



Progressive image dehazing network based on dual feature extraction modules

Yong Yang¹ · Wei Hu² · Shuying Huang³ · Weiguo Wan⁴ · Juwei Guan²

Received: 29 July 2021 / Accepted: 13 December 2022 / Published online: 24 December 2022
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract

Image dehazing is of great importance and has been widely studied, as haze severely affects many high-level computer vision tasks. In this paper, by considering the gradual dissipation process of haze, a progressive dehazing network (PDN) is proposed. The proposed approach realizes haze removal step by step by constructing two main modules: the preliminary and fine dehazing modules. In the preliminary dehazing module, a combined residual block is first constructed to extract and enhance features of different levels. Then, an adaptive feature fusion strategy is designed to integrate these features and output the initial dehazing result. Aiming at the residual haze in the initial results, a fine dehazing module is constructed by simulating the last period of the haze dissipation process to further extract a fine haze layer. The final dehazing result is obtained by removing the fine haze layer from the initial dehazing result. Experimental results indicate that the proposed method is superior to some state-of-the-art dehazing methods in terms of visual comparison and objective evaluation.

Keywords Progressive image dehazing · Combined residual block · Adaptive feature fusion strategy · Fine haze layer

1 Introduction

Reflected light is absorbed and scattered by haze in the process of light propagation, leading to weakened object visibility, thus resulting in the degradation of the quality of the captured images [1]. High-quality images are required for high-level computer vision tasks, such as object detection [2], scene understanding [3], and pedestrian reidentification [4]. An effective image dehazing technology can improve the quality of hazy images, as shown in Fig. 1. Numerous image dehazing methods have been developed in the past few decades, which are mainly divided into two categories: traditional and deep learning-based methods.

Traditional dehazing algorithms focus on improving the contrast and color saturation of hazy images to improve their quality. The earlier methods mainly included histogram equalization-based [5], Retinex theory-based [6], and frequency domain-based [7] methods. Mathematically, the imaging theory of haze in traditional dehazing methods can be formulated using the following atmospheric scattering model [8]:

$$I(x) = J(x)t(x) + A(1 - t(x)) \quad (1)$$

where x is the pixel position, $I(x)$ is the hazy image, $J(x)$ is the clear image, A is the atmospheric light value, and $t(x)$ is the transmittance map.

In traditional methods, the evaluation of the transmittance map plays a key role. Therefore, to obtain a haze-free image by solving Eq. (1), the transmittance map of the image must be estimated accurately. This estimation is usually performed by establishing various assumptions [9]. Fattal et al. [10] estimated the transmittance map by calculating the distance between the pixel and color line. He et al. [11] proposed the dark channel prior theory, which is used in the atmospheric scattering model to obtain a rough estimation of the transmittance map. Based on the dark channel theory, Zahid et al. [12] estimated the transmittance map from RGB channels and the intensity channel of the hazy image.

✉ Shuying Huang
shuyinghuang2010@126.com

¹ School of Computer Science and Technology, Tiangong University, Tianjin 300387, China

² School of Information Technology, Jiangxi University of Finance and Economics, Nanchang 330032, China

³ School of Software, Tiangong University, Tianjin 300387, China

⁴ School of Software and Internet of Things Engineering, Jiangxi University of Finance and Economics, Nanchang 330032, China

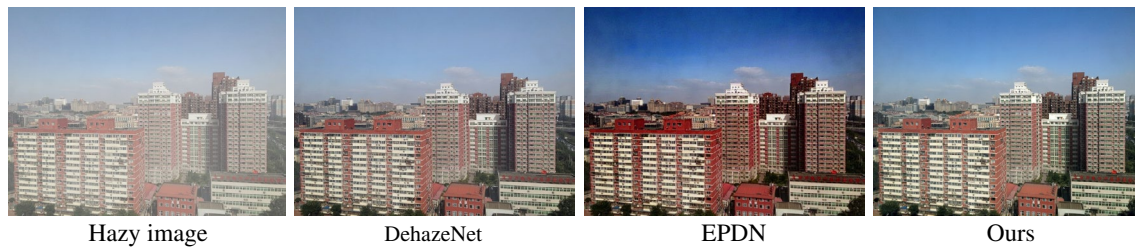


Fig. 1 Image dehazing examples. Our result has rich details and normal color information

Traditional image dehazing methods rely on the evaluation accuracy of the transmittance map to obtain a better dehazing result; otherwise, they encounter inaccurate dehazing problems such as excessive dehazing, incomplete dehazing, and color distortion [13]. In addition, these methods usually consume a significant amount of computing time, which cannot meet the requirements of real-time haze removal.

In recent years, owing to the strong feature representation and nonlinear mapping capacities of deep learning, an increasing number of deep learning-based image dehazing methods have been proposed to solve the limitations of the traditional methods. Chen et al. [14] designed a gated context aggregation network to achieve image dehazing by directly learning the residual between clear and hazy images. Qu et al. [15] proposed an end-to-end generative adversarial network for image dehazing that contains a multi-resolution generator, multi-scale discriminator, and enhancer. Pang et al. [16] proposed a novel binocular image dehazing framework called BidNet, which explores and encodes the correlation between the left and right binocular images by introducing a stereo transformation module. Shao et al. [17] proposed an image dehazing method based on a domain adaptation paradigm, which solved the problem of domain shift between real and synthetic hazy images by constructing a bidirectional translation network. Although these end-to-end dehazing networks can directly restore haze-free images, they do not accurately simulate the image formation of the haze process, and thus may result in artifacts and haze residuals in some special scenes [18].

