

PDR-Net: Perception-Inspired Single Image Dehazing Network with Refinement

Chongyi Li, Chunle Guo, Jichang Guo, Ping Han,
Huazhu Fu, *Senior Member, IEEE*, and Runmin Cong *Member, IEEE*,

Abstract—During recent years, we have witnessed a rapid development of wireless network technologies which have revolutionized the way people take and share multimedia content. However, images captured in the outdoor scenes usually suffer from limited visibility due to suspended atmospheric particles, which directly affects the quality of photos. Despite the recent progress of image dehazing methods, the visual quality of dehazed results still needs further improvement. In this paper, we propose a deep convolutional neural network (CNN) for single image dehazing called PDR-Net, which includes a perception-inspired haze removal subnetwork that reconstructs the latent dehazed image and a refinement subnetwork that further enhances the contrast and color properties of the dehazed result by joint multi-term loss optimization. Compared to the previous methods, our method combines the advantages of existing indoor and outdoor image dehazing training data, which makes the proposed PDR-Net generalized to various hazy images and effective for improving the visual quality of the dehazed results. Extensive experiments demonstrate that the proposed method achieves comparable and even better performance on both real and synthetic images in qualitative and quantitative metrics. Additionally, the potential usage of our method in high-level vision tasks is discussed.

Index Terms—Image dehazing, perceptual loss, end-to-end, image refinement.

I. INTRODUCTION

MORE and more people are taking photos or videos to share their experience anytime and anywhere, which benefits from the popularity of mobile devices equipped with various cameras and the rapid development of wireless network technologies [1], [2]. Multimedia data play significant roles in our daily life; however, photos and videos captured in the outdoor environment often suffer from noticeable

This work was supported in part by the National Natural Science Foundation of China under Grant 61771334 and Grant 61571442, in part by the Fundamental Research Funds for the Central Universities under Grant 2019RC039, and in part by the National Key Research and Development Program of China under Grant 2016YFB0502405. (*Corresponding author: Chunle Guo*)

Chongyi Li is with the School of Electrical and Information Engineering, Tianjin University, Tianjin, China, and also with the Department of Computer Science, City University of Hong Kong, Hong Kong, China (e-mail: lichongyi25@gmail.com).

Chunle Guo, and Jichang Guo are with the School of Electrical and Information Engineering, Tianjin University, Tianjin, China (e-mail: guochunle@tju.edu.cn; jcguo@tju.edu.cn).

Ping Han is with College of Electronic Information and Automation, Civil Aviation University of China, Tianjin, China (e-mail: hanpingcauc@163.com).

Huazhu Fu is with the Inception Institute of Artificial Intelligence, Abu Dhabi, United Arab Emirates (e-mail: hzfu@ieee.org).

Runmin Cong is with the Institute of Information Science, Beijing Jiaotong University, Beijing 100044, China, and also with the Beijing Key Laboratory of Advanced Information Science and Network Technology, Beijing Jiaotong University, Beijing 100044, China (e-mail: rmcong@bjtu.edu.cn).

interference from the haze which is a natural atmospheric phenomenon caused by floating particles (e.g., dust, smoke, and liquid droplets). Haze has two main effects on the captured images: attenuation of the light and contamination with an additive component to the image [3], [4]. High-quality images are desired in the multimedia content sharing.

To remove the haze, traditional methods usually follow the similar pipeline of (1) modeling the transmission, (2) refining the coarse transmission, (3) estimating the global atmospheric light, and (4) reconstructing the latent image according to the predicted model parameters. However, this pipeline imposes several limitations. Firstly, transmission is ordinarily estimated based on priors. Nevertheless, the priors relying on statistics are not accurate when hazy images are taken under uncontrolled light conditions, different haze concentrations, and various scene depths. Secondly, conventional global atmospheric light estimation methods often make mistakes when there are white objects, highlight regions, or shadows. Thirdly, the errors in the separate estimation steps will be accumulated and amplified when the separately estimated variables are combined [5]. Besides, the recent deep-learning based methods usually produce the results with low contrast and monotonous colors due to the synthesized training data by using indoor RGB-D data. Moreover, the existing outdoor hazy image training data only have light haze, which results in the limitations of the dehazing models trained on such data.

Inspired by recent advances on big data driven multimedia data processing [6]–[9], we propose an image dehazing network that benefits from pure data-driven and perception-driven learning to improve the visual quality of the hazy images. Observing in Figure 1(a), it is obvious that the presence of haze impairs the visual quality of the image. Compared to the results of recent methods [5], [10] shown in Figure 1(b) and (c), our result has better contrast and more vivid color.

The main contributions are summarized as follows:

- We propose a PDR-Net to deal with image dehazing and quality refinement separately by using a dehazing subnetwork and a refinement subnetwork, which makes the advantages of existing indoor and outdoor training data complementary (*i.e.*, different haze concentrations from indoor data and vivid colors and high contrast from outdoor data). As a result, our method effectively removes various haze and produces visually pleasing results.
- Aiming at the problems of low contrast and monotonous color induced by indoor training data, the proposed refinement subnetwork is optimized by a multi-term loss function which consists of content loss, color loss, and

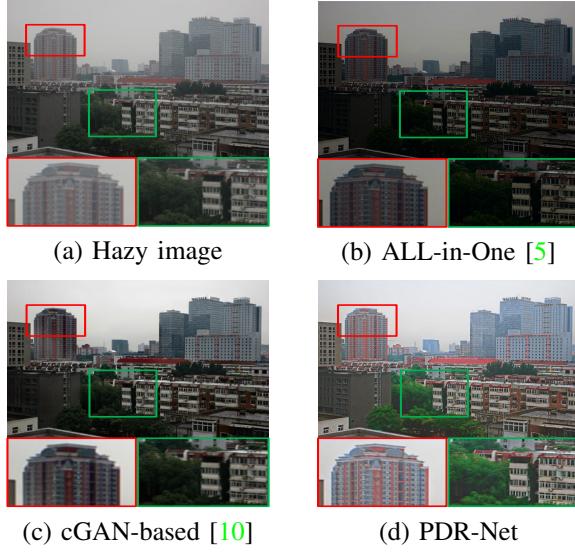


Fig. 1. Sample results. Red and green boxes indicate the obvious differences.

contrast loss to improve the visual quality of our dehazed results.

- Our method achieves comparable and even better results than the state-of-the-art methods on both real and synthetic images. Benefiting from the light-weight and end-to-end manner, the proposed PDR-Net has potential for practical applications.

The remainder of this paper is organized as follows. Section II introduces the related work. Section III presents the problem formulation and explains the proposed PDR-Net. Section IV performs experiments to validate the proposed method. Section V discusses and concludes this paper.

II. RELATED WORK

Image dehazing methods can be categorized as supplementary information-based methods and single image-based methods.

A. Supplementary Information-Based Method

To circumvent the ill-posed nature of image dehazing, supplementary information-based methods usually require additional knowledge such as 3D geographical models [11], scene depth [12], multiple images [13], and polarization filters [14]. Nevertheless, these methods are mostly computationally intensive and not applicable to dynamic scenes.

