



# Prior-combined dehazing network based on mutual learning

Dong Qiao<sup>1</sup> · Xiangtong Kong<sup>1</sup> · Lingjian Kong<sup>1</sup> · Jifang Liu<sup>1</sup> · Wenpeng Mi<sup>1</sup> · Shenghao Meng<sup>1</sup>

Received: 8 August 2022 / Revised: 7 October 2022 / Accepted: 13 November 2022 / Published online: 20 December 2022  
© The Author(s), under exclusive licence to Springer-Verlag London Ltd., part of Springer Nature 2022

## Abstract

Single-image dehazing is an important problem for high-level computer vision tasks since the existence of haze severely degrades the recognition ability of computers. Most recent works tend to combine prior-based dehazing method with a convolutional neural network to improve the dehazing effect in real scenes. However, these methods do not tackle with the color shifts caused by prior-based methods effectively. In this paper, we propose a prior-combined dehazing network based on mutual learning. Specifically, we build two sub-networks to achieve dehazing by both supervised and unsupervised ways. The supervised sub-network is optimized by ground truth, which provides color fidelity but may acquire under-dehazed images when applied to real scenes. The unsupervised sub-network is optimized by the dehazed images of dark channel prior, which improves the generalization ability but introduces some color shifts or artifacts. Since the dehazing of these two sub-networks shows complementary advantages, a mutual learning mechanism is built for the joint optimization. And we propose a feature fusion module based on the perceptual differences to acquire the final results. The experimental results demonstrate that our method surpasses previous state-of-the-arts on both synthetic and real-world datasets.

**Keywords** Prior-combined dehazing · Mutual learning mechanism · Feature fusion

## 1 Introduction

Haze is a commonly natural phenomenon, which makes the captured image degraded, and further hinders the recognition capability of computers. Thus, as a key prerequisite of high-level computer vision tasks, single-image dehazing has been extensively studied in recent years [1].

Single-image dehazing can be roughly divided into model-based methods and model-free methods. The model-based methods achieve dehazing via the following atmospheric scattering mode [2]:

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

where  $I(x)$  denote the hazy images captured in hazy conditions, and  $J(x)$  denote the restored haze-free images.  $A$  and  $t(x)$  denote the atmospheric light and transmission maps, respectively. Moreover, we have  $t(x) = e^{-\beta d(x)}$  with  $\beta$  and  $d(x)$  being the atmosphere scattering parameter and the

scene depth, respectively. Equation (1) is an ill-posed problem, which means we cannot directly acquire a clear image  $J(x)$  from a hazy image  $I(x)$  since both of  $A$  and  $t(x)$  are underdetermined.

To this end, early model-based methods (also called as prior-based methods) use the priors based on the observation of clear images to estimate the atmospheric light and transmission maps. These methods include color-lines prior (CLP) [3], boundary constraint and contextual regularization (BCCR) [4], dark channel prior (DCP) [5], color attenuation prior (CAP) [6], and non-local dehazing (NLD) [7]. Prior-based methods have strong generalization for image dehazing but always cause color shifts and artifacts since the unilateral hypothesis cannot estimate parameters accurately especially in complex scenes. Thus, recent model-based methods tend to utilize a designed convolutional neural network (CNN) to estimate the atmospheric light and transmission map [8–11]. However, as a simplified hazy model, the atmospheric scattering model cannot simulate the process of haze thoroughly and may restrict the final dehazing performance [12].

To this end, more model-free works [13–18] are proposed, which build an end-to-end CNN to directly map the features between the hazy images and the corresponding haze-free images. These end-to-end methods have showed

✉ Lingjian Kong  
2926057894@qq.com

<sup>1</sup> High-tech Institute, Fan Gong-ting South Street on the 12th, Qingzhou 262550, Shandong, China

strong dehazing ability in synthetic scenes, but they always acquire under-dehazed images when applied to real scenes due to domain shifts. In other words, the synthetic hazy images cannot represent uneven haze distribution and complex illumination in natural conditions and making the trained model cannot hold in these scenes. Hence, some more recent works [19–22] combine traditional priors (i.e., dark channel prior) with learning-based methods to achieve better dehazing effect in both synthetic and real scenes. However, the existing prior-combined dehazing methods cannot alleviate the distortions caused by prior-based methods effectively, and a more efficient feature aggregation mechanism should be studied to combine the complementary advantages of these two categories.

In this paper, we resort to knowledge distillation to solve the problem, and propose a prior-combined dehazing network dubbed as PCD. Knowledge distillation [23–25] is a widely used method for parameter reduction, in which a cumbersome network (teacher) is used to guide the learning process of a designed light-weight network (student). Based on it, recent dehazing works [23, 24] adopt the features of ground truths to enhance the image restoration. And paper [25] proposes a mutual distillation mechanism to improve the accuracy of a detection task. Inspired by them, we propose a mutual learning mechanism to combine the complementary merits of prior based methods and learning-based methods. Specifically, we build two sub-networks to achieve dehazing by both supervised and unsupervised learning. The supervised sub-network is optimized by the ground truths, and the unsupervised sub-network is optimized by DCP dehazed images. Hence, the outputs of supervised subnetwork provide color fidelity since the supervision contains completely correct information, and the outputs of unsupervised sub-network provide generalization ability to real scenes since the DCP is a statistical law of clear images. Then a novel mutual learning mechanism can be used to combine the complementary merits adoptively and acquire two preliminary dehazed images. Moreover, since it is highly possible that either of them is better than the other in some local regions, a feature fusion method (FFM) based on perceptual differences is further proposed. The FFM merges the preliminary dehazed images and then achieves better dehazing effect with realness.

