



Deep Dilated Convolutional Network for Single Image Dehazing

S. Deivalakshmi^(✉) and J. Sudaroli Sandana

Department of Electronics and Communication Engineering, National Institute of Technology,
Tiruchirappalli, India
deiva@nitt.edu

Abstract. The visual quality of the images gets decreased due to bad weather conditions. The image captured under hazy weather conditions have serious attenuation in terms of color and saturation. In addition, these hazy images have very low contrast and the visual quality will be drastically poor. Moreover, object detection in hazy environment is too challenging. So, single image dehazing is a demanding, challenging and ill-posed problem. In this paper, we propose a 9-layer convolutional neural network with deep dilated filters of different dilation rates to achieve an end-to-end mapping from haze image to haze free image. Exponential expansion of receptive field is possible with the dilated filters without increasing the model complexity. Furthermore, the dilated convolutional layers help for efficient model compactness. We did experiments on synthetic dataset and on naturally obtained hazy images. The results show that our network achieves outstanding performance over the existing algorithms in terms of PSNR, SSIM and visual quality.

Keywords: dehazing · dilated filters · receptive field

1 Introduction

The visual quality of the image's acquired outdoors is decreased due to dust, smoke and other small particles accumulated in atmosphere. This type of traditional atmospheric situation is known as haze and it attenuates the light reflected by the objects along its path towards the camera. Haze causes a serious problem in terrestrial photography, as it becomes difficult to image distant objects because the light penetration of dense atmosphere is necessary. The light scattered by the hazy particles results in the visual effect of a loss of contrast in the subject. This poor visual quality can inhibit the performance of computer systems which are intended to operate on clear conditions. It also obscures the visual amicability of image contents for the users who use standard cameras. Hazy weather also affects variety of systems such as aerial photography systems [1], image classification [2], satellite remote sensing systems [3] and target recognition systems that depend on optical imaging instruments. In recent years, the need for restoring the quality of the visually degraded images due to poor weather conditions has been increasingly gaining the attention.

It is hard to detect haze because concentration of the haze differs from place to place. Haze is a depth dependent phenomenon. So, image dehazing is thus an ill-posed problem. The existing techniques of image processing which uses a variety of visual information to capture both statistical and deterministic properties of hazy images such as histogram based [4, 5] contrast based and saturation-based image dehazing methods are used by the early researchers to obtain hazy free image from a single image. Later, researchers try to achieve the same performance with multiple images. In [6], polarization-based image dehazing methods are used. In polarization based dehazing, multiple images are taken with different degrees of polarization. In [7], multiple images of the same scene are captured under distinct weather conditions. These are multi constraint based dehazing methods. Dehazing has been carried out based on the depth information from the user inputs. But practically, multiple hazy images or information related to depth are not always available. To overcome this difficulty, single image dehazing approaches were introduced. In recent years, single image dehazing based on the physical model achieved significant progress. It was carried out under the assumption that the local contrast of the haze-free image is very much higher than that in the image captured under haze. Recently, convolutional neural network (CNN) shows promising solutions on many vision tasks, including dehazing. CNNs have shown an explosive popularity because of capability in producing state-of-the-art performance. DehazeNet [8] has been proposed for single image haze removal with bilateral rectified linear unit as nonlinear activation function. The DehazeNet [9] models mapping functions between hazy image and their medium transmissions. The DehazeNet [9] assumed the atmospheric light as a global constant which must be learned along with medium transmission. Recently, Artificial Multiple Exposure Fusion (AMEF) [9] has been proposed for haze removal. The AMEF method [9] used two steps to remove haze. First of all, a sequence of gamma correction operations is used to artificially under-expose the hazy image. Secondly, a multi-scale Laplacian blending scheme is used to merge the resulting set of under-exposed images to yield haze-free image.

The DehazeNet [8] was modeled only using the information of medium transmissions and assumed the atmospheric scattering as a global constant, but in this paper both medium transmission and atmospheric scattering are considered and the network is trained in an end-to-end manner using deep dilated filters. The end-to-end training between hazy image and haze free image learns a single and unique mapping function which is able to remove haze from any real-world hazy image.

The major contributions of the proposed work are as follows.

- A novel 9-layer convolutional neural network is proposed for robust dehazing in an end-to-end manner by considering both medium transmissions and atmospheric scattering.
- The layers are constructed in symmetrical structure with respect to dilation rates. The dilation rate increases the receptive field thereby giving network more scope to exploit contextual information efficiently.
- The network is compactly modeled by selecting proper dilatation rates without causing gridding effect in the produced haze-free images.