B. Single Image-Based Method

Much attention has been devoted to single image dehazing methods [3], [5], [9], [10], [15]–[34]. Tan [16] improved the visibility of image captured in hazy weather by maximizing the contrast. In [18], He *et al.* proposed a Dark Channel Prior (*i.e.*, DCP). Based upon this prior, the transmission is roughly estimated. After that, the dehazed image is achieved based on the refined transmission by guided filter [35] as well as the estimated global atmospheric light. Meng *et al.* [19] used contextual regularization to constrain the estimation

of transmission. In [21], Fattal proposed a local color-lines prior to estimate the transmission. Tang *et al.* [22] trained a regression model to estimate the transmission by hand-crafted haze-relevant features. In [23], Zhu *et al.* proposed a prior (*i.e.*, CAP) which can be used to estimate the scene depth. In [3], Berman *et al.* proposed a non-local prior, which relies on the assumption that colors of a haze-free image are well approximated by a few hundred distinct colors.

In recent years, deep learning based methods have achieved a promising performance in diverse visual tasks [36]–[39]. In [17] and [24], CNNs are utilized to predict the transmission. After that, the guided filtering [35] as post-processing is used to suppress the halo effect in the predicted coarse transmission. With the transmission and global atmospheric light estimated by conventional methods, the haze-free image is reconstructed. In [28], a cascaded CNN was proposed to separately estimate the transmission and global atmospheric light. In [5], Li *et al.* unified the transmission and global atmospheric light into one formula, and then estimated the parameters by a CNN model. To obtain effective features for single image dehazing, Song *et al.* [9] proposed a ranking CNN. Based on the captured haze related features, the dehazed results can be obtained by using a haze density prediction model trained by a random forest regression. Yang and Sun [32] proposed a prior learning-based deep network for image dehazing, which combines the advantages of traditional prior-based dehazing methods and deep learning-based methods. Recently, generative adversarial networks (GANs) [40] have been used in image dehazing. Zhang and Patel [33] proposed a densely connected pyramid dehazing network, which uses a joint-discriminator to decide whether the dehazed result is real or fake. Li *et al.* [10] restored hazy images based on a modified conditional GAN [41] (*i.e.*, cGAN) which introduces the VGG features and ℓ_1 regularized gradient prior. Du and Li [34] proposed a perceptually optimized GAN for image dehazing, which learns a nonlinear mapping from the space of degraded images to that of haze-free ones.

Different from the learning-based methods [17], [24], [28] which separately estimate model parameters, the proposed PDR-Net directly produces a clear image in its end-to-end system and can be seamlessly embedded with other deep network architectures. Compared to the latest GAN-based methods [10], [33], [34], we treat the problem of single image dehazing as two tasks namely dehazing and refinement. Thus, the proposed method generalizes well to various scenes and light conditions and produces realistic results. Besides, to combine the advantages of indoor and outdoor hazy image training data, we propose two task-driven subnetworks, which is a new attempt in the research area of single image dehazing.

III. PROPOSED PDR-NET

We propose a haze removal subnetwork to reveal the underlying correlations between hazy image and haze-free image in an end-to-end manner. To further refine the visual quality of the dehazed results, we propose a refinement subnetwork optimized by a multi-term loss. Next, we explain the PDR-Net in detail. Before that, we first formulate the problem.

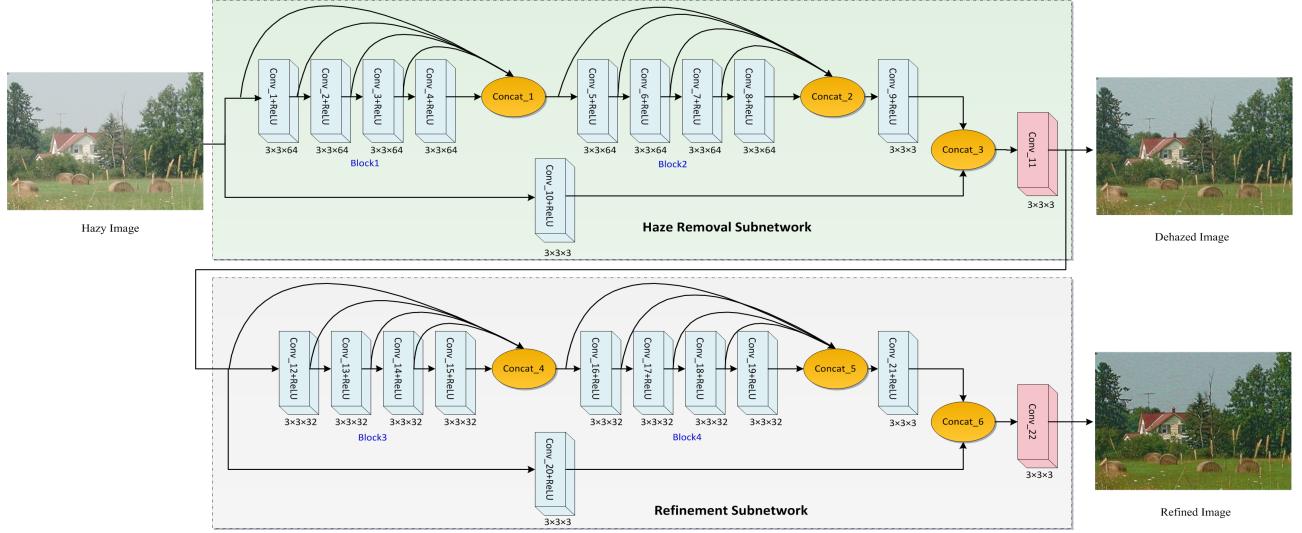


Fig. 2. An overview of the PDR-Net architecture. PDR-Net consists of two subnetworks: haze removal subnetwork and refinement subnetwork. Different color blocks represent different operations. “Conv” is the convolutional layer; “Concat” is the stacked convolutional layers; “ReLU” is the rectified linear unit.

A. Problem Formulation

According to the atmospheric scattering model [42], hazy image formation can be described as

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

where x denotes the pixel coordinates, I is the observed image, J is the haze-free latent image, A is the global atmospheric light which indicates the intensity of ambient light, and $t(x) \in [0, 1]$ is the transmission. When the haze is homogenous, $t(x)$ can be further expressed in an exponential decay term as

$$t(x) = \exp(-\beta d(x)), \quad (2)$$

where β is the atmospheric attenuation coefficient and $d(x)$ is the distance from the scene to the camera.

B. PDR-Net Architecture

An overview of the PDR-Net is shown in Figure 2. Our PDR-Net consists of two subnetworks: haze removal and refinement. The input to haze removal subnetwork is a hazy image while the input to refinement subnetwork is the dehazed result. In what follows, we explain its basic building blocks.

