Multi-scale depth information fusion network for image dehazing

Guodong Fan[1] · Zhen Hua[1] · Jinjiang Li[2]

Accepted: 25 January 2021 / Published online: 4 March 2021

Abstract

According to the atmospheric physical model, we can use accurate transmittance and atmospheric light information to convert a hazy image into a clean one. The scene-depth information is very important for image dehazing due to the transmittance directly corresponds to the scene depth. In this paper, we propose a multi-scale depth information fusion network based on the U-Net architecture. The model uses hazy images as inputs and extracts the depth information from these images; then, it encodes and decodes this information. In this process, hazy image features of different scales are skip-connected to the corresponding positions. Finally, the model outputs a clean image. The proposed method does not rely on atmospheric physical models, and it directly outputs clean images in an end-to-end manner. Through numerous experiments, we prove that the multi-scale deep information fusion network can effectively remove haze from images; it outperforms other methods in the synthetic dataset experiments and also performs well in the real-scene test set.

Keywords Image dehazingd · U-Net · Depth map

1 Introduction

Haze is an everyday weather phenomenon; it is primarily caused by numerous tiny particles in the atmosphere; these particles scatter and absorb light, thereby producing the haze weather effect. Images collected under hazy weather conditions exhibit low-picture-contrast and color-saturation problems. Therefore, haze can be seriously detrimental to the performance of a vial system. Clarifying these hazy images and ensuring the effectiveness of the vision system has been an active topic for researchers and has implications in the fields of computer vision and imaging science. The methods that researchers initially proposed to improve the contrasts of hazy images include Histogram equalization, partial Histogram equalization, Wavelet transform, and Homomorphic filtering. Typically, these image-enhancement methods neglect the mechanism of haze formation; as a result, the display effect of the restored images is poor, and the image details are easily lost. Moreover, such algorithms are highly complex, and it is often difficult to find algorithm parameters suitable for all scenarios.

The introduction of the atmospheric physics model [1–3] considerably accelerated the development of image-dehazing practices. Today, numerous dehazing algorithms effectively apply the theoretical background of the atmospheric physical model. This model is expressed as

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

$$T(x) = e^{-\beta d(x)}. \tag{2}$$

Where $I(x)$ is a hazy image; $J(x)$ is a clean image; A is the global atmospheric light; $T(x)$ is the image transmission rate; x is the position of the pixel in the image; β is the atmospheric scattering coefficient; $d(x)$ is the scene depth.

It can be seen from the model that if an accurate $T(x)$ and A can be obtained, the hazy image can be restored to a clean one. To accurately identify $T(x)$ and A, researchers have proposed several classical methods. He et al. [4] proposed a dark channel prior (DCP) dehazing method by analyzing a large number of hazy images; they achieved a relatively strong dehazing performance. Berman et al. [5] assumed the colors of a clean image to be well-approximated by hundreds of different colors, thereby proposing a dehazing method that uses non-local priors to describe clear image features. It is worth noting that using an priori-based method

✉ Zhen Hua
huazhen66@foxmail.com

Jinjiang Li
lijinjiang@gmail.com

[1] Present address: School of Information and Electronic Engineering, Shandong Technology and Business University, Yantai 264005, China

[2] School of Computer Science and Technology, Shandong Technology and Business University, Yantai 264005, China

for image dehazing in a specific scene is effective; however, these methods are often difficult to adapt to all scenarios. When the real situation violates the priori condition, an inaccurate transmission-map estimation will be produced, resulting in a poor dehazing effect.

The development of deep learning theory [6] has brought new ideas to the field of image dehazing, and methods for applying neural networks to image dehazing [7] tasks have been indicated as a promising research avenue. Cai et al. [8] built a convolutional neural network (CNN) model to learn the relationship between hazy images and transmission rates. Li et al. [9] converted the transmission map and atmospheric light information in the atmospheric physics model into a unified variable, K, predicted K using a CNN, and dehazed the image through a variant of the atmospheric physical model. Atmospheric-physical-model-based dehazing methods are well supported by theory. When accurate transmission maps and atmospheric light information can be found, the model typically exhibits a strong performance; however, when these are inaccurate, the dehazing effect is significantly diminished.

In recent years, a new processing method has emerged in the field of image dehazing. In this method, a neural network model is used to construct an intermediate channel between the hazy and haze-free images, and the haze-free image is directly output in an end-to-end manner. Qu et al. [10] proposed the enhanced Pix2pix dehazing network (EPDN), an end-to-end image dehazing method employing a Generative Adversarial Network. Chen et al. [11] proposed the gated context aggregation network (GCANet) by using gated subnets to integrate different levels of features. Liu et al. [12] proposed GridDehazeNet by combining multi-scale estimation and an attention mechanism. Qin et al. [13] proposed an end-to-end feature-fusion network based on attention. Dong et al. [14] proposed a multi-scale dense feature fusion method based on the U-Net structure, in which the scene depth was intrinsically linked to the haze distribution. However, these end-to-end dehazing methods neglect depth information; thus, although they perform well for composite images, they suffer from limitations when handling real-scene images.

Unlike our previous work, this study focuses on applying image depth information to the field of image dehazing. We propose a multi-scale depth-information fusion network (MSDFN), which is based on the U-Net architecture [15]; thus, it does not rely on atmospheric physical models. We encode the image depth information [16] and then decode it, to obtain a clean image. However, if a depth map is encoded without color information, it may be difficult to obtain a haze-free color image. Therefore, we add an input pyramid branch to the structure of the U-Net. This branch inputs hazy color images and participates in the encoding and decoding of the model; then, it uses skip-connections to integrate hazy image features of different scales into the encoder and different stages of the decoder. The main contributions of the method proposed in this paper are as follows.

1. MSDFN: a U-Net-structure-based method for modelling depth information.
2. Additional branches on the U-Net structure: This enables the network to restore image color, contrast, and other information. It outperforms existing methods in terms of subjective impressions and objective indicators.
3. The method outperforms existing ones in both qualitative and quantitative tests, and it achieves superior results in real scenarios.

2 Related work

In 1992, Bissonnette [17] proposed a new and efficient method of calculating the point-spread and modulation-transfer functions produced by aerosol forward scattering. Nayar and Narashimhan's modeling of atmospheric scattering has led to significant progress in the field of image dehazing; since then, numerous dehazing methods have been constructed on the theoretical bases of atmospheric physical models [2, 3, 18].

Image dehazing is an ill-posed problem. In this section, we introduce in detail the current classical and mainstream image-dehazing methods and—according to their characteristics—classify them into three categories: image-enhancement-based, atmospheric-physical-model-based, and end-to-end model-based.

Image-enhancement-based dehazing These methods perform image enhancement on the degraded image, to improve its quality. These methods offer numerous advantages; for instance, they can apply well-established image-processing algorithms, apply and improve upon popular image-enhancement algorithms in a targeted manner, enhance image contrasts and other characteristics, and highlight features and valuable information in the image. Because of the low contrasts of hazy images, these methods operate upon the form of a narrowly concentrated single-peak histogram. In the early stages of image-dehazing method development, researchers used histogram equalization [19] to enhance the image; this balanced the image's histogram distribution, thereby expanding the image's dynamic range and enhancing its contrast. Global histogram equalization can enhance hazy images to a certain extent; however, its dehazing effect is less than ideal when the haze-density distribution is uneven. Thus, Kim et al. [20] used partially over-lapped sub-block histogram equalization to enhance the local contrast of hazy images. To the same end, generalized local histogram equalization [21] can perform dynamic contrast control through adaptive adjustment of equalization parameters. To better adapt to local enhancement, Khan et al. [22] used a histogram segmentation algorithm, to segment the image and enhance the sub-regions in a targeted manner. On the basis of histogram equalization, Rudolf Richter et al. [23] realized edge-preserving image enhancement through histogram matching and local weighted fusion.

The aforementioned methods can improve the image contrast to a certain extent before applying image dehazing and enhancement. In addition, enhancement algorithms based on homomorphic filters [24], wavelet transforms [25], and curvelet transforms [26] have also been applied in image dehazing. The homomorphic filter-based enhancement method combines frequency filtering and grayscale changes. It can mitigate the effects of uneven illumination and is more suitable for enhancing hazy images. However, this method only considers the frequency domain characteristics of the image, neglecting its spatial information; thus, its enhancement effect on dense hazy images is not ideal. The wavelet transform-based method integrates frequency and spatial-domain information with multi-scale, decorrelation, and local-feature-expression capabilities; it can improve the visibility of images through detail enhancement and is primarily used for dehazing thermal images and enhancing medical ones. The curvelet transform-based method is a high-dimensional generalization of the wavelet transform one; it can capture image characteristics of different scales and angles, and it has a clear dehazing effect on hazy images. Image-enhancement-based dehazing methods offer the advantages of high efficiency and easy implementation. However, because they essentially enhance the contrast of the image, they lack the imaging mechanism and degradation model required to consider hazy images. Therefore, image enhancement can improve visual effects to a certain extent but cannot remove the influence of haze in the image. Moreover, due to the lack of research on image-degradation mechanisms, color distortion can arise through the contrast enhancement process.

Atmospheric-physical-model-based image dehazing These methods restore hazy images by directly or indirectly estimating the image transmission rate and atmospheric light via the atmospheric physical model. He et al. [4] proposed the DCP model, in which most local pixel blocks in the clear and haze-free images contain some pixels, and at least one color channel of these pixels has a very low value. The rough image transmission rate is solved using a priori information; then, the precise image transmission rate is optimized by soft matting the image, to obtain the final dehazed image. However, when large regions of white or sky feature in the image, a halo effect occurs; this invalidates the prior. Fattal et al. [27] based on the assumption that the transmittance and the image transmission rate are locally uncorrelated, constructing an independent component analysis method and using Gauss Markov random field to solve the scene albedo and finally get a haze-free image. However, when the image color information is missing, the obtained dehazing image will appear image color distortion. Arigela et al. [28] constructed a nonlinear sinusoidal function to modify the DCP and obtain a rough transmission rate; furthermore, they used the gray scale transform function to replace the soft matting algorithm. Compared with DCP, this algorithm more effectively overcomes the halo effect; however, dim and distorted colors still remained in the sky regions.

Cai et al. [8] proposed DehazeNet; this uses CNN to extract haze-related features (e.g., dark primary color, color attenuation, and maximum contrast features) from the image, to optimize transmittance estimation. Empirically preset atmospheric light was used for the atmospheric physical model, to restore the haze-free image. Ren et al. [29] proposed a multi-scale CNN by designing a set of coarse- and fine-scale networks to predict the image transmission rate independently, then using multi-scale fusion to complete image dehazing. Li et al. [9] proposed the all-in-one dehazing network (AOD-Net); this model simplifies the unknown quantity in the atmospheric scattering model to a coefficient K, and it learns the relationship between the hazy images and K using a CNN. Zhu et al. [30] used a residual block to estimate the image transmission rate and global atmospheric light; then, they generated the haze-free image using the atmospheric scattering model and used a single-scale discriminator to perform generative adversarial network [31]; this model is referred to as DehazeGAN. When the transmission map and atmospheric light are accurately estimated, the physical-model-based method can obtain a good dehazing effect, especially in non-uniformly distributed hazy images. Therefore, accurately estimating the transmission map and atmospheric light is a focal challenge for this type of method. However, the physical-model-based dehazing method also suffers problems of excessive dehazing, which produces an overall color deviation and distortion in the restored image.

