IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received June 12, 2021, accepted July 25, 2021, date of publication July 27, 2021, date of current version August 5, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3100604

Haze Relevant Feature Attention Network for Single Image Dehazing

XIN JIANG[1,2], LU LU[1], MING ZHU[1,2], ZHICHENG HAO[1], AND WEN GAO[1]
[1]Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun 130033, China
[2]University of Chinese Academy of Sciences, Beijing 100049, China

Corresponding author: Xin Jiang (xinjiang@zju.ed.cn)

ABSTRACT Single image dehazing methods based on deep learning technique have made great achievements in recent years. However, some methods recover haze-free images by estimating the so-called transmission map and global atmospheric light, which are strictly limited to the simplified atmospheric scattering model and do not give full play to the advantages of deep learning to fit complex functions. Other methods require pairs of training data, whereas in practice pairs of hazy and corresponding haze-free images are difficult to obtain. To address these problems, inspired by cycle generative adversarial model, we have developed an end-to-end haze relevant feature attention network for single image dehazing, which does not require paired training images. Specifically, we make explicit use of haze relevant feature by embedding an attention module into a novel dehazing generator that combines an encoder-decoder structure with dense blocks. The constructed network adopts a novel strategy which derives attention maps from several hand-designed priors, such as dark channel, color attenuation, maximum contrast and so on. Since haze is usually unevenly distributed across an image, the attention maps could serve as a guidance of the amount of haze at image pixels. Meanwhile, dense blocks can maximize information flow along features from different levels. Furthermore, color loss is proposed to avoid color distortion and generate visually better haze-free images. Extensive experiments demonstrate that the proposed method achieves significant improvements over the state-of-the-art methods.

INDEX TERMS Single image dehazing, cycle generative adversarial networks, haze relevant feature, attention module, dense block.

I. INTRODUCTION

In most cases, computer vision systems are very sensitive to changes of environment [1]. However, due to rapid development of industry, the source of air pollution has not been effectively controlled, thereby leading to frequent occurrence of hazy weather. According to the principle of imaging, there will be a large number of particles floating in the air under the condition of hazy weather, which will absorb and scatter light before imaging sensor. At the same time, the light received by imaging sensor is mixed with part of atmospheric scattered light. Under these conditions, image quality is seriously reduced with some degradation phenomena such as desaturation, color distortion, loss of clarity and so on.

The degradation of image quality in hazy weather not only reduces viewing value, but also seriously affects the effectiveness of various computer vision tasks. For example, images obtained by monitoring systems are blurred in hazy weather, which directly affects subsequent abnormal behavior detection, vehicle detection and pedestrian identification tasks. Therefore, in order to reduce degradation impact of hazy weather, study of image haze removal algorithm has become a practical and urgent problem to be solved. Effective image dehazing algorithm can be applied in the image preprocessing stage of computer vision system to improve clarity and quality of the image, and facilitate subsequent feature extraction or other applications.

Image dehazing refers to methods of image enhancement or restoration to minimize coverage and interference of haze on various information in the image [2]. The goal is to highlight details and restore real scene color of images, so that it is easy for human eyes to identify images and computers to extract features of images [3]. Single image dehazing has received a large number of attentions in recent years [4].

The associate editor coordinating the review of this manuscript and approving it for publication was Shenghong Li.

However, some existing image dehazing methods including both learning-based and model-based are heavily limited to atmospheric scattering model by estimating the so-called transmission map and global atmospheric light as medium, which will then restore corresponding haze-free images based on a simplified model. This does not take full advantage of powerful ability of deep neural networks to fit advanced complex functions [5], and inaccuracies in the estimation of atmospheric scattering parameters results in poor dehazed images. Furthermore, image dehazing without paired hazy and corresponding haze-free images is of immense importance, since it is almost impossible to acquire a large number of paired images in reality.

Towards these problems, we propose an end-to-end haze relevant feature attention network for single image dehazing. It does not rely on the traditional physical scattering model, and instead it adopts the image-to-image translation approach without pairs of hazy and corresponding ground truth images for network training.

Our main contributions can be summarized as follows:

- We propose an end-to-end haze relevant feature attention network for single image dehazing, which can directly recover haze-free images without the estimation of atmospheric scattering model parameters. It integrates effective haze relevant features as additional priors into the framework for better image dehazing and does not require paired training images.
- A densely connected encoder-decoder structure is derived as the generator, in which dense blocks are implemented as the backbone to enhance the transmission ability and reusability of features.
- A novel attention module is proposed to provide useful information about haze regions and haze density in a per-pixel manner, which serves as a guidance for the subsequent network.
- Color loss is proposed to avoid color distortion and generate clearer haze-free images associated with adversarial loss, cycle consistency loss, identity loss and perceptual loss.
- Through extensive experiments, our proposed method performs favorably against the state-of-the-art dehazing algorithms on both synthetic and real-world images.

The rest of this paper is organized as follows: In Section 2, we present a brief overview of the related work. The proposed method is given in Section 3. Experimental results are presented and discussed in Section 4. Finally, the conclusions of this paper are provided in Section 5.

II. RELATED WORK

Single image dehazing and generative adversarial network are the two topics that are most relevant to this paper. We will give a brief overview of the related works in this section.

A. SINGLE IMAGE DEHAZING

At present, there are roughly two kinds of research methods in terms of single image dehazing problems, and they have achieved certain results, respectively. One is model-based dehazing method which is based on classical atmospheric scattering model. This kind of method utilizes prior knowledge and physical model to remove haze in hazy image. The other is learning-based method which relies on powerful deep neural networks. This kind of methods learns characteristics of haze through feature extraction, so as to achieve better dehazing effect.

Model-based dehazing method heavily depends on atmospheric scattering model, which can be formulated as

$$I(x) = J(x)t(x) + A\,(1 - t(x))\,, \tag{1}$$

where $I(x)$ is the hazy image, $J(x)$ is the haze-free image or the scene radiance, $t(x)$ is the medium transmission map, and A is the global atmospheric light on each x pixel coordinates. $t(x)$ can be formulated as:

$$t(x) = e^{-\beta d(x)}, \tag{2}$$

where $d(x)$ is the depth of scene point and β is defined as the scattering coefficient of atmosphere. Based on atmospheric scattering model, Tan [6] proposed maximizing local contrast methods for image dehazing, where the color of output images was too saturated, and a lot of details would be lost after haze removing. He *et al.* [7] proposed dark channel prior (DCP) based dehazing method with good dehazing effect and simple implementation, but it is susceptible to influence of bright areas such as sky regions and has problem of high time and space complexity. Color attenuation prior (CAP) theory [8] established a linear model of scene depth and difference between brightness and saturation of scenic spots to obtain scene transmission map. Non-local prior [9] observes that pixels in a given cluster are often spread over the entire image plane and are located at different distances from sensors, which can be translated to different transmission coefficients. Based on hue disparity between original image and its semi-inverse, Ancuti *et al.* [10] proposed a fast semi-inverse approach to detect and remove haze from a single image.