1) Dense Concatenation: These two subnetworks share the same basic structure that makes full use of the multi-level features, which is inspired by the densely connected network (*i.e.*, DenseNet) [43]. Different from DenseNet, we only stack input and all the convolutional layers at the end of each block in each subnetwork (each subnetwork has two stacking blocks), which reduces the computational burden.

2) Network Layers: The PDR-Net architecture consists of three different layers indicated by different color. The first type of the layer is convolutional layer represented by “Conv”. The parameters of the convolutional layer are denoted as “kernel size \times output feature maps”. Similar to the transition layer used in the DenseNet [43] which refines multi-level features from multiple dense blocks and reduces the dimension of features, the introduction of “Conv_9” in our network also

has such purpose. Such a transition layer has been widely employed in the CNNs [43], [44]. Besides, we found the dehazed result usually contains the amplified noise and artifacts which are induced by directly passing the input hazy image to the end of the network. Thus, we empirically added the “Conv_10” to map the input image to the feature space. By doing this, we hope that the input hazy image is not directly used at the end of the network. Since there is no hazy image dataset with ground truth collected for denoising or artifacts removal purpose, it is difficult to quantitatively demonstrate the potential reason for using the “Conv_10” in our network. The last convolutional layer “Conv_11” reconstructs the latent result. Different from the haze removal subnetwork, the refinement subnetwork has fewer numbers of output feature maps. Though the PDR-Net includes two subnetworks, the entire PDR-Net only consists of 22 convolutional layers. Accordingly, our PDR-Net is a lightweight CNN, which makes it computationally inexpensive and highly efficient in training and inference. The second type is the “ReLU” activation layer, which introduces the nonlinearity. The third type is concatenation layer represented by “Concat”, which is used to accumulate the multi-level features.

3) Reducing Boundary Artifacts: To ensure that the size of output is equal to the input, we enforce two strategies (1) we do not use any pooling layers in our network, and (2) we pad zeros before each convolutional layer.

C. Loss function

1) Haze removal subnetwork loss function: Instead of using pixel-wise loss (*e.g.*, ℓ_2) as loss function, we introduce perceptual loss [45] that measures high-level perceptual and semantic differences between the reconstructed image and the ground truth image to our haze removal subnetwork optimization. Perception-inspired learning has two advantages:

- it reduces the impact of artifacts induced by pixel-wise optimization and improves the subjectively visual quality of the dehazed results;

- it avoids the use of time-consuming guided filter which is employed to remove artifacts of estimated transmission.

Following [45], we define the perceptual loss based on the ReLU activation layers of the pre-trained 19 layers VGG network [46]. Let $\phi_j(x)$ be the j th convolutional layer (after activation) of the VGG19 network ϕ pretrained on the ImageNet dataset [47]. We define the perceptual loss as the distance between the feature representations of a dehazed image \hat{y} and the ground truth image y :

$$L_{pl}^{\phi,j} = \frac{1}{C_j H_j W_j} \sum_{i=1}^M \|\phi_j(\hat{y}_i) - \phi_j(y_i)\|, \quad (3)$$

where M is the number of each batch; C_j , H_j , and W_j represent the number, height, and width of feature maps.

2) **Refinement subnetwork loss function:** The haze removal subnetwork can yield a clear image; however, this dehazed output may have unpleasing color range, contrast, and details in some cases, especially for real-world hazy images. The reason is that the haze removal subnetwork is trained using the hazy images synthesized by indoor images. Indoor images usually have low brightness and contrast, and monotonous colors such as NYU-V2 dataset [48]. To remedy this problem, we design a refinement subnetwork to enhance the visual quality of dehazed result by joint optimizing a multi-term loss.

Specifically, the content loss uses the perceptual loss defined in the haze removal subnetwork, which measures the distance between the feature representations of the refined image \hat{z} and the target image z , and can be expressed as

$$L_{con} = \frac{1}{C_j H_j W_j} \sum_{i=1}^N \|\phi_j(\hat{z}_i) - \phi_j(z_i)\|, \quad (4)$$

where N is the number of each batch; $\phi_j(x)$ is the j th convolutional layer of the VGG19 network ϕ .

For the color loss, we compute the color difference between the refined image and target image by 2D Gaussian blurring the images content, and then measure the ℓ_2 distance as

$$L_{col} = \sum_{i=1}^N \|\hat{z}_{bi} - z_{bi}\|^2, \quad (5)$$

where \hat{z}_b and z_b are the blurred \hat{z} and z , respectively.

Inspired by [49], [50], we introduce the GAN to measure the contrast difference between the refined image and target image. We first transfer the inputs to the grayscale images since we only focus on the contrast. For the generator function G (*i.e.*, our refinement subnetwork) and the discriminator D , we minimize the contrast loss as

$$L_{contr} = - \sum_{i=1}^N \log(D(\hat{z}_g, z_g)), \quad (6)$$

where \hat{z}_g and z_g are the grayscale images of \hat{z} and z . G tries to generate image \hat{z} that looks similar to target image z in terms of contrast while D aims to distinguish them.

The total loss for the refinement subnetwork can be expressed as

$$L_{total}^{re} = W_{con} \cdot L_{con} + W_{col} \cdot L_{col} + W_{contr} \cdot L_{contr}. \quad (7)$$

The W_{con} , W_{col} , and W_{contr} are the weights of the corresponding losses. We will show the contributions of each loss in Section IV.

D. Training and Implementation Details

1) **Training Dataset:** To train the haze removal subnetwork, we first synthesize a hazy image dataset according to Eqs. (1) and (2) using indoor data from NYU-V2 dataset [48]. Here, we assume that each channel of an image has the same global atmospheric light and transmission values. Then, we randomly select global atmospheric light A from [0.7, 1.0] and set the atmospheric attenuation coefficient β varying from 0.6 to 2.8. Such parameter settings can generate divers hazy images with different haze concentrations and light intensities.

To be specific, we divide NYU-V2 dataset into two parts: the training part with 1000 images and the validation (test) part with 449 images. For each RGB-D image, we randomly select 5 global atmospheric light and atmospheric attenuation coefficient values to synthesize 5 hazy images. We obtain a training set including 5000 training samples and a validation (test) set including 2245 samples. We resize these samples to size $207 \times 154 \times 3$. We do not train our model using the D-HAZY dataset [51] because the number of samples in this dataset is relatively insufficient. We do not train our model using the RESIDE dataset [52] since most hazy images in it have light haze, which leads to the suboptimal dehazing performance for the images with the heavy haze.

For training the refinement subnetwork, we use the synthetic outdoor hazy image dataset (*i.e.*, the SOTS subset in RESIDE dataset [52]) which includes 500 pairs of outdoor images. We divide these images into two parts: the training part with 400 images and the validation (test) part with 100 images. The target images have vivid colors and good contrast and details, which is suitable for our refinement subnetwork training. Specifically, the inputs of refinement subnetwork are the dehazed results of our haze removal subnetwork for the outdoor hazy images in the SOTS, while the target images are the corresponding clear outdoor images. Since we only consider the local contrast, details, and color of the refined results, we use image patches with size $100 \times 100 \times 3$ obtained by non-overlapped cropping to train our refinement subnetwork. Finally, we obtain a training set including 8000 image patches and a validation (test) set including 1000 image patches.