End-to-end model-based image dehazing Th- ese methods do not rely on the atmospheric physical model; they learn the channel between haze and clean images using a neural network model, which directly outputs a haze-free image. Ren et al. [32] proposed a gated fusion network using white balance, contrast enhancement, and gamma-correction methods to pretreat the hazy image; then, they used a CNN to learn the confidence maps corresponding to three pre-processed images and obtained haze-free images through multi-scale fusion. Qu et al. [10] proposed the EPDN; this method enhances the dehazing effect by using a staged dehazing module. Chen et al. [11] proposed GCANet; this method uses smooth expansion technology to eliminate the gridding artifacts produced by negligible parameters of the expanded convolution kernel, and they implemented a gated sub-network to fuse the features from different levels. Liu et al. [12] proposed GridDehazeNet by integrating the attention mechanism into multi-scale estimation; this solved the bottlenecking of traditional multi-scale estimation. Qin et al. [13] proposed an end-to-end feature fusion attention network (FFA-Net); this method retains the shallow information and transfers it to a deep layer through an attention-based feature fusion structure. Using the U-Net architecture. Zhang et al. [33] proposed the Pyramid Channel-based Feature Attention Network by taking advantage of the complementarity of the pyramidal features at different levels and the channel attention mechanism. Dong et al. [14] proposed the multi-scale boosted dehazing network (MSBDN) with dense feature fusion; this method was based on the two principles of boosting and error feedback, demonstrating that these principles are suitable for dehazing. By adding the "strengthen-operate-subtr- act" boosting strategy to the decoder of the proposed model, a simple and effective boosted decoder was developed, to restore images gradually.

3 Proposed method

In this section, we introduce the proposed MSDFN. As shown in Fig. 1, the model inputs a hazy image and predicts its depth map. This map and the hazy image are processed via different branches of the encoding process, in which the different levels of depth information are incorporated. For decoding, the hazy image is first convolved and pooled multiple times, to generate feature maps of different levels. To decode the encoder-generated feature map, the feature information for different levels of the original hazy image is incorporated, and the feature information for different stages of the encoder is skip-connected to the decoding process, to facilitate haze-free image generation. In the following subsections, we elaborate on the model.

3.1 Network design

According to the atmospheric physical model, if the transmittance and atmospheric light can be determined, the hazy image can be converted into a haze-free one. The transmittance is affected by the depth of the scene: nearby objects are easier to identify than distant objects in real haze weather conditions. Therefore, the image's depth information is critical for image dehazing. To integrate the depth information into the image dehazing process, we designed an MSDFN based on U-Net. This method first encodes the depth map and then decodes it to obtain a haze-free image, concatenating the multi-level features of the hazy image in the process. According to its structural characteristics, we divide the model into four components: depth-map-prediction module, input pyramid branch, encoder branch, and decoder branch.

3.1.1 Depth-map prediction module (DPM)

We use the depth-map-prediction method proposed by Liu et al. [34]. In contrast to previous methods, we estimate the depth by treating it as a continuous conditional random field (CRF) learning problem, without relying on any geometric priors or extra information. First, super-pixel segmentation is performed on the hazy image, to generate a super-pixel image. The subsequent processing is divided into two branches: un- ary item processing and paired item processing. The purpose of unary item processing is to obtain the depth of a single super-pixel. Paired item processing encourages adjacent super-pixels with similar appearances to adopt similar depths. Our goal is to obtain the depth of all super-pixels by uniformly processing unary and pairwise terms in the network model, to finally output a depth map.

Figure 2 shows the depth-prediction module. The input of the module is a hazy image, and its output is a depth one. The model primarily employs single-element processing, pair-wise processing, and a CRF loss layer.

Unary item processing. The hazy image is divided by super-pixels into an image containing n super-pixels; then, n image blocks are formed with each super-pixel as the center, and the size of each image block is set to 224 * 224. These n image blocks are used as the inputs of the CNN for unary item processing; finally, an n-dimensional vector containing depth information is obtained.

The unary item processing component is performed by a small CNN network, which consists of six convolutional layers, three pooling layers, and four fully connected layers. The network parameters are identical for all super-pixels. The convolutional layer and first two fully connected layers of the network use ReLU as the activation function. The third fully connected layer uses a sigmoid activation function. The final layer integrates the network and has no activation function. The final output of a single image block network input is a one-dimensional depth value. The network is constructed by minimizing the following equation:

$$U(y_p, x; \theta) = (y_p - z_p(\theta))^2, \forall_P = 1, ..., n. \quad (3)$$

Here, z_p is the depth value predicted for super-pixel p by parameter θ of the CNN.

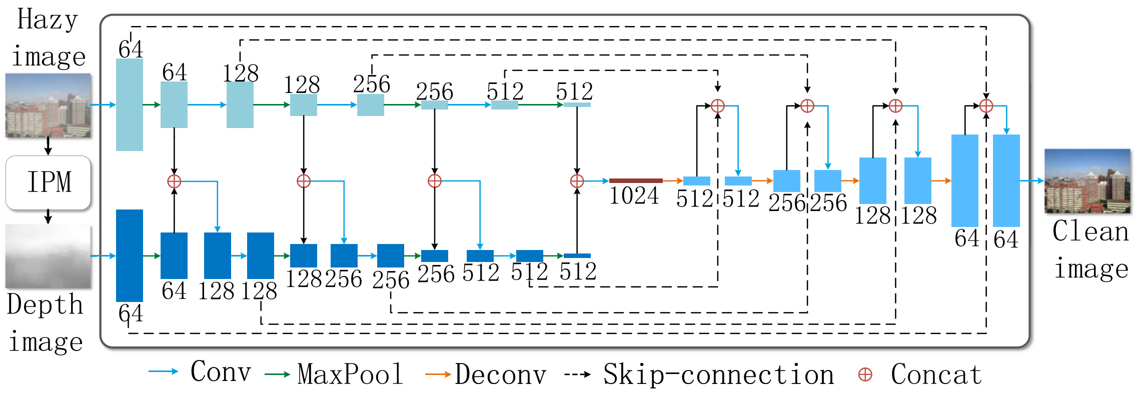


Fig. 1 Structural diagram of multi-scale depth information fusion network

3.1.2 Input pyramid branch (IPB)

Paired item processing. This process is used to form pairs of all adjacent super-pixels as the input of the neural network. The neural network is a combination of fully connected layers and outputs the similarity of adjacent super-pixels. In this section, we use the color difference, color histogram difference, and texture disparity [in terms of local binary patterns (LBP)] to measure the similarity of adjacent elements. The optimization objective of the fully connected layer is

$$V(y_p, y_q, x; \beta) = \frac{1}{2} R_{pq}(y_p - y_q)^2, \forall p, q = 1, ..., n, \quad (4)$$

where S_{pq} denotes the output for the adjacent pair (p, q) in the super-pixel image in the fully connected network. The parameterized representation of the fully connected layer is

$$R_{pq} = \beta^T[S^{(1)}_{pq}, S^{(1)}_{pq}, S^{(3)}_{pq}]^T = \sum_{k=1}^{3} \beta_k S^{(k)}_{pq}, \quad (5)$$

where $\beta = [\beta_1, \beta_2, \beta_3]^T$ contains the network parameters, $S^{(1)}_{pq}$, $S^{(2)}_{pq}$, $S^{(3)}_{pq}$ respectively represent the adjacent super-pixel pairs (p, q) in three different similarity matrices, the calculation method is as follows:

$$S^{(k)}_{pq} = e^{-\gamma} \left\| s^{(k)}_p - s^{(k)}_q \right\|, k = 1, 2, 3, \quad (6)$$

where $s^{(k)}_p$ and $s^{(k)}_q$ are the values calculated by the color, color histogram and LBP of super pixels p and q respectively, $\|\cdot\|$ is the second norm of $s^{(k)}_p$ - $s^{(k)}_q$, γ is a constant.

$$S^{(k)}_{pq} = e^{-\gamma} \left\| s^{(k)}_p - s^{(k)}_q \right\|, k = 1, 2, 3, \quad (7)$$

where $s^{(k)}_p$ and $s^{(k)}_q$ are the values calculated for super-pixels p and q, respectively, using their color, color histogram, and LBP; $\|\cdot\|$ is the second norm of $s^{(k)}_p$ - $s^{(k)}_q$; and γ is a constant.

CRF loss layer. The CRF loss layer receives the outputs from the unary and paired items and seeks to minimize the negative log-likelihood. Similar to a traditional CRF, the CRF layer uses a density function to model the conditional probability distribution of the data, as

$$P_r(y|x) = \frac{1}{Z(x)} \exp(-E(y, x)), \quad (8)$$

where E is the energy function and Z is the partition function, defined as

$$Z(x) = \int_y \exp\{-E(y, x)\} \mathrm{d}y, \quad (9)$$

where y is continuous, which in some cases allows the integral in Eq. 7 to be calculated analytically; this differs from the discrete case, in which approximation methods

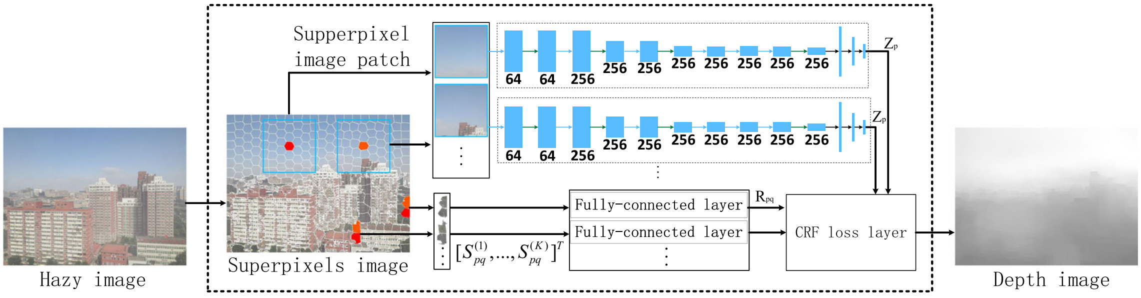


Fig. 2 Depth-prediction module. The input of the module is a hazy image. The output is a depth image

must be applied. To predict the depth of the new image, previous studies have solved the following maximum a posteriori probability inference problem:

$$y^* = \underset{y}{argmax}\ P_r(y|x) \tag{10}$$

The CRF layer expresses the energy function as a typical combination of unary term processing U and binary phase processing V on the node (super-pixel) N of image x and edge S, as

$$\begin{aligned} E(y,x) &= \sum_{p\in N} U(y_p, x) + \sum_{(p,q)\in S} V(y_p, y_q, x) \\ &= \sum_{p\in N} (y_p - z_p)^2 + \sum_{(p,q)\in S} \frac{1}{2} R_{pq}(y_p - y_q)^2. \end{aligned} \tag{11}$$

We used the NYU V2 and Make3D datasets to train the model to predict indoor (Fig. 3) and outdoor (Fig. 4) image depths, respectively. For more details, please refer to Reference [34].