In recent years, there has been a large number of image dehazing methods based on deep neural learning. Cai *et al.* [11] proposed an end-to-end single image dehazing network called Dehaze-Net, which integrates artificial methods into each layer of network and achieves good haze removal effect. However, separate estimation of medium transmission rate and atmospheric light value affects haze removal performance to a certain extent. Chen *et al.* [12] built a network model based on the idea of radial basis function to achieve image dehazing, which can recover more visible edges and details of image. The All-in-One dehazing network (AOD-Net) [13] was proposed with an end-to-end trainable model without relying on any separate and intermediate parameter estimation step. Qu *et al.* [14] implemented two enhancing blocks based on the receptive field model, which reinforced dehazing effect in both color and details. Chen *et al.* [15] adopted the latest smoothed dilation

technique in an end-to-end gated context aggregation network to directly restore haze-free image.

B. GENERATIVE ADVERSARIAL NETWORKS

Generative Adversarial Network (GAN) was proposed by Goodfellow in 2014 [16], which is a new method used to train generator model. GAN creatively combines the idea of confrontation in game theory successfully with deep learning. GAN treats training process as a game between two independent networks: generator network and discriminator network, which attempts to classify samples as coming from true distribution or model generated distribution as accurately as possible. Whenever discriminator network notices a difference between two distributions, generator network slightly adjusts its parameters to make the difference disappear, until finally generator network accurately reproduces real data distribution and discriminator network fails to distinguish between true and false. In recent years, a variety of improved GAN has been widely used in the field of image processing, covering almost all the traditional image processing fields, as well as some new applications, such as image editing [17], image translation [18], style transfer [19], and so on.

Combined with the idea of GAN and duality learning, an unsupervised learning method called cycle generative adversarial network (CycleGAN) was proposed by Zhu *et al.* [20]. The basic idea is to use generator and discriminator to complete the conversion between different image domains, and add cyclic consistency to preserve content information of image [21]. Cycle-Dehaze [22] enhances CycleGAN formulation by combining cycle-consistency and perceptual losses to improve the quality of textural information recovery for single image dehazing. Liu *et al.* [23] proposed a two stage mapping strategy in each transformation path to enhance the effectiveness of haze removal.

III. PROPOSED METHOD

In this section, we introduce the architecture of the proposed haze relevant feature attention network in detail. First, we present an overview of the architecture. Then we provide details of generator module, haze relevant feature attention module and discriminator module. Finally, we describe loss functions that are utilized to train the networks.

A. OVERVIEW OF THE PROPOSED ARCHITECTURE

As shown in Fig.1, the network consists of two generators G_{haze}, G_{dehaze} and two discriminators D_{clear}, D_{haze}. G_{haze} translates haze-free images from clear domain to hazy domain, while G_{dehaze} translates hazy images from hazy domain to clear domain. D_{clear} and D_{haze} classify whether the reconstructed images are clear or hazy. Color loss is proposed to avoid color distortion and generate clearer haze-free outputs together with adversarial loss, cycle consistency loss, identity loss and perceptual loss.

B. GENERATOR

The goal of generator G_{dehaze} is to directly restore a clear image from a hazy image, without estimating atmospheric scattering parameters. To achieve this goal, the generator is designed to not only preserve content and detailed information of origin hazy input image, but also remove haze as much as possible.

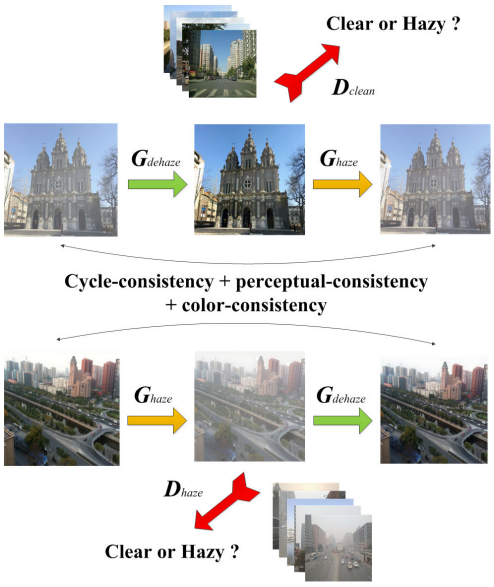


FIGURE 1. The architecture of the proposed haze relevant feature attention network.

In this paper, we propose a densely connected encoder-decoder structure as the generator, where dense blocks are implemented as backbone. Dense block was proposed in [24] which connects each layer to every other layer in a feed-forward mode. It has the potential to enhance transmission ability and reusability of features, while alleviate vanishing-gradient problem. As shown in Fig.2, firstly we concatenate original hazy image and feature maps output by attention block in the channel direction. The attention block derives feature maps from several hand-designed priors to make the following network aware of hazy regions, which will be discussed in detail in the next subsection. The encoder is mainly constructed with three dense blocks, which include a series of bath normalization, ReLU and convolution operations. To make full use of pre-trained weights, we adopt the first three dense blocks with their corresponding down-sampling transition layers from Dense-net121 [24], in which the first dense block contains 6 densely-connected layers, the second dense block contains 12 densely-connected layers and the third dense block contains 24 densely-connected layers. Each dense block is followed by a transition down layer. After that, feature map size of encoding part is 1/32 of input size. Similarly, on the decoder side, three dense blocks containing 16, 24 and 12 densely-connected layers are employed with transition up layers. The size of feature maps is gradually recovered to the input resolution. Furthermore, skip connections are employed with same dimension feature maps to recover multi-level details, and finally features from encoder side are concatenated to decoder side. We present a detailed description of the architecture in Table 1. Note that the growth rate is set to be