The main contributions of this paper are summarized as follows:

1. We introduce a prior-combined dehazing (PCD) network based on mutual learning to combine the merits of prior-based methods and learning-based methods.
2. We propose a novel mutual learning mechanism to achieve the joint optimization of the supervised sub-network and unsupervised sub-network.
3. We propose a feature fusion module based on perceptual differences to aggregate the outputs of the sub-networks, which acquires the final dehazed images with clearer textures.

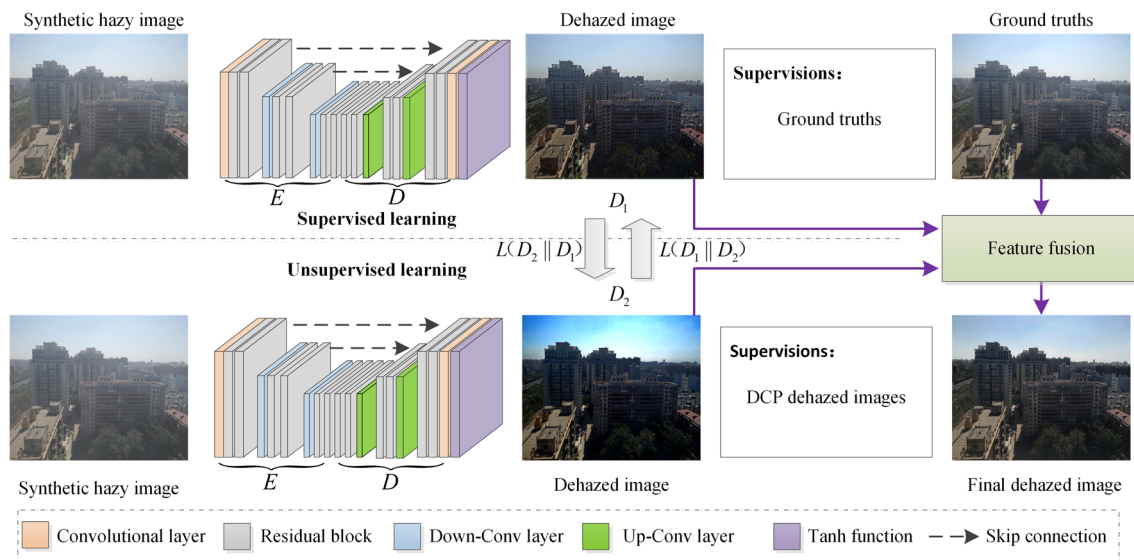
## 2 Network architecture

### 2.1 General architecture

Since it is hard to collect a large number of hazy images and their haze-free images, the existing learning-based methods still train the model by synthetic images. The synthetic hazy images exist apparent differences from real hazy images w.r.t haze distribution, and the CNN model lacks of the knowledge to natural scenes, which results in that learning-based methods acquire under-dehazed images in real scenes. Considering that traditional priors (i.e., dark channel prior) are statistical laws of clear images, recent works tend to combine the prior-based method with CNN-based methods to achieve better dehazing effect in real scenes. However, since the prior dehazed images contain severe distortion information such as artifacts, illumination changes, color shifts and halos, these prior-combined methods also suffer from image distortions due to insufficient feature aggregation mechanism. Thus, as shown in Fig. 1, a prior-combined dehazing network based on mutual learning is proposed to solve the problem, which consists of a supervised sub-network, an unsupervised sub-network and a feature fusion module.