3.1.3 Input pyramid branch(IPB)

In this section, we use convolution + pooling layers to step-wise transform the depth map into feature images of different scales. First, we convolve the depth image into a 64-channel C_{IPB}^{64} (Eq. 11). Then, through maximum pooling, we down-sample C_{IPB}^{64} to obtain M_{IPB}^{64} (Eq. 12). Repeating the previous process (Eqs. 13 and 14), we obtain C_{IPB}^{64}, C_{IPB}^{128} , C_{IPB}^{256} , and C_{IPB}^{512} in turn.

$$C_{IPB}^{64} = \sigma(W_{IPB}^{1} * Input\_image + b_{IPB}^{1}), \tag{12}$$

$$M_{IPB}^{64} = maxpool(C_{IPB}^{64}), \tag{13}$$

$$C_{IPB}^{128} = \sigma(W_{IPB}^{2} * M_{IPB}^{64} + b_{IPB}^{2}), \tag{14}$$

$$M_{IPB}^{128} = maxpool(C_{IPB}^{128}). \tag{15}$$

Here, C represents the convolution result, M represents the maximum pooling result, the superscripts C and M indicate the number of channels, the subscript IPB denotes the input pyramid branch, W represents the weight parameter, b represents the bias, the superscripts of W and b denote the i-th convolutional layer (i = 1,2,3,4), and represents the nonlinear activation function. The window size of the max pooling layer is 2 * 2, and the size of the convolution kernel is 3 * 3.

3.1.4 Encoder branch(EB)

The input of the EB is the depth map, which is converted into a 64-channel feature map through one convolution (Eq. 15). Then, using maximum pooling to reduce the dimensionality of the feature map, we obtain M_{EB}^{64} (Eq. 16). By connecting the 64-channel feature map C_{IPB}^{64} of the input image in the IPB to Eq. 17 and concatenating it with M_{EB}^{64}, we obtain S_{EB}^{128} (Eq. 18). Then, we convolve S_{EB}^{128} to obtain a 128-channel feature map C_{EB}^{128} (Eq. 18). This process combines the 64-channel depth-map features and input-image features to generate a 128-channel feature map. The general steps run as convolution–pooling–concatenation–convolution, respectively. By performing this operation multiple times, we finally obtain C_{EB}^{64} , C_{EB}^{128}, C_{EB}^{256} , C_{EB}^{512} and C_{EB}^{1024}.

$$C_{EB}^{64} = \sigma(W_{EB}^{1} * Depth\_image + b_{EB}^{1}), \tag{16}$$

$$M_{EB}^{64} = maxpool(C_{EB}^{64}), \tag{17}$$

$$S_{EB}^{128} = concat(M_{EB}^{64}, C_{IPB}^{64}), \tag{18}$$

$$C_{EB}^{128} = \sigma(W_{EB}^{2} * S_{EB}^{128} + b_{EB}^{2}). \tag{19}$$

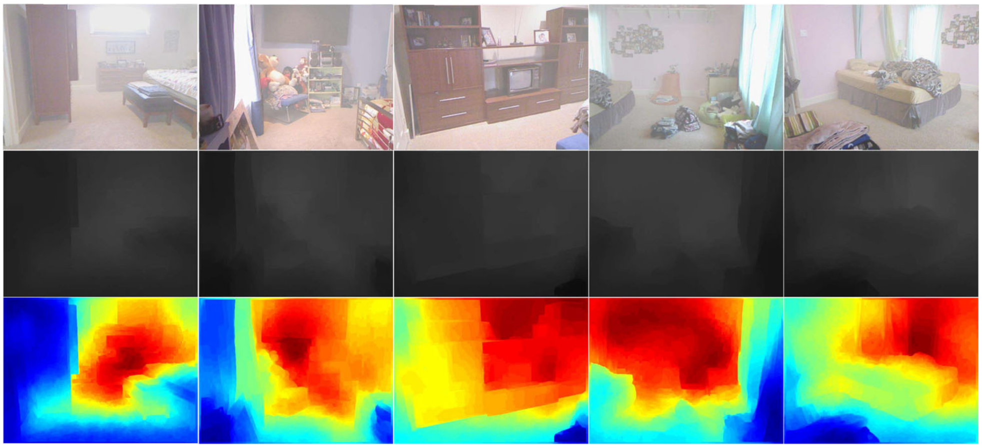


Fig. 3 Indoor image depth maps. The second row displays the gray depth maps, and the third row displays the color depth maps

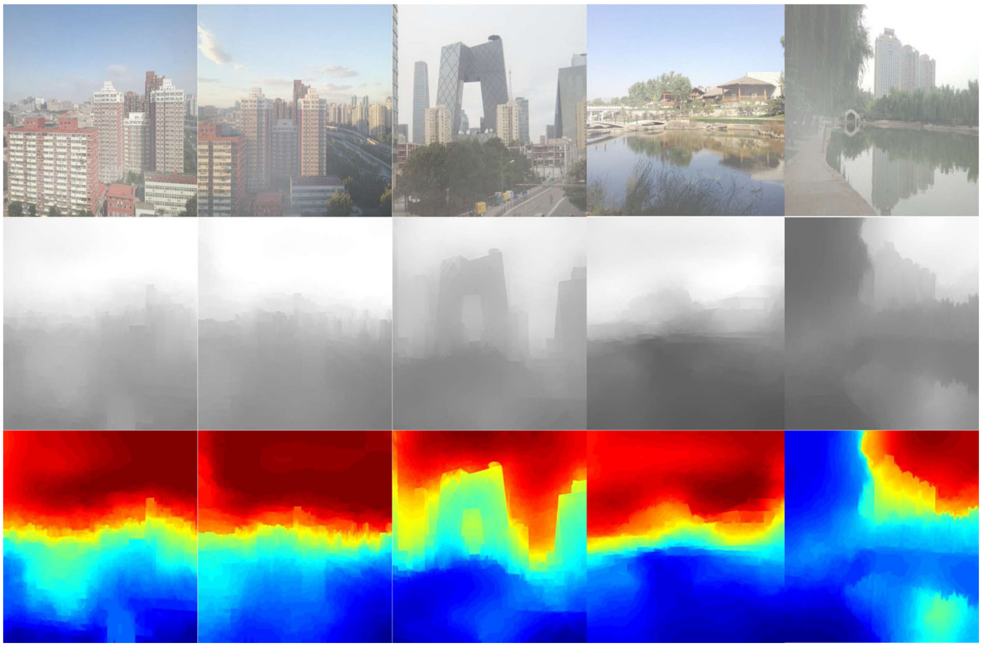


Fig. 4 Outdoor image depth maps. The second row displays the gray depth maps, and the third row displays the color depth maps

Here, C represents the convolution result, M represents the maximum pooling result, S represents the concatenation result, the superscripts C, M, and S represent the number of channels, the subscript EB denotes the encoder branch, W represents the weight parameter, b represents the bias, and the superscripts of W and b represent the i-th convolutional layer (i = 1,2,3,4,5,6,7,8), and σ represents the nonlinear activation function. The window size of the max pooling layer is 2 * 2, and the size of the convolution kernel is 3 * 3.

3.1.5 Decoder branch(DB)

The input of the DB is the C_{EB}^{1024} output of the EB. First, we obtain C_{DB}^{512} through deconvolution (Eq. 19). Then, we skip-connect the feature maps C_{EB}^{512} and C_{IPB}^{512} of the corresponding dimensions in EB and IPB and concatenate them in series with C_{DB}^{512} to obtain S_{DB}^{1536} (Eq. 20). Then, we performing another convolution operation on S_{DB}^{1536} to obtain C_{DB}^{512} (Eq. 21). C_{DB}^{256} can be obtained by performing the aforementioned steps on C_{DB}^{512}. Then, we obtain C_{DB}^{128} and C_{DB}^{64} . Finally, C_{DB}^{64} is convolved once, to output a three-channel-color haze-free image (Eq. 23).

$$D_{DB}^{512} = \sigma(W_{DB}^{1} * C_{EB}^{1024} + b_{DB}^{1}), \tag{20}$$

$$S_{DB}^{1536} = concat(C_{DB}^{512}, C_{EB}^{512}, C_{IPB}^{512}), \tag{21}$$

$$C_{DB}^{512} = \sigma(W_{DB}^{2} * S_{DB}^{1536} + b_{DB}^{2}), \tag{22}$$

$$D_{DB}^{256} = \sigma(W_{DB}^{3} * C_{DB}^{512} + b_{DB}^{3}), \tag{23}$$

$$Clean\_image = \sigma(W_{DB}^{8} * C_{DB}^{64} + b_{DB}^{8}). \tag{24}$$

Here, D represents the deconvolution result, M represents the maximum pooling result, S represents the concatenation result, the superscripts D, M, and S represent the number of channels, the subscript DB represents the encoder branch, W represents the weight parameter, b represents the bias, and the superscripts W and b represent the i-th convolutional layer (i = 1,2,3,4,5,6,7,8), and σ represents the nonlinear activation function. The window size of max pooling is 2 * 2, the sizes of the convolution kernels for deconvolution and convolution are 2 * 2 and 3 * 3, respectively. After the number of channels reaches 64, the convolution component's convolution kernel size is changed to 1 * 1.

3.2 Activation function

In neural network modelling, activation functions are typically used to increase the network model's nonlinear modeling capabilities. The commonly used activation functions are Sigmoid [35], Tanh [36], and Rectified Linear Unit (ReLU) [37]. Among them, Sigmoid and Tanh suffer problems of gradient disappearance and slow convergence. In CNNs, this problem is further amplified. Nair et al. proposed ReLU, which avoids the vanishing gradient problem simply and efficiently. This advantage renders ReLU suitable to a wide variety of neural networks. Experiments have shown that the speed of ReLU is six times that of Tanh. However, ReLU also exhibits shortcomings. For instance, when the input gradient is too large, neuron death occurs, which leads to the failure of model training. To overcome this phenomenon, Leaky ReLU [38], Parametric

ReLU [39], ELU [40], and Exponential Linear Unit have been developed. These ReLU variants alleviate the problems of ReLU to a certain extent. The aforementioned activation functions are static; however, if the parameters of ReLU can be adjusted according to the input characteristics, the model performance may improve. Based on this idea, Chen et al. [41] proposed Dynamic ReLU.

Dynamic ReLU is a piecewise function, $f_{\theta(x)}(x)$. The parameters are obtained from super-function $\theta(x)$ with respect to input x. $\theta(x)$ synthesizes the input context of each dimension to adapt the activation function; this can significantly improve the expressivity of the network with a small number of additional calculations (Fig. 5).

Dynamic ReLU is an extension of ReLU; which develops this piecewise linear function from a static to a dynamic one, by adapting a_c^k and b_c^k for all input elements $x = \{x_c\}$ as follows:

$$y_c = f_{\theta(x)}(x_c) = \max_{1 \leq k \leq K} \left\{ a_c^k(x)x_c + b_c^k(x) \right\}, \tag{25}$$

where the coefficients (a_c^k, b_c^k) are the output of a hyper function $\theta(x)$, expressed as

$$[a_1^1, ..., a_C^1, ..., a_1^K, ..., a_C^K, b_1^1, ..., b_C^1, b_1^K, ..., b_C^K]^T = \theta(x), \tag{26}$$

where K is the number of functions and C the number of channels. Note that the activation parameters (a_c^k, b_c^k) are not only related to the corresponding input x_c but to the other input elements $x_j \neq c$.