To fit the scene information more accurately, deep learning methods combining the atmospheric scattering model and scene depth information have been proposed. By learning the mapping between hazy images and the corresponding transmission maps, a multi-scale depth neural network [19] for single-image dehazing was proposed, which includes coarse- and fine-scale networks to predict the overall transmission maps and optimize the results locally. Cai et al. [20] proposed a trainable end-to-end system called DehazeNet, the layers of which were specially designed to embody the established principles of image dehazing. Based on a reformulated atmospheric scattering model, for image dehazing, an all-in-one dehazing network (AOD-Net) [21] was

proposed to generate a haze-free image directly through a lightweight convolutional neural network (CNN). Inspired by the idea of dense connection, Zhang et al. [22] proposed a densely connected pyramid dehazing network by directly embedding the atmospheric scattering model into the network. The network can simultaneously learn the transmission map, atmospheric light, and dehazing result. Guo et al. [23] proposed a depth-aware dehazing network using a reinforcement learning system; their technique gradually realizes the removal of haze in a near-to-far manner by utilizing the depth information from the scene. Lee et al. [24] proposed a CNN-based image dehazing network for both dehazing and depth estimation, and this method estimates a dehazing image and a full-scale depth map from a single hazy image. Wang et al. [25] proposed an end-to-end CNN-based color-shift-restraining dehazing network, which is termed as front white balance network (FWB-Net), to address the color shift problem in image dehazing. Ullah et al. [26] proposed an efficient lightweight convolutional neural network called Light-DehazeNet (LD-Net), which uses a transformed atmospheric scattering model and a color visibility restoration method for image dehazing. These deep learning-based dehazing methods can effectively combine the atmospheric scattering model and depth information and remove haze according to the haze concentration at different depths of the scene. However, owing to the change in scene depth, the haze concentration in the image will be different. This will lead to inaccurate evaluation of depth information and incomplete dehazing in the prospective area. Recently, ViT [27] has performed particularly well in high-level vision tasks, which has led researchers to try to apply it in the field of dehazing. Song et al. [28] introduced some improvements for Swin Transformer [29] and proposed DehazeFormer for image dehazing, which achieves superior performance on several datasets. However, the Transformer-based methods depend on extensive training data and require relatively high computational cost [30, 31].

In view of the above-mentioned problems, we propose a novel progressive dehazing network (PDN) for haze removal based on dual feature-extraction modules. By simulating the process of haze dissipation, two main modules in the network are constructed: preliminary and fine

dehazing modules. In the preliminary dehazing module, different levels of features are extracted and enhanced by constructing a combined residual block. These features are then integrated using an adaptive feature fusion strategy to output the initial dehazed image. For the initial dehazing result, a fine dehazing module is constructed to extract the depth information, that is, the residual haze in the initial result, which is then used to further refine the dehazing result. Compared to state-of-the-art methods, the proposed PDN can achieve better image dehazing results with more accurate details and color information. Mathematically, our PDN, as shown in Fig. 2, can be expressed as follows:

$$J(x) = dec(P(Enc(I(x)))) - F(dec(P(Enc(I(x)))))) \quad (2)$$

where $I(x)$ denotes a hazy image; $Enc(\cdot)$ and $dec(\cdot)$ denote the encoder and decoder layers, respectively; and $P(\cdot)$ and $F(\cdot)$ denote the preliminary and fine dehazing modules, respectively.

The main contributions of this study are summarized as follows:

- (1) A progressive image dehazing network composed of a preliminary and a fine dehazing module is proposed to gradually realize haze removal by simulating the progressive dissipation process of haze.
- (2) The preliminary dehazing module is designed to obtain the initial dehazing results. In this module, a combined residual block is constructed to extract and enhance the features at different levels. Simultaneously, an adap-

tive feature fusion strategy is designed to integrate the features of different levels.

- (3) For the residual haze, following the initial dehazing process, a fine dehazing module is constructed to further extract a fine haze layer, which is then removed from the initial dehazing result to obtain the final dehazing result.

2 The proposed method

Based on the fact that the dissipation of haze is a gradual process, this paper proposes a PDN with two modules to simulate the dissipation of haze and realize the gradual removal of haze. These two modules are named the preliminary dehazing module and fine dehazing module, respectively, as shown in Fig. 2. The detailed construction of the network is as follows.

First, an encoder structure with two convolution layers is added to extract the initial shallow features and reduce the spatial dimension of the feature maps. Then, the feature maps are fed into the preliminary dehazing module to remove haze and enhance the image features. Next, after the preliminary dehazing module, a decoder structure similar to the encoder structure is added to recover the spatial dimension of the feature maps and output the initial dehazing result. Finally, because residual haze will still be present in the deeper scene of the initial result, a fine dehazing module is designed to remove the residual haze in the deeper scene and obtain the final dehazed image. In the network, the size of the convolution kernels used in all convolution layers is

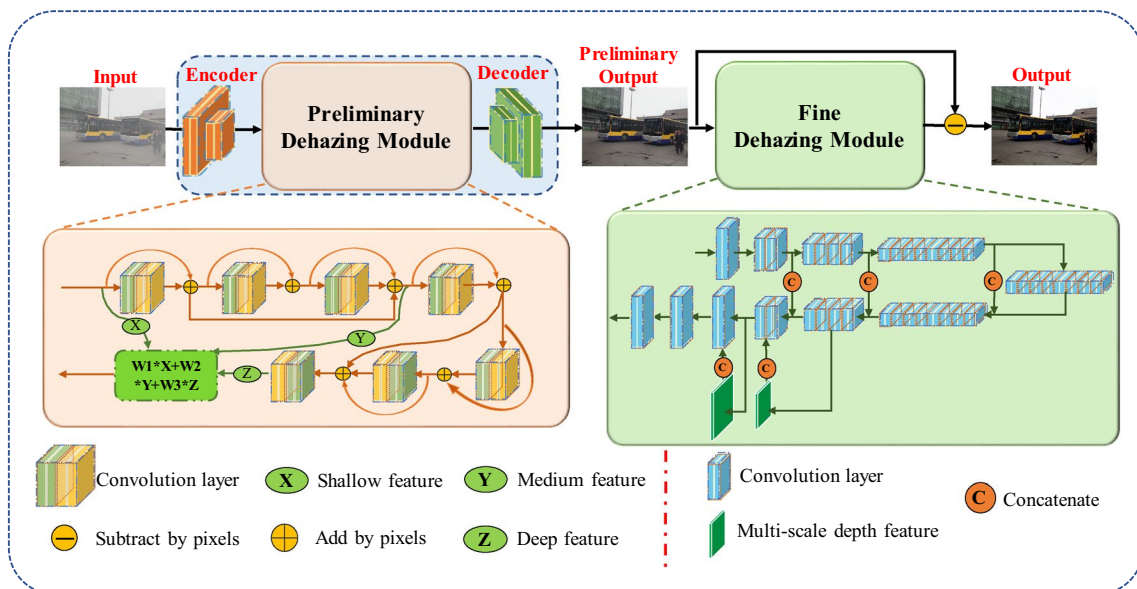


Fig. 2 The overall framework of the proposed PDN

3×3 , with a step size of 1 or 2, and each convolution layer contains 32 convolution channels. The structure of each module is described in detail below.