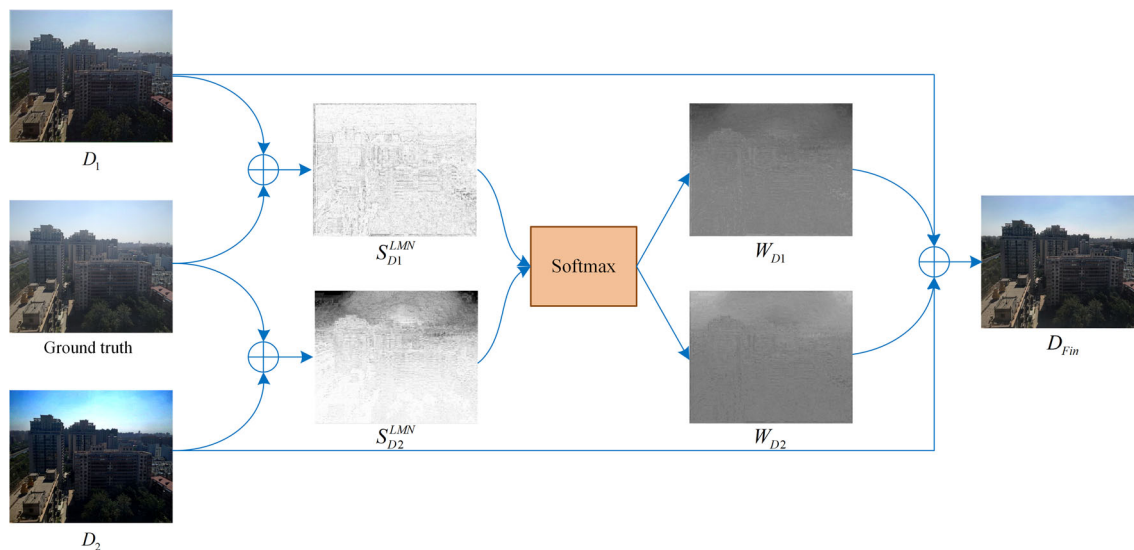
### 2.2 Supervised sub-network

The supervised sub-network achieves dehazing by end-to-end strategies, which directly builds the mapping between synthetic hazy images and ground truths. As shown in Fig. 1, the supervised sub-network is based on a three scales autoencoder structure. Differently, we replace traditional convolutions with residual blocks for feature extraction since the residual structure has been proved efficient for feature flow. Specifically, we first extract the features of synthetic hazy images by a convolutional layer, which changes the channel number from 3 to 64. Then, two residual blocks enhance the feature representation and a Down-Conv layer downsamples feature maps to high-level semantic space. We downsample features by two times and form three scales feature maps, and the features of the bottleneck layer are sent to the decoder  $D$  to restore high-resolution results. The decoder  $D$  contains the structures symmetric to the encoder  $E$ , in which an Up-Conv is used to restore resolution, and then, two residual blocks enhance unsampled features. During the decoding process, the features of encoder  $E$  are sent to the corresponding layer of decoder by skip connection to avoid



**Fig. 1** The general architecture of our PCD, which consists of a supervised sub-network and an unsupervised sub-network optimized by the ground truths and DCP dehazed images. The outputs of sub-networks

show complementary merits, and they are fused by perceptual differences to acquire final dehazed images



**Fig. 2** The architecture of the perceptual feature fusion. The dehazed images of two sub-networks are converted to LMN color space to estimate the similarity of ground truths, and then, two weighed maps  $W_{D1}$

and  $W_{D2}$  are generated by softmax function to assign the dehazed images  $D_1$  and  $D_2$  adoptively

the loss of spatial information. Thus, except for the first scale in the decoder  $D$ , the features from the encoder  $E$  and the previous scale features of  $D$  are concatenated as the input of current scales until restoring the resolution same as the input hazy images. Finally, the outputs of decoder are sent to a convolutional layer with the Tanh activation function to acquire the dehazed images. Since the supervisions of supervised sub-network are ground truths, the dehazed images achieve

high information fidelity although some regions are under-dehazed.

### 2.3 Unsupervised sub-network

To generate the features of DCP method, we build an unsupervised sub-network. As shown in Fig. 1, the unsupervised sub-network has the same structures of supervised sub-network. Differently, we train the unsupervised sub-network

with the supervisions of the dehazed images of dark channel prior (DCP) rather than ground truths. Since the supervisions are fake ground truths (DCP dehazed images), we call it as unsupervised learning in this paper. The DCP method has been proved efficient in real haze removal although it may introduce some distortions, especially in the sky regions. As shown in Fig. 1, the outputs of unsupervised sub-network have similar features of DCP dehazed images, which acquire more discriminative textures although the sky regions suffer from illumination oversaturations. Since the output images of two sub-networks achieve complementary advantages, we apply a mutual learning mechanism to optimize them adoptively by two extra distillation losses. The details of the distillation losses can be seen in Sect. 2.5.

### 2.4 Feature fusion

In our method,  $D_1$  and  $D_2$  are dehazed images under the supervisions of ground truths and DCP dehazed images, respectively. Since the dehazed images  $D_1$  and  $D_2$  are acquired by their own ways, it is highly possible that either of them is better than the other in some local regions. Hence, if better regions from either of  $D_1$  and  $D_2$  are assigned with larger weights, a better result will be acquired. Since  $D_1$  are dehazed images with good fidelity and  $D_2$  are dehazed images with visibility, the fusion should consider the realness of  $D_1$  and maintain the same visibility of  $D_2$ . Thus, we fuse them based on the ground truths. The process can be divided into the follows:

(1) Feature Extraction: Recent IQA research [22, 26] has shown that images in LMN color space can be easily estimated the color distortions and chrominance shifts. Consequently, to objectively estimate the realness of dehazed images  $D_1$  and  $D_2$ , we first transform the images into the LMN color space, which can be expressed as:

$$\begin{bmatrix} L \\ M \\ N \end{bmatrix} = \begin{bmatrix} 0.06 & 0.63 & 0.27 \\ 0.30 & 0.04 & -0.35 \\ 0.34 & -0.6 & 0.17 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{2}$$

(2) Similarity Calculation: We calculate the similarity in LMN space to evaluate the realness of the dehazed images. Taking the similarity value between dehazed image  $D_1$  and ground truth as example. Supposing that  $L_1(x)$ ,  $M_1(x)$  and  $N_1(x)$  are computed from dehazed images  $D_1$ ;  $L_2(x)$ ,  $M_2(x)$  and  $N_2(x)$  are acquired from the ground truth. The similarity  $S_{D_1}^{LMN}$  at pixel  $x$  can be expressed as:

$$S_{D_1}^{LMN}(x) = \frac{2L_1(x)L_2(x) + C_1}{L_1^2(x) + L_2^2(x) + C_1} \times \frac{2M_1(x)M_2(x) + C_1}{M_1^2(x) + M_2^2(x) + C_1} \times \frac{2N_1(x)N_2(x) + C_1}{N_1^2(x) + N_2^2(x) + C_1} \tag{3}$$

where  $C_1$  is a constant set to 130 as suggested in [26].