The rest of this paper is summarized as follows. Section 2 provides the work related to dehazing. Section 3 presents the proposed methodology adopted to achieve image dehazing. Experimental analysis is presented in Sect. 4. The concluding remarks of the proposed work are provided in Sect. 5.

2 Related Work

Although wide variety of Image dehazing literature is available, it is still an open topic to investigate. There are three types of dehazing methods seen in current research: image enhancement-based methods, image fusion-based methods and image restoration methods based on physical modelling. The image enhancement-based method does not consider the specific cause of the image degradation. Fog is not removed from the image to restore the original appearance. So, these methods cannot be adaptable for different scenes and images. Image restoration methods based on physical modelling analyses the specific causes of image degradation. This method also establishes a deteriorate model of images degraded under fog. The physical modelling image restoration methods are based on the atmospheric scattering theory. According to this theory, the scattering of atmosphere is divided into two parts: one is due to the attenuation of light reflected from the surface of the object to the camera; and the second one is due to the airlight scattering reaching to the camera. To describe how a hazy image is formed, the atmospheric scattering model, which is proposed in 1976 [10], is mainly used in image processing and in computer vision [7]. Image fusion based methods maximize beneficial information from multiple sources to finally form a high quality image. These methods do not need a physical model, but the fusion strategy for multiple sources of information is complex. Image fusion is the process of combining relevant information from multiple source channels into a high quality image. Fusion strategies extract the information from each channel to improve the utilization of image information. These methods have also been used in image dehazing in recent years. A single image dehazing algorithm that removes the visual degradation due to haze is described in [9]. It does not depend on the

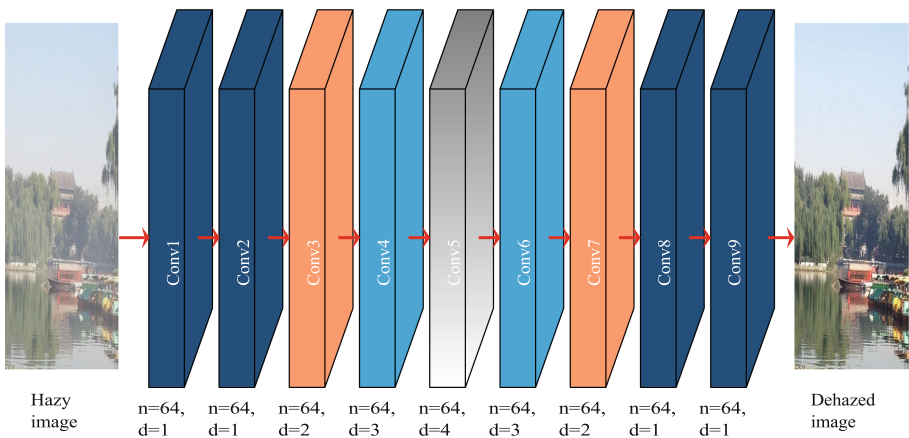


Fig. 1. Architecture of the proposed network

inversion of a physical model of haze formation but considered a few main underlying assumptions to get a haze-free image. A trainable end-to-end model to estimate medium transmission called DehazeNet has been proposed in [8]. It takes a degraded image due to haze as its input, and outputs its transmission matrix. It estimates the global atmospheric light by some empirical rules and haze-free image is recovered using the atmospheric scattering model.

3 Method

In this section, we present the architecture of the deep learning model used for dehazing and the benefits of dilated convolutions. The training procedure and model complexity are also explained.

3.1 Architecture

The deep dilated network used for dehazing is shown in Fig. 1. The proposed network contains 9 convolutional layers. The first two and the last two convolutional layers perform standard convolutions. All remaining layers have specified dilation rates (d) as shown in Fig. 1. Starting layers learns low-level features especially edges and blobs from the hazy image. Dilated convolution with rate d introduces $(d - 1)$ spaces between the adjacent pixels while performing convolution operation. The early layers especially, first layer will learn high frequency information like edges. While applying dilated convolution some of the information about edges may be lost. So, we used normal convolutional filters in the first layer. Similarly, to reconstruct dehazed image accurately, we used normal convolution in reconstruction layer. After each convolutional layer the haze is removed step-by step and at the final layer the dehazed image is produced. Different from the works in [8, 9], our proposed network learns an end-to-end mapping function (F) directly from hazy image to haze-free image. As the deep learning model learns features automatically, there might be a possibility of feature redundancy. Same features may be learned by network in different layers. To avoid feature redundancy, we used dilated filters with different dilation.

