Progressive encoding-decoding image dehazing network

Wang Li[1] · Guodong Fan[1] · Min Gan[1,2]

Abstract

In this paper, we propose a progressive encoding-decoding network (PEDN) for image dehazing. First, we built a basic dehaze unit to progressively process the image to achieve image dehazing in stages. The basic dehaze unit is composed of a feature memory module and an encoding-decoding network. The feature memory module is used to transfer features at different progressive stages. The encoding-decoding network is responsible for feature extraction, encodes and decodes images by fusing different levels of pyramid features. The basic dehaze unit shares parameters during the progressive process, which effectively reduces the difficulty of network training and improves the fitting speed. The proposed model is an end-to-end image dehazing network, which does not depend on the atmospheric scattering model. In addition, we extracted the depth information of the hazy image and obtained its pyramid features, and incorporated the depth information into the feature extraction to guide the network to restore clear images more accurately. Experiments show that the our method not only performs well on synthetic datasets, but also has excellent performance on real-world hazy images. It is superior to current image dehaze methods in quantitative indexes and visual perception. Code has been made available at https://github.com/LWQDU/PEDN.

Keywords Image dehazing · Deep convolutional network · Scene depth · Real scene

1 Introduction

Atmospheric light will be absorbed or scattered by particles in the air, resulting in reduced visibility. Therefore, the contrast and color saturation of the image decrease in the process

---

Wang Li and Guodong Fan contributed equally to this work.

✉ Guodong Fan
  fgd96@outlook.com

  Wang Li
  liwangliman@163.com

  Min Gan
  aganmin@aliyun.com

1 School of Computer Science and Technology, Qingdao University, 266071 Qingdao, Shandong, China

2 College of Computer and Data Science, 350108 Fuzhou, Fujian, China

of degradation. In particular, automatic systems such as surveillance, smart vehicles and object recognition essentially rely on high-quality images, so image dehazing becomes a key preprocessing step [38, 50]. Therefore, how to restore a hazy image to a clean image is an important issue in the field of computer vision and image processing.

Nayar and Narasimhan [39, 40] made a detailed derivation and summary of the hazy imaging mechanism, and proposed an atmospheric scattering model. The current main image dehazing methods are mostly based on the atmospheric scattering model, which converts the hazy image by directly or indirectly estimating the transmission map and atmospheric light. The atmospheric scattering model can be 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 clear image, $T(x)$ is the transmission map, A is the atmospheric light, and x represents the coordinates of the pixel in the hazy image, β represents the atmospheric scattering coefficient, and $d(x)$ represents the depth information.

It can be clearly obtained from formula (1) and (2) that transmission map and atmospheric light must be accurately estimated if a hazy image is to be restored to a clear one. Because image dehazing is an ill-posed problem and its solution space is gigantic, it is very difficult to estimate the transmission map and atmospheric light accurately. Atmospheric light can be easily estimated from the original hazy image, for example by dark channel prior algorithm [20]. But the transmission map is not easy to estimate because it is highly dependent on the scene depth.

Some previous studies aimed to estimate transmission map and atmospheric light by prior methods. For example, He et al. [20] estimated transmission map by exploring dark channel prior and believed that the local minimum value of dark channel in haze-free images was close to zero. Zhu et al. [61] proposed color attenuation prior and pointed out that in hazy images, the higher the fog concentration, the greater the scene depth, and the greater the difference between brightness and saturation of the image. But the prior does not perform well at close range. Therefore, it can be seen that the prior based approach is feasible in some cases, but once the scene does not meet the prior, it will lead to inaccurate estimation of the transmission image and poor dehazing effect.

Recently, deep learning has achieved great success in the field of image processing. There are many image dehazing methods [31, 39, 40] based on convolutional neural network (CNN). Some of these methods aim to estimate transmission map or atmospheric light by using CNN, and then to obtain clean images by using atmospheric scattering model. For example, DehazeNet [5] extracts the relevant features of hazy images through CNN to optimize the estimation of transmission map, and restores clean images by using atmospheric scattering model based on empirical assumptions of atmospheric light. Multi-scale convolutional neural network (MSCNN) [46] predicts the transmission map of images by designing a group of coarse-scale and fine-scale networks respectively. Finally, the transmission map after multi-scale fusion is substituted into the atmospheric scattering model to complete image dehazing. AOD-Net [33] fuses atmospheric light and transmission image into a variable K, and then learns the relationship between K and hazy image through CNN. It can be seen that the above methods depend on the atmospheric scattering model. However, the method based on the atmospheric scattering model is deeply depend on the estimation of transmission map and atmospheric light. Once the estimation of these two factors is inaccurate, the dehazing effect will be affected.

According to the review of previous work above, it can be found that both the earlier prior based haze removal methods and the latest deep-learning based methods mostly rely on atmospheric scattering model. The transmission map or atmospheric light is estimated first and then substituted into the atmospheric scattering model to restore the haze-free image. The disadvantage of this kind of method is that once the transmission map or atmospheric light estimation is not accurate, it will greatly affect the quality of the dehazing image.

Some recent methods directly learn the mapping from hazy images to clean images without relying on the atmospheric scattering model and are called end-to-end learning methods. For example, EPDN [44] uses GAN embedded in the architecture and a well-designed enhancer to remove haze. GridDehazeNet [36] improves network performance through three modules: pre-processing, backbone and post-processing. MSBDN [10] achieves haze removal through dense fusion of multi-scale features. Although these methods do not rely on the atmospheric scattering model, they do not take into account the key factors affecting image dehazing in the model design, so the output image has the problem of blurred details or missing colors. It can be seen from formula (2) that scene depth is a key factor affecting the quality of haze removal. Therefore, it is necessary to consider the influence of this key factor in the design of haze removal model. With the continuous progress of depth camera technology, it is no longer difficult to obtain depth information of images. Therefore, this encouraged us to integrate the depth information of the scene into the dehazing model.

In this paper, we propose a progressive encoding-decoding network to improve dehazing performance. Firstly, we propose the PEDN, which decouples image dehazing from atmospheric scattering model. This network does not depend on atmospheric scattering model, and enables the neural network to directly learn the mapping function from hazy image to clean image. The basic dehaze unit (BDU) is the core of the model. Given a hazy image, the image is progressively processed by constructing a BDU, and finally a clean image is output. The BDU consists of a feature memory module (FMM) and an encoding-decoding network. The function of FMM is to transfer the features of different progressive stages, and the encoding-decoding network is used to extract the pyramid features of images and encode and decode the features. According to the atmospheric scattering model, depth information is crucial for image haze removal. Therefore, we obtain the scene depth information of the image and extract pyramid features, which are integrated into feature learning.

Experimental results show that the proposed method is superior to other methods in both PSNR and SSIM for synthetic data sets and hazy images of real scenes. The main contributions of this paper can be summarized as follows:

(i) We propose the PEDN, which can produce clean images directly, independent of the atmospheric scattering model. This method can produce higher quality dehazing images.

(ii) By using the progressive idea, the model can share parameters and avoid the difficulty of training or slow fitting speed caused by too many parameters.

(iii) In order to enhance the model performance, the depth information of images is extracted and integrated into feature extraction to better assist network learning.

(iiii) Extensive experiments show that our method is superior to other methods in overall perception and quantitative indicators, whether in synthetic data sets or real scenes.

2 Related work

Image dehazing is an ill-posed problem. Current image dehazing methods can be divided into three categories: image-enhancement based dehazing methods, prior based dehazing methods [20][14][4] and deep-learning based dehazing methods [11, 15, 32, 37, 45, 51, 59].