(3) Weight generation and feature fusion:

To make the final result contains more realistic information, we convert the similarity into weights for the feature fusion process. Supposing that  $S_{D_1}^{LMN}$  denotes the similarity value between dehazed image  $D_1$  and the ground truth at pixel  $x$ ;  $S_{D_2}^{LMN}$  denotes the similarity value between dehazed image  $D_2$  and the ground truth. The weights of dehazed images  $D_1$  and  $D_2$  at pixel  $x$  can be expressed as:

$$\begin{bmatrix} W_{D_1}(x) \\ W_{D_2}(x) \end{bmatrix} = \text{Softmax} \left( \begin{bmatrix} S_{D_1}^{LMN}(x) \\ S_{D_2}^{LMN}(x) \end{bmatrix} \right) \tag{4}$$

where Softmax denotes the softmax function that generating weighs adoptively based on the similarity  $S_{D_1}^{LMN}$  and  $S_{D_2}^{LMN}$ . Note that  $W_{D_1}(x) + W_{D_2}(x) = 1$ .

In the end, we aggregate the preliminary dehazed image  $D_1$  and  $D_2$  based on their weights, and the final result can be expressed as:

$$D_{Fin} = W_{D_1} \otimes D_1 + W_{D_2} \otimes D_2 \tag{5}$$

where  $D_{Fin}$  denotes the final dehazed image, and  $W_{D_1}$  and  $W_{D_2}$  are the generated weights of dehazed images  $D_1$  and  $D_2$ , respectively.  $\otimes$  denotes the pixel-wise product.

### 2.5 Loss function

Paper [27] has shown that the combination of pixel-wise loss and feature-wise loss can effectively mimic feature differences between two images. Thus, we use L1 loss, perceptual loss and distillation loss to train the proposed PCD, and the total loss function of supervised sub-network and unsupervised sub-network can be expressed as:

$$L_{Sup} = L_1 + \lambda_1 L_{Per1} + \lambda_2 L(D_1 \| D_2) \tag{6}$$

$$L_{Uns} = L_2 + \lambda_1 L_{Per2} + \lambda_2 L(D_2 \| D_1) \tag{7}$$

where  $L_{Sup}$  and  $L_{Uns}$  are losses of the supervised sub-network and unsupervised sub-network, respectively.  $L_1$  and  $L_2$  denote the L1 loss, and  $L_{Per1}$  and  $L_{Per2}$  denote the perceptual loss.  $\lambda_1$  and  $\lambda_2$  are the weight coefficients equal to 1. Moreover,  $L(D_1 \| D_2)$  and  $L(D_2 \| D_1)$  denote the distillation losses, which make the supervised sub-network (unsupervised sub-network) mimic the features of unsupervised sub-network (supervised sub-network), respectively.

#### 2.5.1 L1 loss

L1 loss (mean absolute error) can rapidly minimize the feature differences between hazy images and clear images by

per-pixel comparison, and thus, we add L1 loss for network training. Different from L2 loss (mean squared error), L1 loss trains network more stably, which can be expressed as:

$$L_1 = \|J_{dcp} - D_1\|_1 \quad (8)$$

$$L_2 = \|J - D_2\|_1 \quad (9)$$

where  $J_{dcp}$  and  $J$  represent DCP dehazed images and ground truths, respectively.  $D_1$  and  $D_2$ , respectively, represent the output of supervised sub-network and unsupervised sub-network.  $\|\cdot\|_1$  denotes the L1 loss.

### 2.5.2 Perceptual loss

Perceptual loss [28] compares two images by perceptual and semantic differences, which effectively helps the network restore more vivid images. In this paper, we pretrain VGG19 network [29] on the ImageNet [30] and extract features of the convolutions in number 2, 7, 12, 21 and 30 to calculate losses. The perceptual losses used in supervised sub-network and unsupervised sub-network are, respectively, denoted as  $L_{Per1}$  and  $L_{Per2}$ , which can be expressed as:

$$L_{Per1} = \sum_{i=1}^5 \frac{1}{C_i H_i W_i} \|\Phi_i(GT) - \Phi_i(D_1)\|_1 \quad (10)$$

$$L_{Per2} = \sum_{i=1}^5 \frac{1}{C_i H_i W_i} \|\Phi_i(J_{dcp}) - \Phi_i(D_2)\|_1 \quad (11)$$

where  $\Phi_i(\cdot)$  ( $i = 1, 2, 3, 4, 5$ ) denotes the five scales perceptual features extracted from the trained VGG19 network.  $C_i$ ,  $H_i$  and  $W_i$  represent the number of channel, height, and width of feature maps.

### 2.5.3 Distillation loss

Due to a limited feature aggregation, recent prior-combined methods cannot deal with the distortions caused by prior-based methods. Since the supervised sub-network and unsupervised sub-network show complement merits about image dehazing, we aggregate the features by the mutual learning mechanism with two designed distillation losses:

$$L(D_1 \| D_2) = L(D_2 \| D_1) = \|D_2 - D_1\| \quad (12)$$

## 3 Experiment and analysis

In Sects. 3.1 and 3.2, we introduce the used datasets and the experimental details, respectively. In Sect. 3.3, we compare

the proposed PCD with some state-of-the-arts on both synthetic dataset and real-world dataset and analyze the results by both qualitative and quantitative ways.

