the image caused by the dehazing operation, effectively improving the restoration of the image's details after dehazing. The Defog-SN algorithm is proposed to improve the structure of the discriminator so that the entire discriminant network satisfies the 1-Lipschitz continuum, thus enhancing the stability of the model and avoiding the problem that the GANs model is prone to collapse. Furthermore, experimental results produced using natural image datasets demonstrate the generality of the present defogging method for images of different scenes.

Data Availability

The data used to support the findings of this work are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (No.61906097) and a project funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD).

References

[1] Y. Chen, D. Li, and J. Q. Zhang, "Complementary color wavelet: a novel tool for the color image/video analysis and processing," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 29, no. 1, pp. 12–27, 2017.

[2] M. Yamakawa and Y. Sugita, "Image enhancement using Retinex and image fusion techniques," *Electronics and Communications in Japan*, vol. 101, no. 8, pp. 52–63, 2018.

[3] A. Faed, E. Chang, M. Saberi, O. K. Hussain, and A. Azadeh, "Intelligent customer complaint handling utilising principal component and data envelopment analysis (PDA)," *Applied Soft Computing*, vol. 47, pp. 614–630, 2016.

[4] D. J. Jobson, Z. Rahman, and G. A. Woodell, "A multiscale retinex for bridging the gap between color images and the human observation of scenes," *IEEE Transactions on Image Processing*, vol. 6, no. 7, pp. 965–976, 1997.

[5] R. T. Tan, "Visibility in bad weather from a single image," in *Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8, IEEE, Anchorage, AK, USA, June 2008.

[6] Y. T. Kim, "Contrast enhancement using brightness preserving bi-histogram equalization," *IEEE Transactions on Consumer Electronics*, vol. 43, no. 1, pp. 1–8, 1997.

[7] K. Zuiderveld, "Contrast limited adaptive histogram equalization," *Graphics Gems*, pp. 474–485, 1994.

[8] J. Kopf, B. Neubert, B. Chen et al., "Deep photo," *ACM Transactions on Graphics*, vol. 27, no. 5, pp. 1–10, 2008.

[9] Y. Y. Schechner, S. G. Narasimhan, and S. K. Nayar, "Instant dehazing of images using polarization," in *Proceedings of the 2001 IEEE computer Society Conference on computer vision and Pattern recognition, CVPR 2001*, IEEE, Kauai, HI, USA, December 2001.

[10] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 12, pp. 2341–2353, 2010.

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

[12] C. O. Ancuti, C. Ancuti, and C. Hermans, "A fast semi-inverse approach to detect and remove the haze from a single image," in *Proceedings of the Asian Conference on Computer Vision*, pp. 501–514, Springer, Berlin, Heidelberg, December 2010.

[13] D. Berman and S. Avidan, "Non-local Image dehazing," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1674–1682, IEEE, Las Vegas, NV, USA, June 2016.

[14] E. J. McCartney, "Optics of the atmosphere: scattering by molecules and particles," *IEEE Journal of Quantum Electronics*, vol. 14, no. 9, pp. 698-699, 1976.

[15] J. P. Oakley and B. L. Satherley, "Improving image quality in poor visibility conditions using a physical model for contrast degradation," *IEEE Transactions on Image Processing*, vol. 7, no. 2, pp. 167–179, 1998.

[16] Y. Y. Schechner, S. G. Narasimhan, and S. K. Nayar, "Polarization-based vision through haze," *Applied Optics*, vol. 42, no. 3, pp. 511–525, 2003.

[17] S. G. Narasimhan and S. K. Nayar, "Interactive (de) weathering of an image using physical models," in *Proceedings of the Methods in Computer IEEE Workshop on Color and Photometric Vision*, IEEE, Nice, France, October 2003.

[18] J. P. Tarel and N. Hautiere, "Fast visibility restoration from a single color or gray level image," in *Proceedings of the 2009 IEEE 12th International Conference on Computer Vision*, pp. 2201–2208, IEEE, Kyoto, Japan, September 2009.

[19] R. Fattal, "Single image dehazing," *ACM Transactions on Graphics*, vol. 27, no. 3, pp. 1–9, 2008.

[20] S. C. Pei and T. Y. Lee, "Nighttime haze removal using color transfer pre-processing and dark channel prior," in *Proceedings of the 2012 19th IEEE International Conference on Image Processing*, pp. 957–960, IEEE, Orlando, FL, USA, September 2012.