Chen et al. [41] presents three forms of Dynamic ReLU: DY-ReLU-A: The activation function is spatial and channel-shared. DY-ReLU-B: The activation function is spatial-shared and channel-wise. DY-ReLU-C: The activation function is spatial and channel-wise. Dynamic ReLU is a new form of activation function. In Section 4.2, we compare the performances of Dynamic ReLU and ReLU in the MSDFN. However, we do not use DY-ReLU-C because it renders the model expensive and difficult to train.

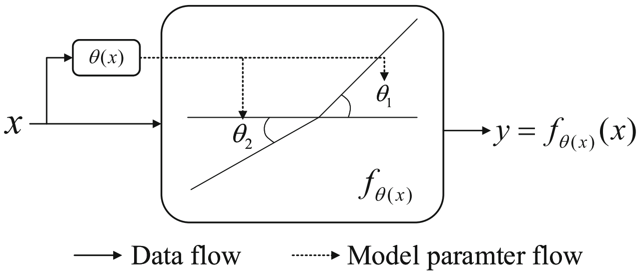


Fig. 5 Dynamic ReLU. The piecewise linear function is determined by input x

3.3 Loss function

Following the previous image-dehazing method [42], we use the negative structural similarity index measure (SSIM) as the model's loss function. This is expressed as

$$L = -\frac{1}{2M} \sum_{m=1}^{M} \frac{\left(2\mu_{Y_m}\mu_{Y'_m} + \theta_1\right)\left(2\sigma_{Y_m Y'_m} + \theta_2\right)}{\left(\mu_{Y_m}^2 + \mu_{Y'_m}^2 + \theta_1\right)\left(\sigma_{Y_m}^2 + \sigma_{Y'_m}^2 + \theta_2\right)}, \tag{27}$$

where Y_m , $\mu_{Y'_m}$ is the average of Y'_m, σ_{Y_m} is the variance of Y_m, $\sigma_{Y'_m}^2$ is the variance of Y'_m, $\sigma_{Y_m Y'_m}$ is the covariance of Y_m and Y'_m, θ_1 and θ_2 are constants. θ_1 and θ_2 are used to prevent the system instabilities produced by a zero denominator. The value range of -SSIM is [-1, 0]. When the structural similarity between the network output and reference image is increased, the -SSIM loss decreases.

4 Experimental results

To verify the effectiveness and advantages of the method proposed in this paper, we performed a series of experiments, as elaborated upon in this section. First, we introduce the dataset, the relevant details of model training, and the evaluation indicators employed. Second, we compare different activation functions. Third, we conduct ablation experiments on the network model structure. Finally, the MSDFN is compared against seven advanced classical image dehazing methods: DCP [4], Hazy-lines [5], AOD-Net [9], MSCNN [29] GCANet [11], FFA-Net [13] and MSBDN [14]. Code has been made available at https://github.com/CCECfgd/MSDFN.

4.1 Experiment settings

Dataset Because it is difficult to collect the hazy and corresponding haze-free images for real haze scenes, we used a synthetic image dataset to verify our method. Our dataset was divided into two parts: indoor and outdoor images. For indoor synthetic images in the training set, we used the NYUv2 depth dataset [43]; we extracted 20,000 synthetic images as the training set and 200 for validation. For outdoor synthetic images in the training set, we used OTS from the RESIDE dataset [44]. OTS contains 8970 clear outdoor images and 313950 hazy images, which are generated for different parameters of the atmospheric physical model. We used 20,000 synthetic images as the training set and 200 as the validation set. The test set was divided into indoor synthetic images, outdoor synthetic images, and real-scene images. For composite images, we used the SOTS test set provided by RESIDE. For real-scene images, we used 500 images from RTTS (in RESIDE)

Table 1 The models using different activation functions are trained on the indoor and the outdoor synthetic image datasets, and the influence of different activation functions on the model is quantitatively evaluated

Config	Indoor		Outdoor	
	PSNR	SSIM	PSNR	SSIM
ReLU	30.8	0.9952	31.9685	0.9977
DY-ReLU-A	30.5374	0.9892	32.0844	0.99695
DY-ReLU-B	30.8817	0.9965	33.7365	0.998

and 100 hazy real-scene images as the test set, to verify the dehazing ability of the model for real scenes. In order to ensure fairness, all convolutional neural networks use synthetic data sets for training, and the trained models are used for testing in O-HAZE.

Training Details We trained the indoor and outdoor synthetic image models separately. The epoch was set to 20, the batch size was set to 4, and the initial learning rate was 0.0001. The training method adopted the attenuation learning rate such that the learning rate was attenuated every two epochs, and the final learning rate was attenuated to 0.000001. The model used the Adam optimizer and was trained on the NVIDIA TITAN RTX.

Quality Measures To reasonably evaluate the effectiveness of the proposed method, we used the peak signal-to-noise ratio (PSNR) and SSIM as objective evaluation indicators. PSNR and SSIM are full reference evaluation indexes; as such, it is necessary to refer to the haze-free image when evaluating the dehazing results. The larger the PSNR, the lower the distortion of the image, the higher its quality, and the better the dehazing effect. The closer the SSIM is to 1, the higher the similarity of the structure, brightness, and contrast between the evaluated and haze-free images.

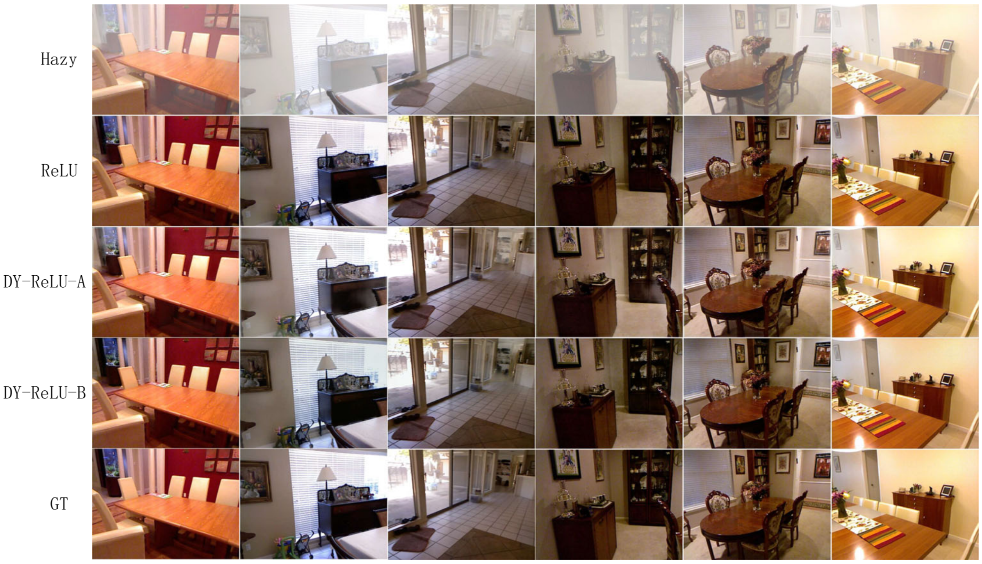
4.2 Activation function

To verify the effectiveness of DY-ReLU-A and DY-ReLU-B, we compared them against ReLU. Thus, we used ReLU, DY-ReLU-A, and DY-ReLU-B as the activation functions of the MSDFN and applied the resulting models to the indoor and outdoor synthetic image datasets for training; then, we selected the most suitable through objective index evaluations and subjective visual impressions.

Table 1 shows the test results—for the SOTS test set—for the models using different activation functions. By comparing the objective indicators, we can see that the model using the DY-ReLU-B activation function achieved the highest scores in terms of both PSNR and SSIM. The models using ReLU and DY-ReLU-A achieved lower objective evaluation scores.

Figure 6 shows the test results—for the SOTS indoor testing set—of several models using different activation functions. From Fig. 6, all models can be seen to achieve good results, and the haze in the pictures is almost entirely

Fig. 6 Selection of test results using different activation functions for the outdoor synthetic image dataset



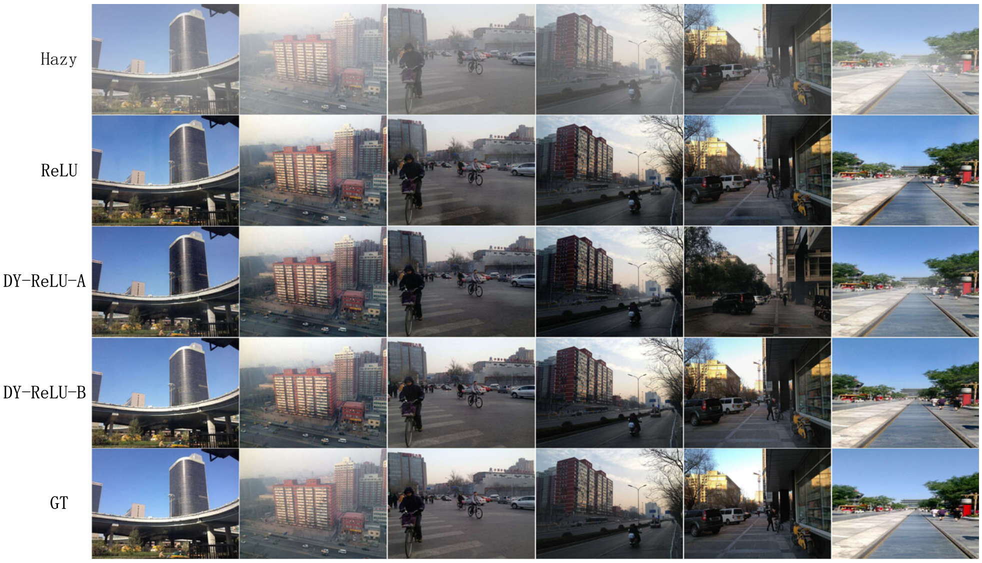


Fig. 7 Selection of test results using different activation functions for the outdoor synthetic image dataset

removed. By considering the picture details, we can see that the DY-ReLU-B-based model outperforms the ReLU-based one in terms of texture and color restoration. This can be seen in the floor area, wall area, and desktop of the third, fifth, and sixth images, respectively. The DY-ReLU-A-based model exhibits a certain residual haze phenomenon, which is especially pronounced in the cabinet, ground, and chair areas of the second, third, and fourth images, respectively. Using objective evaluation indicators and subjective impressions, we conclude that the DY-ReLU-B-based model outperforms the other models proposed in this article.