### 3.1 Datasets

For training, we use the Indoor Training Set (ITS) in Realistic Single Image Dehazing (RESIDE) [31], which contains 13,990 synthetic indoor hazy images and the corresponding haze-free images. For testing, we use the Synthetic Objective Testing Set (SOTS) in RESIDE, which contains 500 paired images captured in indoor and outdoor scenes, respectively. To compare the results quantitatively, the peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) [32] are used. Moreover, to show the generalization in natural scenes, some real-world images in URHI dataset [19] are further used. Since the real-world images do not contain their ground truths, we only qualitatively compare the results.

### 3.2 Implementation details

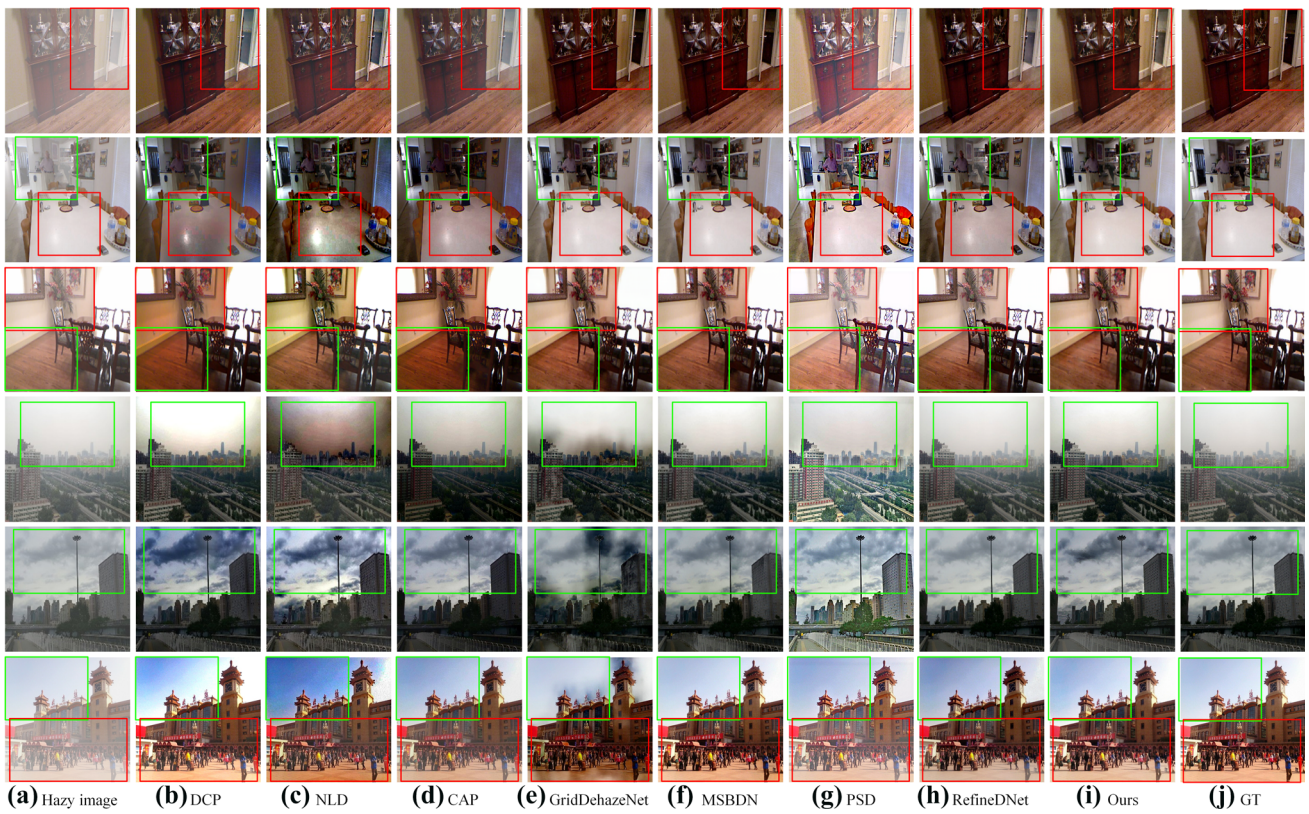
We achieve our PCD by PyTorch framework. For training, we reshape all the training images as  $256 \times 256$ . Moreover, we set the training batch size as 4 and set the total epochs as 20. To optimize the network and accelerate the training process, the Adam [33] optimizer is used with the attenuation coefficient being  $\beta_1 = 0.9$  and  $\beta_2 = 0.999$ , respectively. In addition, the initial learning rate is set as 0.0002, and we decrease it to half after every two epochs.

### 3.3 Experimental results

#### 3.3.1 Results on synthetic images

To show the effectiveness of our PCD, we test on both the indoor and outdoor images in SOTS. Figure 3 shows the results, we can find that prior-based methods (DCP, NLD and CAP) dehaze effectively in both indoor and outdoor scenes but suffer from halos, color shifts and artifacts. By contrast, learning-based methods (GridDehazeNet and MSBDN) dehaze effectively in indoor scenes due to trained by indoor dataset. But unfortunately, GridDehazeNet introduces many artifacts for outdoor scenes, which shows its poor generalization ability. Better than these methods, prior-combined methods (PSD, RefineDNet and Ours) perform stably in both indoor and outdoor scenes. However, there is some residual haze in the results of PSD. And the results of RefineDNet suffer from some color shifts. Only our PCD can effectively dehaze and provide color fidelity.

