Multi-feature Fusion Network for Single Image Dehazing

Jie Luo, Tong Chang, and Qirong Bo[(✉)]

Northwestern University, Xi'an, Shaanxi, China
`boqirong@nwu.edu.cn`

Abstract. Existing image dehazing methods consider the learning-based methods as the mainstream. Most of them are trained on synthetic dataset, and may not be able to efficiently transfer to real outdoor scenes. In order to further improve the dehazing effect of the model in real outdoor scenes, this paper proposes an end-to-end Multi-Feature Fusion Network for Single Image Dehazing (MFFN). The proposed network combines the prior-based methods and learning-based methods. This paper first uses the method of supporting backpropagation in order to directly extract the dark channel prior and color attenuation prior features. It then designs a Multi-Feature Adaptive Fusion Module (MFAFM) which can adaptively fuse and enhance the two prior features. Finally, the prior features are added to the decoding stage of the backbone network in a multi-scale manner. The experimental results on the synthetic dataset and real-world dataset demonstrate that the proposed model performs favorably against the state-of-the-art dehazing algorithms.

Keywords: Single Image Dehazing · Prior-based methods · Learning-based methods

1 Introduction

In haze weather, the increase of suspended particles in the air absorbs and scatters light, which results in poor visibility, reduced contrast and color distortion of the taken images. This process can be modeled as [1, 2]:

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

where I(x) denotes the hazy image, J(x) is the corresponding clear image, t(x) represents the transmission map, A is the global atmospheric light, and x represents the pixel location. Image dehazing aims at making the image clear. That is, given I(x), in order to get J(x), we focus on the solution of t(x) and A.

The commonly used dehazing methods can be divided into two categories: the prior-based methods [3–6] and learning-based methods [7–13]. The prior is generally based on data statistics, which is often very efficient in real outdoor scenes. However, it still has limitations. For instance, dark channel prior will fail in the sky region. The learning-based methods can estimate t(x) or A using a neural network [7, 8], and then synthesize a clear image according to Eq. (2). However, it will cause error superposition, and increase the final error. Therefore, the recent methods for the estimation of clear images directly from hazy images using a neural network [9–13] are the mainstream.

However, these methods lead to some problems. More precisely, training such a neural network requires a large number of hazy/clear image pairs, and it is very difficult to obtain such data. Therefore, the currently used training images are generally synthetic images, while the hazy images are formed by hazing real and clear images according to Eq. (3). As the neural network is trained on synthetic dataset, the effect of dehazing in real scenes is often not satisfactory (see Fig. 1). Although NTIRE has organized several dehazing challenges and introduced several small-scale real-world datasets, these datasets are rare and incomplete. Several studies have been proposed to solve this problem [22, 23]. In fact, we believe that, in order to improve the effect of the model in real scenes, we should extract as many real image features as possible from the hazy images that are suitable for dehazing tasks, especially the prior features. This is due to the fact that the prior features are very efficient in real outdoor scenes. However, they have some limitations. In addition, deep learning is versatile. However, it relies too much on the training set. Therefore, this paper uses the fusion of prior features and deep learning features to further improve the performance of the network in complex real outdoor scenes.

| (a) | (b) | (c) | (d) |

Fig. 1. Dehazing results on synthetic and real images using FFA-Net [13]. (a) synthetic hazy image, (b) dehazed image for (a), (c) real hazy image, (d) dehazed image for (c).

An end-to-end Multi-Feature Fusion Network for Single Image Dehazing (MFFN) is proposed. Note that the proposed network is based on our previous study [14]. The baseline is a global feature fusion attention network based on encoder-decoder architecture, which can extract global context information and fully fuse it. Through experiments, two prior features are selected for extraction then fused into the network: the Dark Channel Prior (DCP) [3] and color attenuation prior (CAP) [15]. According to the definition of two priors, a simple and direct extraction method is designed using tensor calculation and maxpooling, in order to make the extraction process support back-propagation. The Multi-Feature Adaptive Fusion Module (MFAFM) is proposed to selectively fuse the

two prior features using the attention mechanism, and enhance the features using residual connections. Finally, the fusion of two scales is performed in the decoder stage of the baseline.