Figure 7 depicts the test results—for the SOTS outdoor testing set—of the models using different activation functions. From the figure, we can see that all models exhibit dehazing effects; however, certain differences exist between these effects. In contrast to its performance upon the indoor synthetic image test set, the ReLU-based model performs worst in outdoor synthetic image processing. The processed images exhibit distinct black patches, and the surface textures of the objects are poorly restored. The DY-ReLU-A-based model shows some improvements; though dark spots still appear in the image and the texture recovery is poor, it outperforms the DY-ReLU-A-based model and achieves the best performance. The post-dehazing image quality is higher, and the texture restoration is superior. After the second image is processed, the image quality exceeds that of the ground-truth image. Through

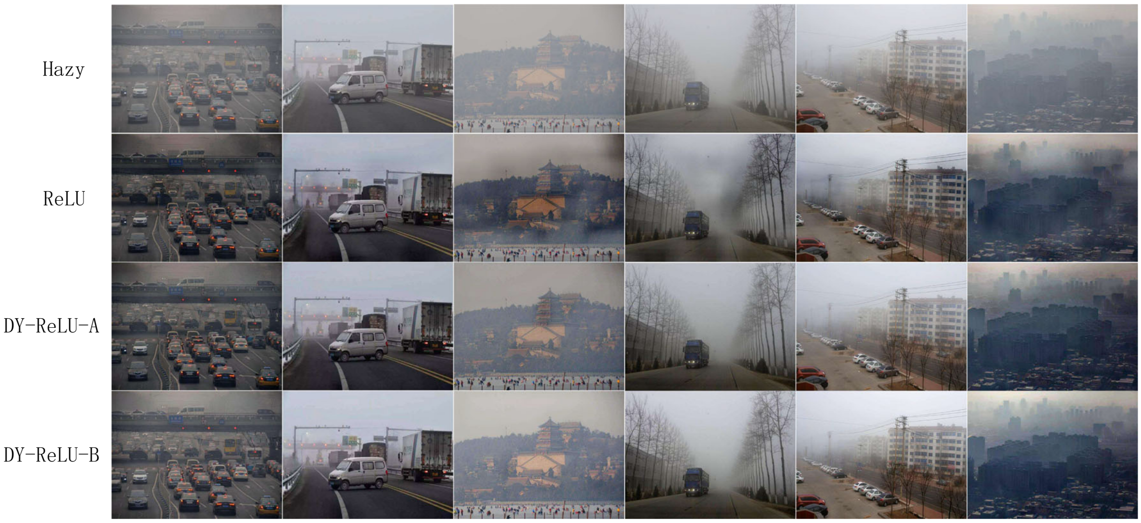


Fig. 8 Part of the test results of models using different activation functions in real scene image dataset

Table 2 The models using different activation functions are trained on the indoor synthetic image dataset and the outdoor synthetic image dataset, and the influence of different activation functions on the model is quantitatively evaluated

Config	Indoor		Outdoor	
	PSNR	SSIM	PSNR	SSIM
E-IPB	30.8	0.9952	31.9685	0.9977
D-IPB	30.5374	0.9892	32.0844	0.99695
E-D-IPB	30.8817	0.9965	33.7365	0.998

objective evaluation indicators and subjective impressions, we conclude that the DY-ReLU-B-based model outperforms the other models proposed in this article.

To test with hazy real-scene images, we used the model trained on the outdoor synthetic image dataset. Figure 8 shows the image results obtained for the real-scene test set by the models using different activation functions. The results show that DY-ReLU-A and DY-ReLU-B are more suitable for processing real-scene images. The ReLU-based model produces impure processing results, and its texture restoration is poor. The DY-ReLU-A- and DY-ReLU-B-based models exhibit fewer differences when dealing with real scenes. However, in terms of details, the DY-ReLU-B-based model produces smoother and more natural results for hazy real-scene images.

Using comprehensive objective evaluation indicators and subjective visual impressions, we conclude that the DY-ReLU-B-based model is the most suitable for real-scene images. However, when using DY-ReLU-B as the activation function, the computational complexity exceeds that of DY-ReLU-A and ReLU, and the training time is significantly longer.

4.3 Ablation study

We conducted an ablation experiment upon the network, to better investigate its structure and components. In this experiment, indoor and outdoor synthetic image datasets were used for training. The ablation experiment involved the following three models:

E-IPB: In this model, we retain the skip-connection between EB and IPB in the original model and remove the skip-connection between DB and IPB.

D-IPB: In this model, we retain the skip-connection between DB and IPB in the original model and remove that between EB and IPB.

E-D-IPB: The original model.

Table 2 shows the test results of E-IPB, D-IPB, and E-D-IPB for the SOTS test set. ED-IPB exhibits a superior performance, especially in the outdoor synthetic images;

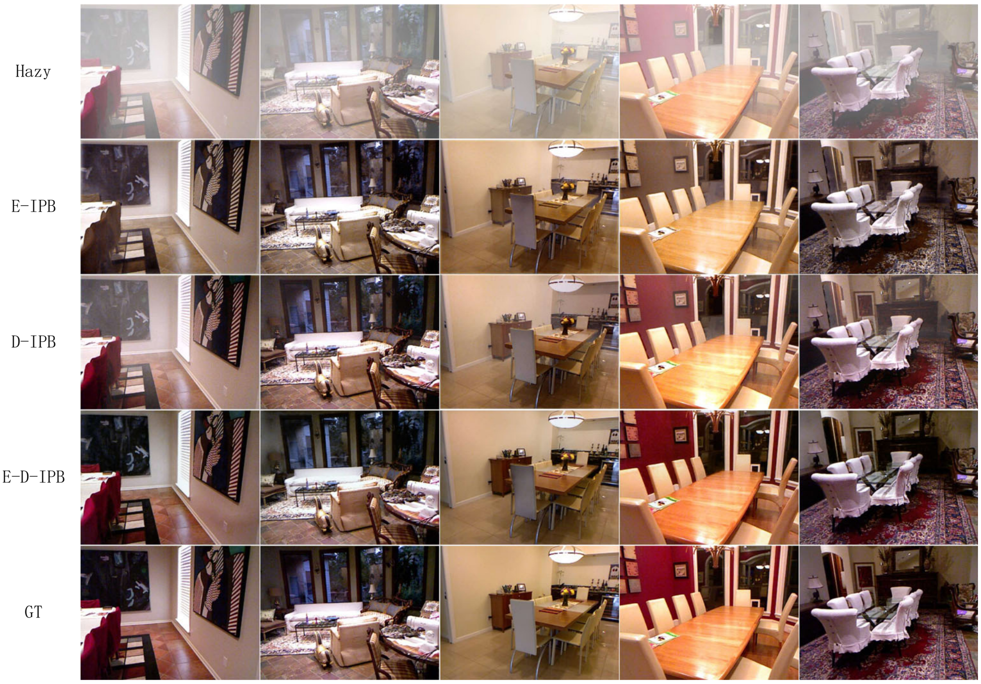


Fig. 9 Selection of test results obtained by E-IPB, D-IPB, and E-D-IPB for indoor synthetic image dataset

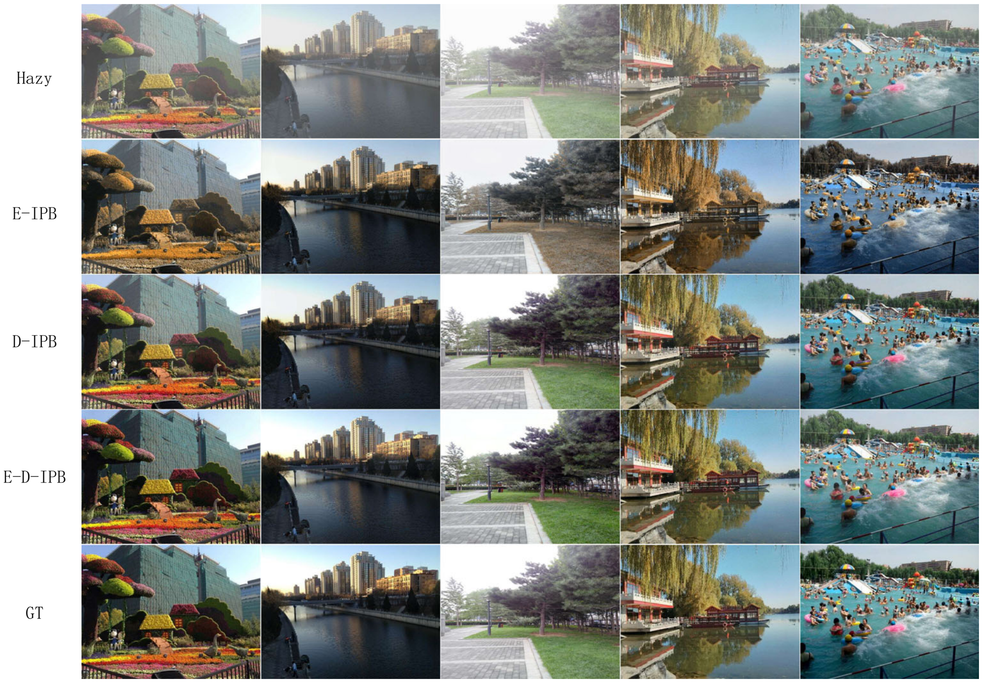


Fig. 10 Selection of test results obtained by E-IPB, D-IPB, and E-D-IPB for outdoor synthetic image dataset

this is perhaps due to the lack of IPB participation in the decoding process, which can lead to poor color recovery. E-IPB achieves the lowest objective evaluation index value.

Figure 9 shows several of the test results obtained by E-IPB, D-IPB, and E-D-IPB for the SOTS indoor synthetic image dataset E-IPB does not involve the IPB during decoding, resulting in poor color restoration. D-IPB does not involve the IPB during encoding; this improves the color recovery, though the residual haze in the picture can be clearly observed. E-D-IPB outperforms E-IPB and D-IPB: its color recovery is superior and the haze is almost entirely removed.

Figure 10 shows several test results obtained by E-IPB, D-IPB, and E-D-IPB for the SOTS outdoor synthetic image dataset. E-IPB suffers the same problem it exhibited for the outdoor composite image test: it struggles to restore the color of the image. D-IPB removes the haze from the image, and the image colors are better restored. However, a slight difference is observed between its surface texture restoration and that of E-D-IPB;

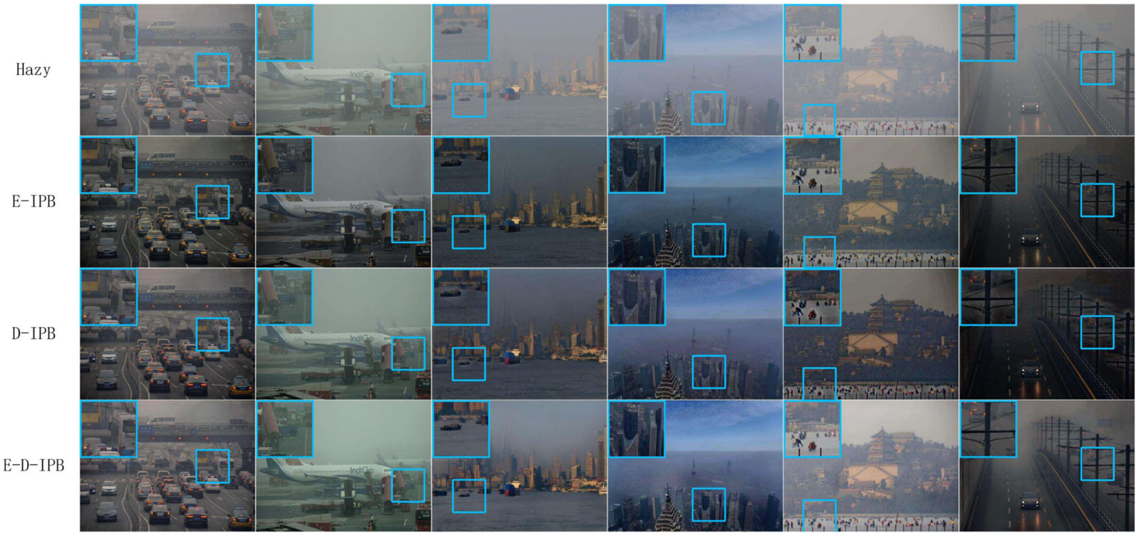


Fig. 11 Selection of test results obtained by E-IPB, D-IPB, and E-D-IPB for real-scene image dataset

Table 3 The average score obtained by different models of indoor and outdoor synthetic image datasets

Method		DCP	Hazy-lines	AOD-Net	MSCNN	GCANet	FFA-Net	MSBDN	Ours
PSNR	Indoor	28.804	28.1599	28.7184	28.4762	28.5638	29.1523	28.0113	30.8817
	Outdoor	28.5063	28.0783	28.2475	28.2134	28.8829	30.1502	27.9608	33.7365
SSIM	Indoor	0.9803	0.8157	0.9596	0.9312	0.9667	0.9649	0.9109	0.9965
	Outdoor	0.9467	0.8804	0.9701	0.9634	0.9813	0.9676	0.9465	0.998
TIME		30.03s	6.53s	0.54s	2.93s	0.6s	0.68s	0.088s	0.64s
VIDEO MEMORY		–	–	875MB	–	1201MB	1563MB	1175MB	5951MB

The average time consumption of the model to process images and computing resource consumption

this can be seen in the building and grass areas in the first and third pictures, respectively. The processing result of E-D-IPB is superior to those of E-IPB and D-IPB. The image has almost no residual haze, and the restored image presents better color saturation and contrast.