To stabilize our training, we adopt a stage-wise learning scheme, which separately optimizes two subnetworks.

2) **Implementation:** PDR-Net was implemented on a computer with Nvidia GTX 1080Ti GPU and Intel I7 6700 CPU using the TensorFlow framework. We separately trained these two subnetworks using ADAM. We assigned the learning rate to 0.0001 and fixed the learning rate in the entire training procedure. The batch size for the haze removal subnetwork was 16 while it was 64 for the refinement subnetwork. We computed perceptual loss at layer `relu5_4` of the VGG19 network and utilized the same network architecture with [50] for the discriminator. The amplitude, mean, and variance for the 2D Gaussian filter are set to 0.053, 0, and 3, respectively. The weights W_{con} , W_{col} , and W_{contr} are set to 10, 0.5, and 1 based on our training data, respectively.

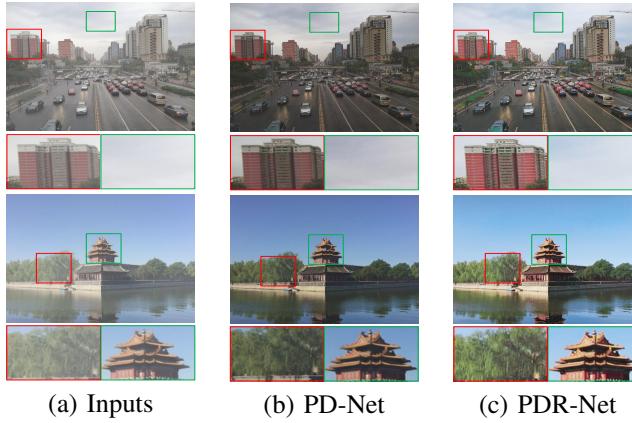


Fig. 3. Visual results after haze removal subnetwork (*i.e.*, PD-Net) and refinement subnetwork (*i.e.*, PDR-Net). Red and green boxes indicate the obvious differences and amplified details.

IV. EXPERIMENTS

To evaluate the proposed PDR-Net, we use both real and synthetic data and compare with several state-of-the-art single image dehazing methods. We mainly compare our method with deep learning based methods such as DehazeNet (TIP'16) [17], MSCNN (ECCV'16) [24], All-in-One (ICCV'17) [5], ProximalNet (ECCV'18) [32], and cGAN-based method (CVPR'18) [10], but also present the results of prior based methods such as DCP (TPAMI'11) [18] and CAP (TIP'15) [23], and a regularization-based method (ICCV'13) [19] as baselines. We set the gamma values of MSCNN [24] to 1 when the accurate gamma values are unknown. To demonstrate the improvement introduced by the proposed refinement subnetwork, we first present the step-wise results of our method in Figure 3.

In Figure 3, our haze removal subnetwork effectively removes the haze; however, the results are not bright or vivid due to the low contrast and monotonous colors of indoor training data. Besides, similarly to [5] and [24] trained by indoor data, our haze removal subnetwork also tends to over enhance the regions of the cloud (*e.g.*, the cloud in the green box in Figure 3(b)). These methods usually treat the cloud as the haze since the indoor training data do not include the real-world sky and cloud. By contrast, after processing by our refinement subnetwork, the refined results improve the brightness and contrast and have vivid color (*e.g.*, the building and tree in the red boxes in Figure 3(c)). Besides, our refinement subnetwork can correct the over-enhanced sky regions, which makes the results realistic and vivid. For the limited space, we only show our final result (*i.e.*, **PDR-Net**) in the following experiments.

A. Experiments on Real Data

We collect several challenging hazy images, including the hazy image with heavy haze and low-light, the hazy image with medium haze, the hazy image with light haze, the haze-free image, and the hazy image with various scene depths, and compare with other methods on these images in Figures 4-8.

Hazy images with heavy haze and low-light are significantly challenging for haze removal methods since the priors are

usually ineffective and the learning based methods do not consider the training sample like this. In Figure 4, the methods such as CAP [23], DehazeNet [17], and MSCNN [24] leave the haze on the results while the All-in-One [5] method makes the result darker. The methods such as DCP [18], Regularization-based [19], and ProximalNet [32] can remove the haze; however, they do not work well for improving the brightness and color of the hazy image. Compared with other methods, the results of cGAN-based [10] method and the proposed method have less haze and obtain better color.

For the hazy images with medium or light haze as shown in Figures 5 and 6, the methods such as DCP [18] and Regularization-based [19] are sensitive to the sky regions, which results in artifacts (*e.g.*, the results in Figure 5(b) and (c) and Figure 6(b) and (c)). All-in-One [5] method tends to decrease the brightness and contrast of the results (*e.g.*, the details in the green boxes in Figures 5(g) and 6(g)). In Figure 6(i), the cGAN-based [10] method blurs the details of the result (*e.g.*, the tower). In comparison, our results have clear structure and vivid color (*e.g.*, the regions in the red and green boxes in Figures 5(j) and 6(j)).

It is expected that image dehazing methods should have less effect on a haze-free image. In Figure 7, our method causes a little negative impact on the haze-free image. The compared methods produce over-enhancement for the haze-free image, particularly apparent in the results of DCP [18], Regularization-based [19], and MSCNN [24] methods, or degrades the color of the haze-free image such as the result of the cGAN-based [10] method shown in Figure 7(i).

For the hazy image with various scene depths shown in Figure 8(a), the methods such as DCP [18], Regularization-based [19], MSCNN [24], and ProximalNet [32] which estimate transmission usually obtain good performance, because the estimated transmission is related to the scene depth. Though the proposed method does not consist of a transmission estimation network, it still has good performance, even for some regions which are far away from the camera.

In summary, although our method is trained by synthesized data, it generalizes well to real-world hazy images even captured in challenging scenes. This benefits from the advantages of indoor and outdoor training data.

B. Experiments on Synthetic Data

We carry out experiments on two synthetic datasets:

- our validation (test) set for haze removal subnetwork including 2245 hazy images, denoted as **Test set 1**;
- our validation (test) set for refinement subnetwork including 1000 image patches denoted as **Test set 2**.

We report the qualitative and quantitative results on the test sets. Ground truth is denoted as **GT**. We use Peak Signal-to-Noise Ratio (PSNR, dB) and Structural Similarity (SSIM) [53] for quantitative evaluations. A higher SSIM value indicates a result that is closer to the ground truth in terms of structural properties. A higher PSNR value indicates the similarity in terms of pixel-wise values. The best result is in bold, whereas the second one is underlined. Figure 9 first shows the visual comparisons for the synthesized hazy image from **Test set 1**.