The experiments show that the proposed algorithm has higher performance than other state-of-the-art dehazing algorithms. The contributions of this paper include:

By combining the advantages of the prior-based methods and learning-based methods, the proposed MFFN fuses the two prior features and deep learning features. This model has a better performance in real outdoor scenes.

DCP and CAP are directly and efficiently extracted, while supporting backpropagation in order to make the model end-to-end.

The MFAFM is proposed to select the effective feature from the two prior features for fusion, so as to avoid excessive features that affect the network performance.

2   Proposed Method

In this section, the proposed MFFN is detailed. The latter consists of three parts: extraction of two prior features, MFAFM and basic network (see Fig. 2).
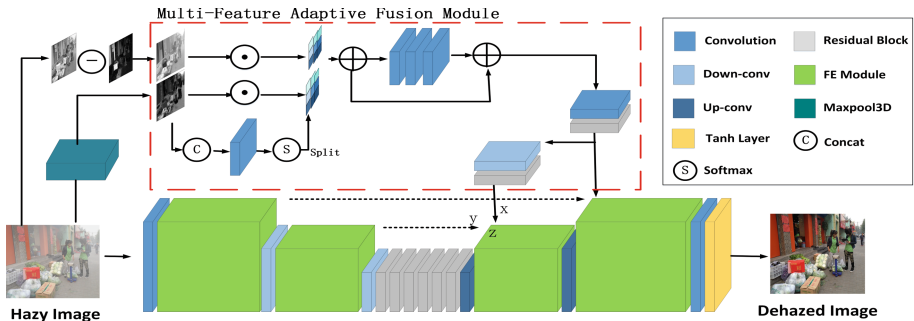

Fig. 2.   The architecture of the Multi-Feature Fusion Network for Single Image Dehazing (MFFN).

2.1   Extraction of Two Prior Features

Dark Channel Prior. He et al. [3] made statistics of a large number of outdoor hazy-free images and determined a rule: in most of the local areas of the outdoor hazy-free image, there are some pixels that have very low values (approaching 0) in at least one color channel. It is referred to as the dark channel prior, expressed as:

$$J^{dark}(x) = \min_{y \in \Omega(x)} (\min_{c \in \{r,g,b\}} J^c(y)) \qquad (4)$$

The input of the neural network is the hazy image. Due to the presence of haze, the white area in the image increases, which makes the dark channel value of the image not approaching 0. Therefore, the DCP feature map, obtained from the hazy image

I(x), can represent the concentration and hazy area to a certain extent. In this paper, three-dimensional maxpooling is used to perform DCP feature map extraction:

$$I^{dark}(x) = 1 - \max pool3D(1 - I(x)) \tag{5}$$

The obtained result is shown in Fig. 3 (b). It can be seen that, in the near non-hazy area, Idark(x) is almost all black, and it is possible to clearly distinguish between the hazy area and the non-hazy area. Due to the dark channel value of each local area (of 7 ◇ 7 size) is the same, it lacks detailed information.

Color Attenuation Prior. Hu et al. [15] found that the difference between brightness and saturation is positively correlated with the haze density, using statistics of outdoor hazy images. The CAP feature map is directly computed as:

$$sv(x) = HSV(I(x))_v - HSV(I(x))_s \tag{6}$$

The hazy image is converted to the HSV format. The value of the s channel minus that of the v channel is then used as the color attenuation prior feature map (sv(x)). It can be seen from Fig. 3(c) that sv(x) has a larger pixel value in the area where the hazy density is greater, and it contains lot of detailed information due to the direct extraction method.