Figure 11 shows several of the test results obtained by E-IPB, D-IPB, and E-D-IPB for the real-scene image dataset. The primary problem of E-IPB is still its poor color recovery. Because of the high haze density of the displayed real image, D-IPB features artifacts on object contours. E-D-IPB contains no artifacts and shows a strong performance when processing real-scene images.

Through the ablation experiment, we conclude that the IPB's participation in the encoding and decoding processes is highly effective. Its participation in the decoding and

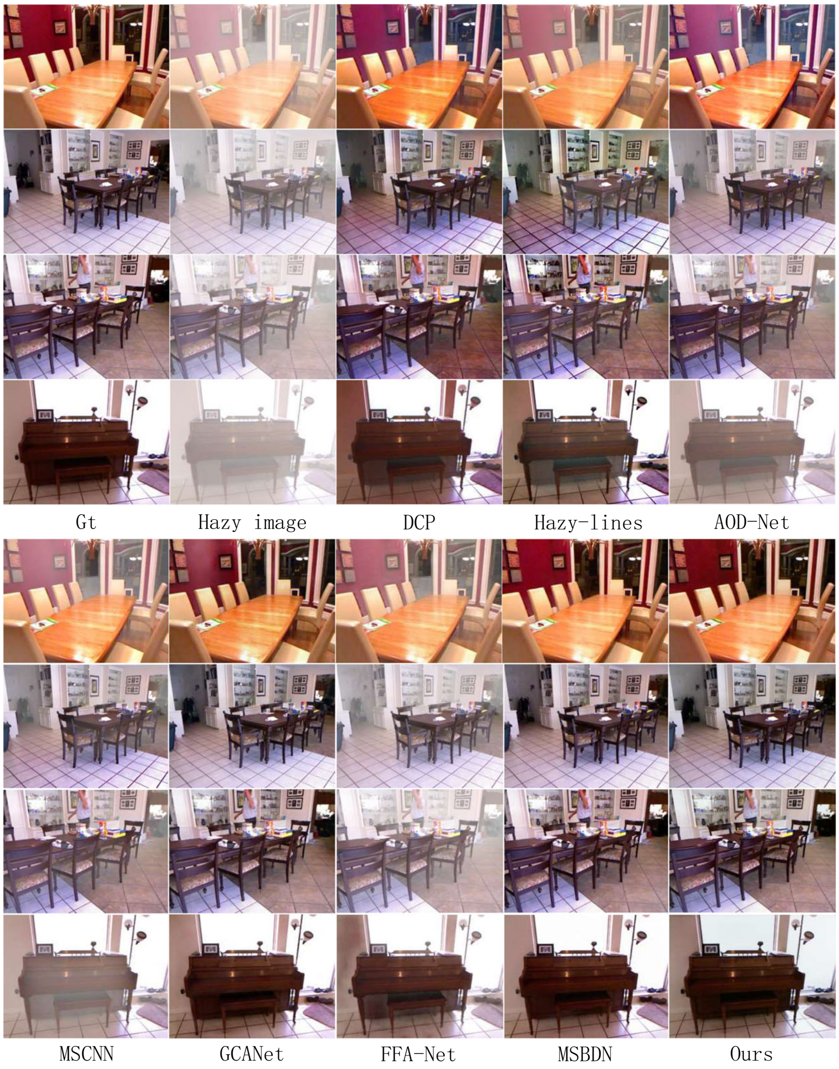


Fig. 12 Indoor synthetic image samples, to compare the subjective impressions of different model results

encoding processes can effectively restore the image colors and enhance the model's dehazing ability, respectively.

4.4 Comparisons with state-of-the-art methods

We tested the proposed method on the RESIDE test set, and compared it with DCP, Hazy-lines, AOD-Net, MSCNN, GCANet, FFA-Net and MSBDN using objective indicators and subjective visual perception. To ensure fairness, we retrained the data-driven dehazing methods using the same training set. As shown in Table 3, our proposed method achieves excellent results for both the indoor and outdoor synthetic images of SOTS, and its PSNR and SSIM scores exceed those of other methods in quantitative evaluations. The method proposed in this paper is at an average level in the calculation time consumption of a single image. The test platform is i7-10700k, TITAN RTX, and the test data size is 640 * 480. This method takes up more resources due to more parameters.

Figure 12 shows the performances of all methods for indoor synthetic images. The DCP presents a significant dehazing effect; however, the restored image quality is low, a serious color shift has occurred, and the object surface textures are poorly restored. Hazy-lines produces a poor dehazing effect; subjectively, its processing results can be seen to contain more haze residues than those of DCP; furthermore, the problems of color shift and poor object surface texture restoration observed for DCP and also be seen. AOD-Net achieves a certain dehazing effect, though considerable haze remains. Compared with DCP and Hazy-lines, AOD-Net improve its object surface texture restoration. However, owing to its relatively simple network model, the dehazing effect does not yet achieve satisfactory results. MSCNN is effective in processing light haze images, but the performance is poor when the haze density is high. GCANet achieves a strong dehazing performance. Most of the haze in the image has been removed well, though a small amount remains. FFA-Net's performance in

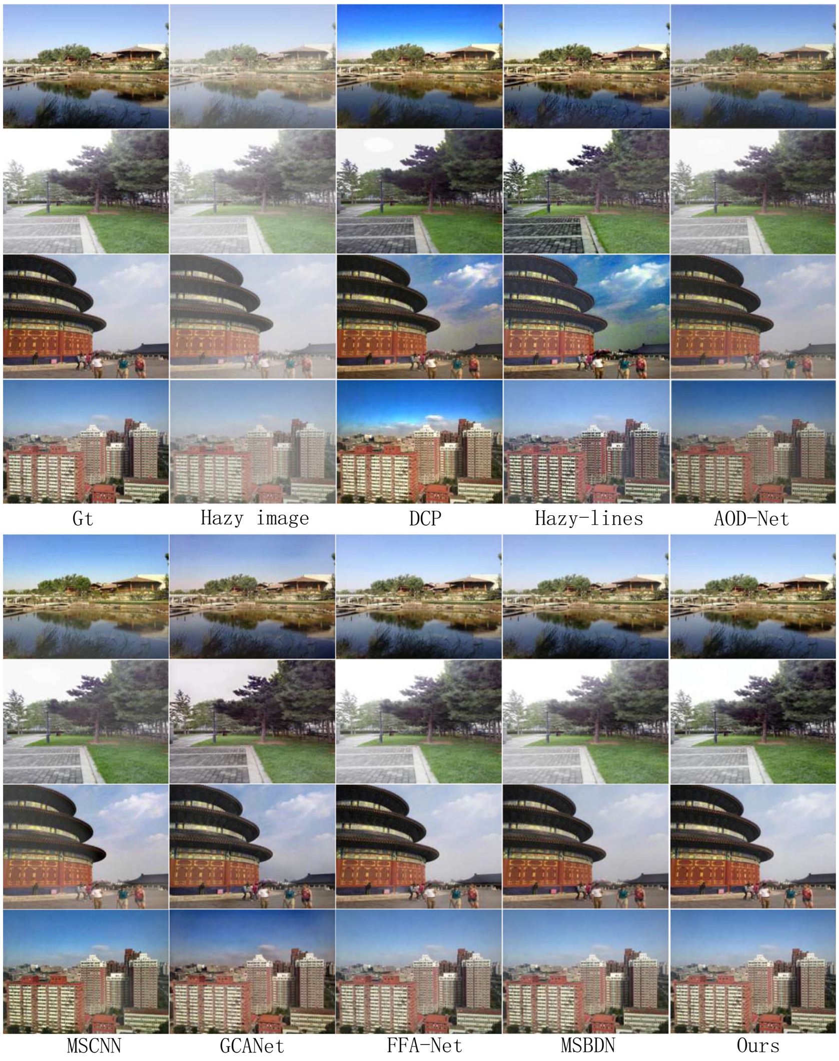


Fig. 13 Outdoor synthetic image samples, to compare the subjective visual impressions of different model results

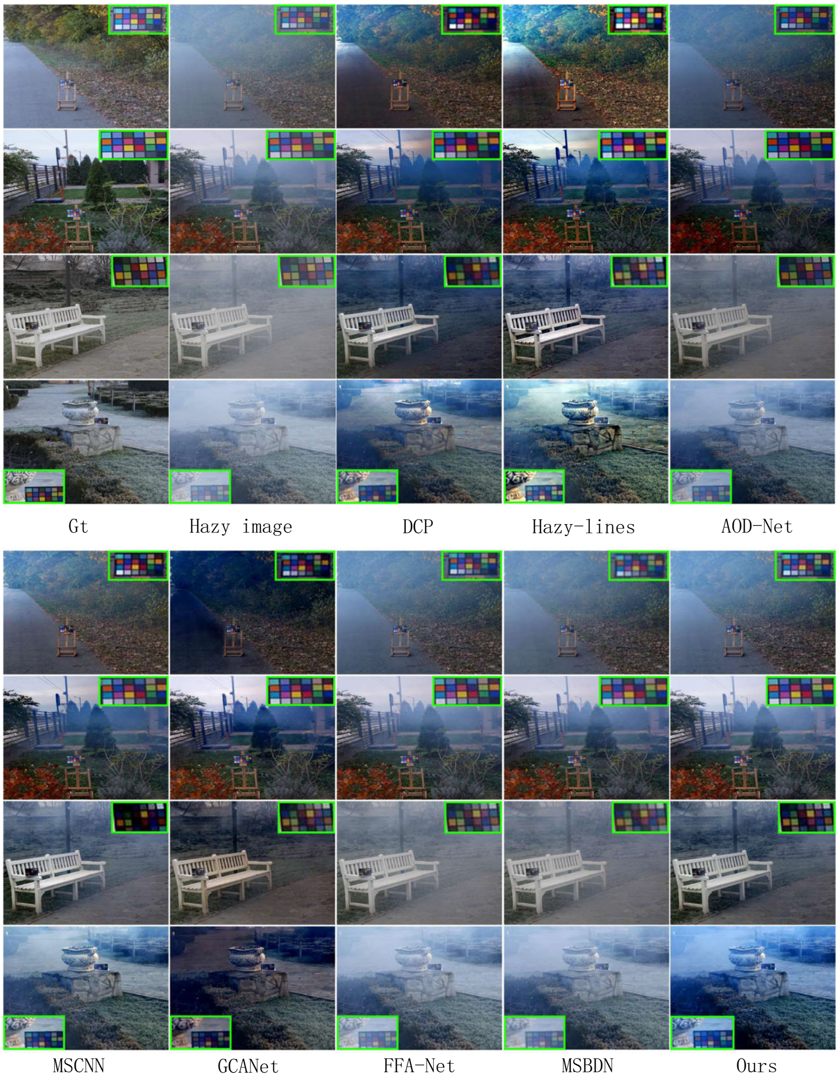


Fig. 14 Different methods are used to process the comparison of hazy images in O-HAZE

