Machine Vision and Applications (2020) 31:33
https://doi.org/10.1007/s00138-020-01082-5

ORIGINAL PAPER

Photo-realistic dehazing via contextual generative adversarial networks

Shengdong Zhang[1] · Fazhi He[1] · Wenqi Ren[2]

Received: 29 April 2019 / Revised: 27 February 2020 / Accepted: 14 April 2020

Abstract
Single image dehazing is a challenging task due to its ambiguous nature. In this paper we present a new model based on generative adversarial networks (GANs) for single image dehazing, called as dehazing GAN. In contrast to estimating the transmission map and the atmospheric light separately as most existing deep learning methods, dehazing GAN restores the corresponding hazy-free image directly from a hazy image via a generative adversarial network. Extensive experimental results on both synthetic dataset and real-world images show our model outperforms the state-of-the-art algorithms.

Keywords Generative adversarial networks · Dehazing · Image restoration · Contextual network

1 Introduction

Haze reduces the image contrast and color. Floating particle in the atmosphere absorbs and scatters the light. Mathematically, we can express this process as the following formulation [19]:

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

where $I(x)$ denotes the hazy image, $J(x)$ represents the corresponding haze-free image, A denotes the global atmospheric light, and $t(x)$ is the transmission map, which can be expressed as:

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

where β is the scattering coefficient of the atmosphere, and $\mathrm{d}(x)$ is the distance between the object and the camera. We can recover the final haze-free image by the following equation:

$$J(x) = \frac{1}{t(x)} I(x) - A\frac{1}{t(x)} + A. \tag{3}$$

✉ Fazhi He
fzhe@whu.edu.cn

[1] School of Computer Science, Wuhan University, Wuhan, China

[2] Institute of Information Engineering, Chinese Academy of Sciences, Beijing, China

Based on the atmospheric scattering model, most dehazing methods share the same methodology estimating the transmission map and the atmospheric light separately.

Traditional single image dehazing methods based on physical model (2) often estimate the key parameters A and $t(x)$ by exploiting visual cues or statistical properties of hazy images [1,8,17,34,51]. He et al. [8] propose the dark channel prior (DCP) to estimate the atmospheric light and calculate the transmission map. Meng et al. [17] extend the boundary constraint and contextual regularization for sharper edges. Instead of using hand-crafted features, CNN-based methods [3,24,26] are proposed to learn the relation between transmission map and hazy image automatically. However, these methods still need to estimate transmission maps and atmospheric lights and then recover clear images using Eq. (3). Thus, if the transmissions or atmospheric lights are not estimated accurately, they will lead to poor-quality dehazed results. To overcome this problem, Li et al. [11] propose an end-to-end dehazing method by combining the two parameters transmission map $t(x)$ and atmospheric light A into a new variable $K(x)$. However, this method still needs to compute the new variable $K(x)$ which combines both transmission $t(x)$ and atmospheric light A. Though DCPDN [42] achieves end-to-end dehazing by directly embedding image degradation model (1) into the optimization framework via math operation modules, the generalization ability of DCPDN is limited by model (1) and results on natural images are far from optimization. Furthermore, it is not easily scaled up/down to different input image sizes. GFN

Published online: 16 May 2020

[27] is a deep end-to-end trainable neural network based on a novel fusion-based strategy. However, GFN needs to derive three inputs from an original hazy image, which is a preprocessing. Furthermore, if the three derivations do not contain enough information for removing haze, it will generate low-quality dehazing result. Preprocessing may result information loss. If intermediate variables (transmission map and atmospheric light) are not well estimated, it will affect the final dehazing result. We utilize end-to-end trainable network to avoid preprocessing and immediate parameters by learning the mapping from hazy input to haze-free output directly. Prior CNN-based methods [11,26,27,42] are trained on synthetic indoor haze dataset, and we find the generalization ability of these methods is limited and these methods cannot deal real hazy image well. In order to overcome this problem, we synthesize an outdoor hazy image dataset.

To address the above issues, we propose a generative dehazing network that directly restores the final dehazing result from the input hazy image. Inspired by the recent success of GANs, which have been applied in low-level vision tasks such as image generation [7], image inpainting [37] and image super-resolution [10], our network consists of two models: generator model (G) and discriminator model (D). The generator model contains dilation layers with a features fusion layer, so that it can capture more contextual information and generate more Haze distribution is dependent on depth, near areas tend to show haze-free appearance and far areas tend to show dense haze. In order to consider the nonuniform distribution, we design a contextual network to capture the large-scale haze distribution information. CNN-based methods tend to blur the final recover results using $L1$ loss. In order to overcome this problem, we propose to incorporate GAN loss into our network to recover photo-realistic result. The contributions of this paper are listed as follows:

* We present a novel fully end-to-end trainable neural network for single image dehazing based on GAN. Our model learns more effective features for dehazing, which helps our model overcome the limitation of the image degradation model.
* We propose a new local-based discriminator to generate much better dehazed results from generator. Compared with traditional discriminator, our patch-based discriminator focuses more on the image detail, which is helpful for our model to recover a photo-realistic dehazing result.
* We propose a method to eliminate artifact incurred by dilation convolution.
* We conduct extensive experiments on both synthetic dataset and real-world images, and the quantitative and qualitative comparison with the state-of-the-art single image dehazing algorithms demonstrates the effectiveness of the proposed method.

2 Related work

2.1 Priors-based dehazing

Based on model (3), there exist two important factors for single image dehazing: (1) accurate estimation of transmission map and (2) accurate estimation of atmospheric light. However, estimating transmission map or atmospheric light is an ill-posed problem. To solve the ill-posed dehazing problem, researchers have explored various hand-crafted priors [2,18,43,48] for single image dehazing. Tan [31] removes haze from image by maximizing the contrast. The dehazed results of this method often tend to show color distortions. Nishino et al. [20] model an image as two independent latent layers, which are the scene scattering and depth. A factorial Markov random field is used to solve the model. However, this method tends to be over-estimated the haze concentration and results in over-dehazed result. Based on several assumptions Fattal proposes a single image dehazing method [5]. First, the albedo is assumed to be a constant vector in a local patch, and then the clean image and the depth are assumed to be independent in a local patch. Although this approach could generate a high-quality dehazing result, it would lose effect when there are low variation or signal-to-noise ratio, especially in the very dense haze regions. Dark channel prior is discovered by He et al. [8] for single image dehazing. Based on the observation that most local patches in haze-free image without sky contain at least one channel with low intensity, He et al. propose a method to estimate transmission map. When dark channel prior holds for hazy image, it will generate a high-quality dehazed result. However, this prior is invalid and the dehazing result tends to be over-enhanced for the transmission being over-estimated. The dark channel prior has been proved to be effective for image dehazing, and thus, numerous dehazing methods based on the dark channel prior have been developed [4,17]. Fattal [6] proposes a method which uses color-line (pixels in a local clean patch will form a line in RGB color space) to estimate the transmission. Recently, Berman et al. [1] propose a nonlocal prior, based on the assumption that each color cluster in the clear image became a haze line in RGB space. All these methods share the same problem that it is a separate way to solve the haze removal problem, and the inaccurate of transmission map will affect the clear image estimation. Apart from estimating transmission map, some other works [2,30] focus on estimating atmospheric light. The problem with estimating atmospheric light is the case in estimating transmission map.