(a) hazy   (b) $I^{dark}(x)$ (c) sv(x)   (d) GT

(e) p1   (f) p2   (g) f   (h) df
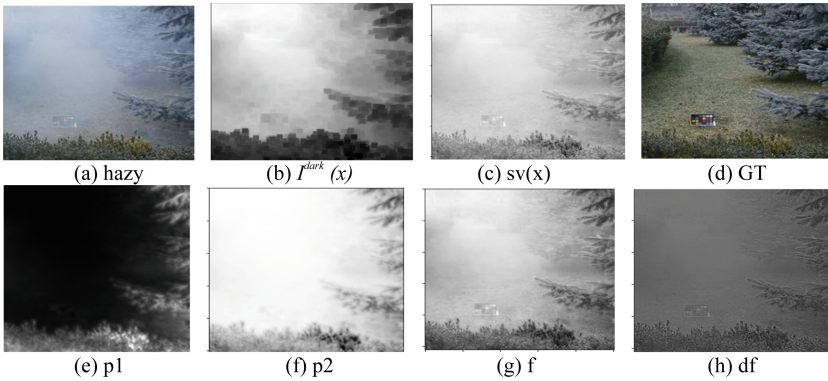
Fig. 3. Results of prior feature extraction, and intermediate results of MFSFM.

2.2 Multi-feature Adaptive Fusion Module

The two priors are based on statistics of real outdoor images. Therefore, their addition will allow the model to capture features that are more suitable for real outdoor scenes. In this paper, the extraction of prior feature maps is straightforward. The most primitive prior features can then be extracted. However, these two types of prior feature maps have some shortcomings. More precisely, the DCP feature map will be invalid in the white or sky area, and the CAP feature map will also show white color in the close-range hazy-free area. The direct introduction of these features to the network will affect the

performance of the network. Therefore, this paper designs the MFAFM (see Fig. 2) using the attention mechanism to adaptively and selectively fusion the two prior feature maps, in order to obtain the most efficient features:

$$p1, p2 = split(soft\max(conv(concat(I^{dark}(x), sv(x))))) \tag{7}$$

$$f = (p1 \otimes I^{dark}(x)) \oplus (p2 \otimes sv(x)) \tag{8}$$

$$df = f \oplus conv(conv(conv(f))) \tag{9}$$

The two prior feature maps are first concatenated. A 2-channel attention feature is then obtained using a 3×3 convolution and softmax function. Afterwards, the feature map of each channel is treated as an attention map of a prior feature map. The corresponding multiplication and addition are then performed to obtain the fusion feature f, which is gone and added using three convolutions. Finally, the residual connection is used to enhance the feature of f, and therefore the enhanced feature ef is obtained.

In Fig. 3, p1 and p2 represent the attention maps of $I^{dark}(x)$ and sv(x), respectively. It can be seen that for $I^{dark}(x)$, the close-range non-hazy area is mainly reserved, while for sv(x), the hazy area and the detailed information of the close-range area are reserved. In f, the recovery effect is better in the close-range non-hazy area. In addition, a certain dehazing effect is achieved in the hazy area. Moreover, e f removes more haze while retaining the detailed features. Finally, e f will be fused to the two scales of the decoder.