To compare the results quantitatively, we calculate the average PSNR and SSIM. Table 1 shows the results; for indoor scenes, our PCD achieves the third-best results with the PSNR and SSIM being 27.34 dB and 0.971, respectively.



**Fig. 3** Results on images of SOTS. The upper three rows are the results of synthetic indoor images, and the bottom three rows are the results of synthetic outdoor images

**Table 1** Quantitative comparison on SOTS

Dataset	SOTS				FLOPs (KM)
	Indoor		Outdoor		
	PSNR	SSIM	PSNR	SSIM	
DCP	19.85	0.872	<b>20.44</b>	<b>0.898</b>	–
NLD	17.29	0.801	18.11	0.871	–
CAP	18.87	0.848	19.01	0.868	–
GridDehazeNet	<i>32.16</i>	<b>0.984</b>	16.21	0.783	247.1
MSBDN	<b>32.67</b>	<i>0.983</i>	<i>20.46</i>	<i>0.916</i>	89.2
PSD	17.23	0.74	15.14	0.77	289.1
RefineDNet	26.14	0.962	18.84	0.764	96.1
Ours	<b>27.34</b>	<b>0.971</b>	<b>23.72</b>	<b>0.934</b>	<b>79.2</b>

Number in bold, italic and bold–italic denote the best, second-best and third-best results, respectively. KM represents the number of FLOPs setting as  $1 \times 10^9$

But for outdoor scenes, our PCD improves PSNR from 20.46 to 23.72 dB and increases SSIM by 0.018 when compared with the second-best method MSBDN. The results show that learning-based methods (PSD and RefinedNet) drop the performance when applied to outdoor scenes. But our PCD alleviates it by adopting the efficient mutual learning mechanism to combine prior-based and learning-based ways. In addition, the comparison of FLOPS shows our PCD achieves dehazing with the minimum computational overhead.

### 3.3.2 Results on real hazy images

To verify the generalization ability to natural scenes, we further evaluate the dehazing performance on real-world images in URHI [19]. As shown in Fig. 4, prior-based methods still dehaze effectively and restore most textures for these scenes. But unfortunately, some color shifts, artifacts or residual haze may exist. By contrast, learning-based methods fail to these scenes and there are a large amount haze in the results of GridDehazeNet and MSBDN. This further verifies that prior-based method has more stable dehazing performance than learning-based methods since the performance of learning-based methods is restricted by training data. For prior-combined methods, PSD acquires visually pleasing results with some illumination changes. RefinedNet darkens the images and also removes most haze. Better than them, our PCD removes more dense haze existing in the sky regions.

## 3.4 Ablation study

### 3.4.1 Ablation study on the overall architecture

To show the effectiveness of the overall architecture, we conduct some ablation studies to explore the influences of the following three key factors: supervised learning (SL), unsupervised learning (UL), mutual learning mechanism (MLM) and feature fusion module (FFM). Thus, we construct the following variants: (1) SL, only train the network by supervised learning; (2) SL + UL, train the network by both supervised learning and unsupervised learning, and combine the outputs by channel-wise concatenation; (3) SL + UL + MLM, train the network by both supervised learning and unsupervised learning with mutual learning mechanism; (4) SL + UL + MLM + FFM (Ours), replace the channel-wise concatenation by feature fusion module. We train these variants on the

