Hindawi
Mathematical Problems in Engineering
Volume 2020, Article ID 4945214, 11 pages
https://doi.org/10.1155/2020/4945214

Research Article

FAOD-Net: A Fast AOD-Net for Dehazing Single Image

Wen Qian,[1,2] **Chao Zhou**,[1,2] **and Dengyin Zhang**[2,3]

[1]*College of Telecommunications & Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210003, China*
[2]*Jiangsu Key Laboratory of Broadband Wireless Communication and Internet of Things, Nanjing University of Posts and Telecommunications, Nanjing 210003, China*
[3]*School of Internet of Things, Nanjing University of Posts and Telecommunications, Nanjing 210003, China*

Correspondence should be addressed to Dengyin Zhang; zhangdy@njupt.edu.cn

Received 24 November 2019; Revised 14 January 2020; Accepted 22 January 2020; Published 24 February 2020

Academic Editor: Łukasz Jankowski

In this paper, we present an extremely computation-efficient model called FAOD-Net for dehazing single image. FAOD-Net is based on a streamlined architecture that uses depthwise separable convolutions to build lightweight deep neural networks. Moreover, the pyramid pooling module is added in FAOD-Net to aggregate the context information of different regions of the image, thereby improving the ability of the network model to obtain the global information of the foggy image. To get the best FAOD-Net, we use the RESIDE training set to train our proposed model. In addition, we have carried out extensive experiments on the RESIDE test set. We use full-reference and no-reference image quality evaluation indicators to measure the effect of dehazing. Experimental results show that the proposed algorithm has satisfactory results in terms of defogging quality and speed.

1. Introduction

Many cities are shrouded in smog due to waste incineration, construction dust, and automobile exhaust. Images taken in smog weather are not clear due to contrast and color saturation, which affects the use of target detection and traffic monitoring. Therefore, there is an urgent theoretical and practical need to improve the image quality of foggy days. With the development of computer technology, image dehazing technology is widely used in civil and military fields, such as remote sensing, target detection, and traffic monitoring. At present, image dehazing algorithms can be mainly divided into three types. The first type is an image enhancement-based defogging algorithm. This algorithm does not consider the imaging mechanism of degraded images and turns the image dehazing problem into a contrast-enhanced problem to highlight the details of the image and enhance the overall contrast of the image. Commonly used image enhancement-based defogging algorithms are included in [1–3]. The second type is an image dehazing algorithm based on the physical model. This algorithm analyzes the causes of foggy image formation, establishes an imaging model, and then performs inversion and calculation according to the model to obtain the image before degraded. The image dehazing algorithm based on the physical model is mainly included in [4–6]. In recent years, deep learning techniques have been widely used in the field of image processing, such as image classification, object recognition, and face recognition. The third type is that the existing image dehazing algorithm in studies such as [7–9] based on deep learning mostly estimates the transmittance of foggy images through neural network model, then estimates the atmospheric light value separately, and finally obtains the fog-free image according to the atmospheric scattering model. However, the estimation is not always accurate. In DCP [5], the estimation of atmospheric illumination is based on prior knowledge of dark channels. First, the pixel values of the solved dark channel map are sorted, and the top one-thousandth of the pixels in the sorted pixels are selected as candidate points. These points are mapped to the response positions in the original image, then the brightness values of the corresponding positions in the original image are obtained, and the largest of these candidate values is used as the atmospheric light value. In DehazeNet [8] and MSCNN [9],

the transmission map is estimated by a convolutional network model, and the atmospheric light value is estimated based on the prior knowledge of the dark channel. However, when the colors of the objects in the image are close to the atmospheric light, such as when the image contains a large number of white objects or other light sources, the estimated atmospheric illumination may be biased, making the image after defogging overexposed. Moreover, the nonjoint estimation of two critical parameters, transmission matrix and atmospheric light, may further amplify the error when applied together. Li et al. proposed an efficient end-to-end dehazing convolutional neural network (CNN) model, called all-in-one dehazing network (AOD-Net) [10]. AOD-Net [10] is designed based on a reformulated atmospheric scattering model. The output of the model is a clean image, rather than the transmittance map. Experiments demonstrate the superiority of AOD-Net [10] over several state-of-the-art methods.

With the development of cloud computing and Internet of Things technologies, convolutional neural network models are widely used in various terminals and platforms, including autonomous driving and augmented reality. These application scenarios require high memory and training speed of the model, so the convolution needs to be developed in a lightweight direction. At present, the classic lightweight convolution has SqueezeNet [11], ShuffleNet [12], and MobileNets [13]. SqueezeNet [11] adopts a different method than the traditional convolution method, and the fire module has two parts: the squeeze layer and the expand layer. ShuffleNet [12] arbitrarily scrambles the channels of the feature maps of each part to form a new feature map to solve the problem of poor information flow caused by group convolution. MobileNet [13] uses a convolution method called depthwise separable convolution instead of the standard convolution to achieve the purpose of reducing network weight parameters. MobileNet [13] is a network model that can be applied to the mobile side.

One of the keys to image defogging in complex scenes is to obtain global information about foggy images in complex scenes. The pyramid pooling module combines the features of different pyramid scales to fully extract the global information of the foggy image, making the image after defogging clearer and more natural. At present, the pyramid pooling module is mainly included in [14, 15]. He et al. proposed SPPnet [14], which solves the problem that the input of the deep convolutional neural network must require a fixed image size and improve the efficiency of extracting features. Zhao et al. proposed a pyramid pooling module (PSP) [15] that combines multiscale pooling features, which can aggregate context information from different regions to improve the ability to obtain global information.

The contributions of this paper are summarized as follows:

(1) We propose a new end-to-end network model called FAOD-Net for image dehazing. This model uses a lightweight convolution depthwise separable convolution instead of the standard convolution in AOD-Net [10]. Moreover, we analyze the advantages of using depthwise separable convolution instead of standard convolution.

(2) In order to aggregate the context information of different areas of the foggy image, we added a pyramid pooling module to FAOD-Net. This module combines the features of 4 different pyramid scales, which can improve the ability of the network model to obtain global information.