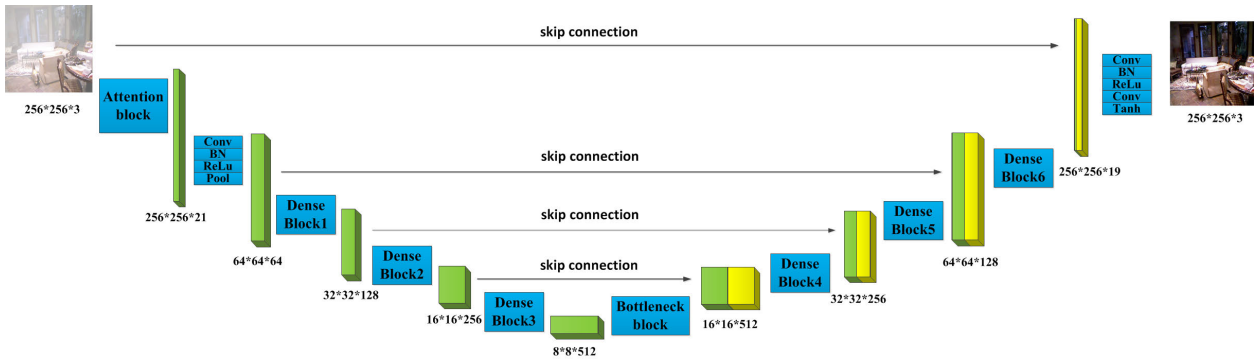


FIGURE 2. The densely connected encoder-decoder structure as the generator.

TABLE 1. The detailed description of the generator.

Layers	Output Size	Operation
Input	256 × 256	
Attention block	256 × 256	refer to haze relevant feature attention module
Convolution	128 × 128	7*7 conv, stride 2
Normalization	128 × 128	BatchNorm 128
ReLU	128 × 128	Relu
Pooling	64 × 64	3*3 max pool, stride 2
Dense Block1	64 × 64	$\begin{bmatrix} 1\times1 & \text{conv} \\ 3\times3 & \text{conv} \end{bmatrix} \times 6$
Transition Layer1	64 × 64	1*1 conv
	32 × 32	2*2 average pool, stride 2
Dense Block2	32 × 32	$\begin{bmatrix} 1\times1 & \text{conv} \\ 3\times3 & \text{conv} \end{bmatrix} \times 12$
Transition Layer2	32 × 32	1*1 conv
	16 × 16	2*2 average pool, stride 2
Dense Block3	16 × 16	$\begin{bmatrix} 1\times1 & \text{conv} \\ 3\times3 & \text{conv} \end{bmatrix} \times 24$
Transition Layer3	16 × 16	1*1 conv
	8 × 8	2*2 average pool, stride 2
Bottleneck Block	8 × 8	$\begin{bmatrix} 1\times1 & \text{conv} \\ 3\times3 & \text{conv} \end{bmatrix} \times 1$
Transition Layer4	8 × 8	1*1 convtranspose
	16 × 16	2*2 upsample nearest
Dense Block4	16 × 16	$\begin{bmatrix} 1\times1 & \text{conv} \\ 3\times3 & \text{conv} \end{bmatrix} \times 16$
Transition Layer5	16 × 16	1*1 convtranspose
	32 × 32	2*2 upsample nearest
Dense Block5	32 × 32	$\begin{bmatrix} 1\times1 & \text{conv} \\ 3\times3 & \text{conv} \end{bmatrix} \times 24$
Transition Layer6	32 × 32	1*1 convtranspose
	64 × 64	2*2 upsample nearest
Dense Block6	64 × 64	$\begin{bmatrix} 1\times1 & \text{conv} \\ 3\times3 & \text{conv} \end{bmatrix} \times 12$
Transition Layer7	64 × 64	1*1 convtranspose
	128 × 128	2*2 upsample nearest
Transition Layer8	128 × 128	1*1 convtranspose
	256 × 256	2*2 upsample nearest
Convolution	256 × 256	3*3 conv, stride 1
Normalization	256 × 256	BatchNorm 256
ReLU	256 × 256	Relu
Convolution	256 × 256	3*3 conv, stride 1
Tanh	256 × 256	Tanh

$k = 32$, each "conv" layer shown in the dense block contains sequence BN-ReLU-Conv, and each "convtranspose" layer includes sequence BN-ReLU-Convtranspose.

Since the role of generator G_{haze} is to generate a hazy image from a clear image, the attention block is not employed in G_{haze}, while the other part is basically the same as generator G_{dehaze}. Specifically, the numbers of input channel for the convolution operation on the fourth row of Table 1 are 18 for G_{dehaze}, and 3 for G_{haze} since original hazy image is convolved directly from clear domain to hazy domain, while the other parameters remain consistent.

C. HAZE RELEVANT FEATURE ATTENTION MODULE

The domain knowledge of hazy and haze-free images are asymmetrical, therefore it is hard to directly utilize CycleGAN framework to solve the dehazing problem. Specifically, hazy images contain hazy information and background, while haze-free images only contain background. In this network, we integrate some haze relevant features as basic priors into design of the generator, since hazy regions and clear regions differ from each other [25]. We extract several significant features that are high correlated to the properties of hazy images, detailed as follows.

1) COLOR ATTENUATION FEATURE

The color attenuation prior observes that the concentration of haze is positively correlated with difference between brightness and saturation, since hazy regions in the image are characterized with high brightness and low saturation. According to [8], the simple and powerful prior can be expressed as

$$h(x) = \theta_0 + \theta_1 v(x) + \theta_2 s(x), \tag{3}$$

where x is position within input hazy image, h denotes concentration of haze, v is brightness component and s is saturation component. $\theta_0 = 0.121779$, $\theta_1 = 0.959710$ and $\theta_2 = -0.780245$ are linear coefficients where the best results are obtained after 120 million scene points training [8].

Fig.3 investigates the color attenuation feature of input hazy image. As expected, denser haze results in higher brightness, lower saturation and larger difference between brightness and saturation. We also observe that brightness and saturation are important pixel characters to estimate the concentration of haze. Based on these, we utilize brightness, saturation and color attenuation feature to build the bride

(a) Input (b) Brightness (c) Saturation (d) Color attenuation

FIGURE 3. The color attenuation feature of hazy images.

(a) Input (b) 3*3 patch (c) 7*7 patch (d) 11*11 patch (e) 15*15 patch

FIGURE 4. The multi-scale dark channel feature of hazy images.

between hazy image and its corresponding concentration map.