2.2 Learning-based dehazing

Learning-based methods have many advantages over traditional approaches [9,13,15,21,23,38,41] in computer science. Learning-based image dehazing methods [33,51] aim to auto-

(a) input **(b)** FEED-Net [47] **(c)** Ours

Fig. 1 Visual comparison on real-world images with FEED-Net

matically learn relationships between image features and transmission maps in a regression framework. These methods still follow the three-step methodology: transmission estimation, computing atmospheric light and image dehazing. In addition, the estimated transmission maps by the regression framework still need to be refined by guided filter to suppress the blocky artifacts [3,26,33,46]. Zhu et al. [51] create a linear model for modeling the scene depth of the hazy image under a color attenuation prior and learning the parameters of the model in a supervised manner. However, these methods are still developed based on hand-crafted features.

Convolutional neural networks (CNNs) have been successfully applied in image restortion [25,28,35] and saliency detection [32]. Recently, CNN-based image dehazing methods [3,26,27,44,45,49] have been proposed. DehazeNet [3] proposes a trainable model as well as a Bilateral Rectified Linear Unit (BReLU) layer to estimate the transmission map from a hazy image. Ren et al. [26] propose a multi-scale CNN (MSCNN) that firstly generated a coarse-scale transmission map and gradually refined it by a fine-scale CNN.

In order to bypass the transmission and atmospheric light estimation step, Li et al. [11] propose to integrate both transmission and atmospheric light in a new variable and propose a new dehazing method (AOD-Net) to generate the clean image based on the introduced intermediate variable. The dehazing ability of this method is limited by small receptive size and cannot deal heavy hazy image well. DCPDN [42] is a novel end-to-end jointly optimization dehazing network by directly emending hazing model (1) into the optimization framework via math operation modules. Thus, it allows the network to optimize the transmission map, atmospheric light and dehazed image jointly. However, its dehazing ability is limited by the atmospheric scattering model, which cannot describe the natural hazy image well. GFN [27] is a fusion-based dehazing method. GFN can generate visual pleased results for most cases. However, GFN cannot handle hazy images with very large haze area. The dehazing ability of GFN is also limited by its derivations from inputs, which assumed containing clear cues to reconstruct haze-free image. In order to overcome the above limitations, we propose using progressively downsampling to increase receptive size and capture more context information to reconstruct the final dehazed result.

3 Proposed algorithm

In this section, we detail the architecture of our model including generator and discriminator as well as training loss. To generate clean images from hazy inputs, we modify GAN framework by adding $L1$ and gradient loss as supervised train signal.

3.1 Motivation

Our method is motivated by the traditional end-to-end image restoration problems such as image denoising [16] and image inpainting [22]. These methods directly use the degraded input to predict clean output using an end-to-end trainable network. All of these algorithms achieve great performance. However, existing dehazing methods, including CNN-based algorithms, need to first estimate transmission map and atmospheric light [3,26], or a newly introduced variable [11], and then dehaze image using physical model (2). As mentioned above, if the intermediate variables are not estimated well, they will interfere the following dehazing results. This observation motivates our fully end-to-end architecture for single image dehazing. Therefore, the main idea that is used for designing the network contains three steps: features extraction, features fusion and image dehazing under the guidance of the extracted features.

To this end, we propose a novel end-to-end framework, using the contextualized dilated network for extracting haze-relevant features and use the dense skip links to fuse features for guiding the refinement of the dehazed images. We concatenate all the features from the shallow dilated convolutional layers (whose features contain edges and finer details information) to generate more useful features for image dehazing. By training the network in an end-to-end

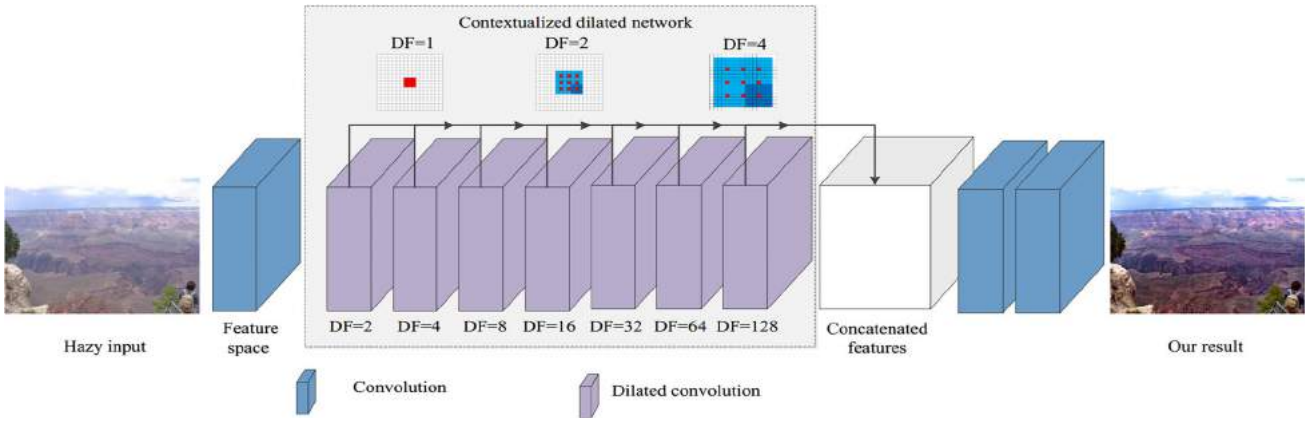


Fig. 2 The architecture of the generator

manner, the proposed network is able to learn useful features for haze removal.

Dilated convolution has been proved effective for dehazing [47]. Gridding artifacts [40] appear when a feature map has higher-frequency content than the sampling dilated rate, it will introduce unwanted checkerboard or gridding artifacts as shown in Fig. 1. In order to remove the gridding artifacts, Yu et al. [40] develop a scheme to remove gridding artifacts from output activation maps produced by dilated convolution. In this paper, we use GAN and gradient loss for removing gridding artifacts.