(3) We use the classic RESIDE training set [16] to train FAOD-Net. Then, we test our proposed algorithm on the synthetic objective testing set (SOTS) and hybrid subjective testing set (HSTS) [16]. Moreover, we use the full-reference image quality evaluation indicators PSNR and SSIM to measure the algorithm's defogging effect on the synthetic foggy test set. For the real-world foggy test set, we use the no-reference image quality assessment indicators spatial-spectral entropy-based quality (SSEQ) [17] and blind image integrity notator using DCT statistics (BLIINDS-II) [18] to measure the dehazing effect.

2. Background

2.1. Atmospheric Scattering Model. To describe the formation of a hazy image, the atmospheric scattering model is first proposed by McCartney [19], which is further developed by Narasimhan and Nayar [20, 21]. The atmospheric scattering model can be formally written as

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

where $I(x)$ is the observed intensity, $J(x)$ is the intensity of light coming from the scene objects and before getting scattered, $t(x)$ is the scene transmittance denoting the amount of light that reaches the observer after getting scattered, and A denotes the global environmental illumination. Moreover, $t(x)$ is the middle transmission matrix defined as

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

where β is the atmospheric scattering coefficient, and the uniform concentration of the fog can be approximated as a constant; γ is the wavelength of the reflected light; and $d(x)$ is the depth of the scene, that is, the distance between the corresponding object in the scene and the imaging device.

2.2. Deformation Formula of Atmospheric Scattering Model. From the atmospheric scattering model, the key to restoring a fog-free image is to estimate the transmittance of the fog map and the corresponding atmospheric light value. Li et al. introduced a new variable $K(x)$ by deforming the atmospheric scattering model [10], so the neural network model can directly estimate the joint value of transmittance and atmospheric light. Formula (1) is modified as follows:

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

$$J(x) = K(x)I(x) - K(x) + b, \tag{4}$$

$$K(x) = \frac{(1/t(x))(I(x) - A) + (A - b)}{I(x) - 1}. \tag{5}$$

By jointly estimating the transmittance and the atmospheric light value, it is possible to avoid the problem that the atmospheric light is estimated to be large due to the influence of the white area or the sky area.

2.3. Depthwise Separable Convolution. The MobileNets paper [13] proposed a depthwise separable convolution, which is a form of decomposed convolution. It solves standard convolution integrals into depthwise convolution and pointwise convolution. This decomposition has the effect of greatly reducing calculations and model size. Figure 1(a) shows the standard convolution. The input of the standard convolution is a feature map P of $H \times W \times C$. Among them, H is the height of the input feature map, W is the width of the input feature map, and C is the number of channels of the input feature map. We use N filters with a convolution kernel size $k \times k$ to perform standard convolution on the input feature map P and use appropriate stride and padding to ensure that the output feature map F is of size $H \times W \times N$. Figure 1(b) shows a depthwise convolution, which groups the same input feature map P according to the number of channels and then convolves each group of feature maps, where the convolution kernel size is $k \times k$. The output is a depthwise feature map. Figure 1(c) shows the pointwise convolution. It performs N convolutions with a kernel size of 1×1 on the depthwise feature map. The final output is the same as the output of the standard convolution.

3. The Proposed Method

The key to defogging based on the atmospheric scattering model is to estimate the transmission rate and atmospheric illumination. However, accurately estimating the transmission rate and atmospheric illumination value is a difficult task. In the DCP [5], the atmospheric illumination value is estimated through the dark channel prior knowledge. When the image contains a large number of white objects or other light sources, it will cause the estimated atmospheric light value to be too high, making the image after defogging overexposed. This phenomenon was demonstrated in the experiments in Section 4.3. It can be seen from Section 2 that it is possible to recover fog-free images by building an end-to-end model without having to estimate the transmission rate and atmospheric illuminance values separately. In addition, image defogging is often used for advanced computer vision tasks such as foggy target recognition and video defogging. Therefore, the performance of the defogging model has higher requirements on the calculations and model size. To this end, based on the conversion formula of the atmospheric scattering model, this paper uses the depthwise separable convolution instead of the standard convolution and increases the pyramid pooling model to extract the global information. The model (FAOD-Net) can be divided into two parts: the first part is a neural network model based on depthwise separable convolutions of different scales and pyramid pooling model. The model can estimate the joint value $K(x)$ of transmittance and atmospheric light from multiple channels; the second part substitutes the output of the first part into the atmospheric scattering model deformation formula to recover corresponding fog-free images. The architecture of FAOD-Net is shown in Figure 2.

3.1. FAOD-Net for Estimating $K(x)$. The task of this section is to estimate the combined value $K(x)$ of the transmittance and atmospheric light of the input foggy image. The model includes input layer, depthwise separable convolution layers (DS-Conv) of different scales, excitation layers, and different combinations of connection layers, where the input layer is a foggy image $I(x)$, and the depthwise separable convolution layer DS-Conv1 is divided into a depthwise convolution (DW-Conv) and a pointwise convolution (PW-Conv).

The depthwise convolution first divides the input image $I(x)$ into three groups according to the RGB color channel and uses a Gaussian filter to convolve each group of channels separately. The result after depthwise convolution is $\mathbf{F}_{1a}^{c}$:

$$\mathbf{F}_{1a}^{c} = \mathbf{W}_1 * \mathbf{I}^{c} + \mathbf{B}_1, \quad c \in [\mathrm{R}, \mathrm{G}, \mathrm{B}], \tag{6}$$

where $\mathbf{I}^{c}$ represents a matrix of pixel values representing a color channel of the input image R, G, and B color spaces and $\mathbf{W}_1$ and $\mathbf{B}_1$ represent the weight coefficient matrix and the deviation matrix of the corresponding convolution network, respectively.

The pointwise convolution uses Gaussian filters to simultaneously convolve all channels of $\mathbf{F}_{1a}^{c}$, where the number of filters is k and the convolution kernel size is $1*1$. The result after pointwise convolution is $\mathbf{F}_{1b}$:

$$\begin{aligned} \mathbf{F}_{1b} &= \mathbf{W}_2 * \mathbf{F}_{1a} + \mathbf{B}_2, \\ \mathbf{F}_{1a} &= \bigcap\{\mathbf{F}_{1a}^{c}\}, \quad c \in [\mathrm{R}, \mathrm{G}, \mathrm{B}], \end{aligned} \tag{7}$$

where $\mathbf{F}_{1a}$ represents a matrix of pixel values after fusion of all color channels in $\mathbf{F}_{1a}^{c}$ and $\mathbf{W}_2$ and $\mathbf{B}_2$ represent the weight coefficient matrix and the deviation matrix of the corresponding convolution network, respectively.