2.1 Preliminary dehazing module

In the preliminary dehazing module, a combined residual block is first constructed to extract and enhance the features of the input image. The output features of the encoder structure are the shallow features, and the features extracted by the two combined residual blocks can be regarded as the middle and deep features. Then, according to the characteristic, i.e., the human eyes pay different attention to each kind of features, so the contribution of different features to the output result of the network is also different. Therefore, this paper proposes an adaptive feature fusion strategy to evaluate the contribution and realize the fusion of three types of different-level features. The specific descriptions of each structure in the preliminary dehazing module are as follows.

(1) Combined residual block

The combined residual block is composed of two convolution layers: local residuals and an external residual. The convolution layers and local residuals are constructed for feature extraction, and an external residual is designed for feature reuse and enhancement. The combined residual block, as illustrated in Fig. 3, can be represented as

$$\begin{cases} F_1 = \text{conv}(F_0) + F_0 \\ F_2 = \text{conv}(F_1) + F_1 \\ F_3 = \text{conv}(F_2) + F_1 + F_2 \end{cases} \quad (3)$$

where F_0 and F_3 are the input and output of the residual block, respectively. F_1 and F_2 are the intermediate results. $\text{conv}(\cdot)$ represents convolution operations. To obtain more abundant feature information, in the convolution operations, we use dilation convolution with dilation coefficients of 1, 2, and 4 to expand the receptive field. Moreover, a convolution layer with a dilation

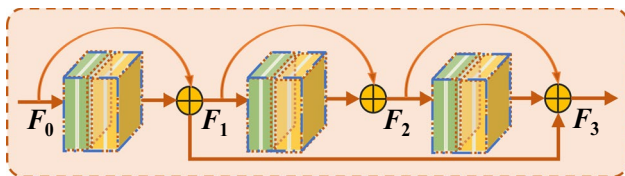


Fig. 3 Illustration of the combined residual block

coefficient of 1 is adopted to avoid the associated gridding effects.

(2) Adaptive feature fusion strategy

To better integrate the features from different levels, we propose an adaptive feature fusion strategy, which is defined as the weighted sum of the previous three types of feature maps. Assume that the shallow features obtained by the encoder structure are X , and the medium and deep features obtained from the two combined residual blocks are expressed as Y and Z , respectively. The fusion strategy of the shallow, medium, and deep features can be represented as

$$\bar{F} = W_1 \times X + W_2 \times Y + W_3 \times Z \quad (4)$$

where \bar{F} denotes the fusion feature maps. W_1 , W_2 , and W_3 are the weighted coefficients, which are learned adaptively during network training. In our work, these weighted coefficients are first initialized randomly, and then updated iteratively by gradient descent method with the loss function.

2.2 Fine dehazing module

The previous preliminary dehazing module can achieve the initial dehazing result by fusing and enhancing the image features from the hazy image. However, because the distance between the deeper scene and camera is significant, the distant areas in the image will show a thicker haze than the areas that are closer. Therefore, the initial dehazing results display the problem of incomplete dehazing in distant areas. To remove the residual haze in the distant areas, a fine dehazing module is designed, as shown in Fig. 2.

Because the features of haze contain very few high-frequency components, the fine dehazing module is designed as an encoding and decoding network, which can better extract high-level semantic information by reducing the spatial dimension of the feature maps. In addition, to supplement the loss of information caused by dimension reduction, the concatenation operations between the same spatial dimension layers are added. At the end of the decoding part, to better extract the depth features, two scale depth-feature layers are designed by expanding the size of the feature maps and reducing the channel dimension. The depth features of the two scales are fused to extract the depth information, as shown in the green block in Fig. 2. The depth features are named multiscale depth features and are transferred to the subsequent convolution layers to obtain fine depth information. The output of the fine dehazing module is regarded as the fine haze layer and subtracted from the initial dehazing result to obtain the final dehazed image.

2.3 Loss functions

To restore good structural information and realistic color for the resulting dehazed image, we design a joint loss function that combines the L_1 , structural similarity (SSIM), and color-difference losses to train the proposed network.

(1) L_1 loss

This loss provides a quantitative method for measuring the pixel-level difference between the dehazed and ground-truth images. Therefore, we use the following L_1 normal formula as part of the total loss function:

$$L_1 = \sum ||y - \bar{y}||_1 \tag{5}$$

where L_1 denotes the value of L_1 loss, y represents the ground-truth, \bar{y} denotes the output of the network, and $|| \cdot ||_1$ represents the L_1 -norm.

(2) Structural similarity loss

The SSIM index is a measure of the structural similarity between two images [32]. To better evaluate the structural similarity between the dehazed and ground-truth images at different scales, a multiscale SSIM (MS_SSIM) quality evaluation index [33] is utilized as one term of the joint loss function, and it is defined as follows:

$$MS_SSIM(y, \bar{y}) = l(y, \bar{y})_m^{\alpha m} \prod_i^m c(y, \bar{y})_i^{\beta i} s(y, \bar{y})_i^{\gamma i} \tag{6}$$

where l is the luminance, c is the contrast, s is the structural similarity, i is the index of the pixel, m is the total number of pixels, and α , β , and γ are the parameters that adjust the importance of l , c , and m , respectively. The larger the value of MS_SSIM , the more complete is the structural information. For the convergence of the network training, the SSIM loss function is defined as follows:

$$L_s = 1 - MS_SSIM(y, \bar{y}) \tag{7}$$

where L_s denotes the value of SSIM loss.