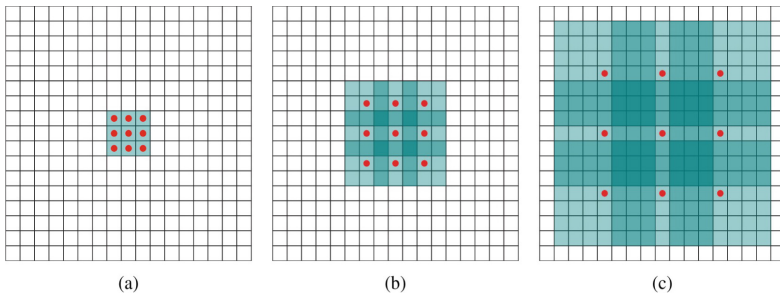


Fig. 2. Dilated convolution operation

rates (d). All the convolutional layers in the proposed network have 64 filters. Each filter has a size of 3×3 . The first layer learns 64 feature maps from the hazy color

image. In the second layer these 64 features are mapped to another feature space. In the third layer, the d of 2 is used. So, the effective filter size (f_e) becomes 5. Therefore, the third layer can exploit more information by expanding receptive field.

3.2 Dilated Convolution

The benefit of using dilated convolution is three folds. First, the dilated convolution can exploit more contextual information. Secondly, by using dilated filters computational complexity will not increase but receptive field can be increased. Thirdly, model can be designed as compactly as possible with a smaller number of convolutional layers. Basically, in low-level vision, exploiting more contextual information is essential to yield better performance. This is possible only with filters of large sizes. Employing large size filters will increase the computational complexity of the model. The dilated filters will consider spaces between each pixel in the image while doing convolution operation. The dilated convolution operation is shown in Fig. 2.

Figure 2(a) represents the standard convolution with filter of size 3×3 . Figure 2(b) represents dilated convolution with $d = 2$. A single space is considered while doing convolution operation between each pixel. So, the dilated convolution introduces corresponding spaces ($d - 1$ spaces) between pixels depending upon d and performs standard convolution thereafter. Similarly, Fig. 2(c) represents dilated convolution with $d = 7$. So, the effective filter size becomes 15. We generalize the effective filter size (f_e) with d as follows:

Table 1. Proposed Network Specifications

| Layer Index (l) | F | d | f_e | RF |
|-----------------|---|---|-------|----|
| 1 | 3 | 1 | 3 | 3 |
| 2 | 3 | 1 | 3 | 5 |
| 3 | 3 | 2 | 5 | 9 |
| 4 | 3 | 3 | 7 | 15 |
| 5 | 3 | 4 | 9 | 23 |
| 6 | 3 | 3 | 7 | 29 |
| 7 | 3 | 2 | 5 | 33 |
| 8 | 3 | 1 | 3 | 35 |
| 9 | 3 | 1 | 3 | 37 |

$$f_e = f + (f - 1)(d - 1) \quad (3)$$

where f is the standard filter size. Similarly, the receptive field (RF) of the network can be formulated based on the depth (D) of the network as follows:

$$RF = f + (f - 1)(D - 1) \quad (4)$$

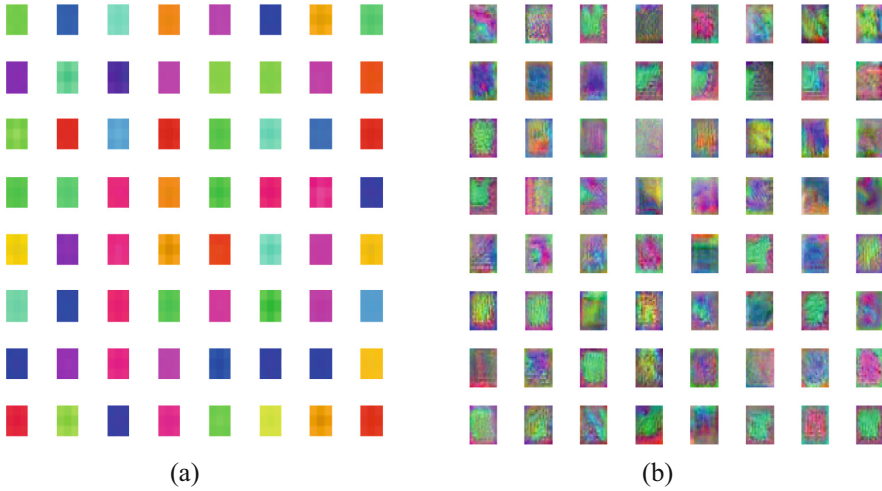


Fig. 3. Visualization of different layers

Equation 4 represents the RF for fully convolutional neural network. If dilated filters are used in the network, then the Eq. 4 is modified as follows:

$$RF = f_e + (f_e - 1)(D - 1) \tag{5}$$

In the proposed network, total 9 convolutional layers are used. The dilation rate (d), filter size (f) effective filter size (f_e) and receptive field (RF) are tabulated as shown in Table 1. The total receptive field of the proposed network is 37. So, we cropped patches of size 37×37 from the hazy and haze free images to generate training samples. From Table 1, 9 layers are enough to exploit the whole content in the 37×37 image patch. Adding extra layers will increase the model complexity. Moreover, performance improvement is also marginal. So, we used only 9-layers in the proposed network for dehazing.

3.3 Computational Complexity

Computational complexity is one of the key factors while designing any network. It will decide the real time deployment of the network. Theoretically, computational complexity is the number of parameters that the network must be optimized.

The computational complexity of any fully convolutional neural network is calculated by using the following formulation:

$$C = \sum_{l=1}^D n_{l-1} \times f^2 \times n_l \times S_l^2 \tag{6}$$

where l is the layer index, D is the depth of the network and n_1 is the number of filters in l^{th} layer. Here, n_0 is equals to number of channels (3 color channels) in input patch. S is the size of the image to be dehazed. The practical computational complexity varies from the theoretical complexity as it depends on the hardware and type of framework used for implementation.

3.4 Loss Function

Let $\{H^i, D^i\}$ represents the training samples. Where i stands for the i^{th} training sample. Similarly, let $\Theta = \{W^l, B^l\}$ represents the network parameters of individual layers. We minimize the Mean Squared Error (MSE) $L(\Theta)$ as follows:

$$L(\Theta) = \frac{1}{K} \sum_{i=1}^K \|f(H^i, \Theta) - D^i\|^2 + \lambda \|\Theta\|^2 \quad (7)$$

where λ denotes the regularization factor. We used Adam optimizer [12] with the parameters β_1, β_2 and ϵ are set to 0.9, 0.999 and 10^{-8} respectively. We used standard mini-batch gradient descent with each batch size of 64. The gradients with respect to l^{th} layer are computed as:

$$g_t = \nabla_{\Theta_t} L(\Theta_t - 1) \quad (8)$$

The momentum vectors are formulated as below:

$$\begin{aligned} m_t &= \beta_1 m_{t-1} + (1 - \beta_1) g_t \\ v_t &= \beta_2 v_{t-1} + (1 - \beta_2) g_t^2 \end{aligned} \quad (9)$$

Finally, the parameter update is as follows:

$$\Theta_t = \Theta_{t-1} - \frac{\eta \hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon} \quad (10)$$

where \hat{m}_t and \hat{v}_t are used to counteract the occurrence of zero vectors due to zero initialization of m_t and v_t and computed as:

$$\begin{aligned} \hat{m}_t &= \frac{m_t}{1 - \beta_1^t} \\ \hat{v}_t &= \frac{v_t}{1 - \beta_2^t} \end{aligned} \quad (11)$$

3.5 Visualization of the Network

The proposed network adopts standard and dilated convolution operations in different layers, we analyzed the features learned by different convolutional layers in this subsection. In Fig. 3(a), we show the features learned by first convolutional layer. It contains mostly the edges and colors which are required for later convolutional layers in removing haze efficiently. Figure 3(b) represents the features learned in convolutional layer with $d = 4$. These features are combination of lower layer features, and the representations are very complicated as shown in Fig. 3(b). The high-level features can be seen as features that are encoded by higher levels of visual cortex. The size of the features learned in both the layers is same (i.e., 3×3) but the layer with dilated filter exploited contextual information of size 9×9 by inserting $d - 1$ spaces (i.e., 3) between each pixel. The number of parameters of each filter in both the layers are the same.