The excitation layer uses the modified linear unit ReLU activation function to perform nonlinear regression on the output result $\mathbf{F}_{1b}$ of the depthwise separable convolutional layer to obtain $\mathbf{F}_1$:

$$\mathbf{F}_1 = \max(0, \mathbf{F}_{1b}). \tag{8}$$

Similarly, $\mathbf{F}_1$ is used as the input of DS-Conv2, and the output of DS-Conv2 is used as the input of the excitation layer to obtain $\mathbf{F}_2$. Concat1 can be obtained by splicing $\mathbf{F}_1$ and $\mathbf{F}_2$ by channel dimension. Concat1 is used as the input of DS-Conv3, and the output result is passed through the excitation layer to obtain $\mathbf{F}_3$. Concat2 can be obtained by splicing $\mathbf{F}_2$ and $\mathbf{F}_3$ by channel dimension. Concat2 is used as

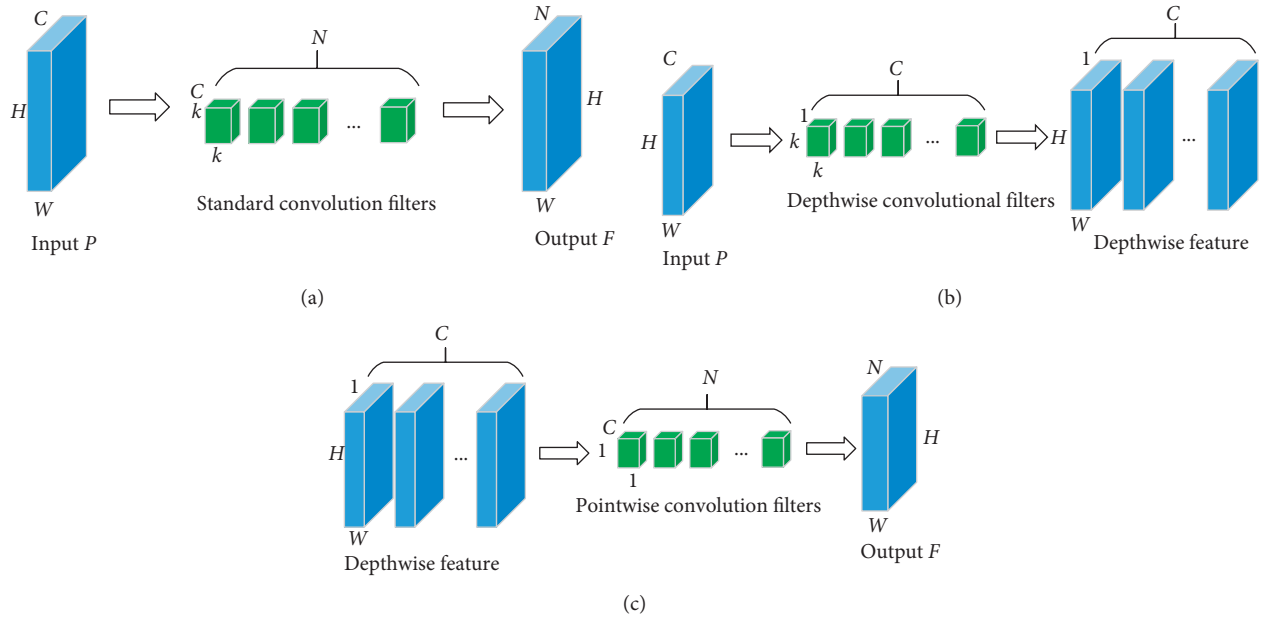

Figure 1: (a) The process of standard convolution. (b) The process of depthwise convolution. (c) The process of pointwise convolution. H, W, and C are the height, width, and number of channels of the input feature map P, respectively. N is the number of filters, and k is the convolution kernel size.

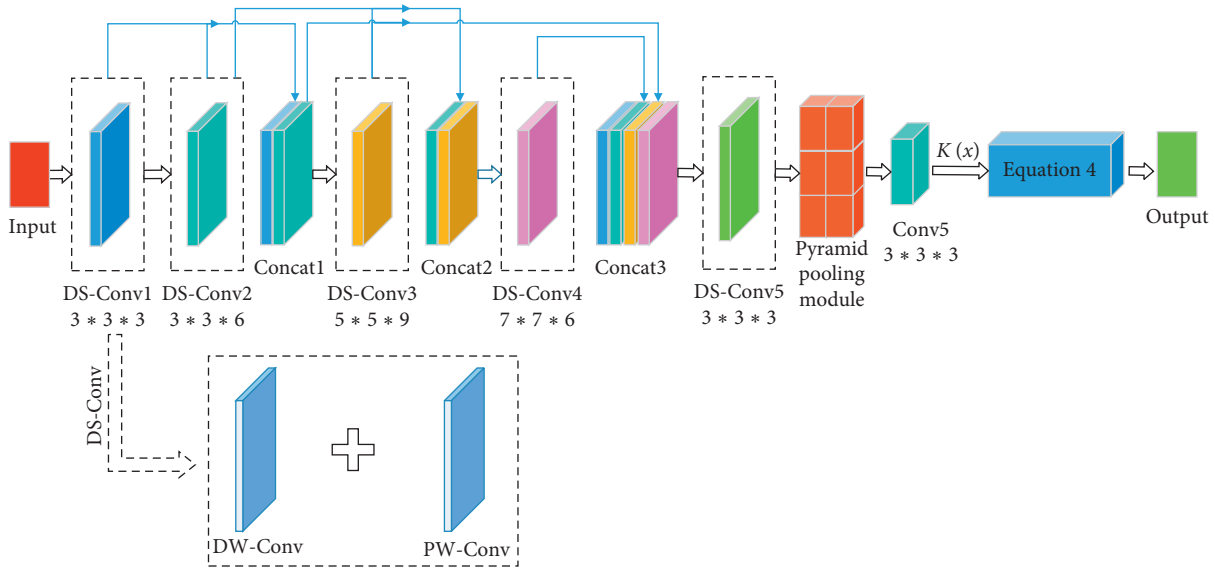

Figure 2: The architecture of FAOD-Net. DS-Conv is the depthwise separable convolution layers. DW-Conv is the depthwise convolution. PW-Conv is the pointwise convolution. k in $k * k * N$ is the size of the convolution kernel, and N is the number of filters.

the input of DS-Conv4, and the output result is passed through the excitation layer to obtain $\mathbf{F}_4$. Finally, $\mathbf{F}_1$, $\mathbf{F}_2$, $\mathbf{F}_3$, and $\mathbf{F}_4$ are spliced by channel dimension to get Concat3. Concat3 is used as the input of DS-Conv5.