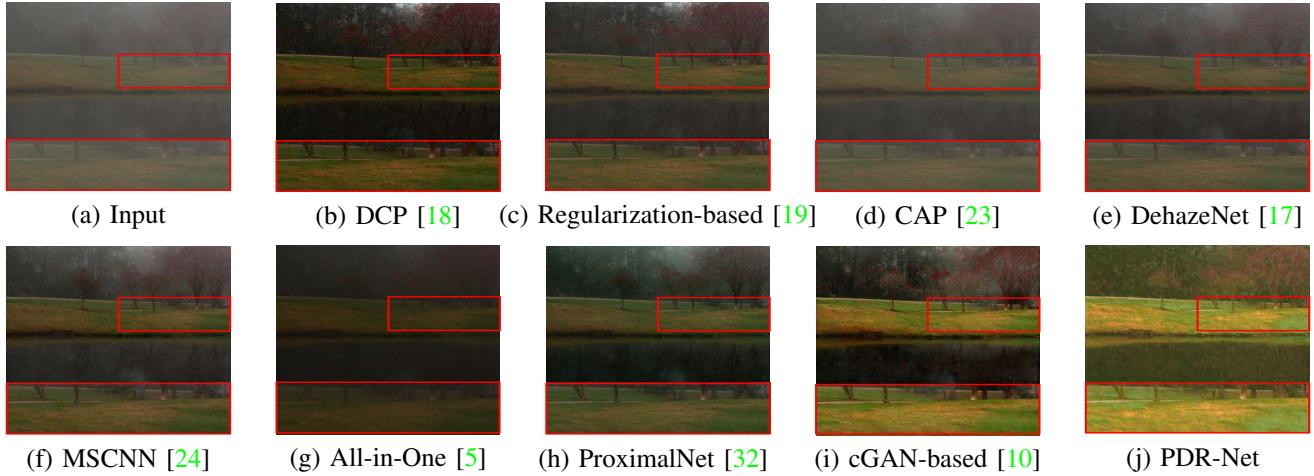


Fig. 4. Visual results on the hazy image with heavy haze and low-light. From (a) to (j) are the input hazy image and the results of DCP [18], Regularization-based [19], CAP [23], DehazeNet [17], MSCNN [24], All-in-One [5], ProximalNet [32], cGAN-based [10], and our PDR-Net.

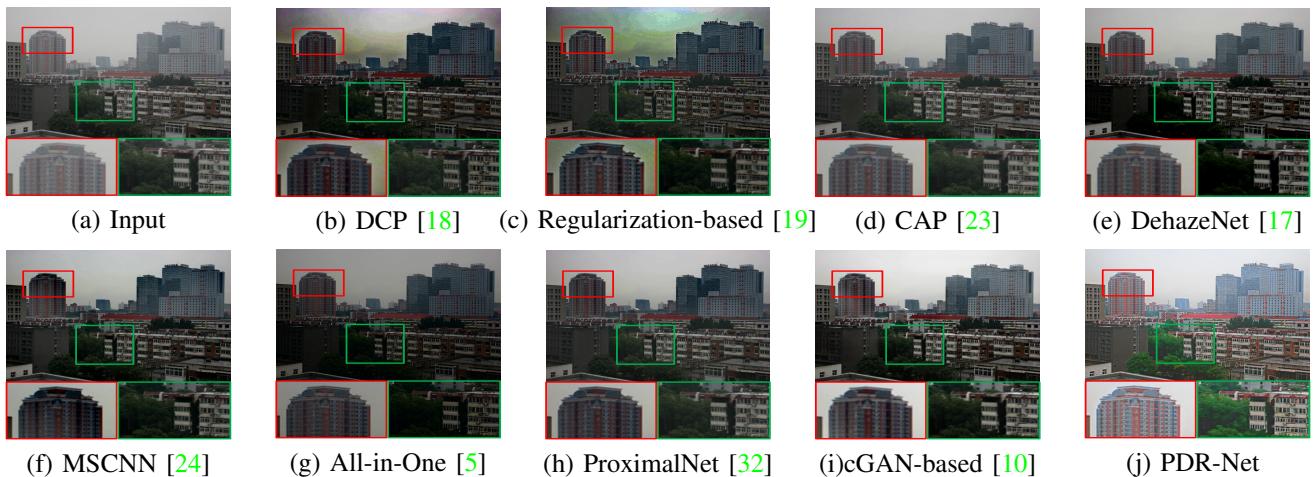


Fig. 5. Visual results on the hazy image with medium haze. From (a) to (j) are the input hazy image and the results of DCP [18], Regularization-based [19], CAP [23], DehazeNet [17], MSCNN [24], All-in-One [5], ProximalNet [32], cGAN-based [10], and our PDR-Net.

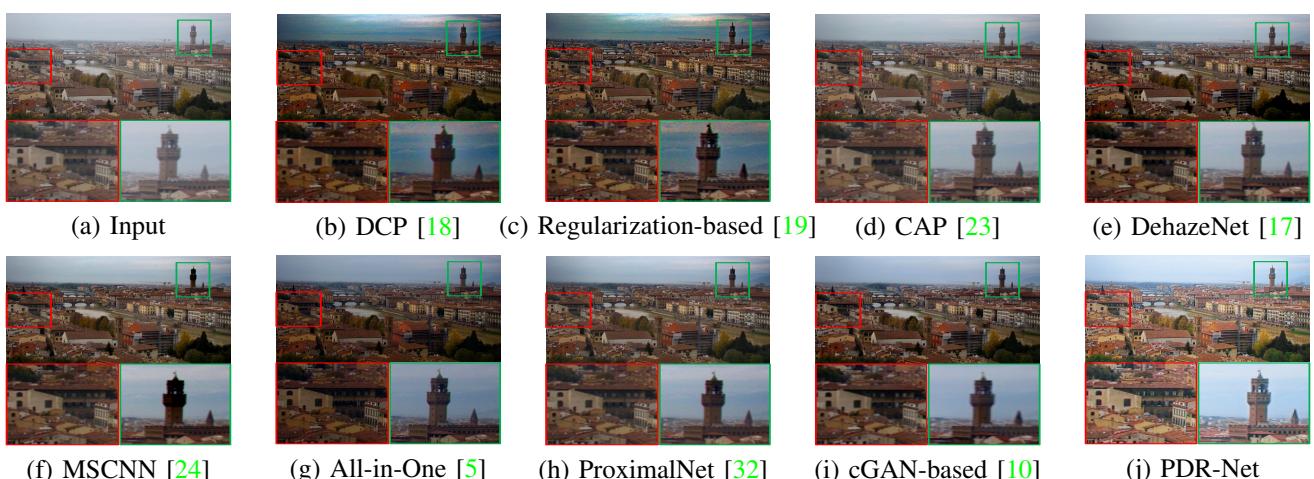


Fig. 6. Visual results on the hazy image with light haze. From (a) to (j) are the input hazy image and the results of DCP [18], Regularization-based [19], CAP [23], DehazeNet [17], MSCNN [24], All-in-One [5], ProximalNet [32], cGAN-based [10], and our PDR-Net.