(3) Color difference loss

In hazy images, the color information of objects is usually hidden or distorted owing to insufficient illumination or light refraction. Therefore, the loss of color information leads to color deviation of the dehazing results. The CIEDE2000 index [34] is commonly used to calculate the color difference between two images. To ensure that the color of the dehazed result is more realistic, we introduce the CIEDE2000 index to construct the color difference loss, which is used to measure the color similarity between the dehazed and

ground-truth images. The calculation processes are as follows: (i) the dehazed and ground-truth images are converted from the RGB to the Lab color space; (ii) the color difference ΔE between each pixel of the ground-truth, y , and dehazing image, \bar{y} , is calculated according to the CIEDE2000 index. The color difference loss between the two images is defined as follows:

$$L_c = \frac{1}{N} \sum \Delta E(y, \bar{y}) \tag{8}$$

where ΔE measures the color difference of each pixel between two images in the Lab space, and N is the total number of pixels.

(4) Joint loss

Based on the aforementioned definitions, the joint loss function is defined as follows:

$$L = \rho L_1 + \sigma L_s + \delta L_c \tag{9}$$

where ρ , σ and δ are the weights of the loss terms, which are set to 1, 1, and 0.5, respectively, by trial and error.

3 Experiments and analysis

In this section, the experimental settings and quantitative metrics used to evaluate the reconstruction results are introduced. Then, in both subjective and objective aspects, this study compares the proposed method with nine mainstream dehazing methods, including DCP [11], DehazeNet [20], MSCNN [19], AOD [21], DCPDN [22], EPDN [15], Grid-Dehaze [35], KNND [36], and GMAN [37]. The source codes of all the compared methods were provided by the authors of the related literature. All experiments were performed on the same machine to ensure consistency of the test environment. Finally, three ablation studies were designed to demonstrate the effectiveness of the key components in the proposed network.

3.1 Experimental setting and quantitative metrics

To ensure fairness in the comparison of algorithms, all the deep learning methods were retrained on the RESIDE dataset [38]. In the experiment, 10,500 outdoor images in this dataset are used for training, 1050 outdoor images are used for verification, and 200 outdoor images of SOTS are used for testing.

During the training, the proposed network was optimized using the Adam optimizer, with the learning rate set to 0.0005 and the number of training epochs set to 40. All

experiments were carried out on an NVIDIA GTX 2080ti GPU and Pytorch framework.

To evaluate the performance of the comparison methods, two metrics, namely SSIM and peak signal-to-noise ratio (PSNR), were used to quantitatively analyze the experimental results from an objective point of view. The higher the values of these two metrics, the better are the results of image dehazing.

3.2 Subjective results

To demonstrate the effectiveness of the proposed network, numerous experiments were performed to compare the dehazing results of various methods. Three groups of dehazing results are shown in Figs. 4, 5, 6. Figures 4 and 5 display the dehazing results of the synthetic outdoor hazy images, and Fig. 6 shows the dehazing results of a real outdoor hazy image.

As shown in Fig. 4, the results of DCP, DCPDN, EPDN, and KDDN have serious color deviation problems. For example, the color of the sky is deeper or brighter than that of the ground truth. The AOD result shows a haze residue on the image surface. Although DehazeNet, MSCNN, GridDehaze, and GMAN can obtain better dehazing results, they encounter some problems such as uneven color or low color saturation in the sky region. Compared with other methods, the proposed method can remove haze more effectively and

restore realistic colors closer to the ground-truth. In particular, the result of our method is observed to be more uniform in the sky region. Furthermore, to better observe the effect of the detailed restoration, we selected a small local area to zoom in, as shown in the green boxes in the figures. As such, the proposed method can restore clearer edges than the other methods.

In Fig. 5, the dehazing result of DCP shows serious color distortion and artifacts. The results of DehazeNet, MSCNN, and AOD show haze residue and obvious edge information loss, which can be observed in the magnified area in the green box. DCPDN obtains brighter result than the ground-truth and experiences the problem of over-enhancement. The results of EPDN and KDDN have higher contrast than that of the ground-truth, resulting in problems, for example, some areas become darker and the color is slightly distorted. Compared with other methods, the results of GridDehaze and GMAN are closer to the ground-truth; however, the enlarged areas show that the edges are slightly blurred. Compared with GMAN and GridDehaze, the result of the proposed method is closer to the ground-truth, and the edge texture in the enlarged area is more abundant. In addition, the proposed method can avoid the problems of color distortion and over-enhancement compared with those observed in the other methods.

To further prove the effectiveness of the proposed method, experiments were carried out on real scene images,

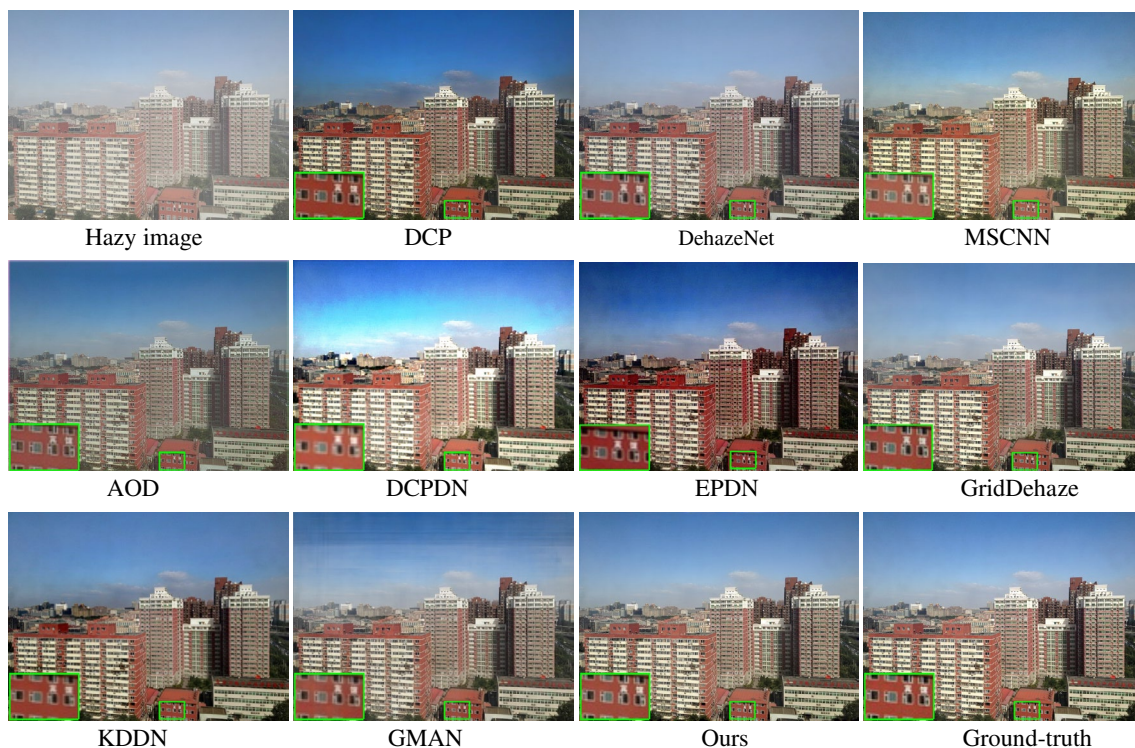


Fig. 4 The first subjective comparison of the dehazing results on synthesis hazy image by different methods