The output result of DS-Conv5 is the input to the pyramid pooling module (PSP). The PSP architecture is shown in Figure 3. In this paper, the context information of different regions is aggregated through four different scale pooling layers. The pooling kernel sizes are $4 * 4$, $8 * 8$, $16 * 16$, and $32 * 32$. In order to guarantee the weight of the global features, if the pyramid has N levels, then using a $1 * 1$ convolution after each level will reduce the level channel to the original $1/N$. Then, by upsampling, the size before the pool is obtained and is finally concat together.

Finally, the result of the output of the pyramid pooling module is convolved, where the convolution kernel size is $3 * 3$ and the number of filters is three. The result after convolution is the estimated value of $K(x)$.

3.2. FAOD-Net for Recovering Fog-Free Images $J(x)$. According to the atmospheric scattering model deformation equation (4) mentioned in 2.2, the fog-free image can be

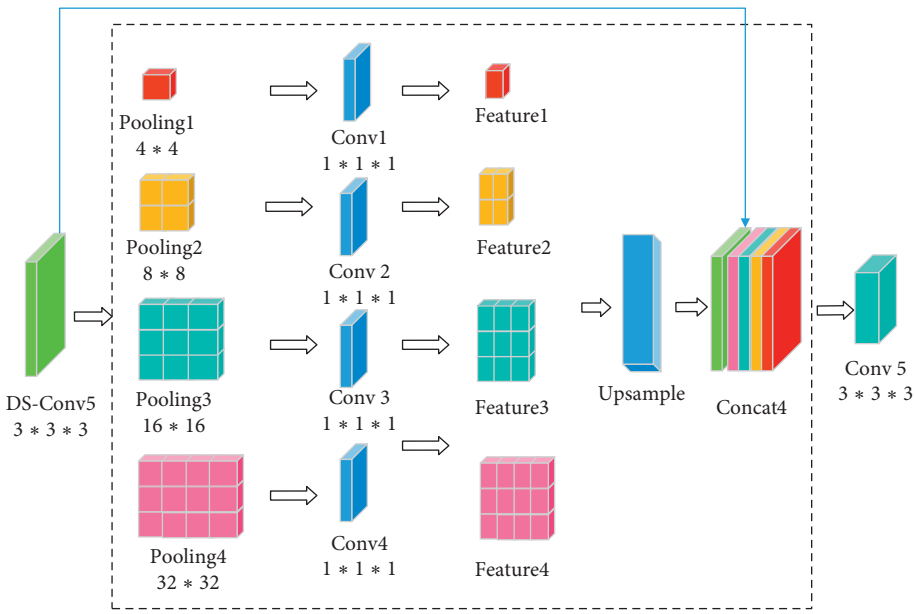

FIGURE 3: The architecture of PSP. DS-Conv is the depthwise separable convolution layers. k in $k * k * N$ is the size of the convolution kernel, and N is the number of filters.

restored by substituting the output result $K(x)$ of FAOD-Net in 3.1 into equation (4), as shown in Figure 2. Based on the depthwise separable convolution and pyramid pooling module, we designed an end-to-end image dehazing neural network model with the input as foggy images and the output as clear images. Through extensive experiments, the model has satisfactory performance in the defogging effect.

3.3. Advantages and Innovation of FAOD-Net. The FAOD-Net model is based on depthwise separable convolution. This convolution divides the standard convolution into a depthwise convolution and a pointwise convolution, convolving each channel of the foggy image. In this way, the separation of the image channel and the image space region is realized, and the fog characteristics in the image are better extracted. Also, using a depthwise separable convolution instead of a standard convolution can speed up the calculation. Acceleration calculation is mainly reflected in the reduction of the parameter quantity. Assume that the number of input image channels is 3. If the number of output channels required is 240, there are two implementations. In the first method, the input image is convoluted using a filter with a convolution kernel size of $3 * 3$, and the number of filters is 240, and the required parameter amount for training is $3 \times 3 \times 3 \times 240 = 6480$. The second way uses depthwise separable convolution, divided into depthwise convolution and pointwise convolution. The depthwise convolution does a $3 * 3$ convolution for each channel of the input image. The pointwise convolution convolves the result of the depthwise convolution with a convolution kernel size of $1 * 1$ and filter number of 240. The total required parameter amount is $3 \times 3 \times 3 + 3 \times 1 \times 1 \times 240 = 747$. From the comparison of the required parameter quantities, it can be concluded that the use of depthwise separable convolution instead of standard convolution can greatly reduce the amount of parameters. We test AOD-Net [10] and FAOD-Net on the Pytorch 0.4.1 framework. We found that the number of training parameters required for AOD-Net was 1761. After using depthwise separable convolution instead of standard convolution, the number of training parameters reduced to 717.

3.4. Training of FAOD-Net. In the FAOD-Net, learning the mapping relationship between hazy images and corresponding clean images is achieved by minimizing the loss between the training result $J_i(x)$ and the corresponding ground truth image $J_i^*(x)$. We use mean squared error (MSE) as the loss function:

$$L(J_i(x), J_i^*(x)) = \frac{1}{n}\sum_{i=1}^{n}\left\|J_i(x) - J_i^*(x)\right\|^2, \tag{9}$$

where n is the number of hazy images in the training set. We minimize the loss function using the stochastic gradient descent method with the backpropagation learning rule [22–24]. By training the FAOD-Net, we can directly get clear images corresponding to the foggy images.