3.2 Generator

The generator is designed to recover a clean image from an input hazy image. Therefore the generator must remove the haze as much as possible and keep the structure and detail information well. In this paper we adopt the similar model structure as **FEED-Net** [47]. In order to improve the speed of dehazing, we make some modifications. Firstly, we change the last filter size from $3 * 3$ to $1 * 1$. Secondly, we use $TANH$ as activation function for last layer. Based on the motivation in Sect. 3.1, we develop a neural network using an end-to-end scheme which is shown in Fig. 2. There are totally 11 dilated convolutional layers followed by 3 convolutional layers in the proposed network. The kernel size of all the convolutional layers is 3×3. The feature maps are with the same size as the input hazy image. In order to add nonlinearity into the network, Rectified Linear Unit (ReLU) layer is added after each convolution layers except for the last convolution layer. As discussed in Sect. 3.1, we add the skip link to every dilated convolutional layer to collect the feature maps in shallow layers and help preserve the edge information of the dehazed results.

For haze removal problem, contextual information from an input image has been proved to be useful for automatically identifying and removing the haze [26]. The dilated convolution layer [39] weights pixels with a side of the dilated factor and thus increases model's receptive field without losing any resolution. Thus, we propose a contextualized dilated network to collect multi-scale context information for learning the haze-relevant features. The network collects contextual information in two ways: (1) it provides an increasingly larger receptive field for the following layers; (2) convolutional layers with different dilated factors have their own receptive fields. As shown in the gray region of Fig. 2, the network firstly transforms the input hazy image into feature space via the first convolution. Then, the network boosts the features by 7 dilated convolutional layers progressively. Each dilated convolution with different dilated factors and all the dilated convolutional layers are collected via the identity forward. We gradually enlarge the dilated factors in each layer in the proposed model. Specifically, the dilated factors from the 1st to the 7th dilated convolution layers are of 2, 4, 8, 16, 32, 64, 128, respectively.

3.3 Discriminator

The discriminator has been used to distinguish whether an image is real or fake in traditional GAN framework. We note that the results of FEED-Net lose some high frequencies and contain some artifacts. Based on the mentioned observations, we design a patch-based discriminator, which can be used to distinguish whether a patch is real or fake. It is well known that $L1$ loss generates blurry results; however, it accurately captures the low frequencies. So we propose to use GAN and $L1$ loss to restore the low and high frequencies of the image. This discriminator tries to classify whether each $N \times N$ patch in an image is from the generator or not. In order to design an efficient discriminator, we design a fully convolutional discriminator. We run the discriminator on full input image and then determine a patch is real or fake according to the responses of last layer. We show the architecture of our dis-

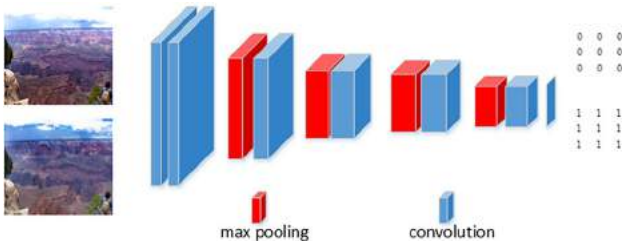


Fig. 3 The architecture of the discriminator

criminator in Fig. 3. The details of the each discriminator structure and parameter settings are shown in Table 1.

3.4 Loss function

Following the standard GAN framework, we adopt a traditional GAN objective, which can be expressed as

$$\mathcal{L}_{GAN}(G, D) = \mathrm{E}_J[1 - D(J_i)] + \mathrm{E}_I[D(G(I_i))], \tag{4}$$

where G denotes the generator, D represents the discriminator, I_i and J_i are the ith input image and ground truth clean image.

$L2$ loss has been proved to lose high-frequency details such as texture [47]. As shown in FEED-Net [47], $L2$ loss can recover low frequencies; however, we note that the final dehazing results are blurry and contain artifacts. It has been proved that it is benefitted to combine GAN loss with transitional loss ($L2$ loss) [22]. We adopt the same strategy to combine GAN loss with $L1$ loss, which can be expressed as:

$$\mathcal{L}_1(G) = \frac{1}{n}\sum_{n=1}^{N} ||\mathcal{G}(I_i, \Theta) - J_i||_1. \tag{5}$$

In order to remove artifacts in the dehazed result, we introduce an gradient loss, which can be defined as follows:

$$\mathcal{L}_g(G) = || \bigtriangledown \mathcal{G}(I_i)||_1, \tag{6}$$

where $|| \bigtriangledown \mathcal{G}(I_i)||_1$ denotes the total variation regularization.

In order to train the generator with GAN framework, we define the adversarial loss as

$$\mathcal{L}_{\text{adv}}(G, D) = \mathrm{E}_I[1 - D(G(I_i))], \tag{7}$$

Based on the above loss, our final objective loss can be written as:

$$\mathcal{L}(G) = \mathcal{L}_1(G) + \lambda_1 \mathcal{L}_{\text{adv}}(G, D) + \lambda_2 \mathcal{L}_g(G), \tag{8}$$

we use λ_1 and λ_2 to control the importance of adversarial loss and gradient loss, respectively.

3.5 Training and testing data

It is critical to prepare a vast amount of labeled data for training deep models. It is a natural way to synthesize training data based on the physical haze formation model. We generate a depth dataset using Liu et al [14]; then, we generate a training dataset. We collect 1100 haze-free images from **SYSU-Scene Dataset**. To show the details of how to obtain the training data, we show an example in Fig. 4. For a haze-free image and the corresponding depth map, we generate a hazy image via (1). First, we generate two random atmospheric lights $A = [k, k, k]$, where $k \in [0.7, 1]$. Second, we choose $\beta \in [0.01, 0.25]$ to generate a transmission map. Finally, we synthesize the hazy image using the haze-free image, the atmospheric light and the transmission map. Therefore, we have 22,000 images in training dataset. We resize hazy images to (500×500) and then use them to train our model. Although the synthesized hazy images can be used to train a deep learning model, small gap can be found between real hazy image and synthesized hazy image. Figure 5 shows one problem of [14]. The first one is that the result by [14] cannot keep the ordinal depth. As shown, the football should be farther than the dog and the background. Due to this problem, the synthesized hazy image is different from the real hazy image.

In order to validate the proposed network, we generate a testing dataset as [26]. We synthesize the hazy image set by fixing the atmospheric light A as $[0.78, 0.78, 0.78]$ and use different scattering coefficient $\beta \in \{0.06, 0.3, 0.54\}$, referred to as **HIFA**.