Fig. 5 The second subjective comparison of the dehazing results on synthesis hazy image by different methods

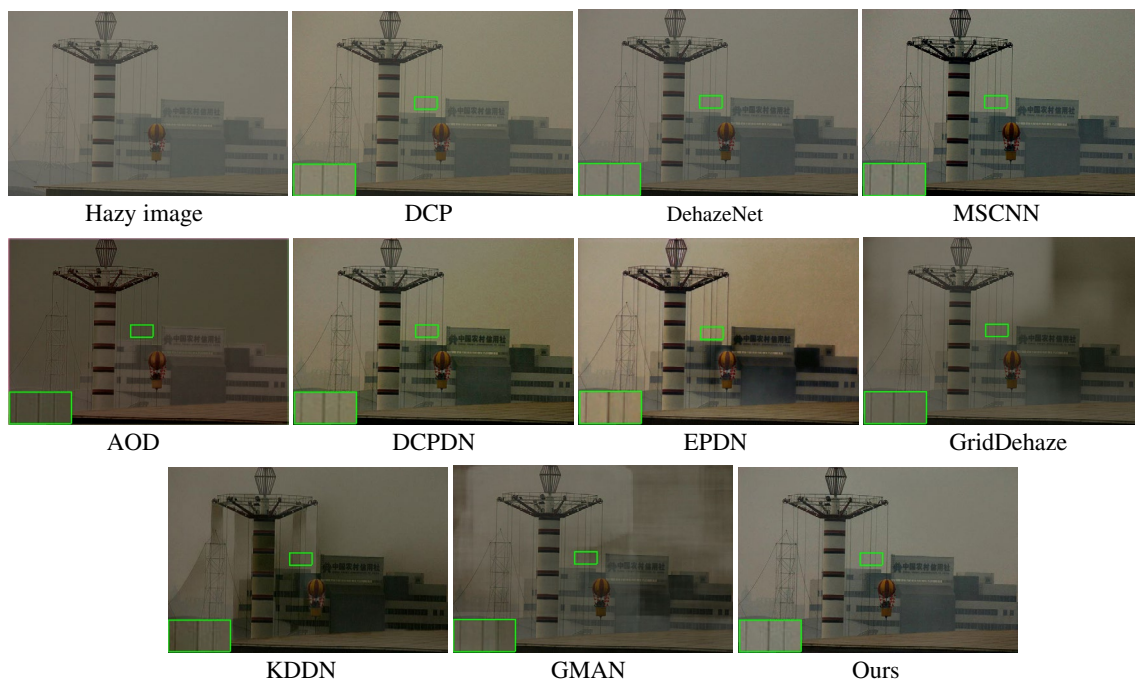


Fig. 6 The third subjective comparison of the dehazing results on real hazy image by different methods

and the results are shown in Fig. 6. As observed, the results of DCP, DehazeNet, and MSCNN all have haze residue, whereas the results of AOD, DCPDN, and EPDN show

color distortion and a small amount of noise. GridDehaze, KDDN, and GMAN produce serious artifacts. In comparison, the proposed method can effectively remove haze while

maintaining color and image details, which are more suitable for human vision.

3.3 Objective results

To objectively evaluate the performance of various methods, we also tested the average PSNR and SSIM values of image dehazing results on the SOTS to further compare the dehazing effects of various methods. The average objective

Table 1 Quantitative comparison of dehazing results by different methods on SOTS

Method	PSNR	SSIM
DCP (2011)	21.16	0.853
DehazeNet (2016)	28.83	0.870
MSCNN (2016)	23.92	0.929
AOD (2017)	23.08	0.920
DCPDN (2018)	18.97	0.875
EPDN (2018)	24.63	0.840
GridDehaze (2019)	32.66	0.988
KDDN (2020)	24.03	0.873
GMAN (2020)	29.07	0.969
Ours	34.60	0.988

evaluation indexes of the test data are shown in Table 1, and the highest scores of the objective indicators are indicated in bold. Table 1 shows that compared to other methods, the proposed method obtains the highest values of SSIM and PSNR, which illustrates that the proposed image dehazing network has better performance.

3.4 Generalization verification

To further prove the effectiveness and generalization ability of the proposed method, experiments were conducted on another dataset, HazeRD [39], which includes 75 test images. Taking one image as an example, the dehazing results of the proposed method as well as the comparison methods are shown in Fig. 7. As can be seen from the figure, the results of DCP, Dehaze-Net, and EPDN are obviously very dark, the results of MSCNN, DCPDN, and GMAN have observable hazy residues, the result of AOD has obvious color distortion, and the result of GridDehaze has serious gridding effect and artifacts, especially in the sky area. Although the result of KDDN looks close to ours, the former is relatively dark as can be seen from the tree area, resulting in unclear details. Overall, the result of the proposed method is closer to the ground-truth compared to the results of the