4. Experiments and Results

In this section, the proposed FAOD-Net is tested with both qualitative and quantitative analysis. The full-reference image quality evaluation indicators PSNR and SSIM and no-reference image quality evaluation indicators SSEQ and BLIINDS-II are considered for the quantitative analysis. Furthermore, we compare it with the state-of-the-art methods, including boundary constrained context

regularization (**BCCR**) [25], dark-channel prior (**DCP**) [5], color attenuation prior (**CAP**) [6], **MSCNN** [9], **DehazeNet** [8], and **AOD-Net** [10].

4.1. Training Data and Experimental Settings. It is difficult to obtain a blurred image and its corresponding blurred image in a natural environment. We use the classic RESIDE dataset [16] to train and verify FAOD-Net. The RESIDE training set contains 13,990 synthetic foggy images. In this dataset, the atmospheric light A of each channel is set between [0.7, 1.0], and β is randomly selected at [0.6, 1.8]. Among them, the number of samples used for training is 13,000, and the number of samples used for verification is 990. The RESIDE testset [16] consists of synthetic objective testing set (SOTS), hybrid subjective testing set (HSTS), and real-world task-driven test set (RTTS) to support a wide range of experiments.

We implement our model using Pytorch 0.4.1. Detailed configurations and main parameter settings of our proposed FAOD-Net (as shown in Figure 2) are summarized in Table 1, which includes 5 depthwise separable convolutional layers, 5 ReLU activations for the back of the convolutional layer, and 3 concat layers. The FAOD-Net is initialized with random weight parameters and trained using stochastic gradient descent (SDG) back-propagation algorithm and learning rate of 0.001. The weight parameters of FAOD-Net are updated in 10 epochs on NVIDIA TITAN Xp 12 GB GPU and CUDA version: 10.0.

Table 1: The parameter settings of the FAOD-Net model.

Type	Input size ($C * H * W$)	Kernel size	Groups
DS-Conv1	$3 * 460 * 620$	$3 * 3$(DW-Conv) $1 * 1$(PW-Conv)	1 —
ReLu	$3 * 460 * 620$	—	—
DS-Conv2	$3 * 460 * 620$	$3 * 3$(DW-Conv) $1 * 1$(PW-Conv)	1 —
ReLu	$6 * 460 * 620$	—	—
Concat1	$3 * 460 * 620$ $6 * 460 * 620$	—	—
DS-Conv3	$9 * 460 * 620$	$5 * 5$(DW-Conv) $1 * 1$(PW-Conv)	1 —
ReLu	$9 * 460 * 620$	—	—
Concat2	$6 * 460 * 620$ $9 * 460 * 620$	—	—
DS-Conv4	$15 * 460 * 620$	$7 * 7$(DW-Conv) $1 * 1$(PW-Conv)	1 —
ReLu	$6 * 460 * 620$	—	—
Concat3	$9 * 460 * 620$ $15 * 460 * 620$	—	—
DS-Conv5	$24 * 460 * 620$	$3 * 3$(DW-Conv) $1 * 1$(PW-Conv)	1 —
ReLu	$3 * 460 * 620$	—	—

4.2. Results on Synthetic Objective Testing Set. To verify the effectiveness of the proposed algorithm, we performed experiments on SOTS synthetic dataset [16] to illustrate the performance of our method compared to other state-of-the-art methods. We use the synthesized fog image as the input of FAOD-Net and compare the output of the ground truth images. We have tested our method on SOTS indoor and outdoor datasets [16], and Table 2 shows the results of comparing algorithm [5, 6, 8–10, 25] based on the no-reference indicators SSEQ [17] and BLIINDS-II [18]. In order to be consistent with the full-reference index PSNR and SSIM, we reverse the result of the no-reference index so that the larger the result, the better the effect. As can be seen from Table 2, FAOD-Net and AOD-Net [10] have the best results of no-reference indicators. The effect of BCCR [25] is suboptimal. By zooming in on the details, we can see from Figure 4 that the effects of FAOD-Net on the desktop and the background outside the window are more consistent with human visual perception.

The above part uses the no-reference image quality evaluation index to compare our proposed algorithm with the other state-of-the-art algorithm on the synthetic data set. In this part, we use the full-reference image quality evaluation index PSNR and SSIM to evaluate the algorithm in the image. The results of average PSNR and SSIM are shown in Table 3. From Table 3, we can see that DehazeNet [8] has the highest PSNR on the synthesized SOTS test set [16]. FAOD-Net and AOD-Net [10] are suboptimal. FAOD-Net has a slightly higher SSIM value than AOD-Net [10] and DehazeNet [8]. From Figure 5, we can see that the traditional DCP algorithm [5] is constrained by prior knowledge, and when defogging the sky area, the sky area color of the defogged image is too bright.

4.3. Results on Real-World Testing Set. In order to verify the effectiveness of the FAOD-Net proposed in this paper in the real world, we performed experiments on real-world foggy images in the HSTS dataset [16]. We compare the algorithm proposed in this paper with the other state-of-the-art algorithms [5, 6, 8–10, 25]. Since we cannot obtain the actual fog-free images corresponding to the real-world foggy images, we cannot use the full-reference index PSNR and SSIM to evaluate the dehazing effect. We use the no-reference indicators SSEQ and BLIINDS-II to compare our proposed algorithm with the other state-of-the-art algorithms. Table 4 shows the average SSEQ [17] and BLIINDS-II [18] obtained by defogging on real-world foggy images in HSTS dataset [16]. Also, we reverse the result of the no-reference index SSEQ and BLIINDS-II. From Table 4, we can see that the results obtained by FAOD-Net of the no-reference image quality evaluation index are the best and the results obtained by AOD-Net [10] are close to those of FAOD-Net. Because FAOD-Net and AOD-Net are end-to-end models, the error between the foggy image and the corresponding fog-free image is directly minimized during the model training process. In addition, the pyramid pooling module added in FAOD-Net enables the network model to fully extract the global information, so the fogging effect obtained using FAOD-Net is better. Figure 6 shows the comparison of the defogging effect on real-world foggy images in HSTS [16]. The DCP algorithm [5] estimates the atmospheric illumination value based on the prior knowledge of the dark

Table 2: Quantitative results on SOTS in terms of no-reference image quality assessment.