[21] Y. Shuai, R. Liu, and W. He, "Image haze removal of wiener filtering based on dark channel prior," in *Proceedings of the 2012 Eighth International Conference on Computational Intelligence and Security*, pp. 318–322, IEEE, Guangzhou, China, November 2012.

[22] D. Fan, X. Guo, and X. Lu, "Image defogging algorithm based on sparse representation," *Complexity*, vol. 2020, Article ID 6835367, , 2020.

[23] B. Cai, X. Xu, K. Jia, C. Qing, and D. Tao, "DehazeNet: an end-to-end system for single image haze removal," *IEEE Transactions on Image Processing*, vol. 25, no. 11, pp. 5187–5198, 2016.

[24] W. Ren, S. Liu, and H. Zhang, "Single image dehazing via multiscale convolutional neural networks," in *Proceedings of the European Conference on Computer Vision*, pp. 154–169, Springer, Amsterdam, The Netherlands, October 2016.

[25] X. Zhang, H. Dong, and Z. Hu, "Gated fusion network for joint image deblurring and super-resolution," 2018, https://arxiv.org/abs/1807.10806.

[26] I. Goodfellow, J. Pouget-Abadie, and M. Mirza, "Generative adversarial nets," in *Proceedings of the Advances in Neural Information Processing Systems*, pp. 2672–2680, Montreal, Canada, December 2014.

[27] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," 2015, https://arxiv.org/abs/1511.06434.

[28] M. Mirza and S. Osindero, "Conditional generative adversarial nets," 2014, https://arxiv.org/abs/1411.1784.

[29] P. Isola, J. Y. Zhu, and T. Zhou, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1125–1134, IEEE, Honolulu, HI, USA, July 2017.

[30] J. Y. Zhu, T. Park, and P. Isola, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2223–2232, IEEE, Venice, Italy, October 2017.

[31] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, https://arxiv.org/abs/1409.1556.

[32] G. Klambauer, T. Unterthiner, and A. Mayr, "Self-normalizing neural networks," in *Proceedings of the Advances in neural Information Processing Systems*, pp. 971–980, IEEE, Long Beach, CA, USA, December 2017.

[33] T. Miyato, T. Kataoka, and M. Koyama, "Spectral normalization for generative adversarial networks," 2018, https://arxiv.org/abs/1802.05957.

[34] S. C. Martin Arjovsky and L. Bottou, "Wasserstein generative adversarial networks," in *Proceedings of the 34 th International Conference on Machine Learning*, Sydney, Australia, August 2017.

[35] A. L. Maas, A. Y. Hannun, and A. Y. Ng, "Rectifier nonlinearities improve neural network acoustic models," in *Proceedings of the International Conference on Machine Learning*, Atlanta, GA, USA, June 2013.

[36] Y. Yoshida and T. Miyato, "Spectral norm regularization for improving the generalizability of deep learning," in *Proceedings of the International Conference on Neural Information Processing Systems*, pp. 1539–1550, Montreal, Canada, December 2018.

[37] H. Sedghi, V. Gupta, and P. M. Long, "The singular values of convolutional layers," 2018, https://arxiv.org/abs/1805.10408.

[38] I. Gulrajani, F. Ahmed, and M. Arjovsky, "Improved training of wasserstein gans," in *Proceedings of the Advances in neural Information Processing Systems*, pp. 5767–5777, Long Beach, CA, USA, December 2017.

[39] M. Abadi, P. Barham, and J. Chen, "Tensorflow: a system for large-scale machine learning," in *Proceedings of the 12th {USENIX} symposium on operating systems design and implementation ({OSDI} 16)*, pp. 265–283, Carlsbad, CA, USA, July 2016.

[40] M. D. Zeiler, "Adadelta: an adaptive learning rate method," 2012, https://arxiv.org/abs/1212.5701.

[41] C. O. Ancuti, C. Ancuti, R. Timofte, and C. De Vleeschouwer, "I-HAZE: a dehazing benchmark with real hazy and haze-free indoor images," 2018, https://arxiv.org/abs/1804.05091.

[42] C. O. Ancuti, C. Ancuti, and R. Timofte, "O-haze: a dehazing benchmark with real hazy and haze-free outdoor images," in *in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 754–762, IEEE, Salt Lake City, UT, USA, June 2018.

[43] B. Li, W. Ren, and D. Fu, "Benchmarking single-image dehazing and beyond," *IEEE Transactions on Image Processing*, vol. 28, no. 1, pp. 492–505, 2018.