2) MULTI-SCALE DARK CHANNEL FEATURE

The dark channel prior is based on the following key observation that most local patches in haze-free images contain some pixels that have very low intensities in at least one color channel [7]. In mathematica, the observation can be expressed as

$$J^{dark}(x) = \min_{c\in\{r,g,b\}} \left(\min_{y\in\Omega(x)} \left(J^c(y)\right)\right), \tag{4}$$

where J^c is a color channel of the input image and $\Omega(x)$ is a local patch centered at x. The value of patch size has an immense impact on dark channel features. If the value is too small, the dark channel will not tend to zero, while a large value will result in over-dehazing. Therefore, we take multi-scale dark channel as priors, in which patch sizes are set to be 3, 7, 11 and 15.

Fig.4 depicts the multi-scale dark channel feature of input hazy image. Since the appearance of haze in an image breaks the statistical law that dark channel tends to zeros, we could regard dark channel as haze-related feature.

(a) Input (b) 3*3 patch (c) 7*7 patch (d) 11*11 patch (e) 15*15 patch

FIGURE 5. The multi-scale maximum contrast feature of hazy images.

3) MULTI-SCALE MAXIMUM CONTRAST FEATURE

Haze significantly reduces the contrast of image, the higher the haze concentration, the lower the image contrast [26]. As such, we take maximum contrast as another haze-relevant feature, which can be derived as

$$J^{con}(x) = 1 - \frac{\min\limits_{c\in\{r,g,b\}}\left(\min\limits_{y\in\Omega(x)}(J^c(y))\right)}{\max\limits_{c\in\{r,g,b\}}\left(\max\limits_{y\in\Omega(x)}(J^c(y))\right)}, \tag{5}$$

where J^c is a color channel of the input image and $\Omega(x)$ is a local patch centered at x. We also take multi-scale maximum contrast feature where patch sizes are set to be 3, 7, 11 and 15.

As shown in Fig.5, the correlation between haze concentration and maximum contrast feature is obvious, which then will help remove haze from a hazy image.

4) ATTENTION MODULE

Based on the observation that haze is usually unevenly distributed in an image, we propose a novel attention module which essentially provides useful information about haze regions and haze density in a per-pixel manner. The central idea is to learn the attention maps based on different haze-relevant features in order to provide guidance for the network, with which the subsequent network can play to its powerful learning ability and autonomously learn appropriate weights for each feature.

As shown in Fig.6, for multi-scale dark channel feature and multi-scale maximum contrast feature, firstly we concatenate multi-scale feature maps as input, then we derive attention map by pixel-wise multiplying original hazy image with adaptive learning weights which are obtained with sequence Conv-BN-ReLu-Conv-Sigmoid operation. From this, the network could pay more attention to haze regions while retain important characteristics of haze-free regions.

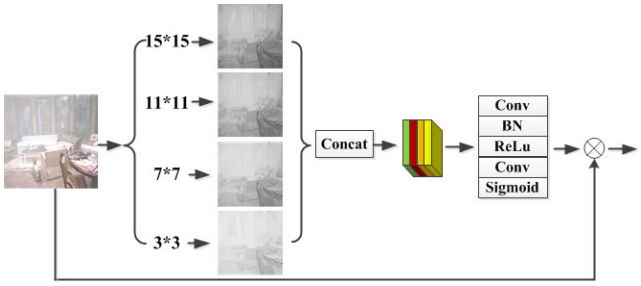


FIGURE 6. Attention module for multi-scale dark channel and maximum contrast feature.

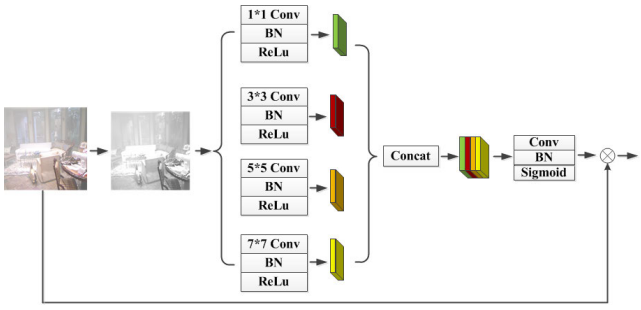


FIGURE 7. Attention module for brightness, saturation and color attenuation feature.

For brightness, saturation and color attenuation feature, the attention module is depicted in Fig.7. Firstly, 1×1, 3×3, 5×5 and 7×7 convolutions are implemented on input feature which is derived from origin hazy image. Convolutions of different scales can provide different receptive fields, which preserve the details of features on various scales [27]. Next, different feature maps are concatenated together before sequence Conv-BN-Sigmoid. Finally, pixel-wise multiplication is implemented to obtain attention map, which will assist the network to pay more attention to important regions.

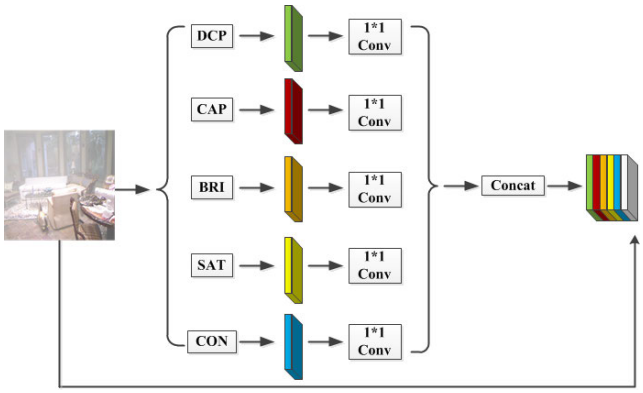


FIGURE 8. The overall attention module.

Finally, each feature map is passed through a 1×1 convolutional layer before concatenation, as described in Fig.8.

The attention maps of different haze relevant features are depicted in Fig.9, which provide some references to the subsequent network about haze distribution of original cropped hazy images.

D. DISCRIMINATOR

The goal of the discriminator is to distinguish whether input image is hazy or clear. Ordinary discriminator outputs a true or false vector to represent the evaluation of whole image, while the discriminator is not suitable for high resolution, high detail preservation image domain. To solve this problem, we employ the same PatchGAN architecture as used in original CycleGAN paper [20]. PatchGAN is a full convolution network, every element of output matrix corresponds to a large receptive field of original image. The final result is an average of all the values in the matrix, which fully takes into account the effects of different parts in the image. During training, the network is able to pay more attention to the details of image. Note that the discriminator D_{clear} and the discriminator D_{haze} share the same architecture.

E. LOSS FUNCTION

The objective loss function contains adversarial loss, cycle consistency loss, perceptual loss, identity loss and color loss. Then we describe these losses in detail.