4 Experiments

In this section, we explain the experimental analysis for dehazing. The datasets, network parameters, comparison with latest dehazing methods and finally, the quantitative and qualitative analysis are also the part of this section.

4.1 Dataset

For training, validation and testing, we used the RESIDE [11] benchmark dehaze dataset. The RESIDE dataset contain synthesized images from depth and stereo datasets as large-scale training and testing hazy image pairs. We considered 100 images for training from the training pairs of RESIDE dataset. A total of 18960 training samples each with a size of 37×37 is cropped from the 100 images. 20 separate images are cropped to produce validation set. Two different datasets test1 and test2 with 100 and 500 images respectively from [11] are considered for testing.

4.2 Training Details

In each layer 64 filters of size 3×3 is used. Initial learning rate is set to 10^{-4} . The learning rate is reduced by 10% after 50 epochs. We trained our network for 100 epochs. The validation error is constant after 50 epochs. So, we reduced learning rate by 10% and terminated learning after 100 epochs. The validation error plateaus after 100 epochs. The regularization factor (λ) is set to 10^{-5} . Gradient clipping is not used as there is no chance of gradient explosion because of using low learning rate. The network is trained on a machine with 32 GB RAM, Intel core I5 processor and NVIDIA GeForce 710 GPU. MATLAB is used for both training and testing.

4.3 Metrics and Compared Methods

The widely used peak signal to noise ratio (PSNR) and Structural SIMilarity (SSIM) index measure are used for quantitative performance comparison. AMEF [9] and DehazeNet [8] methods are used for comparison.

$$PSNR(I_G, I_D) = 10 \log_{10} \frac{I_{G_{max}}^2}{\frac{1}{N} \sum_{i=1}^N (I_{G_i} - I_{D_i})^2} \quad (12)$$

$$SSIM(I_G, I_D) = \frac{(2\mu_{I_G}\mu_{I_D} + C_1)(2\sigma_{I_G I_D} + C_2)}{(\mu_{I_G}^2 + \mu_{I_D}^2 + C_1)(\sigma_{I_G}^2 + \sigma_{I_D}^2 + C_2)} \quad (13)$$

where μ_{I_G} , μ_{I_D} , $\sigma_{I_G}^2$, $\sigma_{I_D}^2$, $\sigma_{I_G I_D}$ are local mean, variance and cross-covariance for images I_G and I_D respectively.

4.4 Quantitative Metrics

Tables 2 and 3 represents the quantitative metrics comparison on two different test sets with test1 and test2 containing 100 and 500 test images respectively. The proposed method is consistently the top performer for both the metrics PSNR and SSIM. An improvement of 1.69 dB PSNR is achieved with the proposed method when compared to the next best DehazeNet [8]. AMEF [9] and DehazeNet [10] are the state-of-the-art methods for dehazing. Moreover, for fair comparison with AMEF [9] and DehazeNet [10], the proposed network is trained with the same training dataset.

4.5 Qualitative Analysis

Figures 4 and 5 represents the visual comparison of dehazed images with different methods. The visual quality of the image produced with the proposed method is good when compared to the images produced with other methods. As one can see from Figs. 4 and 5, the AMEF restored images contain color artifacts.

Table 2. Average Performance Analysis on Test1

| Metrics | Hazy | AMEF [9] | DehazeNet [8] | Proposed Method |
|-----------|--------|----------|---------------|-----------------|
| PSNR (dB) | 15.97 | 17.68 | 23.48 | 25.17 |
| SSIM | 0.7992 | 0.8349 | 0.8915 | 0.9393 |

Table 3. Average Performance Analysis on Test2

| Metrics | Hazy | AMEF [9] | DehazeNet [8] | Proposed Method |
|-----------|--------|----------|---------------|-----------------|
| PSNR (dB) | 14.77 | 17.32 | 22.92 | 24.87 |
| SSIM | 0.7932 | 0.8149 | 0.8715 | 0.9182 |