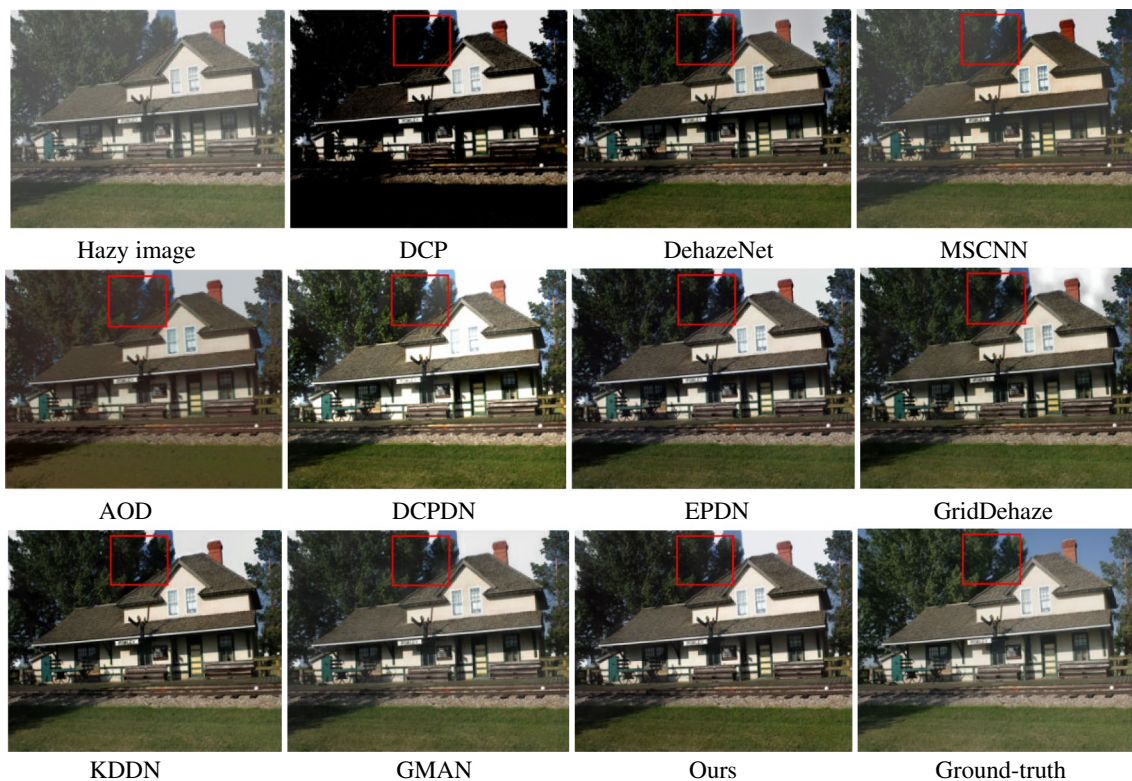


Fig. 7 The subjective comparison of the dehazing results by different methods on HazeRD

Table 2 Quantitative comparison of dehazing results by different methods on HazeRD

Method	PSNR	SSIM
DCP (2011)	11.94	0.539
DehazeNet (2016)	15.81	0.761
MSCNN (2016)	15.30	0.786
AOD (2017)	16.23	0.779
DCPDN (2018)	13.35	0.624
EPDN (2018)	14.98	0.774
GridDehaze (2019)	13.77	0.781
KDDN (2020)	15.16	0.774
GMAN (2020)	15.62	0.809
Ours	15.63	0.819

comparison methods. Furthermore, the average quantitative evaluation results of all the dehazing methods on HazeRD are calculated, as shown in Table 2. From the table, it can be seen that our result is the best on SSIM, and the PSNR value of our method is not much different from the best result. The subjective and objective results presented here can prove that our method is effective and has good generalization.

3.5 Ablation experiments

To verify the effectiveness of the fine dehazing module and proposed adaptive feature fusion strategy, we conducted ablation experiments by testing the dehazing results of the proposed network with and without a fine dehazing module and with and without an adaptive fusion strategy. In addition, we conducted an ablation study to prove the effectiveness of the designed joint loss function. All the ablation experiments

were conducted on the whole test images of SOTS. The average quantitative results of PSNR and SSIM are used to evaluate the performance of the different methods.

(1) Effectiveness of the fine dehazing module

Figure 8 shows the objective comparison of the proposed network with and without a fine dehazing module. In Fig. 8, PDN represents the proposed dehazing network (with both two modules), as indicated by the red line. PDN- represents the network without a fine dehaze module (with only a preliminary dehazing module), as shown by the green line. From Fig. 8, the PSNR and SSIM values of PDN- fluctuate significantly during the training process, and almost all values are lower than those of PDN, especially for SSIM. Compared with the results of PDN-, both PSNR and SSIM indexes of PDN are better, and PDN can reach a stable state in a short time.

In addition, Fig. 9 shows the dehazing results of PDN- and PDN, and gives the PSNR and SSIM values of the two cases. As observed, the PDN- cannot achieve complete dehazing results, and residual haze can still be observed in the results, as shown in the red boxes. After adding the fine dehazing module, the PDN obtains clearer dehazing results, which are closer to the ground-truths. Therefore, the experiments presented here verify that the fine dehazing module is effective in our network.

(2) Effectiveness of adaptive weight-feature fusion

To demonstrate the performance of the adaptive feature fusion strategy in the preliminary dehazing module, another ablation experiment was conducted, as shown in Fig. 10. In the figure, PDN-W and PDN denote the proposed network without and with the

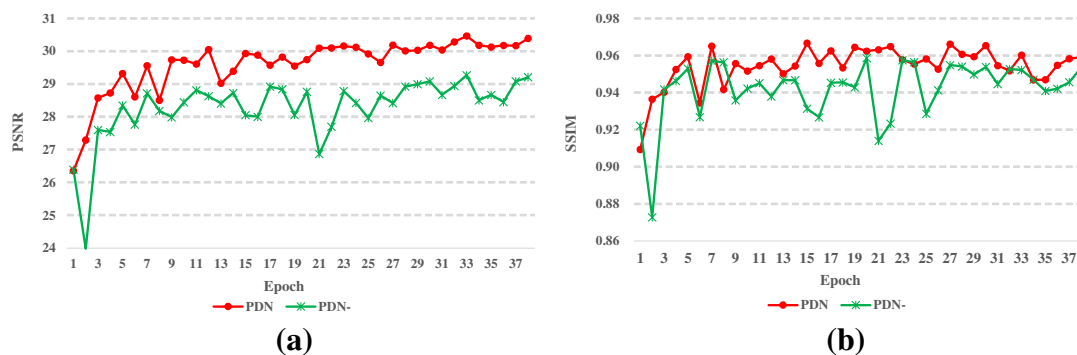


Fig. 8 Ablation study on the effectiveness of fine dehazing module from objective comparison