1) ADVERSARIAL LOSS

Adversarial loss is applied to match the distribution of generated images to data distribution of target domain [20]. For the mapping function $G_{dehaze} : X \rightarrow Y$ and its discriminator D_{clear}, we express dehaze adversarial loss as

$$\mathcal{L}_{Adv\_dehaze} = \mathbb{E}[\log D_{clear}(y)] + \mathbb{E}[\log(1 - D_{clear}(G_{dehaze}(x)))], \quad (6)$$

where $x \in X$ and $y \in Y$ denote the hazy and clear image datasets, respectively. Similarly, for the mapping function $G_{haze} : Y \rightarrow X$ and its discriminator D_{haze}, the haze adversarial loss can be defined as

$$\mathcal{L}_{Adv\_haze} = \mathbb{E}[\log D_{haze}(x)] + \mathbb{E}[\log(1 - D_{haze}(G_{haze}(y)))]. \quad (7)$$

2) CYCLE CONSISTENCY LOSS

Cycle consistency loss is also taken into account to minimize the objective between original image and the corresponding cyclic image [20], which can be expressed as

$$\mathcal{L}_{cyc} = \mathbb{E}[||G_{haze}(G_{dehaze}(x)) - x||_1] + \mathbb{E}[||G_{dehaze}(G_{haze}(y)) - y||_1]. \quad (8)$$

3) PERCEPTUAL LOSS

However, adversarial loss and cycle consistency loss are not able to recover all textural information of heavily-corrupted hazy images. Therefore, perceptual loss is also taken into consideration and the goal of this loss is to compare images in a feature space rather than in a pixel space. To measure the perceptual similarity in feature space, we focus on feature

(a) Input (b) Bright (c) Saturation (d) Color attenuation (e) Dark channel (f) Maximum contrast

FIGURE 9. Attention maps of different haze relevant features.

maps extracted from the 2*nd* and 5*th* pooling layers of a pre-trained VGG16 model [22], which can be expressed as

$$\mathcal{L}_{per} = ||\phi(x) - \phi(G_{haze}(G_{dehaze}(x)))||_2^2 + ||\phi(y) - \phi(G_{dehaze}(G_{haze}(y)))||_2^2, \quad (9)$$

where ϕ is the VGG16 feature extractor.

4) COLOR LOSS

In the process of dehazing, it is necessary to recover color information of original image under hazy scene. However, color distortion is of common occurrence in reality. To solve this problem, inspired by cycle consistency and perceptual consistency, we propose a color consistency achieved by color loss to preserve color features during dehazing. The color loss compares original input image with reconstructed cyclic image at hue and saturation spaces, where hue defines wavelength of color and saturation represents depth of color form the Hue-Saturation-Intensity model. The color loss can be formulated as

$$\begin{aligned}\mathcal{L}_{color} = {} & \lambda_{hue}\mathbb{E}[||\zeta(x) - \zeta(G_{haze}(G_{dehaze}(x)))||_1] \\ & + \lambda_{hue}\mathbb{E}[||\zeta(y) - \zeta(G_{dehaze}(G_{haze}(y)))||_1] \\ & + \lambda_{sar}\mathbb{E}[||\omega(x) - \omega(G_{haze}(G_{dehaze}(x)))||_1] \\ & + \lambda_{sar}\mathbb{E}[||\omega(y) - \omega(G_{dehaze}(G_{haze}(y)))||_1], \quad (10)\end{aligned}$$

where ζ and ω are hue and saturation calculators, λ_{hue} and λ_{sar} are weight parameters.

5) IDENTITY LOSS

In order to avoid hazy image being much enhanced or destroyed by the proposed network, identity loss is introduced to regularize the generator to be near an identity mapping when hazy images are given as the input of G_{haze} and clear images are provided as the input of G_{dehaze} [22]. We express identity loss as

$$\mathcal{L}_{iden} = \mathbb{E}[||G_{haze}(x) - x||_1] + \mathbb{E}[||G_{dehaze}(y) - y||_1]. \quad (11)$$

6) OVERALL LOSS FUNCTION

The overall loss function for training can be defined as

$$\mathcal{L}_{all} = \mathcal{L}_{Adv\_dehaze} + \mathcal{L}_{Adv\_haze} + \lambda_{cyc}\mathcal{L}_{cyc} + \lambda_{per}\mathcal{L}_{per} + \lambda_{color}\mathcal{L}_{color} + \lambda_{iden}\mathcal{L}_{iden}, \quad (12)$$

where λ_{cyc}, λ_{per}, λ_{color} and λ_{iden} are trade-off weights.

IV. EXPERIMENTAL RESULTS

In this section, we evaluate our proposed haze relevant feature attention network on both synthetic datasets and real-world images. We compare our proposed methods against the following state-of-the-art approaches: DCP [7], CAP [8], AODNet [13], EPDN [14], GCANet [15], FFA-Net [28], CycleGAN [20] and Cycle-Dehaze [22]. Furthermore, ablation studies are conducted to demonstrate the effectiveness of our densely connected generator, haze relevant feature attention module and proposed color loss function.

A. DATASETS

Realistic Single Image Dehazing (RESIDE) is a large-scale benchmark consisting of both synthetic and real-world hazy images for fairly evaluating and comparing single image dehazing algorithms [29]. For training, we randomly selects 2000 clear images and 1000 hazy images from outdoor training set (OTS), and 1000 hazy images from real-world task-driven testing set (RTTS), since paired training images are not required in out network. For testing on synthetic datasets, 400 outdoor synthetic hazy images are chosen from synthetic objective testing set (SOTS). For testing on real-world images, 10 realistic hazy outdoor images are selected from hybrid subjective testing set (HSTS) and 390 real-world hazy images are provided by RTTS. We also evaluate the proposed method on NH-Haze dataset [30], which is a novel dataset consisting of 55 pairs of real haze free and nonhomogeneous hazy images that have been recorded outdoor. It is worth noting that the NH-Haze dataset has been employed by the IEEE CVPR 2020 NTIRE workshop associated challenge on image dehazing [31].

B. IMPLEMENTATION DETAILS

In our architecture, all the images are resized to 256×256 and PyTorch framework is utilized with NVIDIA GEFORCE GTX 1080 Ti GPU during training and testing phases. We implement ADAM optimizer with a batch size of 1 to train the network, 200 epochs are performed in order to ensure convergence. The initial learning rate is set to be 0.0001 for the former 100 epochs, with linear decay to 0 over the next 100 epochs. The trade-off weight parameters are empirically set to be $\lambda_{cyc} = 10$, $\lambda_{per} = 6$, $\lambda_{color} = 2$, $\lambda_{iden} = 5$, $\lambda_{hue} = 2$ and $\lambda_{sar} = 3$ during training.