2.1 Image-enhancement based dehazing method

The dehazing method based on image enhancement does not consider the reason for image degradation and improves the visual effect of the image by enhancing the contrast. This kind of algorithm is widely used, but it may cause certain loss or over-enhancement of the salient information in the image. Among them, histogram equalization is to transform the histogram of hazy image into the form of uniform distribution, and improve the image contrast by expanding the range of pixel gray value. Stark [53] and Kim et al. [25] proposed adaptive histogram equalization algorithm and partially overlapping sub-block histogram equalization algorithm respectively. Homomorphic filtering [19] is a technique widely used in signal and image processing, which combines gray transform with frequency filtering to improve image quality. Retinex [28] is a color vision model that simulates what humans see under different light conditions. Based on this model, Adrian et al. [18] proposed an effective hazy image enhancement algorithm. However, none of the above methods consider the essential cause of haze map degradation, so the enhancement effect is limited and the robustness is commonly poor.

2.2 Prior based dehazing method

Dark channel dehazing algorithm is proposed by He et al. [20], which is based on dark channel prior theory (DCP). The transmission map is calculated by the principle of dark channel dehazing, and the soft cut algorithm is used to further optimize the transmission map, and finally the image dehazing is realized. According to the DCP, the minimum channel value of the local map block of haze-free image tends to 0, which is called the dark channel, while the dark channel of hazy image does not approach 0 due to the influence of atmospheric light. Therefore, the thickness of haze can be judged by the dark channel value of hazy image, so as to obtain the atmospheric light and transmission map. Dark channel dehazing algorithm is pioneering and effective in certain scenes. However, if there are objects in the target scene whose color is similar to atmospheric light, such as snow, white wall, sea, etc., it will not achieve satisfactory results. At the same time, its computational complexity is high, which affects the efficiency of dehazing. Zhu et al. [61] proposed a dehazing method based on color attenuation prior, establishing a linear model according to the positive correlation between haze concentration and the difference between image brightness and saturation, and learning model parameters through supervised learning method to recover scene depth information, so as to achieve a single image dehazing. However, this method is difficult to collect samples and lacks theoretical basis. Berman et al. [4] proposed a global evaluation transmission image method based on non-local prior, which could recover both the scene depth and haze-free image. However, when atmospheric light was very intense, this method failed because it could not detect haze lines correctly. Wng et al. [55] proposed a multi-scale deep fusion dehazing algorithm by using Markov random field to mix details of multi-layer chromaticity prior, which can obtain high-quality restored images with rich details, but this algorithm is prone to contain noise. Tarel et al. [54] proposed a fast dehazing algorithm. The algorithm estimates the dissipation function through the deformation of median filter, but with median filter estimation method, improper parameter setting can easily bring Halo effect. Singh et al. [52] proposed gradient profile prior (GPP) to evaluate the depth map of hazy image, which can effectively suppress the visual artifacts of hazy image. Liu et al. [35] proposed a multi-scale correlated wavelet dehazing method, which believes that the haze is usually distributed in the low-frequency spectrum of its multi-scale wavelet decomposition. Based on the priori,

an open dark channel model is designed to eliminate the haze effect in the low-frequency part, and the soft threshold operation is used to reduce the noise. Although the dehazing method based on prior has achieved good results to some extent, once the prior is inconsistent with the actual situation, the image dehazing effect will be significantly reduced.

2.3 Deep-learning based dehazing method

In recent years, with the continuous development of deep learning [6, 7, 9, 16, 17, 23, 30, 42, 56, 57], many methods have applied CNN to image dehazing, and achieved excellent results. Different from the prior based method, the deep-learning based method directly estimates the transmission map and atmospheric light without relying on prior. Cai et al. [5] proposed an end-to-end haze removal model called DehazeNet based on CNN, which directly learned the mapping relationship between hazy images and transmission images through neural network, and then calculated the transmittance. Furthermore, a novel nonlinear activation function BReLU is proposed. The transmission map is obtained by nonlinear regression of CNN, and the atmospheric light is obtained by assuming prior. Finally, the hazy-free image is recovered from the atmospheric scattering model. Ren et al. [46] proposed MSCNN algorithm, which constructed subnets of different granularity through multi-scale network structure to achieve coarse and fine estimation of transmission image. This algorithm effectively suppressed the halo phenomenon in the process of dehazing. Li et al. [33] proposed the network of AOD-Net, believing that predicting the atmospheric light and transmittance independently and then using the atmospheric scattering model to get a clean image would enhance the error. Therefore, they appropriately deformed the formula of atmospheric scattering model, combined transmittance and atmospheric light into a variable, and finally learned the variable through neural network. Mei et al. [37] proposed the progressive feature fusion dehazing network (PFFNet) similar to U-Net structure, which is composed of decoder, feature conversion and decoder, and directly learns the nonlinear transformation function from hazy image to clear image.

Chen et al. [8] proposed an end-to-end image dehazing algorithm based on GAN (gated context aggregation network, GCANet). The key point of this algorithm is to use smooth extended convolution instead of extended convolution to solve the problem of grid artifacts. At the same time, a new fusion network is proposed to fuse the features of different levels and enhance the image dehazing effect. Qin et al. [43] proposed an end-to-end feature fusion attention network (FFA-Net), which can retain shallow information and transmit it to deep layers through an attention-based feature fusion structure. Dong et al. [10] proposed a multi-scale enhanced defogging network (MSBDN) with dense feature fusion based on U-Net architecture. Fan et al. [12] proposed a multi-scale deep information fusion network (MSDFN) based on U-Net architecture. The model extracts the depth information of the hazy image, codes and decodes it together with the hazy image, and generates the haze-free image in an end-to-end manner.

3 Proposed method

In this section, we first introduce the design motivation of the proposed model, then introduce the network architecture and the composition and function of each module, and finally introduce the loss function adopted.

3.1 Design motivation

In recent years, in the field of deep learning, many methods aim to enhance the performance of the network by increasing the depth of the network, but a large number of parameters are introduced, which increases the training difficulty of the network and also increases the risk of over-fitting. How to design an effective network without introducing too many parameters becomes an important problem in the field of image processing. Kim et al. [26] proposed DRCN for image super-resolution, which increased the depth of the network, increased the size of the receptive field of the network and reduced the number of parameters through recursion. Inspired by DRCN, we can use the progressive idea to design an effective dehazing model. By repeatedly executing the BDU many times, the feature learning ability can be continuously improved, and the problem of network capability degradation caused by too many parameters can also be avoided. It can be seen from the atmospheric scattering model that the transmission map is related to the scene depth. Experience has also shown that on hazy days it is easier to distinguish objects close by than objects far away. Therefore, scene depth is essential for image dehazing. Fan et al. [12] proposed MSDFN to introduce depth information into the haze removal process and found that the depth information of the scene can effectively upgrade the haze removal ability of the model in the real scene. Therefore, we consider adding depth information of images into the model to enhance its ability to remove haze.