Fig. 9 Ablation study on the effectiveness of fine dehazing module from subjective comparison

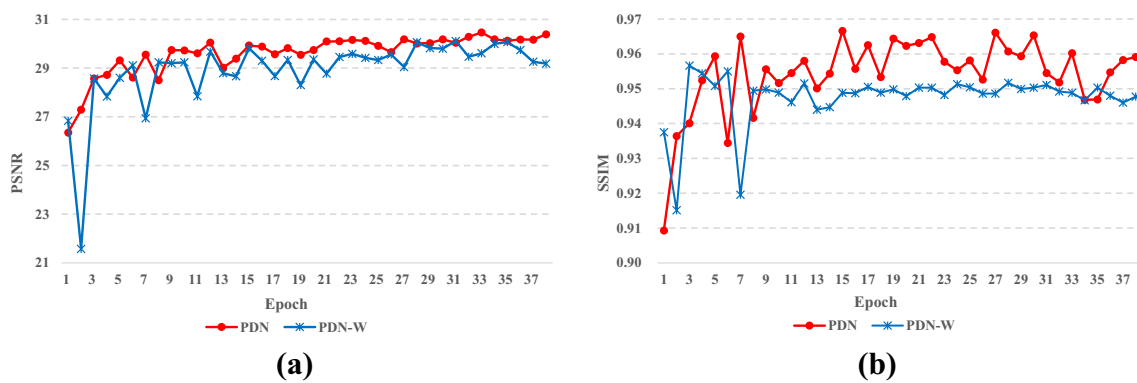


Fig. 10 Ablation study on the effectiveness of adaptive feature fusion from objective comparison

Table 3 Ablation study on the effectiveness of different loss functions

L1	MS_SSIM	Color difference	PSNR	SSIM
✓	×	×	29.78	0.963
✓	×	✓	30.84	0.968
✓	✓	×	31.30	0.979
✓	✓	✓	34.60	0.988

adaptive feature fusion strategy, as indicated by the blue and red lines, respectively. Figure 10 shows that the PSNR and SSIM values of PDN-W are generally lower than those of PDN. Thus, the effectiveness of the adaptive fusion strategy is verified.

(3) Effectiveness of the loss function

In order to prove the effectiveness of the joint loss function, ablation experiments were conducted to compare the performance of the proposed network with only L1 loss, L1 loss and color difference loss, L1 loss and MS_SSIM loss, and L1 loss, MS_SSIM loss and color difference loss. The experimental results are shown in Table 3. As can be seen from the table, compared with the performance of the network with only L1 loss, the performance of adding color difference loss or MS_SSIM loss is much better, which means that each loss is helpful for the method. The performance of the network with three losses is the best, which demonstrates the effectiveness of our designed joint loss function.



Fig. 11 The detection results before and after dehazing

3.6 Object detection application

To demonstrate that image haze might affect the performance of high-level computer vision systems, we did extensive experiments to compare the effects of hazy image and dehazing results in object detection. In the experiments, YOLOv5 [40] trained on COCO dataset [41] was used for object detection. Figure 11 shows a set of experiments as an example. As can be seen from the figure, all the dehazing results detect more objects than the hazy image, and our method detects the objects more correctly. Therefore, the results presented here can verify that image dehazing can improve the performance of high-level computer vision tasks such as object detection, and the proposed method outperforms state-of-the-art dehazing methods.

4 Conclusion

In this paper, according to the principle that the haze gradually dissipates, we proposed a progressive image dehazing network, which includes preliminary and fine dehazing modules. In the preliminary dehazing module, a combined residual block is constructed to extract and enhance different level features, and an adaptive feature fusion strategy is designed to merge the different levels of features and avoid

excessive dehazing. In the fine dehazing module, a network with encoder–decoder structure are constructed to extract multiscale depth features to further obtain the residual haze layer. The final haze-free image can be obtained by subtracting the residual haze layer from the preliminary dehazed result. The experimental results show that the proposed method exhibits outstanding dehazing performance compared to some mainstream dehazing methods.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s13042-022-01753-x>.

Acknowledgements This work is supported by the National Natural Science Foundation of China (No. 61862030, No. 62072218, and 62261025), by the Natural Science Foundation of Jiangxi Province (No. 20182BCB22006, No. 20181BAB202010, No. 20192ACB20002, and No. 20192ACBL21008), and by the Talent project of Jiangxi Thousand Talents Program (No. jxsq2019201056).

Data availability The datasets generated and analyzed during the current study are available from the corresponding author on reasonable request.

References

1. Dudhane A, Biradar K, Patil P, Hambarde P, Murala S (2020) Varicolored image de-hazing. In: Proc IEEE Conf Comput Vis Pattern Recognit, pp 13–19