Metrics	BCCR [25]	DCP [5]	DehazeNet [8]	AOD-Net [10]	MSCNN [9]	CAP [6]	FAOD-Net
SSEQ	65.78	64.89	65.43	67.62	65.27	64.71	67.71
BLIINDS-II	74.42	74.39	71.68	79.01	74.31	73.43	79.04

Figure 4: Result comparison between state-of-the-art methods and proposed FAOD-Net on indoor synthetic dataset in SOTS. (a) Sample hazy images from SOTS indoor dataset [16]. (b) DCP [5]. (c) DehazeNet [8]. (d) AOD-Net [10]. (e) MSCNN [9]. (f) CAP [6]. (g) Proposed method (FAOD-Net).

Table 3: Quantitative results on SOTS in terms of full-reference image quality assessment.

Metrics	BCCR [25]	DCP [5]	DehazeNet [8]	AOD-Net [10]	MSCNN [9]	CAP [6]	FAOD-Net
PSNR (dB)	16.95	16.61	21.23	19.14	17.48	19.12	19.21
SSIM	0.7927	0.8193	0.8495	0.8526	0.8105	0.8374	0.8529

channel. From Figure 6(c), it can be seen that when the scene color of the object is close to the atmospheric light, such as when the picture contains a large number of white objects or other light sources, it may cause the estimated atmospheric light value to be too high and the image after defogging to be over exposed.

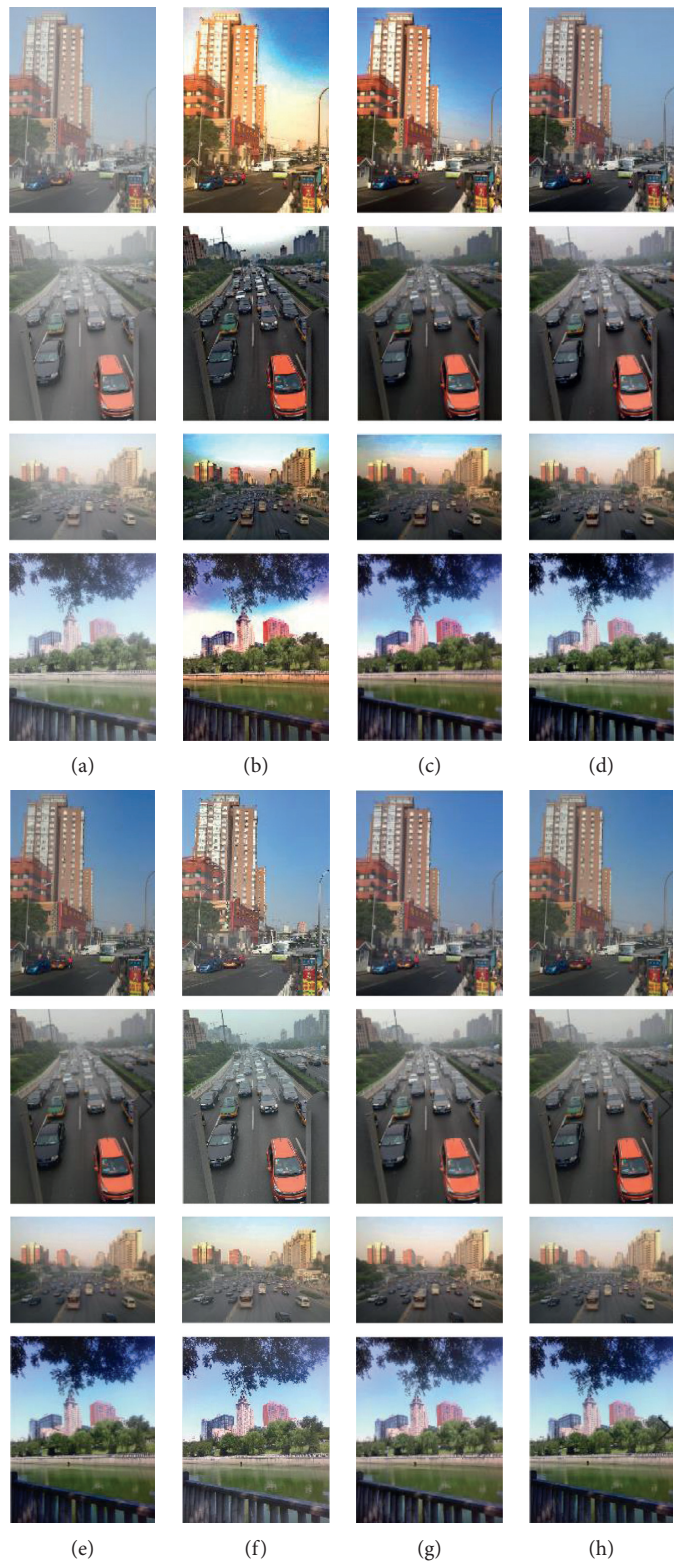

Figure 5: Result comparison between state-of-the-art methods and proposed FAOD-Net on outdoor synthetic dataset in SOTS. (a) Sample hazy images from SOTS outdoor dataset [16]. (b) BCCR [25]. (c) DCP [5]. (d) DehazeNet [8]. (e) AOD-Net [10]. (f) MSCNN [9]. (g) CAP [6]. (h) Proposed method (FAOD-Net).

4.4. Run Time. In order to verify that the depthwise separable convolution mentioned in Section 2.3 can speed up the calculation and reduce the time of image defogging, we compare the FAOD-Net proposed in this paper with other advanced defogging algorithms [5, 6, 8–10, 25]. We performed experiments on the SOTS indoor test set [16], and

Table 4: Quantitative results on the **real-world images** in HSTS in terms of no-reference image quality assessment.