2.3 Baseline

The baseline in this paper is a global feature fusion attention network [14], based on the encoder-decoder architecture. The Feature Enhancement (FE) module is its main module. Figure 4 presents the FE module of the decoder, where x is the information passed by the layer skip connection, y represents the prior features, and z is the information to be up-sampled after decoding. The Global Feature Fusion Attention (GFFA) module is the core of the FE module (see Fig. 5). It can extract the global context features, and fully integrate them with the prior features using the multi-scale and attention mechanism, as well as the residual connection of the FE module, in order to enhance the features. Note that the Mean Square Error (MSE) and perceptual loss are used as the loss function.
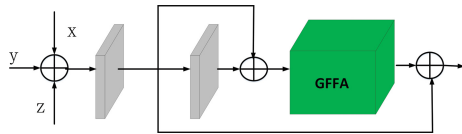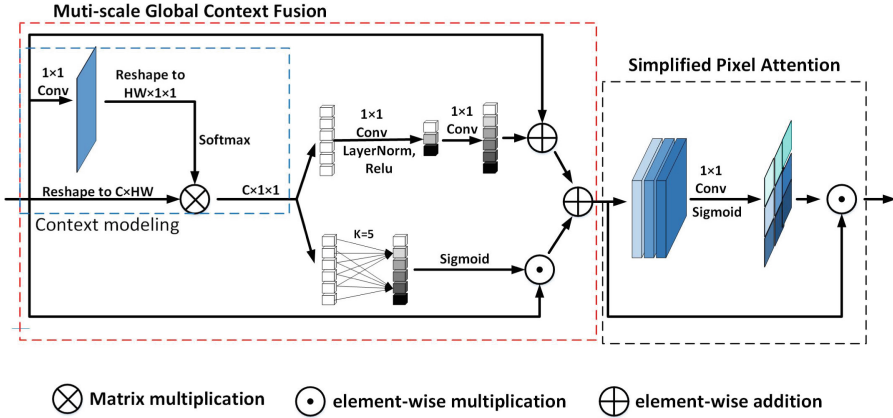

Fig. 4. Architecture of the Feature Enhancement (FE) module [14].

Fig. 5. The architecture of the Global Feature Fusion Attention (GFFA) module [14], including the Muti-scale Global Context Fusion (MGCF) block (red box) and Simplified Pixel Attention (SPA) block (black box). (Color figure online)

3 Experiments

3.1 Datasets

Synthetic Dataset. The synthetic RESIDE [16] dataset contains indoor and outdoor images. The dataset used by MSBDN [17] after data enhancement, is considered as training set. The Outdoor Training Set (OTS) is used as the test set, which contains 500 pairs of outdoor synthetic images.

Real-World Dataset. The O-HAZE dataset [18] from NTIRE2018 Dehazing Challenge and NH-HAZE dataset [19, 20] from NTIRE2020 Dehazing Challenge are used. O-HAZE contains 45 pairs of outdoor hazy and haze-free images, while the first 40 images are used to train the models and the last 5 images are used to test. NH-HAZE contains 55 pairs of outdoor hazy and haze-free images, while the first 50 images are used to train the models and the last 5 images are used to test.

3.2 Implementation Details

A 256×256 patch is cropped from the image and used as input, while the batch-size is set to 8. The initial learning rate is set to $1 \times 10 - 4$, and the cosine annealing strategy [25] is used to adjust the learning rate. The Adam optimizer is used, where β1 and β2 have the default values of 0.9 and 0.999, respectively. The network is trained for 1×106 iterations. PyTorch is used to train the models with an NVIDIA GTX2080 SUPPER GPU.

3.3 Comparison with the State-of-the-Art Methods

In order to more accurately evaluate the proposed MFFN, quantitative and qualitative comparisons with the state-of-the-art methods are conducted on the synthetic dataset

and real-world dataset, respectively. The involved state-of-the-art methods include DCP [3], MSCNN [7], AOD-Net [10], DCPDN [8], GFN [9], GCA-Net [21], GDN [12], FFA [13], MSBDN [17] and MSTN [22].

The comparison results on the three datasets, are presented in Table 1. It can be seen that the proposed model has the highest PSNR and SSIM on the OTS and O-HAZE datasets, where the PSNR values are 0.48 dB and 0.49 dB higher than the sub-optimal models, respectively. On the NH-HAZE dataset, the SSIM of the proposed method is only lower than that of MSTN, but the PSNR is much higher than MSTN.