3.6 Implementation details

In order to speed the dehazing, We choose 8 filters for all layers except the last for generator. We initialize the weights using identity initializer [39] except last two layers, which are initialized using Gaussian random variables. The last layer, used for image dehazing, consists of three filters of size $1 * 1 * 8$. We utilize $LReLU$ neuron as we found it more effective than the $ReLU$ [39]. For the last layer, we utilize $TANH$ as activation function. During training we use ADAM as the optimization algorithm to train our network with learning rate is 0.0001 for generator and discriminator and batch size of 4. Our Dehazing GAN can be converged around 13 epochs. We set λ_1 and λ_2 to 0.01.

3.7 Relation to traditional dehazing methods

Traditional methods often obtain the dehazed result via Eq. (3). These methods firstly estimate the transmission map using hand-crafted priors or convolution neural networks and then follow the conventional method to estimate the atmospheric light to restore the clean images. However, if the

Table 1 Configurations of the discriminator

	conv1	conv2	pool1	conv3	pool2	conv4	pool3	conv5	pool4	conv6	conv7
Size	3	3		3		3		3		3	3
Channel	24	24		32		40		48		56	1
Stride	1	1	↓2	1	↓2	1	↓2	1	↓2	1	1
Pad	1	1		1		1		1		1	1

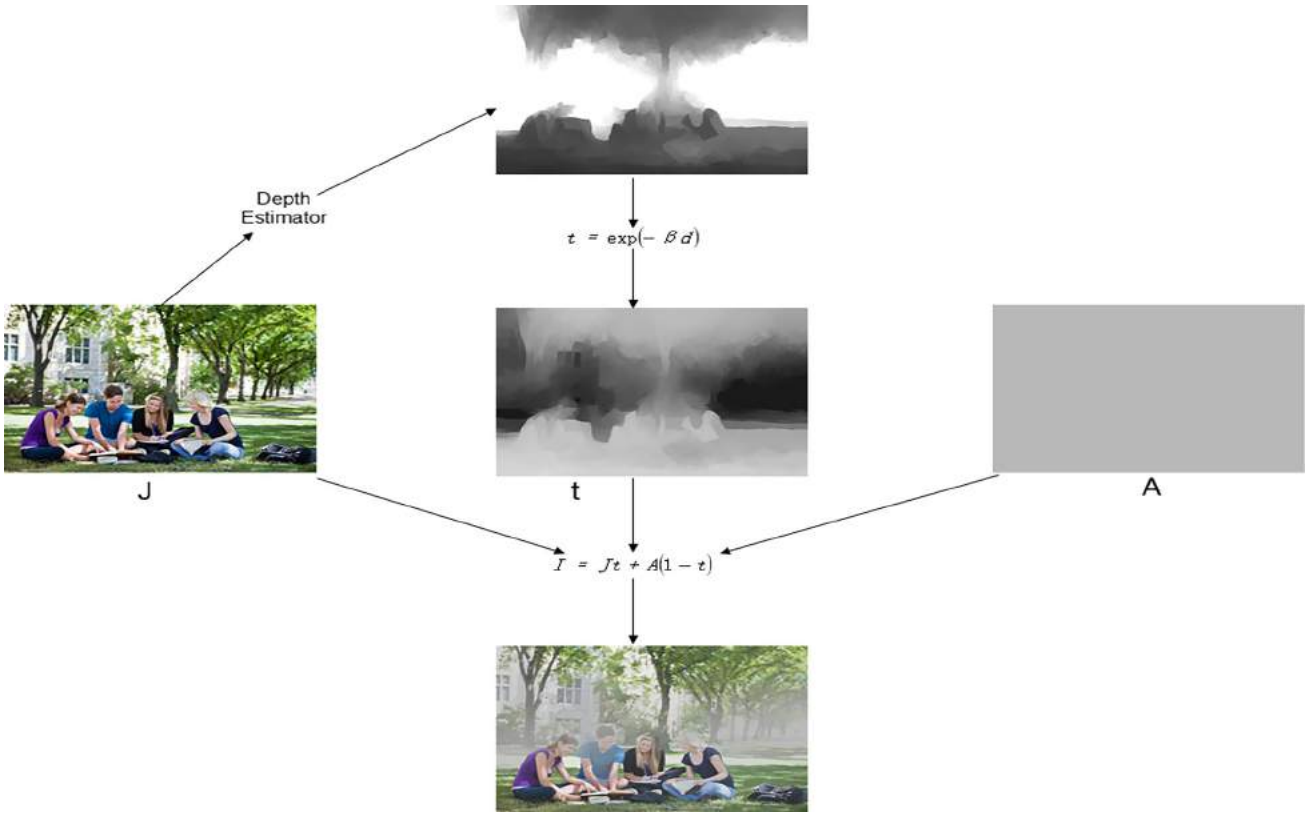
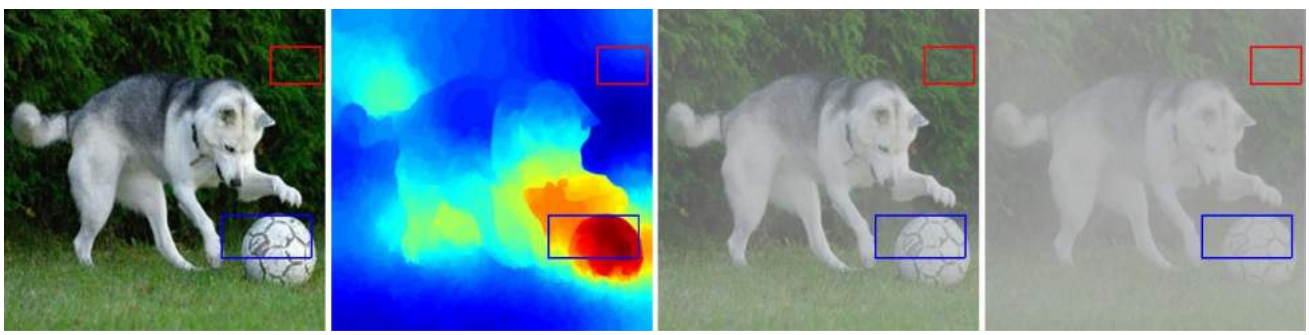


Fig. 4 The flow to generating a synthesized hazy image. A is obtained by a random function, and β is a random value in $[0.01, 0.25]$. The depth estimator is used to obtain depth. In especial, we use CNN-based method [14] to obtain depth for a clean image



(a) Haze-free **(b)** Depth **(c)** Light **(d)** Heavy

Fig. 5 Problem of training data preparation. **a** Haze-free image. **b** The corresponding depth. **c** The corresponding light hazy image. **d** The corresponding heavy hazy image. We find that the depth top logic has been changed in first row

Table 2 Average PSNR/SSIM of dehazed results on the SOTS dataset from **RESIDE**

	DCP	CAP	NLD	MSCNN	DehazeNet	AOD-Net	DCPDN	GFN	Ours
PSNR	16.62	19.05	17.29	17.57	21.14	19.06	15.86	22.30	**22.50**
SSIM	0.82	0.84	0.75	0.81	0.85	0.85	0.82	0.88	**0.90**

Bold values indicate the best result among the compared methods

Table 3 Average PSNR/SSIM of dehazed results on the HazeRD dataset [50]

	DCP	CAP	NLD	MSCNN	DehazeNet	AOD-Net	DCPDN	GFN	PDNet	Ours
PSNR	17.66	18.56	17.47	19.10	19.53	18.13	18.82	19.18	20.14	**20.20**
SSIM	0.8430	0.8256	0.7921	0.8540	0.8498	0.8266	0.8906	0.8601	0.8932	**0.8934**
FID	36.32	**16.81**	34.47	31.19	19.60	18.84	32.14	27.44	22.86	18.41

Bold values indicate the best result among the compared methods

transmission map is not accurate enough, they will accordingly affect the estimation of atmospheric light. Therefore the final restored image usually contains color distortions or haze.