3.2 Network design

The network structure of PEDN is shown in Fig. 1. The core of PEDN is the BDU, whose structure is shown in Fig. 2. After the hazy image is input to the network, the BDU is repeatedly executed several times to finally output the haze-free image. Specifically, the output feature map of each BDU, together with the original hazy image and the corresponding depth map, will serve as the input of the next BDU. In order to better retain the original image information and depth information, the original haze image and its depth map have been run through the whole network learning process. The number of BDU is more critical. In order to balance the relationship between performance and resources, 5 is selected as the number of BDU. The functions of the BDU can be defined as follows:

$$T_t = D(E(FMM(T_{t-1} \oplus I_{hazy}) \oplus I_{depth})), \tag{3}$$

where t represents the number of BDU, and T_t represents the output of the t-th BDU. When $t=0$, T_0 represents a hazy image. I_{hazy} and I_{depth} represent the hazy image and its
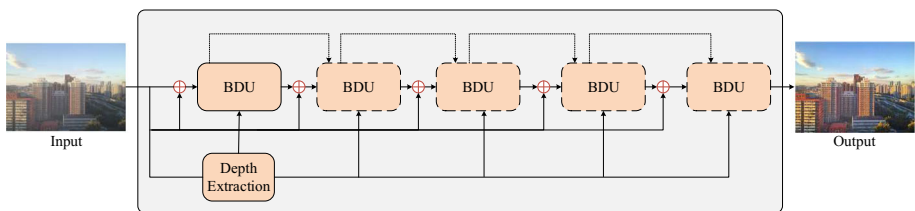
Fig. 1 Architecture of the proposed PEDN. Given a hazy image, the image is progressively processed by constructing a BDU, and finally a clean image is output. The depth information of hazy images is extracted and integrated into network learning
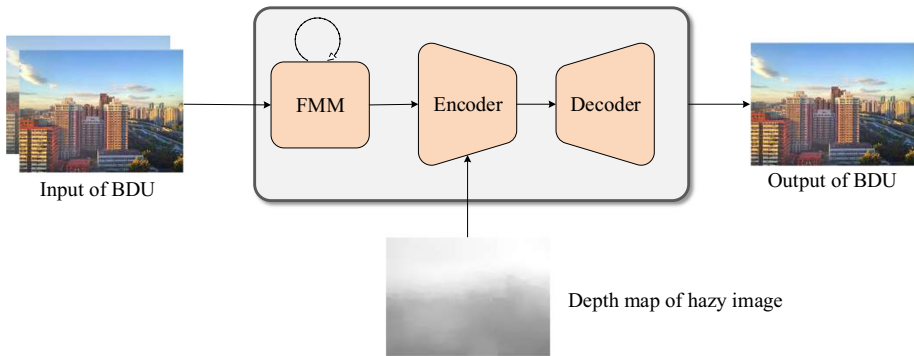
Fig. 2 Internal structure of BDU. The BDU consists of FMM and encoding-decoding network

corresponding depth map respectively. $FMM(\cdot)$, $E(\cdot)$ and $D(\cdot)$ represent FMM, encoding branch and decoding branch respectively. $\oplus$ indicates the connection operation of the channel dimension.

3.2.1 Feature memory module

The FMM is the first part of the BDU, and its structure is shown in Fig. 3. With the increase of executions times of BDU, the original image features will be weakened. Therefore, it is necessary to introduce the FMM to retain the original features. LSTM [21, 24, 41, 49] is commonly used in time series problem modeling. Its unique design can remember and transfer information for a long time, and select features of different moments through a gating mechanism. Therefore, we choose to introduce LSTM into the model as a FMM, which is used to transmit features at different progressive stages. The original hazy image was connected with the output of the last BDU as the input of this FMM, and 32 convolution kernels with a size of 3×3 were used to extract features. Then, the extracted features are merged with the output of FMM of the last BDU, and the merged feature graph needs to perform convolution operation to obtain four gated states for controlling the selection and output of features. FMM can be expressed as:

$$c_t = f \times c_{t-1} + i \times z, \tag{4}$$

$$h_t = o \times Tanh(c_t). \tag{5}$$

3.2.2 Encoding-decoding network

The encoding-decoding network is the second part of the BDU and its structure is similar to U-Net [48]. Its function is to receive the output of FMM and extract the deep features. As shown in Fig. 4, the output of the FMM and the depth map of the original image serve as input to the encoding-decoding network. The purpose of adding the depth map is to make better use of the depth features of the image. In the encoding branch, the depth map and the output of the FMM are coded separately, and feature fusion is performed after the pooling operation. The encoding network uses 64, 128, 256 and 512 convolution kernels respectively to conductconvolution operation on the feature map, and the max pooling operation is
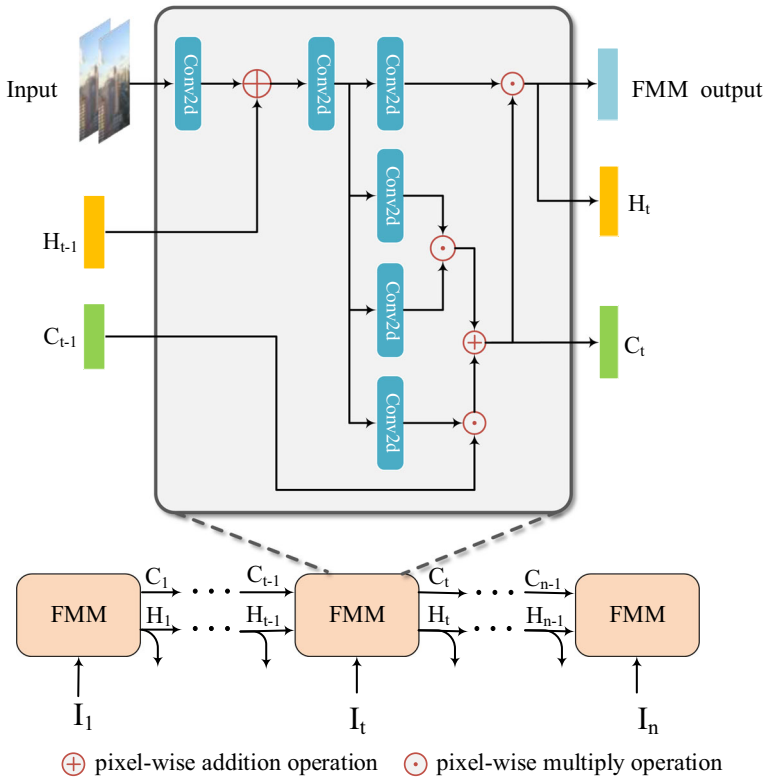
Fig. 3 Internal structure of Feature Momery Module (FMM)

performed after each convolution operation. The decoding network gradually restores the scaled feature map, and introduces skip connections between the same layers to avoid gradient disappearance. The network eventually outputs 3 channels of images.
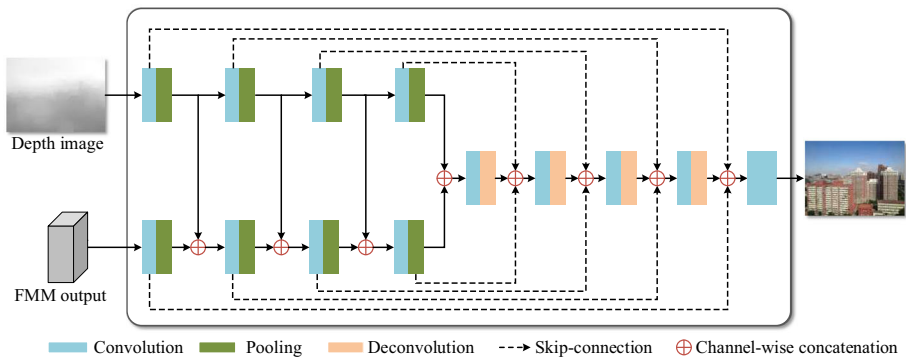
Fig. 4 Internal structure of Encoding-Decoding Network

3.3 Loss function

In the field of image dehazing, L1 and L2 are commonly used loss functions. In addition, there are also some mixed functions, such as L2 loss combined with structural loss [13], and L2 loss combined with adversarial loss [60]. Although some methods improve the image dehazing effect, the calculation is complicated, which affects the efficiency of the network. In this paper, we use -SSIM [22, 58] as the loss function to optimize the network. The reason why -SSIM was chosen was that the recovered image could be more consistent with human visual experience by comprehensively considering brightness, contrast, structure and other factors. The reason for using -SSIM as the loss function is that the closer the output is to the true value, SSIM will be larger, and conversely, -SSIM will be smaller. The loss function can be expressed as:

$$
L = -\frac{1}{2M} \sum_{m=1}^{M} \frac{\left(2\mu_{Y_m}\mu_{Y_m'} + \theta_1\right)\left(2\sigma_{Y_mY_m'} + \theta_2\right)}{\left(\mu_{Y_m}^2 + \mu_{Y_m'}^2 + \theta_1\right)\left(\sigma_{Y_m}^2 + \sigma_{Y_m'}^2 + \theta_2\right)},
\tag{6}
$$

where M is the number of images, Y_M is the hazy image currently processed, Y_m^t is the ground truth. μ is the mean, σ is the variance, θ_1 and θ_2 are constants.