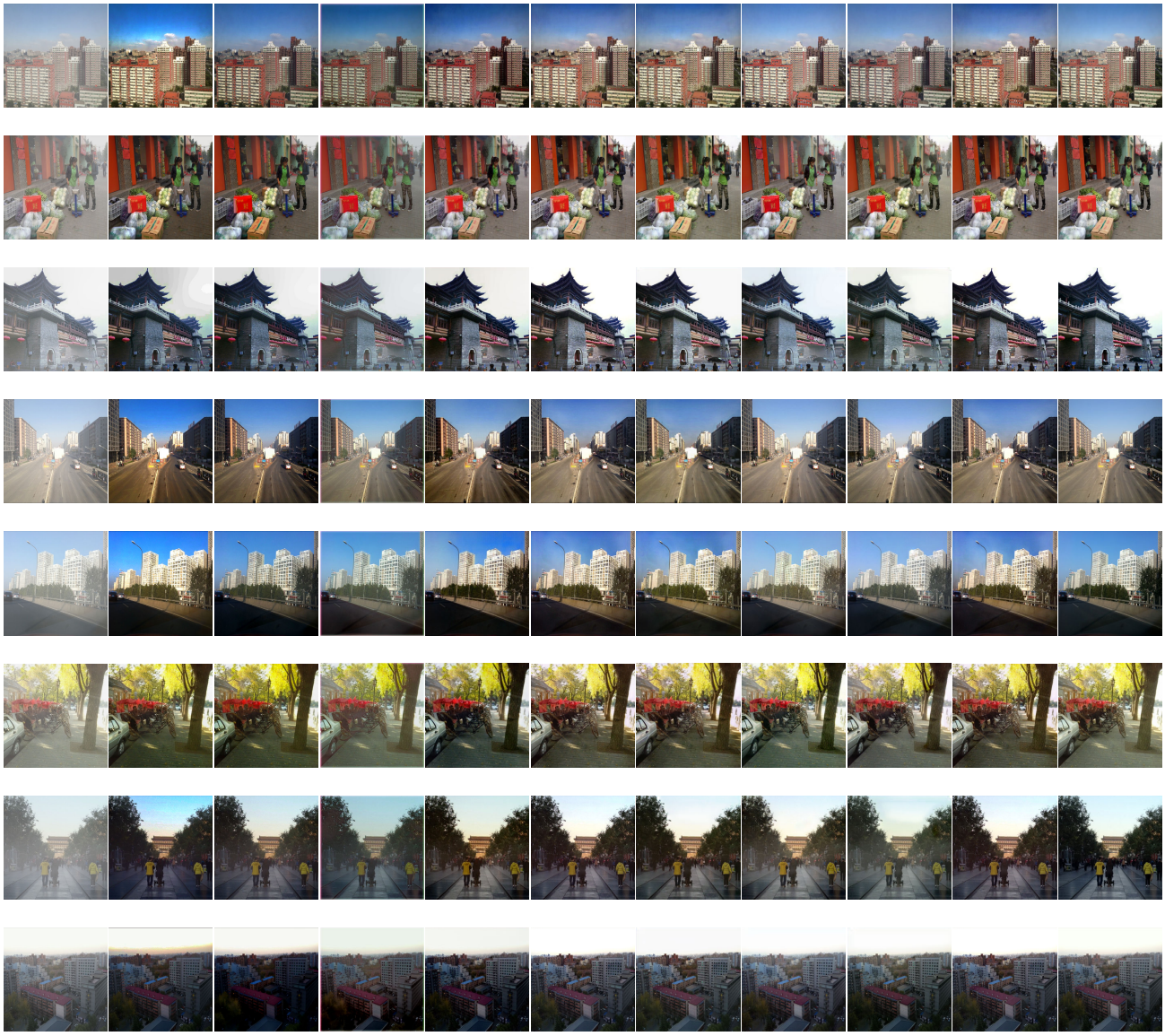
C. EXPERIMENTS ON SYNTHETIC DATASETS

Firstly, we conduct experiments on SOTS to illustrate the performance of our proposed architecture compared to the other state-of-the-art methods including CycleGAN and Cycle-Dehaze which require unpaired training images. Table 2 presents the quantitative results on average peak signal to noise ratio (PSNR) and structural similarity (SSIM) metrics [32]. It can be observed that our haze relevant feature attention architecture reaches the second best performance in terms of both PSNR and SSIM, and achieves the gain with 2.3182dB in PSNR and 0.0206 in SSIM compared with Cycle-Dehaze algorithm, which demonstrates the effectiveness of our proposed module. FFA-Net achieves best scores in terms of PSNR and SSIM metrics since the method has been trained on RESIDE dataset with paired hazy and corresponding haze-free images. However, FFA-Net fails to generate visually pleasing results for nonhomogeneous hazy scenes, which will be demonstrated on the next subsection.

Fig.10 depicts the visual comparisons on the synthetic datasets. We observe that the atmospheric scattering model based methods such as DCP, CAP and AODNet suffer from unsatisfactory results with color distortion and sometimes

TABLE 2. Quantitative results with the other approaches on synthetic datasets.

Metrics	DCP	CAP	AODNet	EPDN	GCANet	FFA-Net	CycleGAN	Cycle-Dehaze	Ours
PSNR	17.6690	18.7848	19.7868	21.4150	23.6187	31.8720	25.9386	26.6483	28.9665
SSIM	0.8743	0.8098	0.8706	0.8833	0.9273	0.9666	0.9191	0.9212	0.9418



(a) Origin (b) DCP (c) CAP (d) AODNet (e) EPDN (f) GCANet (g) FFA-Net (h) CycleGAN (i) Cycle-Dehaze (j) Ours (k) GT

FIGURE 10. Visual comparisons on synthetic datasets.

generate darker images compared to the ground truth. The output from EPDN is fuzzy and not clear enough, while GCANet restores haze-free images without sharper structures and details. FFA-Net generates high contrast, sharper edges and deals better with the variation of haze in the scene. CycleGAN and Cycle-Dehaze methods fail to remove haze thoroughly. Comparing with these methods, our proposed architecture generates better visual haze-free images with less color distortion and clear details, and improves the dehazing results both qualitatively and quantitatively.