We further analyze the influence of dehazing with transmission map. We can apply partial to Eq. (3) and get the following equation:

$$\frac{\partial J(x)}{\partial t} = \frac{(A - I(x))}{t^2(x)}. \tag{9}$$

From Eq. 9 we can see that the small error will boost the dehazed error when the transmission map is small. For a pixel with a small transmission value $t = 0.2$ and pixel intensity value is [0.65, 0.70, 0.84], atmospheric light value is [0.8, 0.8, 0.8], and the corresponding clear pixel value is [0.05, 0.3, 1]. If error is 0.1, we can obtain two transmissions 0.1 and 0.3, we can get the corresponding dehazing pixel value $[-0.7, -0.2, 1.2]$ and [0.3, 0.47, 0.93], and we find that the error is large. To avoid this problem, we recover the haze-free image from hazy image directly.

4 Experimental results

4.1 Synthetic images

HIFA dataset In this subsection, we compare our method with the end-to-end dehazing method proposed by Li et al. [11] and other state-of-the-art methods. In order to show the high performance of our method, we evaluate our method and other state-of-art methods on the **HIFA** dataset. As shown in Table 4, we observe that our method significantly outperforms the state-of-the-art dehazing methods.

RESIDE dataset Public dataset (**SOTS**) is from REalistic Single Image DEhazing (**RESIDE**) [12], which has been used to evaluate the dehazing methods widely. Since the **SOTS** is simulated, the ground truth haze-free images are available, which makes evaluating the performance qualitatively as well as quantitatively possibly. We evaluate our method on the **SOTS** and compare it with several state-of-the-art single image dehazing methods using peak signal-to-noise ratio (**PSNR**) and structural similarity index (**SSIM**). From Table 2, we can see that our method achieved a significant improvement by comparing with state-of-the-art methods.

Table 4 Quantitative comparison on the **HIFA** dataset with the ground truth atmospheric lights A

	Meng [17]	Berman [1]	Cai [3]	Ren [26]	Li [11]	Ours
PSNR	11.46	15.54	16.81	17.68	*18.64*	**19.75**
SSIM	0.527	0.713	0.745	*0.788*	0.747	**0.817**

Bold and italic values indicate the best result among the compared methods

HazeRD dataset HazeRD contains natural outdoor images with corresponding high-quality depth map and therefore encourages us to better synthesize real outdoor hazy images taken under different hazy levels. Proximal Dehaze-Net [36] has created a hazy dataset, which includes 128 images with different haze levels. Since all learning-based methods do not use images in HazeRD as training data, it is a fair way to compare them on HazeRD. We refer proximal Dehaze-Net as PDNet. As shown in Table 3, our deep learning-based method gets highest PNSR and SSIM on HazeRD and exceeds the second best learning-based method PDNet [36] by 0.06 dB in PSNR and 0.0002 in SSIM. We also measure the performance of dehazing methods in FID [29] as the similar measurement in the general generative modeling. The results of each method can be found in Table 3. As shown in Table 3, we can see that our method achieves the second best performance in FID (Table 4).

4.2 Evaluation on challenging natural images

Because all of the state-of-the-art dehazing methods can obtain truly good dehazed results on general outdoor images,