In order to verify the superiority of -SSIM as loss function, we use MSE, L1 and -SSIM as loss function on four datasets, and get their fitting curves. As can be seen from Fig. 5, - SSIM has obvious advantages over MSE and L1. Under the same number of iterations, -SSIM as a loss function can achieve better results. This is because MSE and L1 only consider the distance between pixels, while -SSIM comprehensively considers factors such as brightness, contrast and structure, which can better fit people's visual experience.

4 Experimental results

In this section, we use synthetic datasets and real haze datasets to evaluate the proposed model. We conduct ablation experiments to verify that different numbers of BDU have different effects on the model's dehazing performance. To more clearly demonstrate the advantages of the proposed model, our method was compared with other state-of-the-art dehazing methods, including DCP [20], NLD [4], MSCNN [46], AOD-Net [33], PFFNet [37], GCANet [47], FFA-Net [43], MSBDN [10] and MSDFN [12].
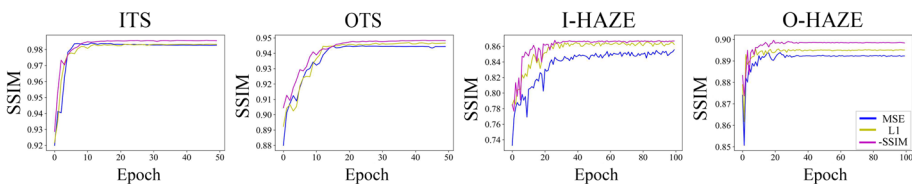

Fig. 5 Fitting curves of different loss functions on four datasets

4.1 Experimental settings

4.1.1 Datasets

We mainly use synthetic dataset and heterogeneous dataset to train our model. The synthetic dataset adopts the RESIDE dataset [29] and D-Hazy dataset [2]. RESIDE dataset is a benchmark dataset for image dehazing proposed in 2018, which contains indoor training set (ITS), outdoor training set (OTS) and test set (SOTS). We use ITS and OTS to train our model, and use SOTS to evaluate our model. D-Hazy dataset contains more than 1400 pairs of images with ground truth reference images and hazy images of the same scene. In order to make our model have stronger generalization ability, we also use heterogeneous datasets to train the model. Therefore, we use I-HAZE [1] and O-HAZE [3] as supplementary datasets. In order to increase the number of images, we perform data enlargement. Because our model requires the depth map of the hazy image, and the selected data set lacks the depth information of the image, we use the DCNF [34] model to generate the depth map of all the datasets. The reason of choosing DCNF as the depth prediction model is the RESIDE dataset adopts the DCNF model to synthesize hazy images, and we keep it consistent with RESIDE. In addition, compared to other depth prediction methods, the DCNF model produces the least visible depth error and causes much less visual artifacts on natural outdoor images. Using the DCNF model leads to more visually plausible results.

4.1.2 Training details

All experiments used the PyTorch framework and were trained using GeForce RTX 3090. ADAM optimizer [27] was used to train the network, and the initial learning rate was set to 0.0001, batchsize to 4, and epoch to 50.

4.1.3 Quality measures

SSIM evaluates the similarity of two images in terms of brightness, contrast and structure. Brightness is expressed as the mean value:

$$\mu_x = \frac{1}{N} \sum_1^N x_i, \tag{7}$$

the brightness difference between the two images is expressed as:

$$l(x, y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2}, \tag{8}$$

where μ_x and μ_y represent the mean of x and y respectively, C_1 is constant. Contrast is represented by variance normalized by the mean:

$$\sigma_x = \left(\frac{1}{N-1} \sum_{i=1}^N (x_i - \mu_x)^2 \right)^{\frac{1}{2}}, \tag{9}$$

the contrast difference between the two images is expressed as follows:

$$c(x, y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}, \tag{10}$$

where σ_x and σ_y represent the variance of x and y respectively, C_2 is constant. Structural differences are expressed by correlation coefficients:

$$s(x, y) = \frac{\sigma_{xy} + C_3}{\sigma_x \sigma_y + C_3}, \tag{11}$$

where σ_{xy} is the covariance of x and y, C_3 is constant. The final SSIM can be expressed as:

$$SSIM(x, y) = \frac{(2\mu_x \mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_1^2 + \mu_2^2 + C_1)(\sigma_1^2 + \sigma_{12}^2 + C_2)}. \tag{12}$$

In the experiment, PSNR and SSIM were used as objective evaluation indexes of the model. In general, a higher PSNR represents a higher resolution image and is usually defined using Mean Square Error (MSE):

$$MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I(i, j) - K(i, j)]^2. \tag{13}$$

PSNR is defined as:

$$PSNR = 20 \cdot lg(\frac{MAX_I}{\sqrt{MSE}}), \tag{14}$$

where I and K represent two images of m×n size. The unit of PSNR is dB. The larger the value is, the closer it is to the original image.

4.2 Performance evaluation

We use SOTS, I-HAZE, O-HAZE, D-Hazy and real scene images to test our method, and compare our method with DCP [20], NLD [4], MSCNN [46], AOD-Net [33] , PFFNet [37], GCA-Net [47], FFA-Net [43], MSBDN [10] and MSDFN [12]. For a fair comparison, we use the same datasets to retrain these methods, and use PSNR and SSIM to evaluate all methods. The Table 1 shows the test results of these methods on the SOTS, I-HAZE, O-HAZE and D-Hazy datasets. It can be seen that our method has greater advantages than other methods on both PSNR and SSIM. Figure 6 can also visually illustrate the superiority of our method. Figure 7 shows the data distribution of different methods on all data sets. The more flat the box is, the more concentrated the data distribution is and the more stable the model is. Compared with other methods, our method has more centralized data distribution, so it has greater advantages.

Figure 8 shows the processing results of each method on indoor hazy images of SOTS dataset [29], and the differences of each method can be felt subjectively. We can see that our method basically eliminates haze in the image and produces an image without distortion. DCP and NLD are considered as baseline methods due to their simplicity of implementation, and when the parameters are more accurate, the haze removal effect is better. However, their shortcomings are also more obvious, especially because color distortion is more serious. MSCNN and AOD-Net cannot effectively remove the haze in the image. It can be seen that the haze in the image generated by MSCNN and AOD-Net still remains to a certain extent. Other methods can restore clear images better, but by comparing details, it can be found that the method proposed by us is closer to the ground truth in detail.