Table 1. Quantitative evaluation (PSNR/SSIM) with some state-of-the-art methods using there datasets

| Methods	OTS	O-HAZE	NH-HAZE
DCP	19.13/0.815	16.78/0.653	10.57/0.520
MSCNN	19.48/0.839	17.56/0.650	13.42/0.489
AOD-Net	20.29/0.877	15.03/0.539	15.41/0.569
DCPDN	19.93/0.845	18.97/0.664	17.42/0.610
GFN Text follows	21.55/0.844	18.16/0.671	15.17/0.520
GCA-Net	28.13/0.945	16.28/0.645	17.58/0.594
GDN	30.86/0.982	18.92/0.672	15.23/0.560
FFA-Net	33.57/0.984	21.62/0.738	19.45/0.612
MSBDN	32.21/0.979	24.36/0.749	18.80/0.590
MSTN	32.61/0.981	-	18.42/**0.630**
Ours	**34.05/0.985**	**24.85/0.762**	**19.59**/0.615

Figure 6 and Fig. 7 present the qualitative comparison results. It can be seen that DCP has clear color distortion, AOD-Net and DCPDN have poor dehazing effects, some areas of FFA-Net are not completely dehazed, and MSBND has insufficient recovery of detailed features. The proposed model has the best performance, and it is efficient for color and details restoration, even in the case of hazy GT images. This proves that the proposed model has a strong dehazing ability, and is suitable for real outdoor environments.

3.4 Ablation Study

Table 2 presents the results of the ablation experiments, performed on the O-HAZE real-world dataset. Both the fusion of $sv(x)$ and $I^{dark}(x)$ are beneficial to the network. Even if MFAFM is not used, the two prior features can be directly added to the decoder, which can highly improve the network performance. This proves the effectiveness of the prior features when dealing with real-world datasets. Furthermore, MFAFM can well fuse the two prior features and further improve the performance of the model.
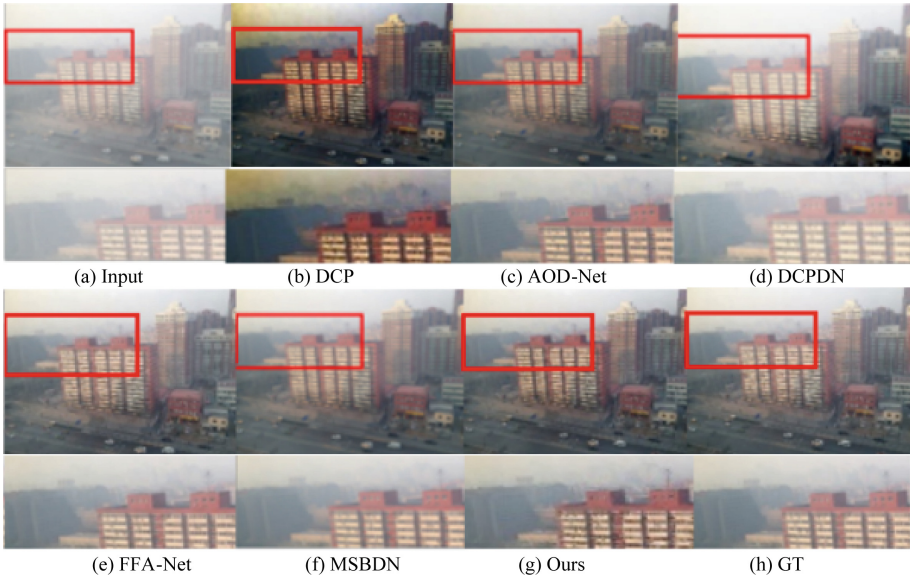
Fig. 6. Qualitative evaluation with some state-of-the-art methods using the OTS synthetic dataset. The bottom row is an enlarged version of the red box area on the top row. (Color figure online)
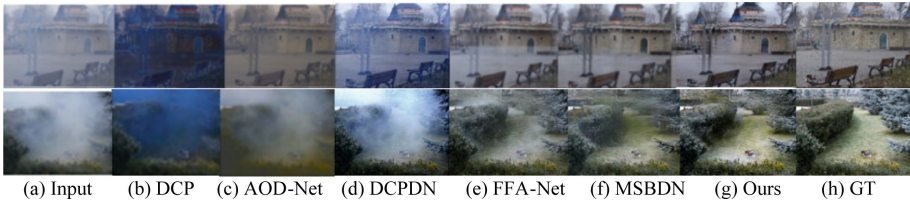
Fig. 7. Qualitative evaluation with some state-of-the-art methods on the O-HAZE and NH-HAZE real-world datasets.