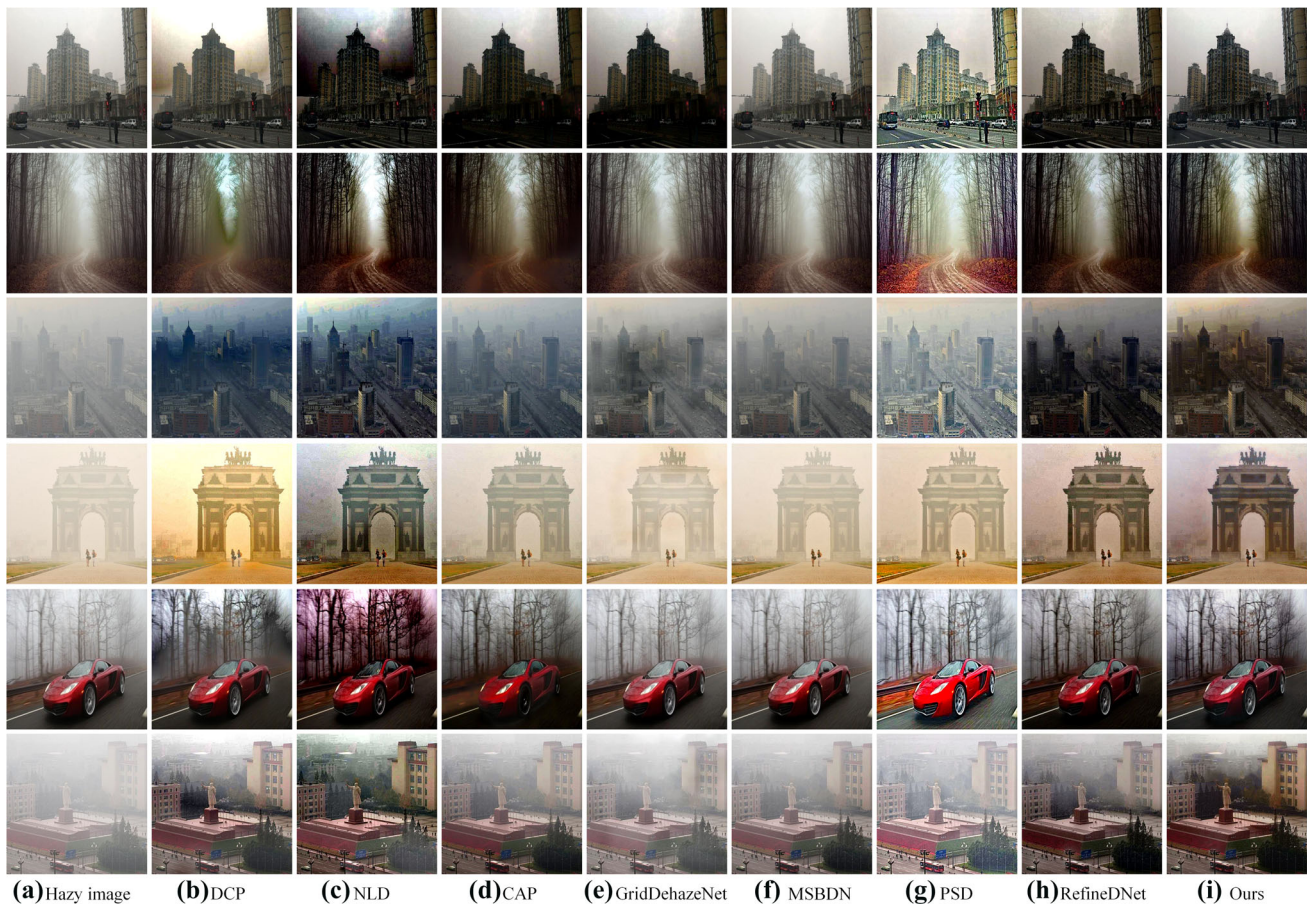
ITS dataset for 20 epoch and test on the outdoor dataset of SOTS. Table 2 shows the results, the proposed PCD achieves the best metrics with PSNR and SSIM being 23.72 dB and 0.934, respectively. Specifically, by adding UL, the proposed PCD significantly improves PSNR from 20.15 to 22.46 dB and increases SSIM by 0.021. Moreover, by adding MLM, the PCD further combines the merits of prior-based method and learning-based method, which improves the metrics by 0.96 dB and 0.013. Finally, adding the FFM acquires a better result and also provides a little gain.

### 3.4.2 Comparison for different prior

In our PCD, we use the DCP dehazed images as fake ground truths to achieve unsupervised learning and improve the generalization ability. Hence, it is important to compare the effectiveness with different prior-based methods. Thus, we acquire the dehazed images of DCP [5], CAP [6] and NLD [7] and use them to train the network, respectively. The quantitative comparisons of 500 outdoor images of SOTS are shown in Table 3, and the DCP combined network acquires better results than NLD combined network and CAP combined network, which shows the DCP method may have the best generalization in various scenes. Due to trained by indoor images, GridDehazeNet acquires poor metrics. Moreover, although PSD combines with multiple prior-based methods, the insufficient fusion mechanism causes severe color shifts and drops the metrics.

## 4 Conclusion

In this paper, we propose a prior-combined dehazing (PCD) network based on mutual learning. The PCD uses two sub-networks optimized by the ground truths and prior dehazed images to acquire two preliminary dehazed images and utilizes a novel mutual learning strategy to further aggregate the complementary features. In addition, a perceptual feature fusion method is proposed to maintain the dehazing ability while alleviate distortions. Experimental results on both synthetic and real-world images have shown that our PCD achieves better results in real scenes although it only acquires the third-best results in synthetic scenes. A more efficient prior-combined strategy will be studied in our further work.



**Fig. 4** Results of images in URHI, our method still dehazes effectively in these natural scenes

**Table 2** Results of different variants about the overall architecture

Metrics	SL	SL + UL	SL + UL + MLM	Ours
PSNR/SSIM	20.15/0.895	22.46/0.916	23.42/0.929	23.72/0.934

**Table 3** Results of our PCD with different prior dehazed images as guidance

Metrics	DCP as guidance	NLD as guidance	CAP as guidance	GridDehazeNet	PSD
PSNR/SSIM	23.72/0.934	23.42/0.927	22.69/0.926	16.21/0.783	15.14/0.77

**Author contributions** DQ was involved in conceptualization, methodology, and writing—original draft. XK contributed to visualization and formal analysis. LK was involved in conceptualization and supervision. JL contributed to conceptualization and validation. WM and SM was involved in review and editing.

**Funding** NO funding.

**Data availability** The data and materials that support the findings of this study are available from the corresponding author for the purpose of academic research.