D. EXPERIMENTS ON REAL IMAGES

To evaluate the performance of our proposed method on real images, we take visual comparisons provided by HSTS and RTTS datasets. As shown in Fig.11, DCP suffers from severe color distortion and CAP generates darker results. AODNet,

(a) Origin (b) DCP (c) CAP (d) AODNet (e) EPDN (f) GCANet (g) FFA-Net (h) CycleGAN (i) Cycle-Dehaze (j) Ours

FIGURE 11. Visual comparisons on real images.

TABLE 3. Quantitative results with the other approaches on NH-HAZE datasets.

Metrics	DCP	CAP	AODNet	EPDN	GCANet	FFA-Net	CycleGAN	Cycle-Dehaze	Ours
PSNR	12.9442	12.7091	12.7581	13.1371	13.0308	12.1778	15.2999	15.9967	18.5397
SSIM	0.4854	0.4542	0.4485	0.5166	0.5367	0.4715	0.5622	0.5868	0.6352

CycleGAN and Cycle-Dehaze methods are not able to generate ideal haze-free images and the effect of haze removal is not obvious. EPDN and GCANet also suffer from color distortion and remain haze artifacts in some areas. FFA-Net generates vivid colors and increased contrast, but it tends to introduce a slight color shifting for this set of images. Comparing with these algorithms, our proposed architecture achieves better visual effect and restores more detailed information with sharper structures.

The proposed method is also evaluated on NH-Haze dataset, which contains paired nonhomogeneous hazy and corresponding haze free images. The quantitative comparisons in terms of PSNR and SSIM are provided in Table 3, and the visual results are presented in Fig. 12. It is observed that DCP, CAP, AODNet and FFA-Net methods perform unsatisfactory dehazing effect with low PSNR and SSIM results. EPDN and GCANet recover quite well the image structure, but amplify color shifting artifacts while removing haze. In contrast, CycleGAN and Cycle-Dehaze perform less color artifacts and increased details, but may not remove the haze completely. The proposed method generates high contrast images with the best PSNR and SSIM values. Although some textures may not be clear as the ground truth, the proposed architecture has a great potential and performs in general better than the other considered techniques.

E. ABLATION STUDY

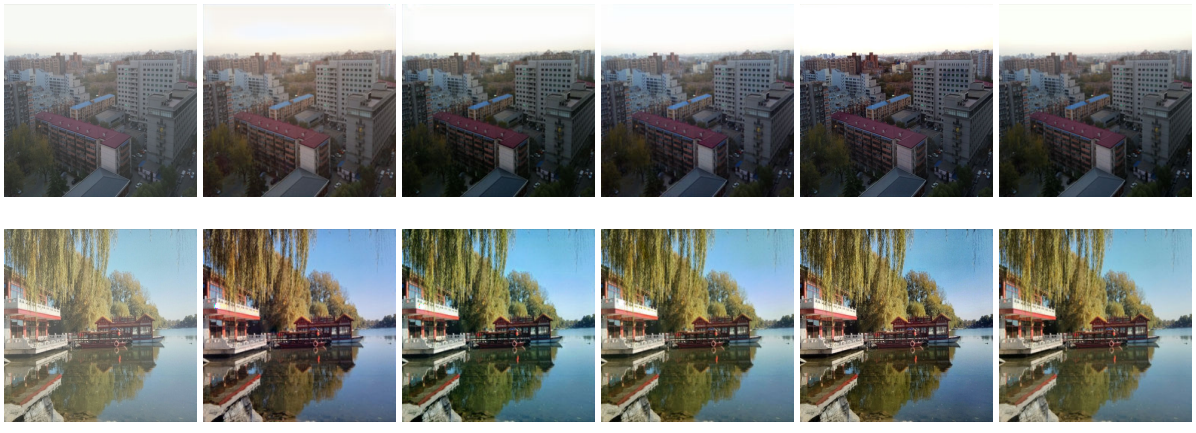
To better demonstrate the effectiveness of our proposed architecture, we conduct a series of ablation studies for analysis. Firstly, we implement the following generator network with different component combinations: 1) GAN: origin CycleGAN model is utilized; 2) GAN+D: the densely connected encoder-decoder structure is implemented in the

(a) Origin (b) DCP (c) CAP (d) AODNet (e) EPDN (f) GCANet (g) FFA-Net (h) CycleGAN (i) Cycle-Dehaze (j) Ours (k) GT

FIGURE 12. Visual comparisons on NH-HAZE datasets.

(a) Input (b) GAN (c) GAN+D (d) GAN+A (e) GAN+D+A (f) GT

FIGURE 13. Ablation studies of different generator networks.



(a) Input (b) Without both losses (c) With perceptual loss (d) With color loss (e) With both losses (f) GT

FIGURE 14. Ablation studies of loss functions.

generator based on CycleGAN; 3) GAN+A: the haze relevant feature attention module is embedded in CycleGAN; 4) GAN+D+A: the whole architecture of our proposed network. The quantitative comparisons in terms of PSNR and SSIM are presented in Table 4, and the visual results are provided in Fig. 13. It is observed that incorporating dense blocks and attention blocks achieves better PSNR and SSIM values with clearer and visually more pleasing haze-free images, since dense blocks are able to extract and transfer image features more thoroughly, and attention blocks give more guidance to the subsequent networks with hazy regions and haze concentration.

TABLE 4. Quantitative results with different generator networks.

Metrics	PSNR	SSIM
GAN	25.9386	0.9191
GAN+D	27.6072	0.9379
GAN+A	27.0183	0.9305
GAN+D+A	28.9665	0.9418

TABLE 5. Quantitative results with different loss functions.

Metrics	PSNR	SSIM
Without both losses	26.6476	0.9284
With perceptual loss	27.2535	0.9369
With color loss	27.5087	0.9391
With both losses	28.9665	0.9418

To verify the effectiveness of loss functions proposed in this work, we train the network without perceptual loss or color loss, and the quantitative and qualitative results are given in Table 5 and Fig. 14. As we can see, the models trained with perceptual loss and color loss achieve better performance in terms of PSNR and SSIM values. Furthermore, color loss makes the restored images more colorful and effectively avoids color distortion, with which more natural dehazed images can be obtained.