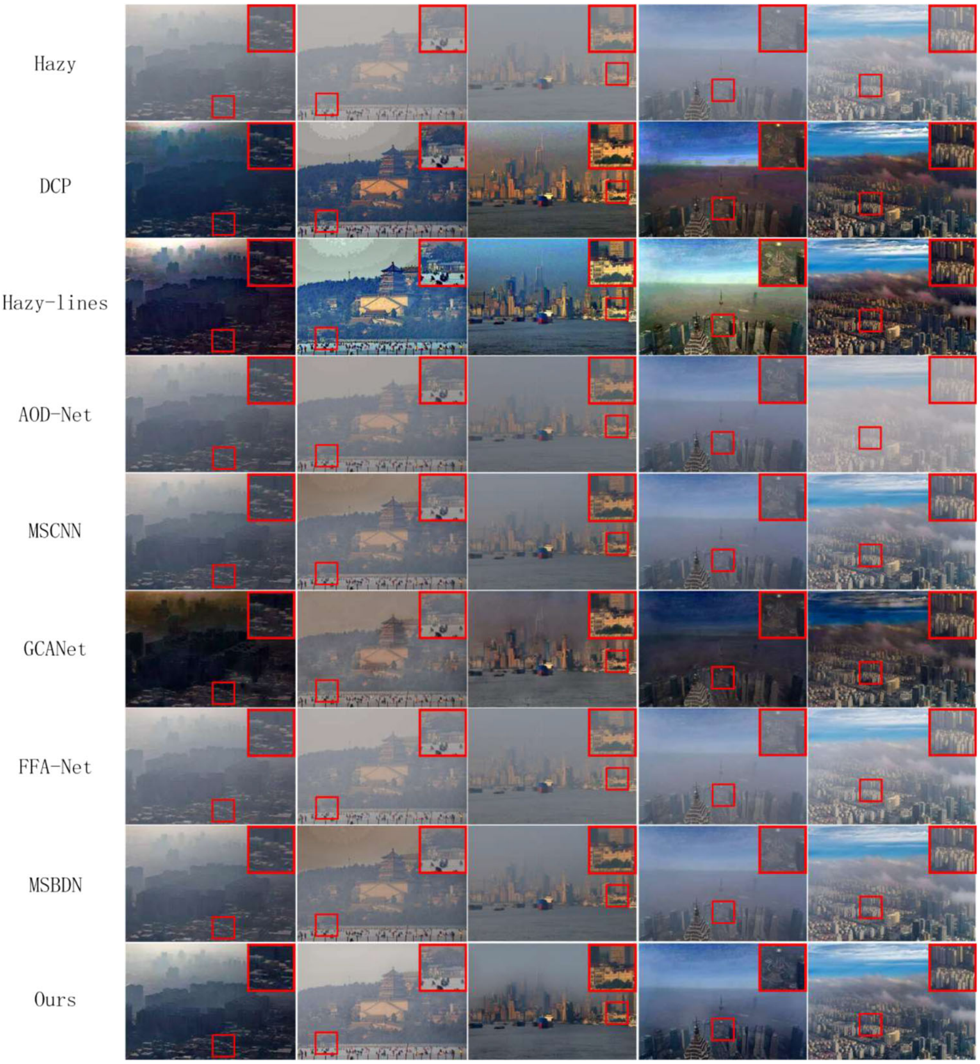
the indoor composite image is unsatisfactory. Haze residues are still present in the image, though the surface textures of the objects are better restored, and no black spots can be observed. MSBDN achieves a significant dehazing effect in the indoor composite images, and the image haze has been almost entirely removed. However, the color is darker in the deeper regions, and the details are not sufficiently clear. The method proposed in this paper achieves a significant dehazing effect: the surface textures of the objects are restored well; no color-shift problems arise; and in the quantitative comparison, it outperforms the other methods.

Figure 13 shows the performance of different methods for outdoor synthetic images. DCP achieves a significant dehazing effect in outdoor composite images; however, problems of high contrast and oversaturated sky colors are present. Hazy-lines also produces high contrasts, though the overall effect is better than that of DCP. The images processed by AOD-Net exhibit darker image tones and overall image colors. The hazy image processed by MSCNN still has a small amount of smog and color distortion. Those processed by GCANet retain a blocky haze, the sky regions are impure, and high contrasts are visible. The effect of

Table 4 The average score obtained by different models of O-HAZE

Method	DCP	Hazy-lines	AOD-Net	MSCNN	GCANet	FFA-Net	MSBDN	Ours
PSNR	28.0369	28.0407	27.9965	28.1234	27.9312	28.1805	28.1258	28.0701
SSIM	0.9589	0.9436	0.9198	0.9545	0.8703	0.9273	0.9197	0.9555

Fig. 15 Comparison of different model results for hazy real-scene image samples



FFA-Net in processing outdoor composite images is better than when processing indoor ones. The dehazing effect of MSBDN is more significant: the haze is almost entirely removed, though the contrast in some detailed regions of the image is low. Our method can effectively remove the haze from the image. Compared with the most advanced methods (FFA-Net and MSBDN), our method restores the color of the images in a more coordinated manner. Its performance in the second and fourth pictures is significantly better than those of MSBDN and the other methods.

4.4.1 Results on a real-world dataset

It is necessary to test the model on real scenes; thus, we chose the O-HAZE and collected images for testing, to compare our method against other methods. Because the hazy real-scene images lack a haze-free reference image, it is impossible to use PSNR and SSIM to compare different methods; hence, it is difficult to quantitatively compare the different methods for a real scene. However, the hazy images feature a relatively concentrated histogram distribution. Therefore, we considered the comprehensive image histograms and subjective visual impressions, to evaluate and compare the models' efficacies for real scenes.

Figure 14 is the comparison of all methods on the O-HAZE test set. Table 4 is the objective evaluation index of the method on O-HAZE. The processing results of DCP and Hazy-line are better, the details are preserved, but the environment color is distorted to a certain extent. AOD-Net performs poorly when processing such images, and the dehazing effect is not obvious. MSCNN is better than AOD-Net when processing such images, and the image clarity has a certain degree of enhancement. The dehazing effect of GCANet is more obvious, the details are clearer, but the overall contrast of the picture is too strong, resulting in distortion of the image. FFA-Net and MSBDN have almost no effect when processing such images, and the haze residue in the processed image is relatively large. The method we propose is the best among all neural network models. This method has a significant dehazing effect, less haze residue, and more obvious details are restored.

Figure 15 illustrates the effects of different methods for processing hazy real-scene images, and Fig. 16 shows the color histograms of all images in Fig. 15. Subjectively, we can see that DCP exhibits a more distinct dehazing effect, considerably altering the color histogram distributions of the hazy images; in fact, the color histograms of DCP are more balanced than those of the original images. Although the dehazing effect is clear, the image quality after dehazing is poor: serious artifacts and color shifts are visible, and the image contrast is too high. Hazy-lines' dehazing effect is better than that of DCP, and the extent of artifacts and color shifts is smaller than observed for DCP. AOD-Net performs poorly in real scenes: the dehazing effect is unclear and the color histogram distribution changes only slightly. The experimental results show that the AOD-Net trained from synthetic images struggles to process the haze in real-scene images. MSCNN removed part of the haze, but there are still many haze residues.

GCANet achieves a reasonable performance for hazy real-scene images, and the color histogram distribution is balanced. However, by observing the color histogram, the distribution of GCANet's color histogram can be seen to differ from that of the original image. Subjectively, we conclude that the image color has changed considerably, the image is insufficiently pure, and some artifacts appear. FFA-Net, trained upon a synthetic image dataset, also achieves a poor performance in processing hazy real-scene images. This problem is also reflected in the color histogram distribution. Subjectively, MSBDN achieves a reasonable effect when processing real-scene images; for instance, its color histogram distribution is more balanced than the original color histogram. However, it is prone to artifacts and other problems. The histogram distribution of the processed image exhibits small differences when compared to the original image.

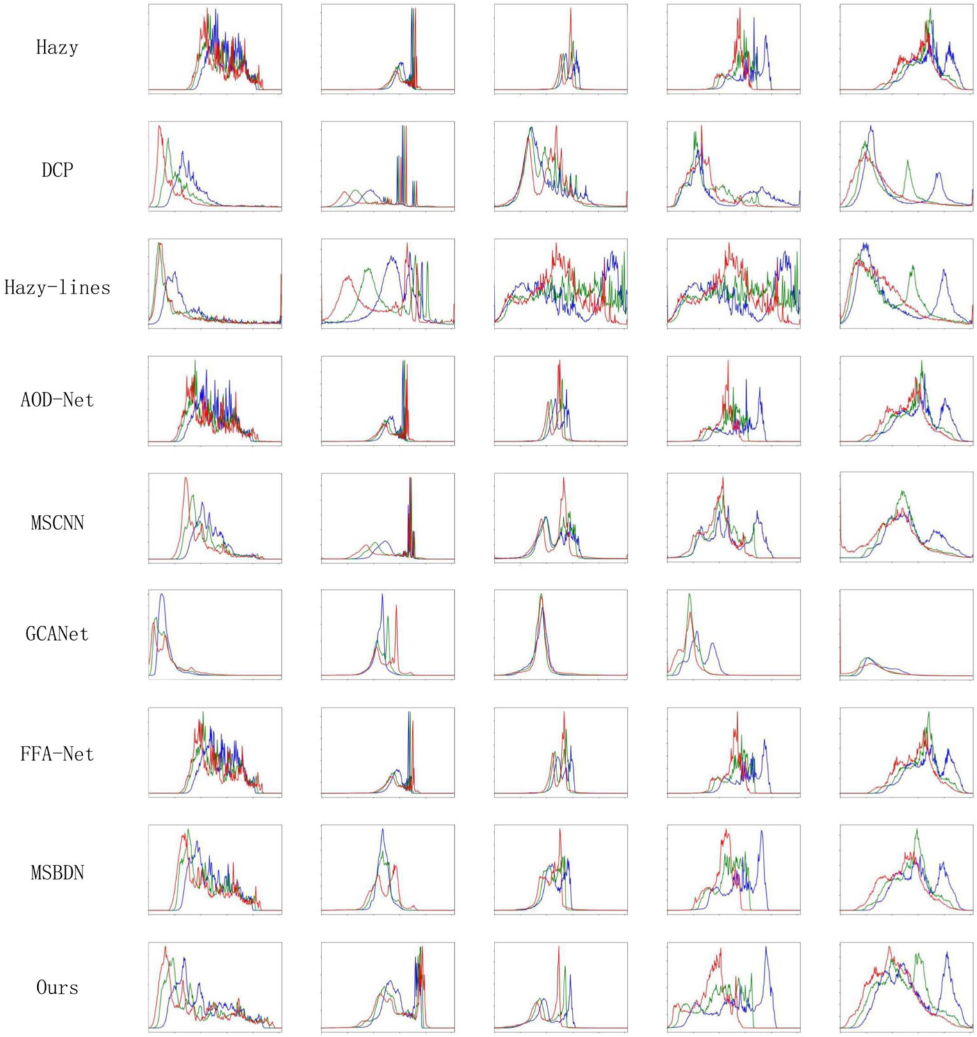


Fig. 16 Color histograms corresponding to the images in Fig. 15. Because the histogram distributions of hazy images are relatively concentrated, the subjective visual impressions and histogram distributions can be combined to compare the image-dehazing effects of different methods

Compared with the above methods, our proposed method offers clear advantages: it improves the image definition but does not produce color shifts or high contrasts, the picture is relatively pure and contains no patchy haze, and the color histogram distribution is more balanced than that of the original image. Our model uses skip-connection the pyramid features of the hazy image to the encoding and decoding parts, and it effectively retains the original color features; as a result, the color histogram retains the same features as the original color histogram. This also explains the absence of color shift after processing using our method.