In order to verify whether the fusion of the two prior features is beneficial for the model trained on the synthetic dataset to better transfer to the real scene, the model is trained for 2*105 iterations on the RESIDE synthetic dataset, and then directly tested on the OTS and O-HAZE datasets. The obtained results are presented in Table 3, where the prior feature fusion uses MFAFM. The color attenuation prior is not applicable on the synthetic dataset. However, the two prior features are applicable to real scenes, which can improve the transfer ability of the model and allow it to directly transfer to real-world images. Finally, when using MFAFM for multi-feature fusion, only a very small number of parameters (0.07M) is increased, which verifies the operating efficiency of the model.

Table 2. Comparison of different types of networks on the O-HAZE

Baseline	✓	✓	✓	✓	✓
MFAFM		✓	✓		✓
sv(x)			✓	✓	✓
Idark(x)		✓		✓	✓
PSNR	24.57	24.73	24,75	24.79	**24.85**
SSIM	0.756	0.760	0.753	0.759	**0.762**

Table 3. Comparison of the transfer ability and parameters of different models

Methods	OTS	O-HAZE	Params
Baseline	30.98/0.975	19.39/0.677	8.49253M
+sv(x)	30.82/0.975	19.49/0.679	8.56696M
+I^{dark}(x)	**31.12/0.977**	19.47/0.680	8.56696M
+sv(x) + I^{dark}(x)	31.06/0.976	**19.61/0.682**	8.56699M

4   Conclusion

This paper proposed an end-to-end Multi-Feature Fusion Network for Single Image Dehazing (MFFN). By combining dark channel prior, color attenuation prior and deep learning, the neural network has a stronger dehazing capacity. A very simple and effective prior feature extraction method is first used. A Multi-Feature Selective Fusion Module (MFSFM) is then designed. It combines the advantages and discards the disadvantages of the two prior features, in order to perform feature enhancement. The experimental results on synthetic and real-world datasets have shown that the proposed MFFN achieved better results than those obtained by the state-of-the-art methods, which proves its effectiveness for real outdoor scenes.

References

1. Narasimhan, S.G., Nayar, S.K.: Vision and the atmosphere. Int. J. Comput. Vis. **48**(3), 233–254 (2002)
2. Cantor, A.: Optics of the atmosphere–scattering by molecules and particles. IEEE J. Quantum Electron., 698–699 (1978)
3. He, K., Sun, J., Tang, X.: Single image haze removal using dark channel prior. IEEE Trans. Pattern Anal. Mach. Intell. **33**(12), 2341–2353 (2011)
4. Liu, Q., Gao, X., He, L., Lu, W.: Single image dehazing with depth aware non-local total variation regularization. IEEE Trans. Image Process, **27**, 5178–5191 (2018)
5. Fattal, R.: Dehazing using color-lines. ACM Trans. Graph. **34**(1), Article no. 13 (2014)
6. Meng, G., Wang, Y., Duan, J., Xiang, S., Pan, C.: Efficient image dehazing with boundary constraint and contextual regularization. In: Proceeding IEEE International Conference Computer Vision, pp. 617–624 (2013)

7. Ren, W., Liu, S., Zhang, H., Pan, J., Cao, X., Yang, MH.: Single image dehazing via multi-scale convolutional neural networks. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9906, pp. 154–169. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46475-6_10