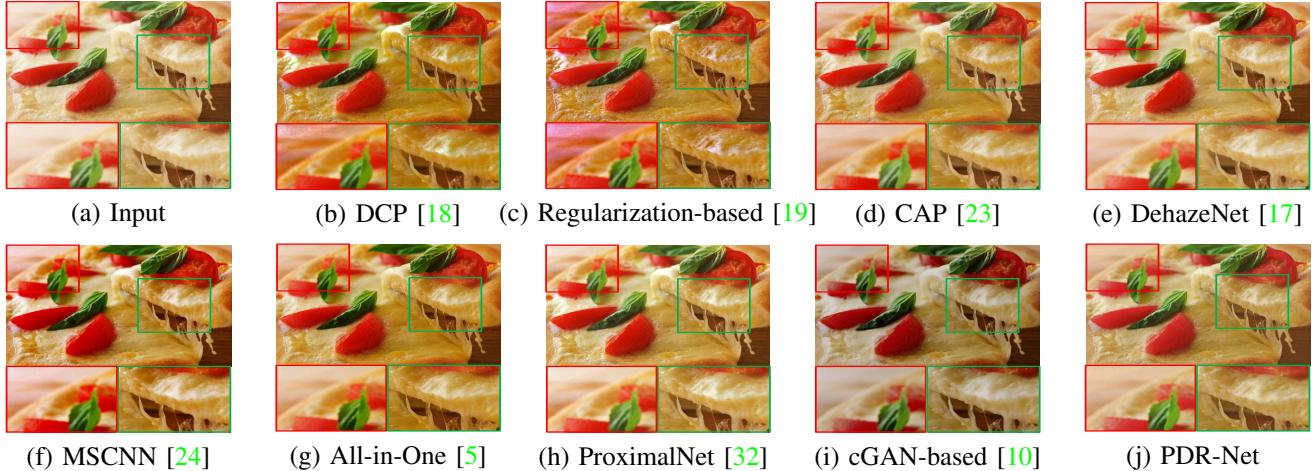


Fig. 7. Visual results on the haze-free image. From (a) to (j) are the input hazy image and the results of DCP [18], Regularization-based [19], CAP [23], DehazeNet [17], MSCNN [24], All-in-One [5], ProximalNet [32], cGAN-based [10], and our PDR-Net.

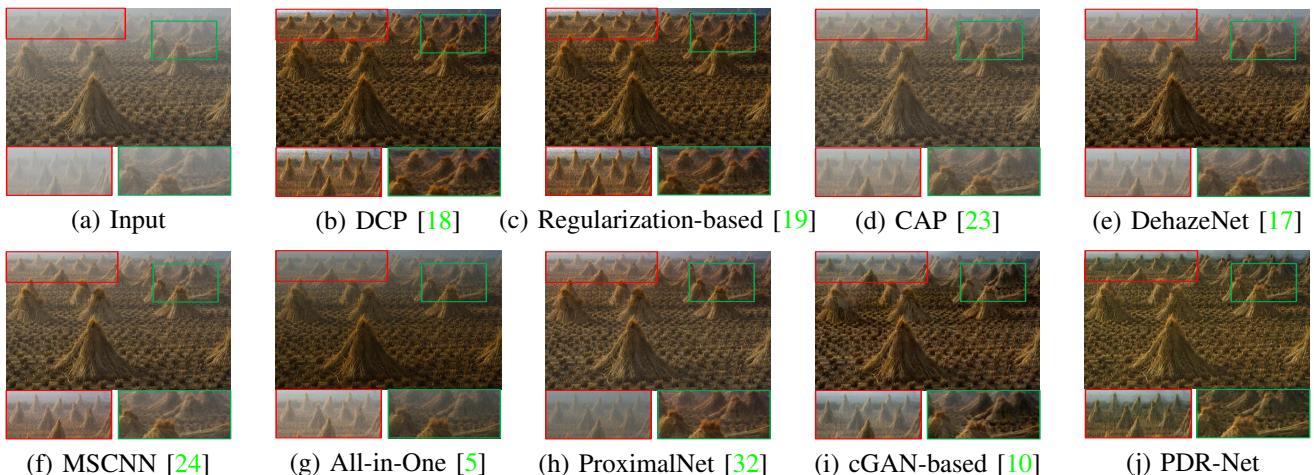


Fig. 8. Visual results on hazy image with various scene depths. From (a) to (j) are the input hazy image and the results of DCP [18], Regularization-based [19], CAP [23], DehazeNet [17], MSCNN [24], All-in-One [5], ProximalNet [32], cGAN-based [10], and our PDR-Net.

In Figure 9, the compared methods leave the haze on the results (*e.g.*, the region in the green box) and tend to introduce color deviation (*e.g.*, the color of the floor in the red box). In comparison, our method is closer to the GT. In Table I, we show the average values of PSNR and SSIM on **Test set 1**.

As visible in Table I, our method achieves the best PSNR and SSIM values. The reasons are 1) our haze removal subnetwork directly minimizes the reconstruction loss, which produces better pixel-wise similarity (*i.e.*, PSNR), and 2) our refinement subnetwork further improves the contrast and details, which leads to better structure similarity (*i.e.*, SSIM). By contrast, except the cGAN-based [10], other methods do not generalize well to the **Test set 1**. We further perform the comparisons on **Test set 2**.

In Figure 10, our method removes the effect of hazy while the methods such as DCP [18], CAP [23], MSCNN [24], All-in-One [5], and ProximalNet [32] introduce over-enhancement and even change the color of original image (*e.g.*, the region of the sky). The results of cGAN-based [10] method and our

method are closest to the GT (*e.g.*, the region of the ground). In Table II, our method does not obtain the best performance, because the hazy images in **Test set 2** only have the light haze. However, our result looks visually pleasing because of the perceptual-driven learning. Note that our training data for hazy removal subnetwork are totally different from **Test set 2** (we take different haze concentrations into account) while the training data of some compared methods (*e.g.*, MSCNN [24] and All-in-One [5]) are similar with the hazy images in **Test set 2**. However, such training data only with light haze show limitations in the real-world hazy images, such as the results in Figures 4-8. In fact, the scores of quantitative comparisons depend on the training data. There are gaps between synthetic and real-world images so that high quantitative scores do not always bring in visually pleasing results. Thus, we conduct a user study on real-world hazy images in the next section.