Figure 9 shows the processing effects of each method on outdoor hazy image of SOTS dataset [29]. DCP and NLD cannot effectively achieve image dehazing, and the

Table 1 Quantitative results on benchmark dehazing datasets

Dataset	Metrics	DCP	NLD	MSCNN	AOD-Net	PFFNet	GCANet	FFA-Net	MSBDN	MSDFN	Ours
SOTS-Indoor	SSIM ↑	0.857	0.7769	0.7942	0.7568	0.8911	0.9589	0.9834	0.8667	0.935	0.9840
	PSNR↑	20.02	17.3055	17.139	14.5171	24.715	30.07	33.5911	25.604	26.544	32.3205
SOTS-Outdoor	SSIM ↑	0.8318	0.8218	0.8543	0.8876	0.8353	0.9179	0.9455	0.927	0.944	0.9465
	PSNR↑	16.9796	17.9118	19.539	22.4536	21.046	22.9659	32.068	29.5252	32.5852	32.8899
I-HAZE	SSIM ↑	0.6512	0.7214	0.7245	0.7223	0.8032	0.684	0.6629	0.7393	0.841	0.8479
	PSNR↑	12.9449	14.214	15.6374	15.7258	20.2778	14.628	16.2809	16.285	20.986	21.4103
O-HAZE	SSIM ↑	0.5752	0.636	0.7435	0.7007	0.8977	0.6831	0.7184	0.6387	0.8343	0.8598
	PSNR↑	16.1744	15.4425	19.1559	18.2019	25.664	16.6301	18.2667	18.377	22.408	24.2086
D-Hazy	SSIM↑	0.8039	0.7641	0.7677	0.6808	0.7662	0.8006	0.5497	0.7922	0.6872	0.8371
	PSNR↑	15.88	14.4	13.86	13.74	15.08	15.78	10.45	16.71	12.48	18.47
Average running time		27.44s	4.32s	1.65s	0.02s	0.05s	0.06s	0.28s	0.18s	0.53s	0.43s

Best and second best results are marked in red and blue, respectively
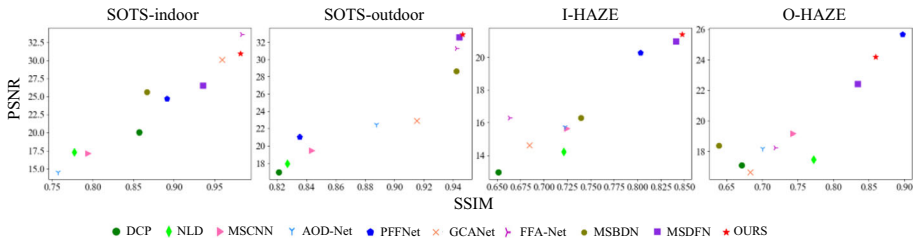
**Fig. 6** Average PSNR and SSIM of different methods on various datasets

generated image color cast is severe, particularly in the sky area. The image processed by MSCNN still has a certain amount of haze. There is still some mist remaining in the image generated by AOD-Net. PFFNet and GCANet reduce the image brightness, the processed image contrast is too high, but also introduces a color shift. Several other methods can restore the image better, but from detailed observation, our proposed method is better than other methods.

Figure 10 shows the dehazing effect of each method on the O-HAZE outdoor dataset [3]. DCP has a certain effect on dehazing the outdoor image, but it can be found that the color intensity of the processed image is low and it is relatively blurry. The image after NLD processing also has a certain color cast. The test results of MSCNN, AOD-Net and MSBDN on the outdoor real data set are not satisfactory, and there are still a lot of haze residues in the image. The color cast of the image processed by GCANet and FFA-Net is more severe and the visual effect is poor. In contrast, the method we propose is closer to the real image in terms of overall perception and details.

Figure 11 shows the dehazing effect of each method on the I-HAZE indoor dataset [1]. The shortcomings of DCP and NLD are apparent. The processed image has color shift and abnormal artifacts. After MSCNN, AOD-Net, and MSBDN, the residual image haze is more serious, and the effect is not ideal. GCANet reduces the color intensity, the image quality is dark, and the visual effect is poor. The color distortion of the image processed by FFA-Net is more severe, and the dehazing effect is not excellent. MSDFN can restore the image better,
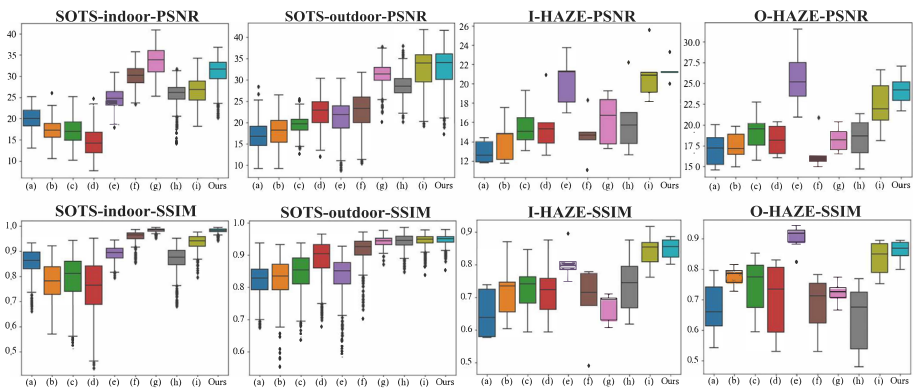
Fig. 7 Numerical distribution of the results of different methods when tested on the dehazing dataset. (a)-(i) represent DCP, NLD, MSCNN, AOD-Net, PFFNet, GCANet, FFA-Net, MSBDN and MSDFN respectively
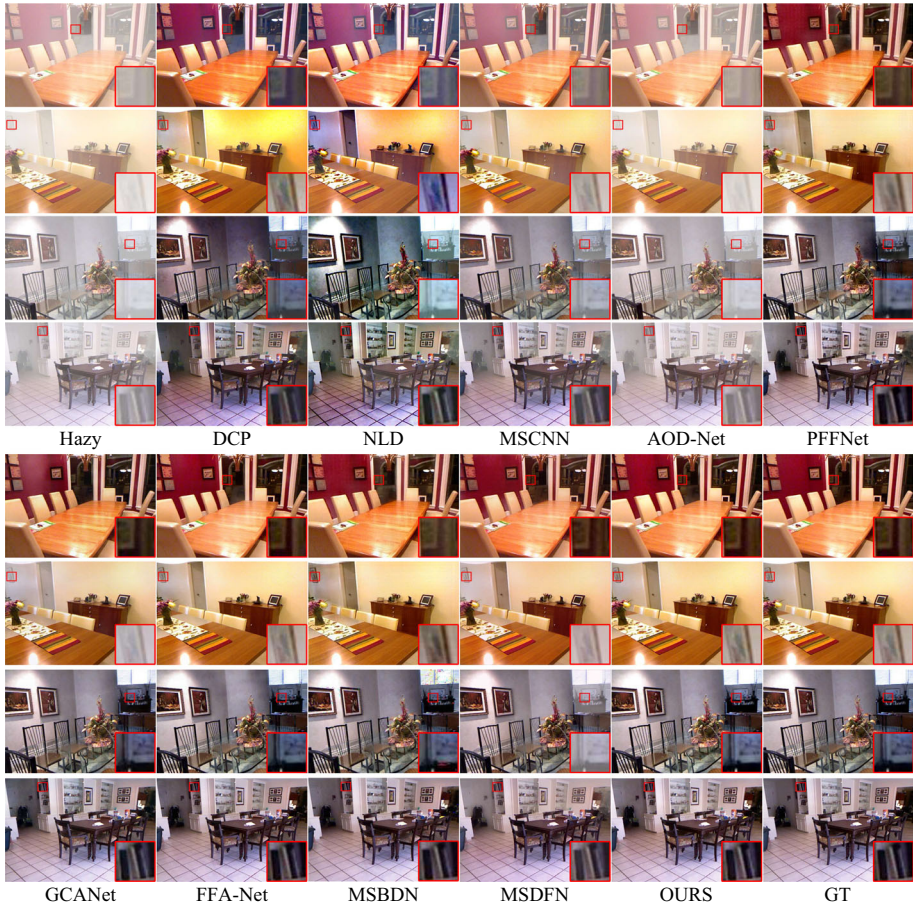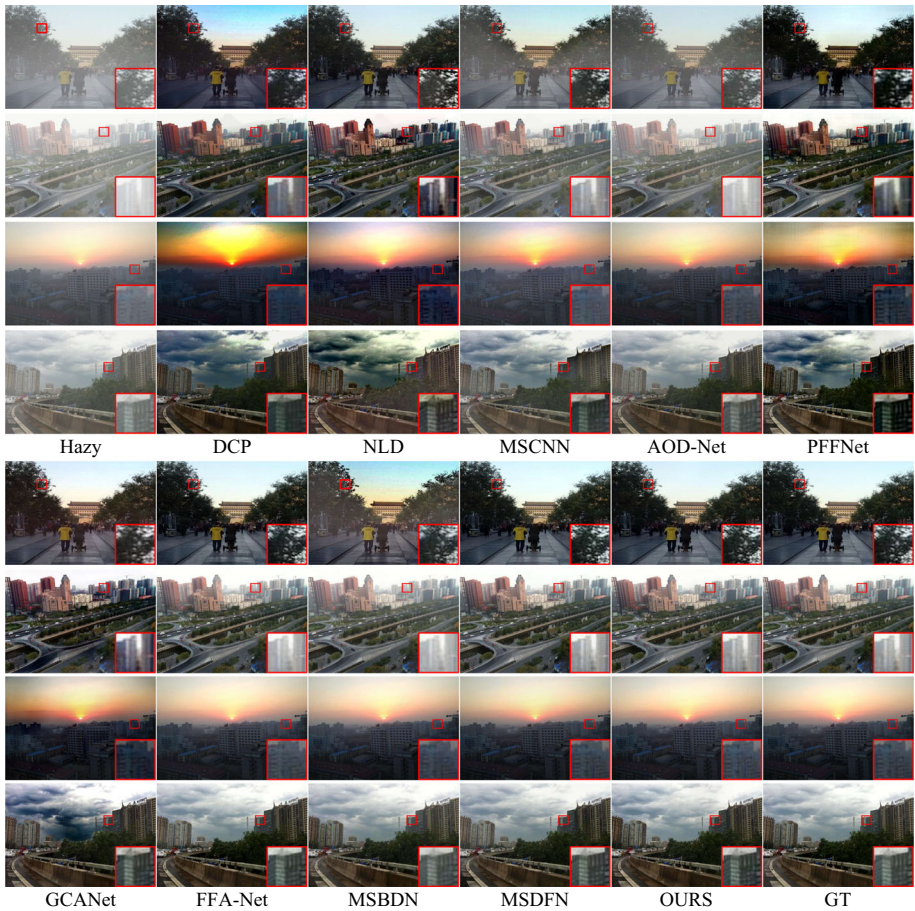
Fig. 8 Visual comparison on the SOTS dataset(indoor)