Metrics	BCCR [25]	DCP [5]	DehazeNet [8]	AOD-Net [10]	MSCNN [9]	CAP [6]	FAOD-Net
SSEQ	66.57	68.62	68.32	70.03	68.39	67.68	70.08
BLIINDS-II	68.51	69.31	60.33	74.71	62.64	63.58	74.73

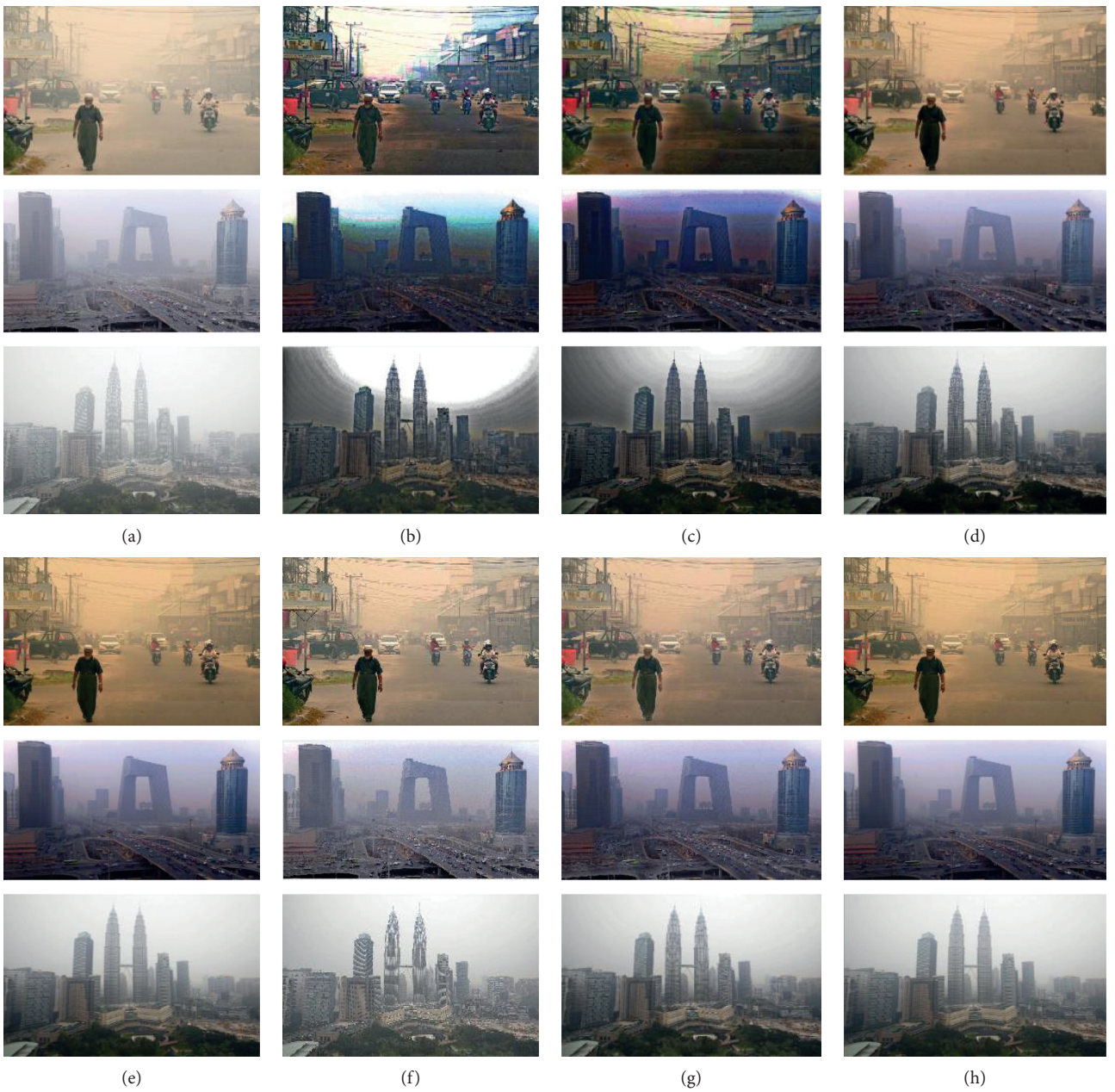

Figure 6: Dehazing results evaluated on the real-world images in HSTS [16]. (a) Sample hazy images from HSTS dataset [16]. (b) BCCR [25]. (c) DCP [5]. (d) DehazeNet [8]. (e) AOD-Net [10]. (f) MSCNN [9]. (g) CAP [6]. (h) Proposed method (FAOD-Net).

the input image size was 620×460. Table 5 shows the average time required for each method to process a single image. As can be seen from Table 5, FAOD-Net has higher efficiency in defogging a single image. Among them, all algorithms run on Matlab, except AOD-Net and FAOD-Net, which are run on Pytorch. Figure 7 shows the loss function graphs of FAOD-Net and AOD-Net [10]. The graph shows the loss function value of 1/2 epoch. During the training process, we found that the training parameters required to train the AOD-Net model were 1761. After using depthwise separable convolution instead of standard convolution, the required training parameters were only 717.

Table 5: Average time taken by various methods to process single image (in seconds).

Method	BCCR [25]	DCP [5]	DehazeNet [8]	CAP [6]	MSCNN [9]	AOD-Net [10]	FAOD-Net
Time (sec.)	3.89	1.65	2.53	0.98	2.67	0.67	0.34

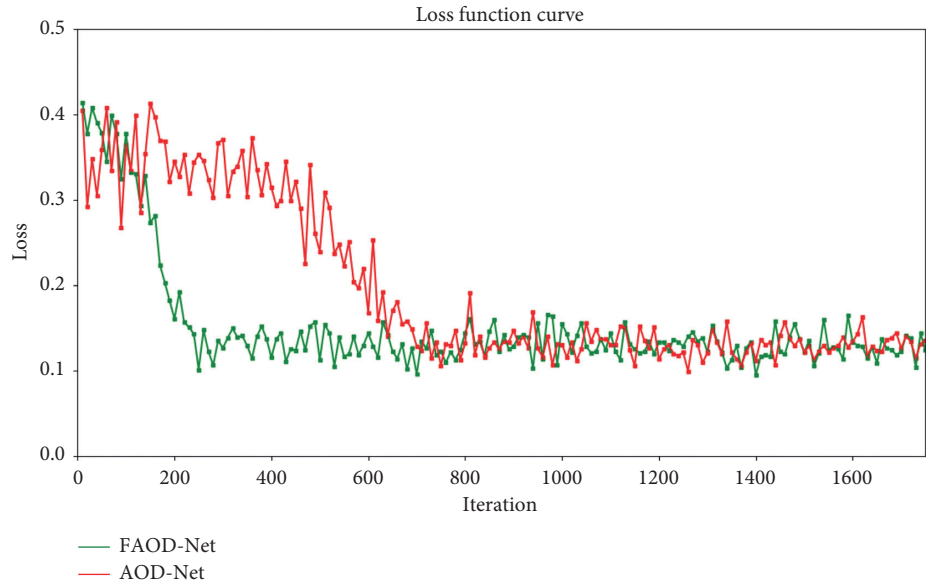

Figure 7: Comparison of loss function between FAOD-Net and AOD-Net [10]. The green curve is the loss function curve of FAOD-Net. The red curve is the loss function curve of AOD-Net [10].