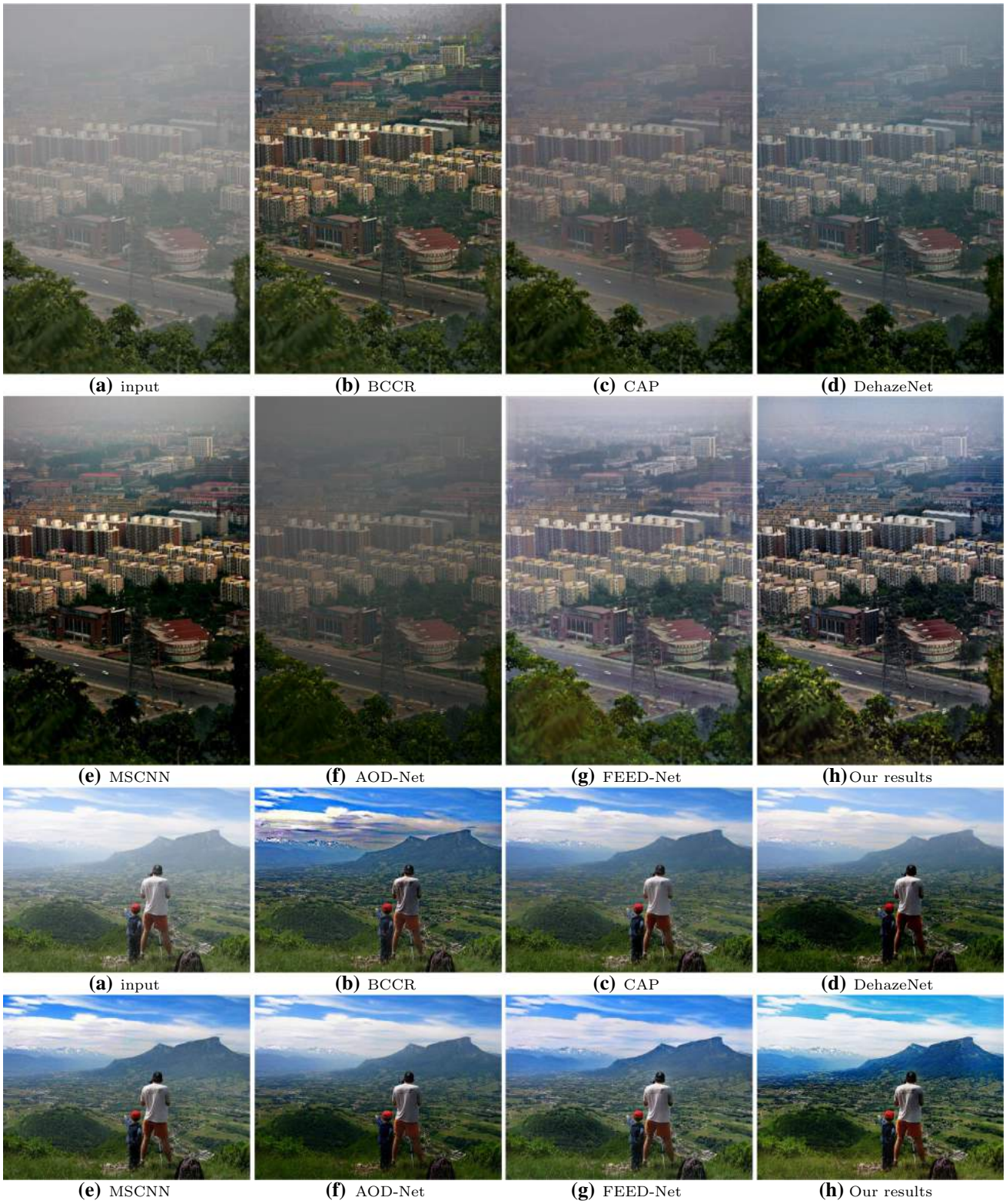


Fig. 6 Visual comparison on real-world images. The proposed method generates much clearer images with finer details

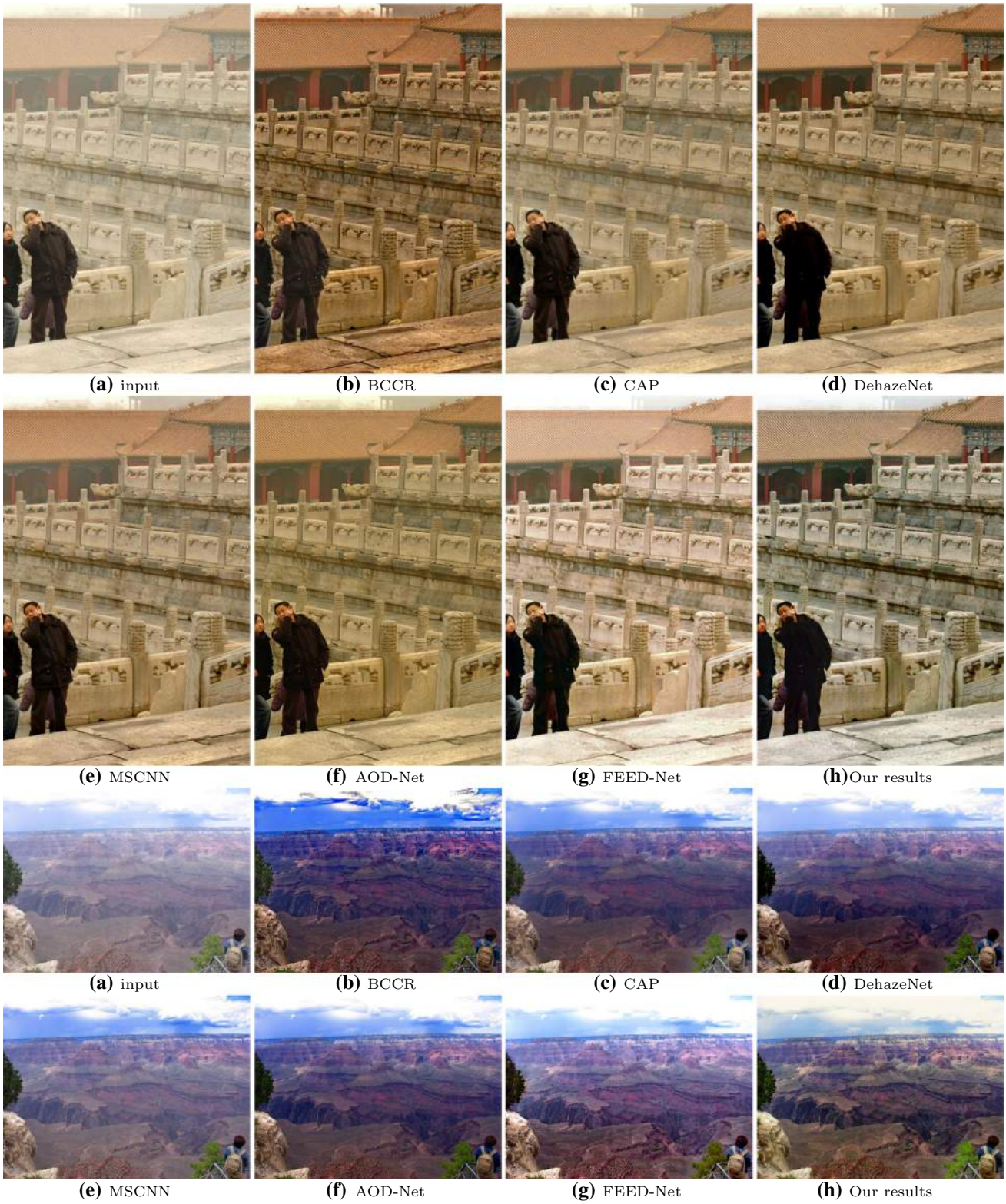


Fig. 7 Visual comparison on real-world images. The proposed method generates much clearer images with finer details