2. Yao L, Xu H, Zhang W, Liang X, Li Z (2020) SM-NAS: structural-to-modular neural architecture search for object detection. In: AAAI Conf Artif Intell, pp 12661–12668
3. Xu Y, Fu M, Wang Q, Wang Y, Chen K, Xia G, Bai X (2021) Gliding vertex on the horizontal bounding box for multi-oriented object detection. *IEEE Trans Pattern Anal Mach Intell* 43(4):1452–1459
4. Deng W, Zheng L, Ye Q, Kang G, Yi Y, Jiao J (2018) Image-image domain adaptation with preserved self-similarity and domain-dissimilarity for person re-identification. In: Proc. IEEE Conf Comput Vis Pattern Recognit, pp 994–1000
5. Al-Sammarai M (2015) Contrast enhancement of roads images with foggy scenes based on histogram equalization. In: Proc. IEEE ICCSE, pp 95–101
6. Yang W, Wang R, Fang S, Zhang X (2010) Variable filter Retinex algorithm for foggy image enhancement. *J Comput Aided Des Comput Graph* 22(6):965–971
7. Jia J, Yue H (2014) A wavelet-based approach to improve foggy image clarity. In: Proc. IFAC, pp 930–935
8. Cantor A (1978) Optics of the atmosphere: scattering by molecules and particles. *IEEE J Quantum Electron* 14(9):698–699
9. Berman D, Treibitz T, Avidan S (2016) Non-local image dehazing. In: Proc IEEE Conf Comput Vis Pattern Recognit, pp 1674–1682
10. Fattal R (2014) Dehazing using color-lines. *ACM Trans Graph* 34(1):1–14
11. He K, Sun J, Tang X (2011) Single image haze removal using dark channel prior. *IEEE Trans Pattern Anal Mach Intell* 33(12):2341–2353
12. Tufail Z, Khurshid K, Salman A, Nizami I, Khurshid K, Jeon B (2018) Improved dark channel prior for image defogging using RGB and YCbCr color space. *IEEE Access* 6(1):32576–32587
13. Chen W, Ding J, Kuo S (2019) PMS-Net: robust haze removal based on patch map for single images. In: Proc IEEE Conf Comput Vis Pattern Recognit, pp 11673–116821
14. 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: Proc IEEE Winter Conf Appl Comput Vis (WACV), pp 1375–1383
15. Qu Y, Chen Y, Huang J, Xie Y (2019) Enhanced Pix2pix dehazing network. In: Proc IEEE Conf Comput Vis Pattern Recognit, pp 8160–8168
16. Pang Y, Nie J, Xie J, Han J, Li X (2020) BidNet: binocular image dehazing without explicit disparity estimation. In: Proc IEEE Conf Comput Vis Pattern Recognit, pp 5930–5939
17. Shao Y, Li L, Ren W, Gao C, Sang N (2020) Domain adaptation for image dehazing. In: Proc IEEE Conf Comput Vis Pattern Recognit, pp 2805–2814
18. Li R, Pan J, He M, Li Z, Tang J (2020) Task-Oriented Network for Image Dehazing. *IEEE Trans Image Process* 29:6523–6534
19. Ren W, Liu S, Zhang H, Pan J, Cao X, Yang M (2016) Single image dehazing via multi-scale convolutional neural networks. In: Proc European Conf Comput Vis (ECCV), pp 154–169
20. 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
21. Li B, Peng X, Wang Z, Xu J, Feng D (2017) AOD-Net: all-in-one dehazing network. In: Proc IEEE Int Conf Comput Vis (ICCV), pp 4780–4788
22. Zhang H, Patel V (2018) Densely connected pyramid dehazing network. In: Proc IEEE Conf Comput Vis Pattern Recognit, pp 3194–3203
23. Guo T, Monga V (2020) Reinforced depth-aware deep learning for single image dehazing. In: Proc IEEE Int Conf Acoust Speech Signal Process (ICASSP), pp 8891–8895
24. Lee B, Lee K, Oh J, Kweon I (2020) CNN-based simultaneous dehazing and depth estimation. In: Proc IEEE Int Conf Robot Autom (ICRA), pp 9722–9728
25. Wang C, Huang Y, Zou Y, Xu Y (2021) FWB-NET: front white balance network for color shift correction in single image dehazing via atmospheric light estimation. In: Proc. IEEE Int Conf Acoust Speech Signal Process (ICASSP), pp 2040–2044
26. Ullah H, Muhammad K, Irfan M, Anwar S, Sajjad M, Imran A, Albuquerque V (2021) Light-DehazeNet: a novel lightweight CNN architecture for single image dehazing. *IEEE Trans Image Process* 30:8968–8982
27. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I (2017) Attention is all you need. In: NeurIPS, pp 1–11
28. Song Y, He Z, Qian H, Du X (2022) Vision transformers for single image dehazing. arXiv preprint [arXiv:2204.03883](https://arxiv.org/abs/2204.03883)
29. Liu X, Lin Y, Cao Y, Hu H, Wei Y, Zhang Z, Lin S, Guo B (2021) Swin transformer: hierarchical vision transformer using shifted windows. In: Proc IEEE Int Conf Comput Vis (ICCV), pp 10012–10022
30. Li K, Yu R, Wang Z, Yuan L, Song G, Chen J (2022) Locality guidance for improving vision transformers on tiny datasets. arXiv preprint [arXiv:2207.10026](https://arxiv.org/abs/2207.10026)
31. Yang Y, Jiao L, Liu X, Fang L, Yang S, Feng Z, Tang X (2022) Transformers meet visual learning understanding: a comprehensive review. arXiv preprint [arXiv:2203.12944](https://arxiv.org/abs/2203.12944)
32. Wang Z, Bovik A, Sheikh H, Simoncelli E (2004) Image quality assessment: From error visibility to structural similarity. *IEEE Trans Image Process* 13(11):600–612
33. Wang Z, Simoncelli E, Bovik A (2003) Multiscale structural similarity for image quality assessment. In: Proc. 37th Asilomar Conf Signals Syst Comput, pp 1398–1402
34. Sharma G, Wu W, Dalal E (2005) The CIEDE2000 color-difference formula: implementation notes, supplementary test data, and mathematical observations. *Color Res Appl* 30(1):21–30
35. Liu X, Ma Y, Shi Z, Chen J (2019) GridDehazeNet: attention-based multi-scale network for image dehazing. In: Proc IEEE Int Conf Comput Vis (ICCV), pp 7313–7322
36. Hong M, Xie Y, Li C, Qu Y (2020) Distilling image dehazing with heterogeneous task imitation. In: Proc. IEEE Conf Comput Vis Pattern Recognit, pp 3459–3468
37. Liu Z, Xiao B, Alrabeiah M, Wang K, Chen J (2019) Single image dehazing with a generic model agnostic convolutional neural network. *IEEE Signal Process Lett* 26(6):833–837
38. Li B, Ren W, Fu D, Tao D, Feng D, Zeng W, Wang Z (2018) Benchmarking single image dehazing and beyond. *IEEE Trans Image Process* 28(1):492–505
39. Zhang Y, Ding L, Sharma G (2017) HazeRD: an outdoor scene dataset and benchmark for single image dehazing. In: Proc IEEE Int Conf Image Process (ICIP), pp 3205–3209
40. Zhu X, Lyu S, Wang X et al (2021) TPH-YOLOv5: Improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios. In: Proc IEEE Int Conf Comput Vis (ICCV), pp 2778–2788
41. Lin T, Maire M, Belongie S et al (2014) Microsoft COCO: common objects in context. In: Proc European Conf Comput Vis (ECCV), pp 740–755

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.