By using different methods to compare results between the synthetic image test set and real-scene dataset, we found that the proposed method outperforms existing methods in terms of both qualitative and quantitative comparisons.

5 Conclusion

In this study, we proposed the MSDFN and, through numerous experiments, proved its effectiveness and superiority in dehazing real-scene images. Existing end-to-end dehazing methods only focus on constructing a channel between the hazy and clean images, and they fail to consider the influence of the depth information of the scene on the imaging effect. In addition to the hazy image itself, depth information is the most important information in the image-dehazing process. Experimental results show that our application of depth information in dehazing images is effective. We conclude that, although promising image dehazing results can be achieved using depth information, room for improvement remains. If more accurate image depth information can be acquired, the effects of image dehazing processes can be further improved.

Acknowledgments The authors acknowledge the National Natural Science Foundation of China (Grant nos. 61772319, 62002200, 61976125, 61873177 and 61773244), and Shandong Natural Science Foundation of China (Grant no. ZR2017MF049).

References

1. McCartney EJ (1977) Optics of the atmosphere: Scattering by molecules and particles. Int J Comput Vis 28(11):521–521
2. Narasimhan SG, Nayar SK (2000) Chromatic framework for vision in bad weather. In: Proceedings IEEE Conference on Computer Vision and Pattern Recognition. CVPR 2000 (Cat. No. PR00662), vol 1. IEEE, pp 598–605
3. Narasimhan SG, Nayar SK (2002) Vision and the atmosphere. Int J Comput Vis 48(3):233–254
4. He K, Sun J, Tang X (2010) Single image haze removal using dark channel prior. IEEE Trans Pattern Anal Mach Intell 33(12):2341–2353
5. Berman D, Avidan S et al (2016) Non-local image dehazing. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 1674–1682
6. LeCun Y, Bottou L, Bengio Y, Haffner P (1998) Gradient-based learning applied to document recognition. Proc IEEE 86(11):2278–2324
7. Li J, Li G, Fan H (2018) Image dehazing using residual-based deep cnn. IEEE Access 6:26831–26842
8. Cai B, Xu X, Jia K, Qing C, Tao D (2016) Dehazenet: An end-to-end system for single image haze removal. IEEE Trans Image Process 25(11):5187–5198
9. Li B, Peng X, Wang Z, Xu J, Feng D (2017) Aod-net: All-in-one dehazing network. In: Proceedings of the IEEE international conference on computer vision, pp 4770–4778
10. Qu Y, Chen Y, Huang J, Xie Y (2019) Enhanced pix2pix dehazing network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp 8160–8168
11. Chen D, He M, Fan Q, Liao J, Zhang L, Hou D, Yuan L, Hua G (2019) Gated context aggregation network for image dehazing and deraining. In: 2019 IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE, pp 1375–1383
12. Liu X, Ma Y, Shi Z, Chen J (2019) Griddehazenet: Attention-based multi-scale network for image dehazing. In: Proceedings of the IEEE International Conference on Computer Vision, pp 7314–7323
13. Qin X, Wang Z, Bai Y, Xie X, Jia H (2020) Ffa-net: Feature fusion attention network for single image dehazing. In: AAAI, pp 11908–11915
14. Dong H, Pan J, Xiang L, Hu Z, Zhang X, Wang F, Yang M-H (2020) Multi-scale boosted dehazing network with dense feature fusion. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp 2157–2167
15. Ronneberger O, Fischer P, Brox T (2015) U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. Springer, pp 234–241
16. Zhu K, Jiang X, Fang Z, Gao Y, Fujita H, Hwang J-N (2020) Photometric transfer for direct visual odometry. Knowl-Based Syst:106671
17. Bissonnette LR (1992) Imaging through fog and rain. Opt Eng 31(5):1045–1053
18. Narasimhan SG, Nayar SK (2003) Contrast restoration of weather degraded images. IEEE Trans Pattern Anal Mach Intell 25(6):713–724
19. Ibrahim H, Kong NSP (2007) Brightness preserving dynamic histogram equalization for image contrast enhancement. IEEE Trans Consum Electron 53(4):1752–1758
20. Kim J-Y, Kim L-S, Hwang S-H (2001) An advanced contrast enhancement using partially overlapped sub-block histogram equalization. IEEE Trans Circ Syst Video Technol 11(4):475–484
21. Stark JA (2000) Adaptive image contrast enhancement using generalizations of histogram equalization. IEEE Trans Image Process 9(5):889–896
22. Khan MF, Khan E, Abbasi ZA (2014) Segment dependent dynamic multi-histogram equalization for image contrast enhancement. Digital Signal Process 25:198–223
23. Richter R (1996) Atmospheric correction with haze removal including a haze/clear transition region. In: Algorithms for Multispectral and Hyperspectral Imagery II, vol 2758. International Society for Optics and Photonics, pp 254–262
24. Seow M-J, Asari VK (2006) Ratio rule and homomorphic filter for enhancement of digital colour image. Neurocomputing 69(7-9):954–958
25. Daubechies I (1990) The wavelet transform, time-frequency localization and signal analysis. IEEE Trans Inf Theory 36(5):961–1005

26. Ma J, Plonka G (2010) The curvelet transform. IEEE Signal Process Mag 27(2):118–133
27. Fattal R (2008) Single image dehazing. ACM Trans Graph (TOG) 27(3):1–9
28. Arigela S, Asari VK (2014) Enhancement of hazy color images using a self-tunable transformation function. In: International Symposium on Visual Computing. Springer, pp 578–587
29. Ren W, Liu S, Zhang H, Pan J, Cao X, Yang M-H (2016) Single image dehazing via multi-scale convolutional neural networks. In: European conference on computer visionSpringer, pp 154–169
30. Zhu H, Peng X, Chandrasekhar V, Li L, Lim J-H (2018) Dehazegan: When image dehazing meets differential programming. In: Proceedings of the twenty-seventh international joint conference on artificial intelligence, IJCAI-18. International Joint Conferences on Artificial Intelligence Organization, pp 1234–1240. https://doi.org/10.24963/ijcai.2018/172
31. Wang T, Zhang X, Jiang R, Zhao L, Chen H, Luo W (2020) Video deblurring via spatiotemporal pyramid network and adversarial gradient prior. Comput Vis Image Underst 203:103135
32. Ren W, Ma L, Zhang J, Pan J, Cao X, Liu W, Yang M-H (2018) Gated fusion network for single image dehazing. In: Proceedings of the IEEE Conference on Computerd Vision and Pattern Recognition, pp 3253–3261
33. Zhang X, Wang T, Wang J, Tang G, Zhao L (2020) Pyramid channel-based feature attention network for image dehazing. Comput Vis Image Underst:103003
34. Liu F, Shen C, Lin G, Reid I (2015) Learning depth from single monocular images using deep convolutional neural fields. IEEE Trans Pattern Anal Mach Intell 38(10):2024–2039
35. Hinton GE, Salakhutdinov RR (2006) Reducing the dimensionality of data with neural networks. Science 313(5786):504–507
36. Glorot X, Bordes A, Bengio Y (2011) Deep sparse rectifier neural networks. In: Proceedings of the fourteenth international conference on artificial intelligence and statistics, pp 315–323
37. Nair V, Hinton GE (2010) Rectified linear units improve restricted boltzmann machines vinod nair. In: Proceedings of the 27th International Conference on Machine Learning (ICML-10), Haifa
38. Maas AL, Hannun AY, Ng AY (2013) Rectifier nonlinearities improve neural network acoustic models. In: Proc. icml, vol 30, pp 3
39. He K, Zhang X, Ren S, Sun J (2015) Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In: Proceedings of the IEEE international conference on computer vision, pp 1026–1034
40. Clevert D-A, Unterthiner T, Hochreiter S (2015) Fast and accurate deep network learning by exponential linear units (elus). arXiv:1511.07289
41. Chen Y, Dai X, Liu M, Chen D, Yuan L, Liu Z (2020) Dynamic relu. arXiv:2003.10027
42. Hua Z, Fan G, Li J (2020) Iterative residual network for image dehazing. IEEE Access 8:167693–167710
43. Silberman N, Hoiem D, Kohli P, Fergus R (2012) Indoor segmentation and support inference from rgbd images. In: European conference on computer vision. Springer, pp 746–760
44. Li B, Ren W, Fu D, Tao D, Feng D, Zeng W, Wang Z (2017) Reside: A benchmark for single image dehazing, vol 1. arXiv:1712.04143

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Guodong Fan received his bachelor's degree from the School of Computer Science and Technology, Shandong Technology and Business University, Yantai, China in 2018. Currently studying for a master's degree in the School of Information and Electronic Engineering, Shandong Technology and Business University, Yantai, Shandong. His research interests include computer graphics, computer vision, and image processing.

Zhen Hua received the B.S. and M.S. degrees in electrical automation from Taiyuan University of Technology, Taiyuan, China, in 1989 and 1992, respectively, the Ph.D. degree in electronic information engineering from China University of Mining and Technology, Beijing, China, in 2008. She is currently a professor at Shandong Technology and Business University. Her research interests include computer aided geometric design, information visualization, virtual reality, and image processing.

Jinjiang Li received the B.S. and M.S. degrees in computer science from Taiyuan University of Technology, Taiyuan, China, in 2001 and 2004, respectively, the Ph.D. degree in computer science from Shandong University, Jinan, China, in 2010. From 2004 to 2006, he was an assistant research fellow at the institute of computer science and technology of Peking University, Beijing, China. From 2012 to 2014, he was a Post-Doctoral Fellow at Tsinghua University, Beijing, China. He is currently a Professor at the school of computer science and technology, Shandong Technology and Business University. His research interests include image processing, computer graphics, computer vision, and machine learning.