**Declarations**

**Competing interests** 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.

**Ethical approval** This article does not deal with research on humans or animals.



## References

1. Xue, Y., Jin, G., Shen, T., Tan, L., Yang, J., Hou, X.: MobileTrack: siamese efficient mobile network for high-speed UAV tracking. *IET Image Process.* **12**, 3300–3313 (2022)
2. Narasimhan, S.G., Nayar, S.K.: Vision and the Atmosphere. *Int. J. Comput. Vis.* **48**, 233–254 (2002)
3. Raanan, F.: Dehazing Using Color-Lines. *ACM Trans. Graph. (TOG)*. **34**, 1–14 (2014)
4. Meng, G., Wang, Y., Duan, J., Xiang, S.: Efficient image dehazing with boundary constraint and contextual regularization. In: 2013 IEEE International Conference on Computer Vision, pp. 617–624 (2013)
5. He, K., Sun, J., Tang, X.: Single image haze removal using dark channel prior. *IEEE Trans. Pattern Anal. Mach. Intell.* **33**, 2341–2353 (2011)
6. Zhu, Q., Mai, J., Shao, L.: A fast single image haze removal algorithm using color attenuation prior. *IEEE Trans. Image Process.* **24**, 3522–3533 (2015)
7. Berman, D., Treibitz, T., Avidan, S.: Non-local image dehazing. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2016)
8. Ren, W., Si, L., Hua, Z., Pan, J., Yang, M.H.: Single image dehazing via multi-scale convolutional neural networks. In: European Conference on Computer Vision (ECCV), pp. 154–169 (2016)
9. Cai, B., Xu, X., Jia, K., Qing, C., Tao, D.: DehazeNet: an end-to-end system for single image haze removal. *IEEE Trans. Image Process.* **25**, 5187–5198 (2016)
10. Li, B., Peng, X., Wang, Z., Xu, J., Feng, D.: AOD-Net: all-in-one dehazing network. In: IEEE International Conference on Computer Vision (ICCV), pp. 4780–4788. IEEE, Venice, (2017)
11. Zhang, H., Patel, V. M.: Densely connected pyramid dehazing network. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 3194–3203. IEEE, Salt Lake City, UT, USA (2018)
12. Liu, X., Ma, Y., Shi, Z., Chen, J.: GridDehazeNet: attention-based multi-scale network for image dehazing. In: IEEE International Conference on Computer Vision (ICCV), pp. 7313–7322 (2019)
13. Chen, D., He, M., Fan, Q., Liao, J., Zhang, L., Hou, D., Yuan, L., Hua, G.: Gated context aggregation network for image dehazing and deraining. In: IEEE Winter Conference on Applications of Computer Vision (WACV), pp. 1375–1383. IEEE, Waikoloa Village, HI, USA (2019)
14. Ren, W., Ma, L., Zhang, J., Pan, J., Cao, X., Liu, W., Yang, M.-H.: Gated fusion network for single image dehazing. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 3253–3261. IEEE, Salt Lake City, UT (2018)
15. Wang, N., Cui, Z., Su, Y., He, C., Li, A.: Multiscale supervision-guided context aggregation network for single image dehazing. *IEEE Signal Process. Lett.* **29**, 70–74 (2022)
16. Cui, Z., Wang, N., Su, Y., Zhang, W., Lan, Y., Li, A.: ECANet: enhanced context aggregation network for single image dehazing. *Signal Image Video Process.* (2020)
17. Wang, N., Cui, Z., Su, Y., Li, A.: RGNAM: recurrent grid network with an attention mechanism for single-image dehazing. *J Electron Imaging* **30**(3), 033026 (2021)
18. Dong, H., Pan, J., Xiang, L., Hu, Z., Zhang, X., Wang, F., Yang, M.-H.: Multi-scale boosted dehazing network with dense feature fusion. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2154–2164. IEEE, Seattle, WA, USA (2020)
19. Shao, Y., Li, L., Ren, W., Gao, C., Sang, N.: Domain adaptation for image dehazing. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2805–2814. IEEE, Seattle, WA, USA (2020)
20. Wang, N., Cui, Z., Su, Y., He, C., Lan, Y., Li, A.: Prior-guided multiscale network for single-image dehazing. *IET Image Process.* **15**, 3368–3379 (2021)
21. Chen, Z., Wang, Y., Yang, Y., Liu, D.: PSD: principled synthetic-to-real dehazing guided by physical priors. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 7176–7185 (2021)
22. Zhao, S., Zhang, L., Shen, Y., Zhou, Y.: RefineDNet: a weakly supervised refinement framework for single image dehazing. *IEEE Trans. Image Process.* **30**, 3391–3404 (2021)
23. Hong, M., Xie, Y., Li, C., Qu, Y.: Distilling image dehazing with heterogeneous task imitation. In: 2020 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3459–3468 (2020)
24. Wu, H., Liu, J., Xie, Y., Qu, Y., Ma, L.: Knowledge transfer dehazing network for nonhomogeneous dehazing. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 1975–1983 (2020)
25. Zhang, Y., Xiang, T., Hospedales, T. M., Lu, H.: Deep mutual learning (2017)
26. Zhang, L., Shen, Y., Li, H.: VSI: a visual saliency-induced index for perceptual image quality assessment. *IEEE Trans. Image Process.* **23**, 4270–4281 (2014)
27. Zhao, H., Gallo, O., Frosio, I., Kautz, J.: Loss functions for image restoration with neural networks. *IEEE Trans. Comput. Imaging* **3**, 47–57 (2017)
28. Johnson, J., Alahi, A., Fei-Fei, L.: Perceptual losses for real-time style transfer and super-resolution. In: European Conference on Computer Vision (2016)
29. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. *Comput. Sci.* (2014)
30. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A.: ImageNet large scale visual recognition challenge. *Int. J. Comput. Vis.* **115**, 211–252 (2015)
31. Li, B., Ren, W., Fu, D., Tao, D., Feng, D., Zeng, W., Wang, Z.: Benchmarking single-image dehazing and beyond. *IEEE Trans. Image Process.* **28**, 492–505 (2019)
32. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.* **13**, 600–612 (2004)
33. Kingma, D., Ba, J.: Adam: a method for stochastic optimization. *Comput. Sci.* (2014)

**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.