8. Zhang, H., Patel, V.M.: Densely connected pyramid dehazing network. In: Proceeding IEEE Conference Computer Vision Pattern Recognition, pp. 3194–3203 (2018)

9. Ren, W. et al.: Gated fusion network for single image dehazing. In: Proceeding IEEE/CVF Conference Computer Vision Pattern Recognition, pp. 3253–3261 (2018)

10. Li, B., Peng, X., Wang, Z., Xu, J., Feng D.: AOD-Net: all-in-one dehazing network. In: Proceeding IEEE International Conference Computer Vision, pp. 4780–4788 (2017)

11. Qu, Y., Chen, Y., Huang, J., Xie, Y.: Enhanced Pix2pix dehazing network. In: Proceeding IEEE/CVF Conference Computer Vision Pattern Recognition, pp. 8152–8160 (2019)

12. Liu, X., Ma, Y., Shi, Z., Chen, J.: GridDehazeNet: attention-based multi-scale network for image dehazing. In: Proceeding IEEE International Conference Computer Vision, pp. 7313–7322 (2019)

13. Qin, X., Wang, Z., Bai, Y., Xie, X., Jia, H.: FFA-Net: feature fusion attention network for single image dehazing. In: Proceeding AAAI Conference Artificial Intelligence, pp. 11908–11915 (2020)

14. Luo, J., Bu, Q., Zhang, L., Feng, J.: Global feature fusion attention network for single image dehazing. In: IEEE International Conference on Multimedia & Expo Workshops, pp. 1–6 (2021)

15. Zhu, Q., Mai, J., Shao, L.: A fast single image haze removal algorithm using color attenuation prior. IEEE Trans. Image Process. **24**(11), 3522–3533 (2015)

16. Li, B., et al.: Benchmarking single-image dehazing and beyond. IEEE Trans. Image Process. **28**(1), 492–505 (2019)

17. Dong, H., Pan, J., Xiang, L., Hu, Z., Zhang, X.: Multi-scale boosted dehazing network with dense feature fusion. In: Proceeding IEEE/CVF Conference Computer Vision Pattern Recognition (2020)

18. Ancuti, C.O., Ancuti, C., De Vleeschouwer, C., Timofte, R.: O-HAZE: a dehazing benchmark with real hazy and haze-free outdoor images. In: Proceeding Conference Computer Vision Pattern Recognition Workshops, pp. 88–97 (2018)

19. Ancuti, C.O., Ancuti, C., Timofte, R.: NH-HAZE: an image dehazing benchmark with non-homogeneous hazy and haze-free images. In: Proceeding IEEE/CVF Conference Computer Vision Pattern Recognition (2020)

20. Ancuti, C.O., Ancuti, C., Vasluianu, F.-A., Timofte, R.: Ntire 2020 challenge on nonhomogeneous dehazing. In: Proceeding Conference Computer Vision Pattern Recognition Workshops (2020)

21. Chen, D., et al.: Gated context aggregation network for image dehazing and deraining. In: IEEE Winter Conference on Applications of Computer Vision, pp. 1375–1383 (2019)

22. Yi, Q., Li, J., Fang, F., Jiang, A., Zhang, G.: Efficient and accurate multi-scale topological network for single image dehazing. IEEE Trans. Multimedia (2021)

23. Li, L., Dong, Y., Ren, W., Pan, J., Gao, C., Sang, N.: Semi-supervised image dehazing. IEEE Trans. Image Process. **29**, 2766–2779 (2019)

24. Shao, Y., Li, L., Ren, W., Gao, C., Sang, N.: Domain adaptation for image dehazing. In: Proceeding IEEE/CVF Conference Computer Vision Pattern Recognition, pp. 2805–2814 (2020)

25. He, T., Zhang, Z., Zhang, H., Zhang, Z., Xie, J., Li, M.: Bag of tricks for image classification with convolutional neural networks. In: Proceeding IEEE/CVF Conference Computer Vision Pattern Recognition, pp. 558–567 (2019)