but false artifacts also appear to a certain extent. Although our method is still inadequate for partial color restoration, compared with other methods, the image restored by our method is closer to the real image in terms of overall perception and clarity.

Figure 12 shows the dehazing effect of each method on the D-Hazy indoor dataset [2]. DCP and NLD can basically remove the haze in the image, but some artifacts will be introduced. There are a lot of haze in the images restored by MSCNN and MSDFN. Several other methods have caused serious color distortion. For example, the style of the image restored by FFA-Net has migrated, and MSBDN makes a large area of black shadow appear in the image. Compared with other methods, the image restored by our method is closer to the real image both in color and clarity.

In addition, the evaluation results on real scene images also illustrate the effectiveness of our proposed method for dehazing real images. As shown in Fig. 13, the results generated by DCP and NLD have serious color distortion. Several other deep-learning based methods have limited generalization capabilities in real situations. The images generated by MSCNN have color shifts and artifacts. AOD-Net cannot remove the haze in the image adequately. The images generated by PFFNet and GCANet are severely distorted and have relatively

| Hazy | DCP | NLD | MSCNN | AOD-Net | PFFNet |

| GCANet | FFA-Net | MSBDN | MSDFN | OURS | GT |

Fig. 9 Visual comparison on the SOTS dataset(outdoor)

severe artifacts. The image processed by FFA-Net still has a lot of haze, such as the second image. MSBDN also has limited dehazing ability in some scenes, such as the mist above the building in the second image is not well removed. Compared with other methods, MSDFN has a better dehazing effect and can remove most of the haze, but our method uses depth map fusion learning, the restored image is more natural and the scene is clearer.

4.3 Running efficiency analysis

To measure the running time efficiency of our model, we separately test the average running time of related methods, as shown in Table 1. The average running time refers to the time required to preprocess the image and forward the network. For our model, the average runtime includes the time to predict the depth map using the DCNF model and the time for the network to restore the image. Among them, it takes about 0.2s to predict the depth map of an image on the GPU platform, and about 0.23s for the network to restore the image, so the average running time of our model is 0.43s. It is worth noting that for a fairer comparison, we uniformly resize
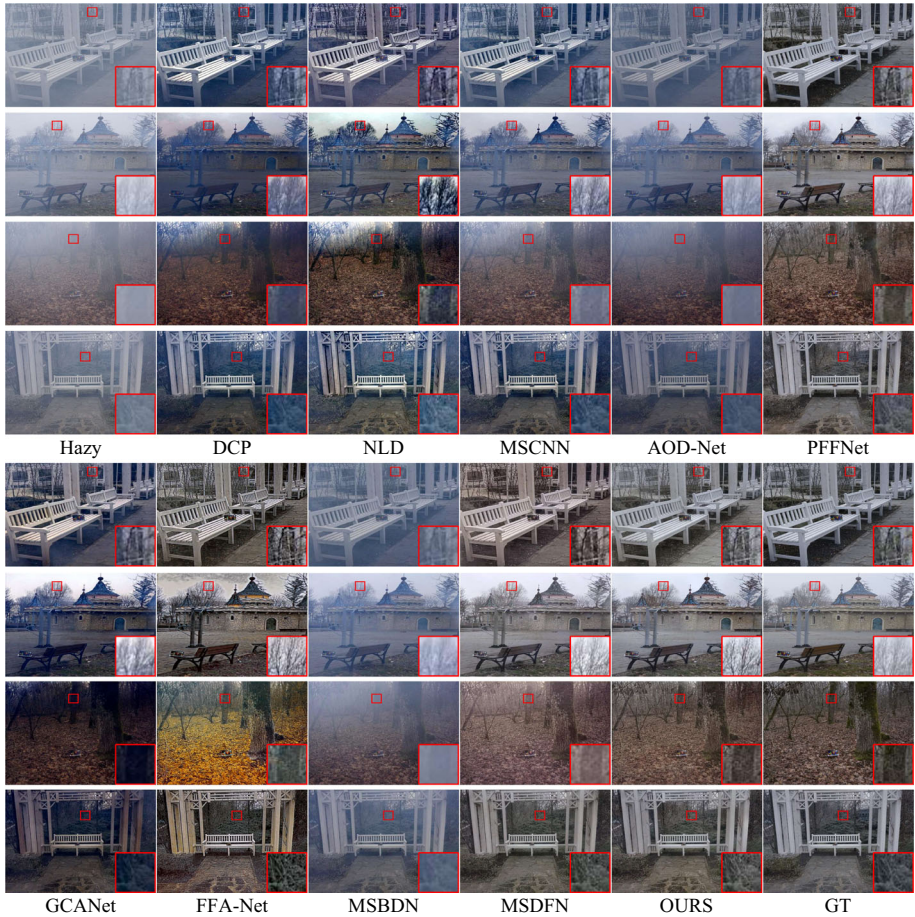
Fig. 10 Visual comparison on the O-HAZE dataset

the input images to 640×480 pixels during testing. From Table 1, it can be found that the average running time of our model is comparable to other methods, and the time loss is acceptable. However, how to continue to improve the running speed of our model and make it lightweight is the focus of our future work.