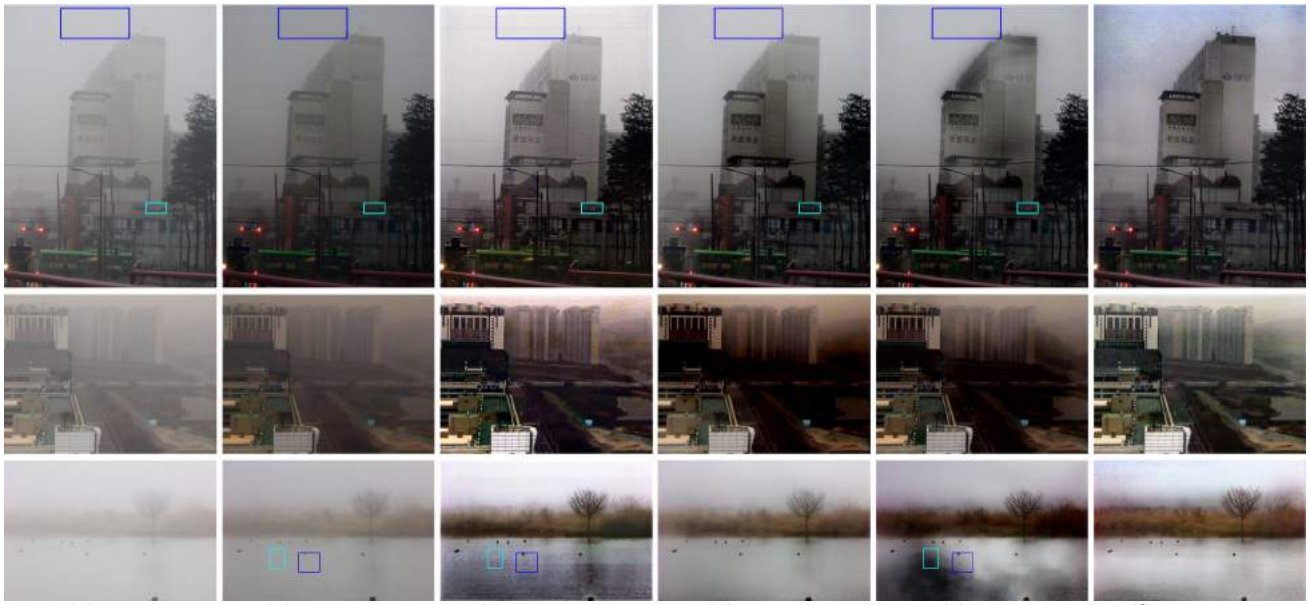


Fig. 8 Visual comparison on real-world dense hazy images

it is hard to distinct the best one visually. In order to overcome this issue, we evaluate the proposed algorithm against the state-of-the-art single image dehazing methods [3,11,17,26,47,51] using seven challenging real-world images in Figs. 6, 7 and 8.

In Fig. 6, we show two dehazing examples. The first one contains very haze distribution that near areas show haze-free appearance and distant areas contain dense haze. The second one contains large white areas, such as white cloud in sky, which are hard for state-of-the-art methods to generate a visual pleasured results. For the first image, the dehazed results in the first row by Meng et al. [17] and Zhu et al. [51] have some color distortions shown in (b) and (c). The dehazing methods of [3,11] tend to underestimate the thickness of the haze and remain some haze in the dehazed results. We also note that the results of [47] exist some halo artifacts. Furthermore, the results of [47] tend to be blurry. MSCNN [26] can deal the dense haze areas well. However, it loses some detail in near areas. For the second image, we can note that the dehazed result of [17] tends to show some color distortion in sky areas. The dehazed results of [3,11,26,51] tend to underestimate the density of haze and leave haze in final dehazed results. Our results can avoid these problems well by capturing large-scale contextual information and recover more image details via GAN.

In Fig. 7, we show two examples which contain gray and large areas inherently similar to the atmospheric light. For the first image containing large gray areas, which is hard for dehazing methods to generate a reasonable result, BCCR [17] tends over-dehazing, which will generate color distortion in dehazed result. Learning-based methods cannot deal well the gray areas well. The reason for this failure is that large gray areas need more contextual information to understand it well. Our model benefits much from the large contextual information, which can provide much useful features to recover a reasonable dehazed result.

Our method is an end-to-end dehazing method, so we also compare our method with four end-to-end dehazing methods [11,27,42,47]. In Fig. 8, we can see that AOD-Net tends to leave haze in the dehazed result, GFN tends to generate color distortion, FEED-Net tends to generate some noise in the dehazed result, and DCPDN loses some details in the dehazed result. In contrast, our method can remove haze well and preserve the image details well.

5 Conclusions

In this paper, we have presented a novel learning-based method (Dehazing GAN) for end-to-end single image dehazing. We address the image dehazing problem via a GAN framework which learns to directly recover the haze-free image from a single hazy image. Our model learns more effective features to deal the cases when the haze imaging model cannot hold. Compared to traditional methods which require estimating transmission map and atmospheric light, the proposed method does not need to compute any intermediate variables. We have confirmed that dehazing GAN generates dehazed results are, by a considerable margin, more

photo-realistic than results obtained by state-of-the-art methods.