V. CONCLUSION

In this paper, we propose an end-to-end haze relevant feature attention network for single image dehazing, which does not require pairs of hazy and corresponding haze-free images, and does not need to estimate intermediate atmospheric scattering parameters. We integrate effective haze relevant features as additional priors into the generator architecture that combines an encoder-decoder structure with dense blocks. A novel attention module is proposed to provide useful information about haze regions and haze density for better image dehazing. Furthermore, color loss is proposed to avoid color distortion. Experiment results demonstrate that our proposed method reaches state-of-the-art performance on both public synthetic datasets and real-world images. Through ablation study, we demonstrate the effectiveness of different factors on the performance of proposed architecture.

REFERENCES

[1] R. Qian, R. T. Tan, W. Yang, J. Su, and J. Liu, "Attentive generative adversarial network for raindrop removal from a single image," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Salt Lake City, UT, USA, Jun. 2018, pp. 2482–2491.

[2] B. Zhang, M. Wang, and X. Shen, "Image haze removal algorithm based on nonsubsampled contourlet transform," *IEEE Access*, vol. 9, pp. 21708–21720, 2021.

[3] S.-C. Huang, D.-W. Jaw, W. Li, Z. Lu, S.-Y. Kuo, B. C. M. Fung, B.-H. Chen, and T. Numnonda, "Image dehazing in disproportionate haze distributions," *IEEE Access*, vol. 9, pp. 44599–44609, 2021.

[4] H. Zhang, V. Sindagi, and V. M. Patel, "Multi-scale single image dehazing using perceptual pyramid deep network," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Salt Lake City, UT, USA, Jun. 2018, pp. 902–911.

[5] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 770–778.

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

[7] 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.

[8] 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.

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

[10] C. O. Ancuti, C. Ancuti, C. Hermans, and P. Bekaert, "A fast semi-inverse approach to detect and remove the haze from a single image," in *Proc. Asian Conf. Comput. Vis.* Berlin, Germany: Springer, 2010, pp. 1–15.

[11] 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.

[12] B.-H. Chen, S.-C. Huang, C.-Y. Li, and S.-Y. Kuo, "Haze removal using radial basis function networks for visibility restoration applications," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 8, pp. 3828–3838, Aug. 2018.

[13] 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)*, Venice, Italy, Oct. 2017, pp. 4780–4788.

[14] Y. Qu, Y. Chen, J. Huang, and Y. Xie, "Enhanced Pix2pix dehazing network," *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Long Beach, CA, USA, Jun. 2019, pp. 8152–8160.

[15] D. Chen, M. He, Q. Fan, J. Liao, L. Zhang, D. Hou, L. Yuan, and G. Hua, "Gated context aggregation network for image dehazing and deraining," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Waikoloa, HI, USA, Jan. 2019, pp. 1375–1383.

[16] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, X. Bing, and Y. Bengio, *Generative Adversarial Nets*. Cambridge, MA, USA: MIT Press, 2014, pp. 2672–2680.

[17] N. U. Din, K. Javed, S. Bae, and J. Yi, "Effective removal of user-selected foreground object from facial images using a novel GAN-based network," *IEEE Access*, vol. 8, pp. 109648–109661, 2020.

[18] C. Mao, L. Huang, Y. Xiao, F. He, and Y. Liu, "Target recognition of SAR image based on CN-GAN and CNN in complex environment," *IEEE Access*, vol. 9, pp. 39608–39617, 2021.

[19] W. Wang, H. C. Wong, S. L. Lo, and G. Zhang, "Uncouple generative adversarial networks for transferring stylized portraits to realistic faces," *IEEE Access*, vol. 8, pp. 213825–213839, 2020.

[20] J. 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)*, Venice, Italy, Oct. 2017, pp. 2242–2251.

[21] Y. Wei, Z. Zhang, Y. Wang, M. Xu, Y. Yang, S. Yan, and M. Wang, "DerainCycleGAN: Rain attentive CycleGAN for single image deraining and rainmaking," *IEEE Trans. Image Process.*, vol. 30, pp. 4788–4801, 2021.

[22] D. Engin, A. Genc, and H. K. Ekenel, "Cycle-dehaze: Enhanced cycle-GAN for single image dehazing," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Salt Lake City, UT, USA, Jun. 2018, pp. 9380–9388.

[23] W. Liu, X. Hou, J. Duan, and G. Qiu, "End-to-end single image fog removal using enhanced cycle consistent adversarial networks," *IEEE Trans. Image Process.*, vol. 29, pp. 7819–7833, 2020.

[24] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Honolulu, HI, USA, Jul. 2017, pp. 2261–2269.

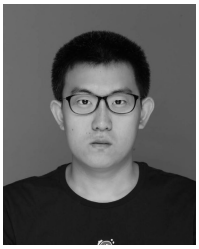
[25] L. K. Choi, J. You, and A. C. Bovik, "Referenceless prediction of perceptual fog density and perceptual image defogging," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3888–3901, Nov. 2015.

[26] K. Tang, J. Yang, and J. Wang, "Investigating haze-relevant features in a learning framework for image dehazing," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Columbus, OH, USA, Jun. 2014, pp. 2995–3002.

[27] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 6230–6239.

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

[29] B. Li, W. Ren, D. Fu, D. Tao, D. Feng, W. Zeng, and Z. Wang, "Benchmarking single-image dehazing and beyond," *IEEE Trans. Image Process.*, vol. 28, no. 1, pp. 492–505, Jan. 2019.
[30] 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)*, Seattle, WA, USA, Jun. 2020, pp. 1798–1805.
[31] C. O. Ancuti, C. Ancuti, F.-A. Vasluianu, and R. Timofte, "NTIRE 2020 challenge on nonhomogeneous dehazing," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Seattle, WA, USA, Jun. 2020, pp. 2029–2044.
[32] A. Hore and D. Ziou, "Image quality metrics: PSNR vs. SSIM," in *Proc. 20th Int. Conf. Pattern Recognit.*, Istanbul, Turkey, Aug. 2010, pp. 2366–2369.



XIN JIANG received the B.S. and M.S. degrees from the College of Information Science and Electronic Engineering, Zhejiang University. He is currently pursuing the Ph.D. degree with Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, China. His research interests include image dehazing and object detection.

LU LU is currently a Research Assistant with Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences. His research interests include super resolution reconstruction and image enhancement.

MING ZHU is currently a Research Fellow and a Ph.D. Supervisor with Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences. His research interests include digital image processing, television tracking, and automatic target recognition technology.

ZHICHENG HAO is currently a Research Fellow with Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences. His research interests include image preprocessing and target tracking.

WEN GAO is currently an Associate Research Fellow with Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences. Her research interests include 3D target detection and multi-target tracking.