4.4 Ablation study

The progressive idea is the core idea of this article. In order to find a more appropriate number of BDU, we use ablation experiments to verify the impact of different numbers of BDU on the model. According to our assumptions, the more BDU, the better the effect of the model. We used 1-8 BDU to train our model. The results show that the more BDU, the larger PSNR and SSIM.

Table 2 shows the performance of the proposed model under different configurations. It can be seen that as the number of BDU increases, both SSIM and PSNR will improve to a certain extent. But memory consumption and running time will also increase. It can be seen
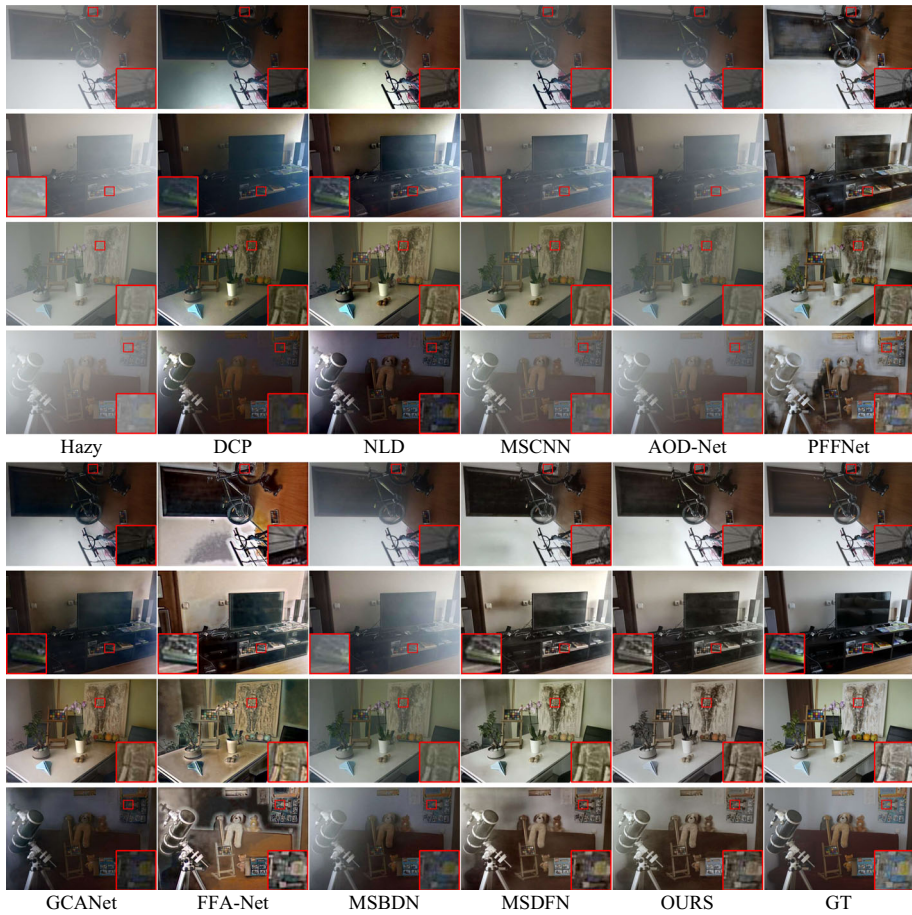
Fig. 11 Visual comparison on the I-HAZE dataset

from Fig. 14 that in various image datasets, as the number of BDU increases, the clarity of the image has been improved to a certain extent. We need to find an appropriate number of BDU in order to balance the relationship between model performance and model complexity. It can be found from the Table 2 that when the number of BDU is less than 5, PSNR and SSIM have a significant improvement with the increase of the number of BDU, but when the number of BDU is greater than 5, PSNR and SSIM tend to be stable. In addition, in order to avoid excessive memory usage and high time consumption, we choose 5 as the execution time of BDU of the model.

5 Conclusions

This paper proposes an end-to-end image dehazing method, which decouples the image dehazing from the atmospheric scattering model. The core of the network is the BDU. The image dehazing is achieved by progressively executing the BDU multiple times. Experiments show that as the number of BDU increases, the dehazing effect becomes more obvious. This

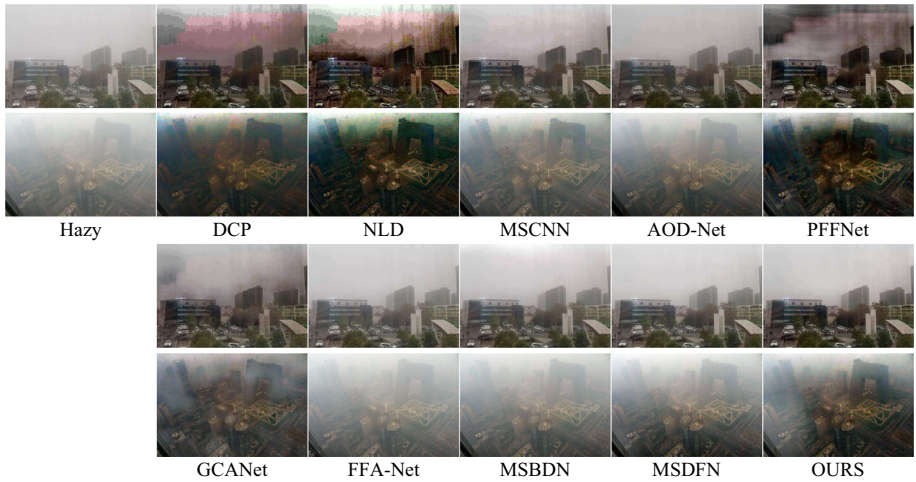**Fig. 12** Visual comparison on the D-Hazy dataset

Fig. 13 Visual comparison on real scene images

Table 2 Use SOTS-outdoor dataset to verify models with different numbers of BDU

Configuration	PEDN(1)	PEDN(2)	PEDN(3)	PEDN(4)	PEDN(5)	PEDN(6)	PEDN(7)	PEDN(8)
SSIM	0.9426	0.9439	0.9458	0.945	0.9465	0.947	0.9472	**0.9503**
PSNR	30.2543	30.7695	31.0727	30.58	32.8903	32.9526	32.9621	**32.9634**
Memory consumption(Mb)	3540	7558	10691	13403	16312	19221	22315	25134
Time consumption(s)	0.08	0.226	0.29	0.35	0.44	0.47	0.51	0.65

The data in the table is the average of all images. Best results are highlighted

Fig. 14 Visual comparison of models with different numbers of BDU on various datasets

method increases the depth of the network and considerably saves the amount of parameters. At the same time, in order to connect the features at different progressive stages, a FMM is added to the BDU. In order to make better use of the depth information of the image, the depth information of the original image is extracted to guide the network learning and avoid the unclear image caused by inaccurate depth prediction. Extensive experiments show that the proposed method has better dehazing effects on synthetic data sets, heterogeneous data sets and real scenes than the current methods.

Declarations

Competing of interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Conflicts of interest All authors declare that there are no conflicts of interest.