Acknowledgements This work is supported by National Key R&D Program of China (Grant No. 2017YFB0503004), National Natural Science Foundation of China (Grant No. 41571436) and China Postdoctoral Science Foundation (Grant No. 2019M662709).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Berman, D., Avidan, S., et al.: Non-local image dehazing. In: CVPR (2016)
2. Berman, D., Treibitz, T., Avidan, S.: Air-light estimation using haze-lines. In: ICCP (2017)
3. Cai, B., Xu, X., Jia, K., Qing, C., Tao, D.: Dehazenet: an end-to-end system for single image haze removal. TIP **25**(11), 5187–5198 (2016)
4. Chen, C., Do, M.N., Wang, J.: Robust image and video dehazing with visual artifact suppression via gradient residual minimization. In: ECCV (2016)
5. Fattal, R.: Single image dehazing. TOG **27**(3), 72 (2008)
6. Fattal, R.: Dehazing using color-lines. TOG **34**(1), 13 (2014)
7. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. In: NIPS, pp. 2672–2680 (2014)
8. He, K., Sun, J., Tang, X.: Single image haze removal using dark channel prior. In: CVPR (2009)
9. Hou, N., He, F., Zhou, Y., Chen, Y.: An efficient gpu-based parallel tabu search algorithm for hardware/software co-design. Front Comput Sci **14**(5), 145,316 (2020)
10. Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., et al.: Photo-realistic single image super-resolution using a generative adversarial network. In: CVPR (2016)
11. Li, B., Peng, X., Wang, Z., Xu, J., Feng, D.: An all-in-one network for dehazing and beyond. In: ICCV (2017)
12. Li, B., Ren, W., Fu, D., Tao, D., Feng, D., Zeng, W., Wang, Z.: Benchmarking single image dehazing and beyond. TIP **28**, 492–505 (2018)
13. Li, H., He, F., Liang, Y., Quan, Q.: A dividing-based many-objectives evolutionary algorithm for large-scale feature selection. Soft Comput **24**(9), 6851–6870 (2020)
14. Liu, F., Shen, C., Lin, G., Reid, I.: Learning depth from single monocular images using deep convolutional neural fields. TPAMI **38**(10), 2024–2039 (2016)
15. Luo, J., He, F., Yong, J.: An efficient and robust bat algorithm with fusion of opposition-based learning and whale optimization algorithm. Intell Data Anal **24**(3), 1–19 (2020)
16. Mao, X., Shen, C., Yang, Y.B.: Image restoration using very deep convolutional encoder–decoder networks with symmetric skip connections. In: NIPS (2016)
17. Meng, G., Wang, Y., Duan, J., Xiang, S., Pan, C.: Efficient image dehazing with boundary constraint and contextual regularization. In: ICCV (2013)
18. Nair, D., Kumar, P.A., Sankaran, P.: An effective surround filter for image dehazing. In: Proceedings of the 2014 international conference on interdisciplinary advances in applied computing, p. 20 (2014)
19. Narasimhan, S.G., Nayar, S.K.: Vision and the atmosphere. Int J Comput Vis **48**(3), 233–254 (2002)
20. Nishino, K., Kratz, L., Lombardi, S.: Bayesian defogging. IJCV **98**(3), 263–278 (2012)
21. Pan, Y., He, F., Yu, H.: Learning social representations with deep autoencoder for recommender system. World Wide Web (2020). https://doi.org/10.1007/s11280-020-00793-z
22. Pathak, D., Krahenbuhl, P., Donahue, J., Darrell, T., Efros, A.A.: Context encoders: feature learning by inpainting. In: CVPR (2016)
23. Quan, Q., He, F., Li, H.: A multi-phase blending method with incrementalintensity for training detection networks. Vis Comput (2020). https://doi.org/10.1007/s00371-020-01796-7
24. Ren, W., Cao, X.: Deep video dehazing. In: Pacific-Rim Conference on Multimedia (2017)
25. Ren, W., Liu, S., Ma, L., Xu, Q., Xu, X., Cao, X., Du, J., Yang, M.H.: Low-light image enhancement via a deep hybrid network. IEEE Trans Image Process **28**(9), 4364–4375 (2019)
26. Ren, W., Liu, S., Zhang, H., Pan, J., Cao, X., Yang, M.H.: Single image dehazing via multi-scale convolutional neural networks. In: ECCV (2016)
27. Ren, W., Ma, L., Zhang, J., Pan, J., Cao, X., Liu, W., Yang, M.H.: Gated fusion network for single image dehazing. In: CVPR (2018)
28. Ren, W., Zhang, J., Ma, L., Pan, J., Cao, X., Zuo, W., Liu, W., Yang, M.H.: Deep non-blind deconvolution via generalized low-rank approximation. In: Advances in neural information processing systems, pp. 297–307 (2018)
29. Ruder, S.: An overview of gradient descent optimization algorithms (2016). arXiv:1609.04747
30. Sulami, M., Glatzer, I., Fattal, R., Werman, M.: Automatic recovery of the atmospheric light in hazy images. In: ICCP (2014)
31. Tan, R.T.: Visibility in bad weather from a single image. In: CVPR (2008)
32. Tan, X., Zhu, H., Shao, Z., Hou, X., Hao, Y., Ma, L.: Saliency detection by deep network with boundary refinement and global context. In: 2018 IEEE international conference on multimedia and expo (ICME), pp. 1–6. IEEE (2018)
33. Tang, K., Yang, J., Wang, J.: Investigating haze-relevant features in a learning framework for image dehazing. In: CVPR, pp. 2995–3000 (2014)
34. Tarel, J.P., Hautiere, N.: Fast visibility restoration from a single color or gray level image. In: ICCV (2009)
35. Yan, Y., Ren, W., Cao, X.: Recolored image detection via a deep discriminative model. IEEE Trans Inf Forensics Secur **14**(1), 5–17 (2018)
36. Yang, D., Sun, J.: Proximal dehaze-net: a prior learning-based deep network for single image dehazing. In: ECCV, pp. 702–717 (2018)
37. Yeh, R.A., Chen, C., Lim, T.Y., Schwing, A.G., Hasegawa-Johnson, M., Do, M.N.: Semantic image inpainting with deep generative models. In: CVPR, pp. 5485–5493 (2017)
38. Yong, J., He, F., Li, H., Zhou, W.: A novel bat algorithm based on cross boundary learning and uniform explosion strategy. Appl Math A J Chin Univ **34**, 482–504 (2019)
39. Yu, F., Koltun, V.: Multi-scale context aggregation by dilated convolutions (2015). arXiv:1511.07122
40. Yu, F., Koltun, V., Funkhouser, T.A.: Dilated residual networks. In: CVPR, vol. 2, p. 3 (2017)
41. Yu, H., He, F.: A scalable region-based level set method using adaptive bilateral filter for noisy image segmentation. Multimed. Tools Appl. **79**, 5743–5765 (2020)
42. Zhang, H., Patel, V.M.: Densely connected pyramid dehazing network. In: CVPR (2018)
43. Zhang, J., He, F.: A new haze removal approach for sky/river alike scenes based on external and internal clues. Multimed Tools Appl **79**, 2085–2107 (2020)

44. Zhang, S., He, F.: Drcdn: learning deep residual convolutional dehazing networks. Vis Comput (2019). https://doi.org/10.1007/s00371-019-01774-8
45. Zhang, S., He, F., Ren, W., Yao, J.: Joint learning of image detail and transmission map for single image dehazing. Vis Comput **2**, 305–316 (2020)
46. Zhang, S., He, F., Yao, J.: Single image dehazing using deep convolution neural networks. In: Pacific rim conference on multimedia, pp. 315–325. Springer (2017)
47. Zhang, S., Ren, W., Yao, J.: Feed-net: fully end-to-end dehazing. In: ICME (2018)
48. Zhang, S., Yao, J.: Single image dehazing using fixed points and nearest-neighbor regularization. In: Asian conference on computer vision, pp. 18–33 (2016)
49. Zhang, S., Yao, J., Garcia, E.B.: Single image dehazing via image generating. In: Pacific-Rim symposium on image and video technology, pp. 123–136. Springer (2017)
50. Zhang, Y., Ding, L., Sharma, G.: Hazerd: an outdoor scene dataset and benchmark for single image dehazing. In: ICIP, pp. 3205–3209. IEEE (2017)
51. Zhu, Q., Mai, J., Shao, L.: A fast single image haze removal algorithm using color attenuation prior. TIP **24**(11), 3522–3533 (2015)

Shengdong Zhang is currently a Ph.D. candidate in School of Computer Science, Wuhan University. His research interests are image processing, computer graphics and deep learning.

Fazhi He received bachelor, master and Ph.D. degrees from Wuhan University of Technology. He was a post-doctor researcher in The State Key Laboratory of CAD&CG at Zhejiang University, a visiting researcher in Korea Advanced Institute of Science & Technology and a visiting faculty member in the University of North Carolina at Chapel Hill. Now he is a professor in School of Computer Science, Wuhan University. His research interests are artificial intelligence, intelligent computing, computer graphics, image processing, computer-aided design, computer-supported cooperative work and co-design of software/hardware.

Wenqi Ren is an Assistant Professor in Institute of Information Engineering, Chinese Academy of Sciences, China. He received his Ph.D. degree from Tianjin University, Tianjin, China, in 2017. During 2015 to 2016, he was supported by China Scholarship Council and working with Prof. Ming-Husan Yang as a joint-training Ph.D. student in the Electrical Engineering and Computer Science Department, at the University of California at Merced. His research interests include image processing and related high-level vision problems.