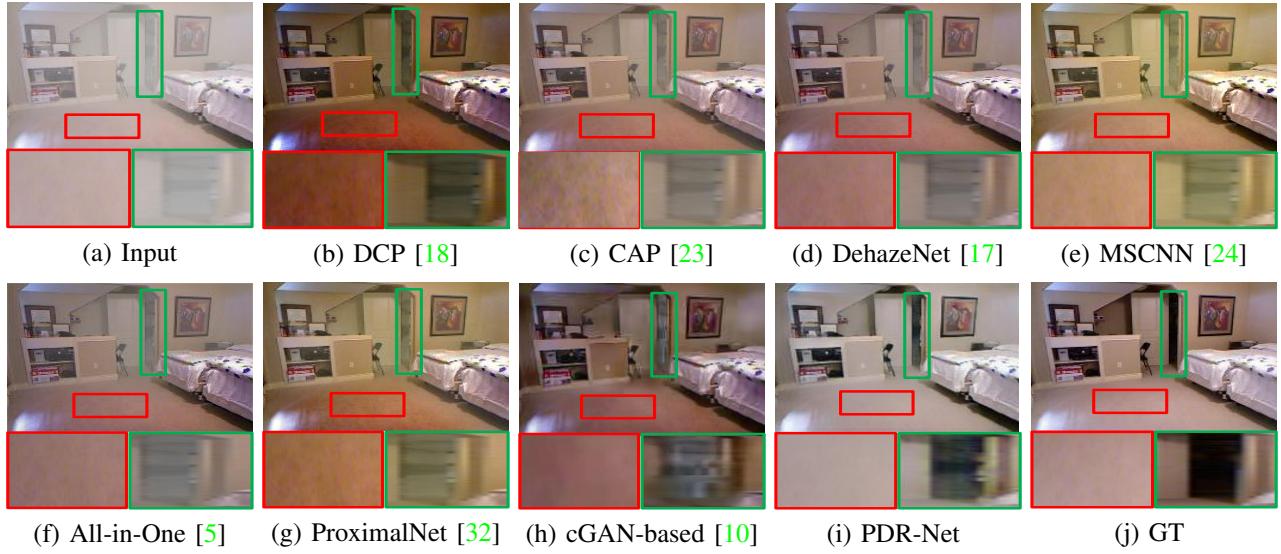


Fig. 9. Results on synthesized hazy image sampled from **Test set 1**. From (a) to (j) are the synthetic input hazy image, the results of DCP [18], CAP [23], DehazeNet [17], MSCNN [24], All-in-One [5], ProximalNet [32], cGAN-based [10], our PDR-Net, and the GT.

TABLE I
QUANTITATIVE EVALUATIONS ON **TEST SET 1**.

Metrics	DCP [18]	CAP [23]	DehazeNet [17]	MSCNN [24]	All-in-One [5]	ProximalNet [32]	cGAN-based [10]	PDR-Net
PSNR	16.71	16.38	17.35	16.66	16.83	17.03	17.79	17.89
SSIM	0.7813	0.7622	0.7812	0.7755	0.7829	0.7711	0.7993	0.8227

TABLE II
QUANTITATIVE EVALUATIONS ON **TEST SET 2**.

Metrics	DCP [18]	CAP [23]	DehazeNet [17]	MSCNN [24]	All-in-One [5]	ProximalNet [32]	cGAN-based [10]	PDR-Net
PSNR	17.34	22.08	22.62	19.47	20.47	19.16	22.01	21.70
SSIM	0.7359	0.8812	0.8677	0.8381	0.9121	0.8515	0.8477	0.8847

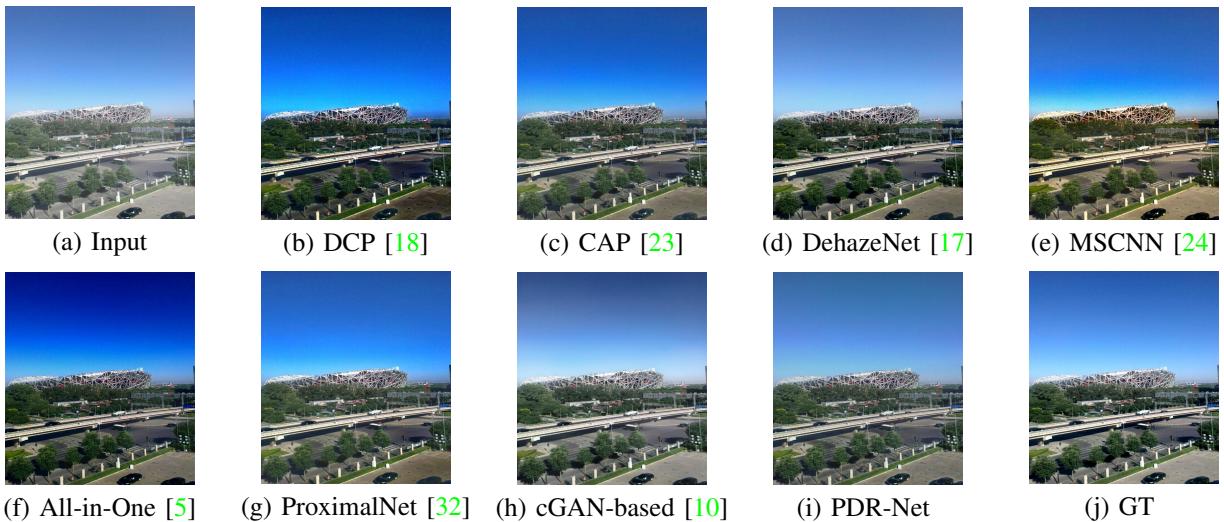


Fig. 10. Results on synthesized hazy image sampled from the testing set of the SOTS in RESIDE dataset [52]. From (a) to (j) are the synthetic input hazy image, the results of DCP [18], CAP [23], DehazeNet [17], MSCNN [24], All-in-One [5], ProximalNet [32], cGAN-based [10], our PDR-Net, and the GT.

C. User Study

We conduct a user study to provide realistic feedback and quantify the subjective visual quality. We collect 40 real-world hazy images from the related papers. Some corresponding results have been presented in Figures 4-8.

For this user study, we randomize the order the results of different methods, and then display them on a screen to human subjects. We invite 10 participants with image processing and computer vision expertise. Each subject ranks the results based on the perceived visual quality from 1 to 5 where 1 is the worst quality and 5 is the best quality (the step size is 1). One expects that the results with high contrast, good visibility, natural color, and authentic texture should receive higher ranks while the results with over-enhancement/exposure, under-enhancement/exposure, color casts, and artifacts should have lower ranks. The average subjective scores are given in Table III. The proposed method receives the best ranking, which indicates that our method can produce better performance on real-world hazy images from the subjective visual perspective.

TABLE IV

AVERAGE RUNNING TIME (IN SECONDS) AND FLOPs FOR AN IMAGE WITH SIZE $640 \times 480 \times 3$.

Method	Running Time (s)	Platform	FLOPs
DCP [18]	17.211	Matlab	-
Regularization-based [19]	1.564	Matlab	-
CAP [23]	0.808	Matlab	-
DehazeNet [17]	1.767	Matlab	$5.1e^9$
MSCNN [24]	1.675	Matlab	$4.8e^9$
All-in-One [5]	0.018	Pycaffe	$1.1e^9$
ProximalNet [32]	0.058	MatConvNet	$3.0e^{10}$
cGAN-based [10]	0.061	Torch7	$1.9e^{13}$
PDR-Net	0.055	TensorFlow	$3.0e^{11}$

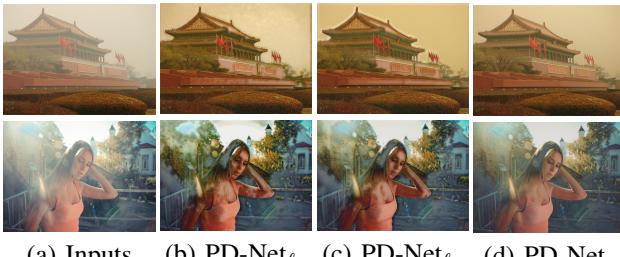


Fig. 11. Results of our haze removal subnetwork trained by different loss functions.