References

1. Ancuti CO, Ancuti C, Timofte R, De Vleeschouwer C (2018) O-haze: a dehazing benchmark with real hazy and haze-free outdoor images. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp 754–762
2. Ancuti C, Ancuti CO, De Vleeschouwer C (2016) D-hazy: A dataset to evaluate quantitatively dehazing algorithms. In: 2016 IEEE International Conference on Image Processing (ICIP), pp 2226–2230. IEEE
3. Ancuti C, Ancuti CO, Timofte R, De Vleeschouwer C (2018) I-haze: a dehazing benchmark with real hazy and haze-free indoor images. In: International Conference on Advanced Concepts for Intelligent Vision Systems. Springer, pp 620–631
4. 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
5. Cai B, Xu X, Jia K, Qing C, Tao D (2016) Dehazenet: An end-to-end system for single image haze removal. IEEE Transactions on Image Processing 25(11):5187–5198
6. Chen C, Wang G, Peng C, Zhang X, Qin H (2019) Improved robust video saliency detection based on long-term spatial-temporal information. IEEE transactions on image processing 29:1090–1100
7. Chen C, Wang G, Peng C, Fang Y, Zhang D, Qin H (2021) Exploring rich and efficient spatial temporal interactions for real-time video salient object detection. IEEE Transactions on Image Processing 30: 3995–4007
8. 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), pp 1375–1383 . IEEE
9. Chen C, Song J, Peng C, Wang G, Fang Y (2021) A novel video salient object detection method via semisupervised motion quality perception. IEEE Transactions on Circuits and Systems for Video Technology
10. 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
11. Engin D, Genç A, Kemal Ekenel H (2018) Cycle-dehaze: Enhanced cyclegan for single image dehazing. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp 825–833
12. Fan G, Hua Z, Li J (2021) Multi-scale depth information fusion network for image dehazing. Applied Intelligence 51(10):7262–7280
13. Fan Z,Wu H, Fu X, Hunag Y, Ding X (2018) Residual-guide feature fusion network for single image deraining. arXiv:1804.07493
14. Fattal R (2008) Single image dehazing. ACM transactions on graphics (TOG) 27(3):1–9
15. Fazlali H, Shirani S, McDonald M, Brown D, Kirubarajan T (2020) Aerial image dehazing using a deep convolutional autoencoder. Multimedia Tools and Applications 79(39):29493–29511
16. Feng X, Li J, Hua Z, Zhang F (2021) Low-light image enhancement based on multi-illumination estimation. Applied Intelligence 1–21
17. Fu X, Liang B, Huang Y, Ding X, Paisley J (2019) Lightweight pyramid networks for image deraining. IEEE transactions on neural networks and learning systems 31(6):1794–1807
18. Galdran A, Vazquez-Corral J, Pardo D, Bertalmio M (2016) Fusion-based variational image dehazing. IEEE Signal Processing Letters 24(2):151–155
19. Gibson KB, Nguyen TQ (2014) An analysis and method for contrast enhancement turbulence mitigation. IEEE transactions on image processing 23(7):3179–3190
20. He K, Sun J, Tang X (2010) Single image haze removal using dark channel prior. IEEE transactions on pattern analysis and machine intelligence 33(12):2341–2353
21. Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural computation 9(8):1735–1780
22. Hua Z, Fan G, Li J (2020) Iterative residual network for image dehazing. IEEE Access 8:167693–167710
23. Jin Y, Sheng B, Li P, Chen CP (2020) Broad colorization. IEEE transactions on neural networks and learning systems 32(6):2330–2343
24. Khorram A, Khalooei M, Rezghi M (2021) End-to-end cnn+ lstm deep learning approach for bearing fault diagnosis. Applied Intelligence 51(2):736–751
25. Kim J-Y, Kim L-S, Hwang S-H (2001) An advanced contrast enhancement using partially overlapped sub-block histogram equalization. IEEE transactions on circuits and systems for video technology 11(4): 475–484
26. Kim J, Kwon Lee J, Mu Lee K (2016) Deeply-recursive convolutional network for image super-resolution. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp 1637–1645
27. Kingma DP, Ba J (2014) Adam: A method for stochastic optimization. arXiv:1412.6980

28. Land EH, McCann JJ (1971) Lightness and retinex theory. Josa 61(1):1–11
29. Li B, Ren W, Fu D, Tao D, Feng D, Zeng W, Wang Z (2019) Benchmarking single-image dehazing and beyond. IEEE Transactions on Image Processing 28(1):492–505
30. Li J, Feng X, Hua Z (2021) Low-light image enhancement via progressive-recursive network. IEEE Transactions on Circuits and Systems for Video Technology
31. Lin H-Y, Lin C-J (2017) Using a hybrid of fuzzy theory and neural network filter for single image dehazing. Applied Intelligence 47(4):1099–1114
32. Ling Z, Fan G, Gong J, Guo S (2019) Learning deep transmission network for efficient image dehazing. Multimedia Tools and Applications 78(1):213–236
33. 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
34. Liu F, Shen C, Lin G, Reid I (2015) Learning depth from single monocular images using deep convolutional neural fields. IEEE transactions on pattern analysis and machine intelligence 38(10):2024–2039
35. Liu X, Zhang H, Cheung Y-m, You X, Tang YY (2017) Efficient single image dehazing and denoising: An efficient multi-scale correlated wavelet approach. Computer Vision and Image Understanding 162:23–33
36. 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
37. Mei K, Jiang A, Li J, Wang M (2018) Progressive feature fusion network for realistic image dehazing. In: Asian Conference on Computer Vision. Springer, pp 203–215
38. Min X, Zhai G, Gu K, Zhu Y, Zhou J, Guo G, Yang X, Guan X, Zhang W (2019) Quality evaluation of image dehazing methods using synthetic hazy images. IEEE Transactions on Multimedia 21(9):2319–2333
39. 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, pp. 598–605 . IEEE
40. Narasimhan SG, Nayar SK (2002) Vision and the atmosphere. International journal of computer vision 48(3):233–254
41. Nguyen H, Tran KP, Thomassey S, Hamad M (2021) Forecasting and anomaly detection approaches using lstm and lstm autoencoder techniques with the applications in supply chain management. International Journal of Information Management 57:102282
42. Pang Y, Sun M, Jiang X, Li X (2017) Convolution in convolution for network in network. IEEE transactions on neural networks and learning systems 29(5):1587–1597
43. 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
44. 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
45. Raj NB, Venketeswaran N (2020) Single image haze removal using a generative adversarial network. In: 2020 International Conference on Wireless Communications Signal Processing and Networking (WiSPNET), pp 37–42 . IEEE
46. 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 Vision. Springer, pp 154–169
47. 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
48. 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
49. Shahid F, Zameer A, Muneeb M (2020) Predictions for covid-19 with deep learning models of lstm, gru and bi-lstm. Chaos, Solitons & Fractals 140:110212
50. Shao Y, Li L, Ren W, Gao C, Sang N (2020) Domain adaptation for image dehazing. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)
51. Sharma T, Agrawal I, Verma NK (2020) Csidnet: Compact single image dehazing network for outdoor scene enhancement. Multimedia Tools and Applications 79(41):30769–30784
52. Singh D, Kumar V, Kaur M (2019) Single image dehazing using gradient channel prior. Applied Intelligence 49(12):4276–4293
53. Stark JA (2000) Adaptive image contrast enhancement using generalizations of histogram equalization. IEEE Transactions on image processing 9(5):889–896
54. Tarel J-P, Hautiere N (2009) Fast visibility restoration from a single color or gray level image. In: 2009 IEEE 12th International Conference on Computer Vision, pp 2201–2208 . IEEE
55. Wang Y-K, Fan C-T (2014) Single image defogging by multiscale depth fusion. IEEE Transactions on image processing 23(11):4826–4837

56. Wiedemann S, Müller K-R, Samek W (2019) Compact and computationally efficient representation of deep neural networks. IEEE transactions on neural networks and learning systems 31(3):772–785
57. Zhang T, Li J, Hua Z (2021) Iterative multi-scale residual network for deblurring. IET Image Processing 15(8):1583–1595
58. Zhang W, Li J, Hua Z (2021) Attention-based tri-unet for remote sensing image pan-sharpening. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 14:3719–3732
59. Zhang H, Patel VM (2018) Densely connected pyramid dehazing network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp 3194–3203
60. Zhu J-Y, Park T, Isola P, Efros AA (2017) Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE International Conference on Computer Vision, pp 2223–2232
61. Zhu Q, Mai J, Shao L (2015) A fast single image haze removal algorithm using color attenuation prior. IEEE transactions on image processing 24(11):3522–3533