5. Conclusion

In this paper, we propose an image dehazing model based on lightweight convolution-depthwise separable convolution. This model has the advantage of the AOD-Net model and replaces the standard convolution with depthwise separable convolution. Therefore, the model can avoid the estimation of the error caused by the fog image transmission rate and the atmospheric light value separately and can significantly reduce the network model training parameters and running time. We add a pyramid pooling module to the model to improve the model's ability to get global information. Extensive experiments demonstrate that the algorithm proposed in this paper can achieve satisfactory results in both the quality and efficiency of defogging. Moreover, the lightweight features of the model make the model widely applicable to mobile terminals, cloud computing, or deeper network models.

Data Availability

The data supporting the research in this paper come from the RESIDE data set proposed by Li et al [16] and can be downloaded from https://sites.google.com/view/reside-dehaze-datasets.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was partially supported by the National Natural Science Foundation of China (Nos. 61571241 and 61872423), Industry Prospective Primary Research & Development Plan of Jiangsu Province (No. BE2017111), the Scientific Research Foundation of the Higher Education Institutions of Jiangsu Province (No. 19KJA180006), and the Postgraduate Research & Practice Innovation Program of Jiangsu Province (No. KYCX19_0891).

References

[1] J. Zhou, D. Zhang, P. Zou, W. Zhang, and W. Zhang, "Retinex-based laplacian pyramid method for image defogging," *IEEE Access*, vol. 7, pp. 122459–122472, 2019.

[2] S.-Y. Yu and H. Zhu, "Low-illumination image enhancement algorithm based on a physical lighting model," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 29, no. 1, pp. 28–37, 2019.

[3] W. Kim, "Image enhancement using patch-based principal energy analysis," *IEEE Access*, vol. 6, pp. 72620–72628, 2018.

[4] Z. Tufail, K. Khurshid Ahmad, and K. Khurshid, "Optimisation of transmission map for improved image defogging," *IET Image Processing*, vol. 13, no. 7, pp. 1161–1169, 2019.

[5] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 12, pp. 2341–2353, 2011.

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

[7] W. Zhang, L. Dong, X. Pan, J. Zhou, L. Qin, and W. Xu, "Single image defogging based on multi-channel convolutional MSRCR," *IEEE Access*, vol. 7, pp. 72492–72504, 2019.
[8] B. Cai, X. Xu, K. Jia, C. Qing, and D. Tao, "Dehazenet: an end-to-end system for single image haze removal," *IEEE Transactions on Image Processing*, vol. 25, no. 11, pp. 5187–5198, 2016.
[9] W. Ren, S. Liu, H. Zhang, J. Pan, X. Cao, and M.-H. Yang, "Single image dehazing via multi-scale convolutional neural networks," in *European Conference on Computer Vision*, pp. 154–169, Springer, Berlin, Germany, 2016.
[10] B. Li, X. Peng, Z. Wang, J. Xu, and D. Feng, "AOD-net: all-in-one dehazing network," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, IEEE, Venice, Italy, October 2017.
[11] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. K. Squeezenet, "Alexnet-level accuracy with 50x fewer parameters and < 0.5 Mb model size," 2016, https://arxiv.org/abs/1602.07360.
[12] X. Zhang, X. Zhou, M. Lin, and J. Sun, "Shufflenet: an extremely efficient convolutional neural network for mobile devices," 2017, https://arxiv.org/abs/1707.01083.
[13] A. G. Howard, M. Zhu, B. Chen et al., "Mobilenets: efficient convolutional neural networks formobile vision applications," 2017, https://arxiv.org/abs/1704.04861.
[14] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," in *Computer Vision—ECCV 2014*, Springer, Berlin, Germany, 2014.
[15] H. Zhao, J. Shi, and X. Qi, "Pyramid scene parsing network," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, July 2017.
[16] B. Li, W. Ren, D. Fu et al., "Benchmarking single image dehazing and beyond," *IEEE Transactions on Image Processing*, vol. 28, no. 1, pp. 492–505, 2019.
[17] L. Liu, B. Liu, H. Huang, and A. C. Bovik, "No-reference image quality assessment based on spatial and spectral entropies," *Signal Processing: Image Communication*, vol. 29, no. 8, pp. 856–863, 2014.
[18] M. A. Saad, A. C. Bovik, and C. Charrier, "Blind image quality assessment: a natural scene statistics approach in the DCT domain," *IEEE Transactions on Image Processing*, vol. 21, no. 8, pp. 3339–3352, 2012.
[19] E. J. McCartney, *Optics of the Atmosphere: Scattering by Molecules and Particles*, Vol. 1, Wiley, New York, NY, USA, 1976.
[20] S. G. Narasimhan and S. K. Nayar, "Contrast restoration of weather degraded images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 6, pp. 713–724, 2003.
[21] S. K. Nayar and S. G. Narasimhan, "Vision in bad weather," in *Proceedings of the Seventh IEEE International Conference on Computer Vision*, vol. 2, pp. 820–827, Kerkyra, Greece, September 1999.
[22] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
[23] S. H. Khan, M. Bennamoun, F. Sohel, and R. Togneri, "Automatic feature learning for robust shadow detection," in *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, OH, USA, June 2014.
[24] D. Eigen, C. Puhrsch, and R. Fergus, "Depth map prediction from a single image using a multi-scale deep network," in *Proceedings of the Conference on Neural Information Processing Systems*, Montreal, Canada, December 2014.
[25] G. Meng, Y. Wang, J. Duan, S. Xiang, and C. Pan, "Efficient image dehazing with boundary constraint and contextual regularization," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 617–624, Sydney, Australia, December 2013.