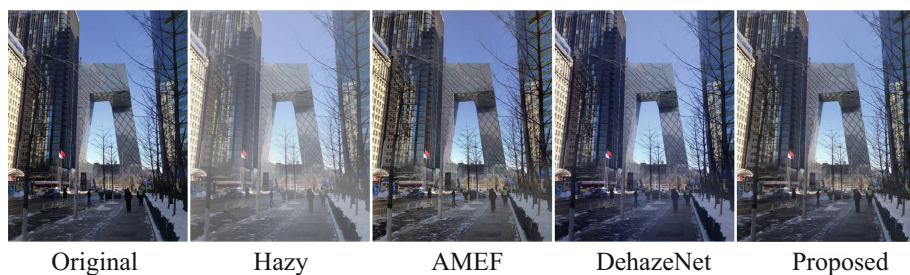


Fig. 4. Qualitative comparison with different methods for sample image 1



Fig. 5. Qualitative comparison with different methods for sample image 2

4.6 Run Time

For real time applications, the run time is as important as performance. In this subsection, we present the testing times of compared methods on an image of size 512×512 in Table 4. The testing is performed on CPU. The proposed method is ≈ 1.8 and 2.57 times faster than the AMEF [9] and DehazeNet [8] respectively.

Table 4. Testing Time (In Seconds) for a 512×512 Image

| Method | AMEF [8] | DehazeNet [7] | Proposed method |
|----------|----------|---------------|-----------------|
| Run time | 2.65 | 3.78 | 1.47 |

5 Conclusion

In this paper, we proposed a novel deep dilated convolutional neural network for single image dehazing. Furthermore, we considered dilated convolutions in different layers to exploit more image content without increasing the network complexity. We avoided adopting higher dilation rates (e.g., 5, 6 etc.) to overcome the gridding effect caused due to leaving more spaces between pixel elements. The proposed method produced superior performance in terms of quantitative metrics and perceptual quality when compared to other competitive methods in terms of quantitative metrics like PSNR and SSIM. In future, the proposed network will be modified with dense skip connections to achieve further improvement in the performance without increasing the network complexity.

References

1. Woodell, G., Jobson, D.J., Rahman, Z., Hines, G.: Advanced image processing of aerial imagery. In: Proceedings of the SPIE 6246, Visual Information Processing XV, 62460E, 12 May 2006
2. Shao, L., Liu, L., Li, X.: Feature learning for image classification via multiobjective genetic programming. *IEEE Trans. Neural Networks Learn. Syst.* **25**(7), 1359–1371 (2014). <https://doi.org/10.1109/TNNLS.2013.2293418>

3. Liu, Q., Gao, X., He, L., Lu, W.: Haze removal for a single visible remote sensing image. *Signal Processing* **137**, 3343 (2017). ISSN 0165-1684, <https://doi.org/10.1016/j.sigpro.2017.01036>
4. Kim, T.K., Paik, J.K., Kang, B.S.: Contrast enhancement system using spatially adaptive histogram equalization with temporal filtering. *IEEE Trans. Consumer Electron.* **44**(1), 82–87 (1998). <https://doi.org/10.1109/30.663733>
5. Stark, J.A.: Adaptive image contrast enhancement using generalizations of histogram equalization. *IEEE Trans. Image Process.* **9**(5), 889–896 (2000). <https://doi.org/10.1109/83.841534>
6. Schechner, Y., Narasimhan, S., Nayar, S.: Polarization-based vision through haze. *Appl. Opt.* **42**, 511–525 (2003)
7. Narasimhan, S.G., Nayar, S.K.: Contrast restoration of weather degraded images. *IEEE Trans. Pattern Analysis and Machine Intell.* **25**(6), 713–724 (2003). <https://doi.org/10.1109/TPAMI.2003.1201821>
8. Cai, B., Xu, X., Jia, K., Qing, C., Tao, D.: DehazeNet: an End-to-End system for single image haze removal. *IEEE Trans. Image Process.* **25**(11), 5187–5198 (2016). <https://doi.org/10.1109/TIP.2016.2598681>
9. Galdran, A.: Image dehazing by artificial multiple-exposure image fusion. *Signal Process.* **149**, 135–147 (2018). ISSN 0165-1684, <https://doi.org/10.1016/j.sigpro.2018.03.008>
10. Cantor, A.: Optics of the atmosphere—scattering by molecules and particles. *IEEE J. Quant. Electron.* **14**(9), 698–699 (1978). <https://doi.org/10.1109/JQE.1978.1069864>
11. Li, B., et al.: Benchmarking single-image dehazing and beyond. *IEEE Trans. Image Process.* **28**(1), 492–505 (2019). <https://doi.org/10.1109/TIP.2018.2867951>
12. Kingma, D., Ba, J.: Adam: a method for stochastic optimization. In: *International Conference on Learning Representations* (2014)