D. Running Time

We compare the running time (s) of different methods on a PC with above-mentioned configuration. The average running time is computed by processing 32 images with size $640 \times 480 \times 3$. Besides, for an image with size $640 \times 480 \times 3$, we compute the number of floating-point operations (FLOPs) of the deep learning-based methods. For the deep learning-based methods which include conventional algorithms, we only compute the FLOPs of their network parts.

In Table IV, the DehazeNet [17], MSCNN [24], and ProximalNet [32] methods which estimate the parameters of models

usually have the fewer FLOPs, while the cGAN-based [10] method has more FLOPs. In addition, the All-in-One [5] method ranks first for the running time and the FLOPs thanks to the all-in-one structure. The proposed method is faster than the latest cGAN-based [10] method.

E. Investigation of the Impact of Loss Function on the End-to-End Dehazing Network

Deep learning based image dehazing methods usually choose pixel-wise loss functions to optimize their networks. The pixel-wise loss functions are effective for single channel transmission estimation; however, they show limitations in the end-to-end system. To analyze the impact of the loss functions on the end-to-end image dehazing network, we fix the default parameter settings of our haze removal subnetwork, and then re-train it using different loss functions. We take the haze removal subnetwork trained by ℓ_1 and ℓ_2 as baselines, and denote them as PD-Net_{ℓ_1} and PD-Net_{ℓ_2} . An example of the results is shown in Figure 11. Besides, we present the quantitative scores in Table V.

TABLE V
QUANTITATIVE EVALUATIONS ON TEST SET 1 FOR DIFFERENT LOSS FUNCTIONS.

Loss	PSNR (dB)	SSIM
PD-Net $_{\ell_1}$	21.47	0.7013
PD-Net $_{\ell_2}$	19.72	0.6579
PD-Net	18.09	0.7997

In Figure 11, the PD-Net_{ℓ_1} and PD-Net_{ℓ_2} produce visible splotchy artifacts, especially on flat regions. The reason is that the ℓ_2 penalizes larger errors, but it is more tolerant to small errors. Although the ℓ_1 does not over-penalize larger errors, it also produces the sub-optimal dehazing performance. In comparison, the perceptual loss measures image similarities more robustly. In Table V, the perceptual loss encourages the output image to be perceptually similar to the ground truth but does not force them to match exactly, which produces visually pleasing dehazed results at the cost of the PSNR value. Compared to the slight decline of the PSNR value, the visual quality is more significant, especially for multimedia content sharing. Besides, the perceptual loss increases the similarity of image structure when compared with the ℓ_1 and ℓ_2 .

F. Ablation Study on the Contribution of Each Loss in the Refinement Subnetwork

To demonstrate the contribution obtained by each loss used in the proposed refinement subnetwork, we perform an ablation study involving the following three experiments:

- the proposed refinement subnetwork without content loss (**PDR-Net w/o contl**)
- the proposed refinement subnetwork without texture loss (**PDR-Net w/o textl**)
- the proposed refinement subnetwork without color loss (**PDR-Net w/o colol**)

In these experiments, we use the same parameters settings with the original refinement subnetwork. Here, the input

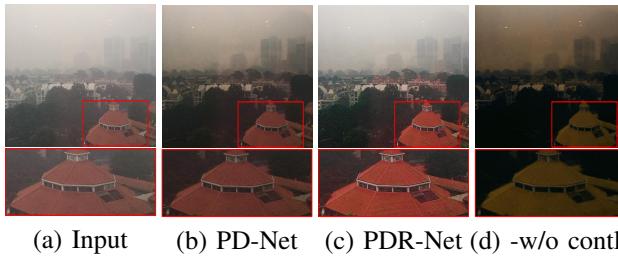
TABLE III
USER STUDY ON REAL-WORLD HAZY IMAGE DATASET.

	DCP [18]	CAP [23]	DehazeNet [17]	MSCNN [24]	All-in-One [5]	ProximalNet [32]	cGAN-based [10]	PDR-Net
Scores	3.54	3.76	3.88	4.02	3.85	3.97	4.21	4.39

TABLE VI

QUANTITATIVE EVALUATIONS ON **TEST SET 2** FOR ABLATION STUDY ON THE CONTRIBUTION OF EACH LOSS IN THE REFINEMENT SUBNETWORK.

Method	PSNR (dB)	SSIM
PDR-Net	21.70	0.8847
PDR-Net w/o contl	14.42	0.5771
PDR-Net w/o textl	21.22	0.8449
PDR-Net w/o colol	21.03	0.8708



(a) Input (b) PD-Net (c) PDR-Net (d) -w/o contl

Fig. 12. Subjective results of the ablation study on the contribution of each loss. From (a) to (e) are the real-world hazy image, the results of our haze removal subnetwork, our original refinement subnetwork, and our refinement subnetwork without content loss. Red boxes indicate the amplified details.

of the refinement subnetwork is the dehazed result by our haze removal subnetwork. Thus, the results of the refinement subnetwork are also the final results of our method (*i.e.*, PDR-Net). The quantitative evaluation is performed on **Test set 2**. As shown in Table VI, we observed that 1) the quantitative values significantly decrease when we remove the content loss, which demonstrates that the content loss determines the performance of our refinement subnetwork; and 2) removing the texture and color losses, the performance of our refinement subnetwork slightly decreases, which indicates these two losses also make contributions to our refinement subnetwork. Since the slight contributions of the texture and color losses cannot be accurately reflected in the color image format, we only present a subjective example to explain the contribution of the content loss in Figure 12.

Compared with the result of original refinement subnetwork shown in Figure 12(c), the result of PDR-Net w/o contl loses the color and details information of the original image, which indicates the importance of the content loss in our refinement subnetwork. In summary, the quantitative and qualitative evaluations demonstrate the importance of content loss and also indicate the improvement introduced by the combination of content loss, texture loss, and color loss.

G. Ablation Study on the Effectiveness of the Refinement Subnetwork

To demonstrate the effectiveness of the proposed refinement subnetwork, we perform an ablation study involving the following four experiments:

- the proposed haze removal subnetwork is trained by the loss function of refinement subnetwork (*i.e.*, Eq. (7)), denoted as **PD-Net-RefineLoss**
- the proposed refinement subnetwork is trained by the loss function of haze removal subnetwork (*i.e.*, Eq. (3)), denoted as **PDR-Net-DehazeLoss**
- the haze removal subnetwork including four blocks, denoted as **PD-Net-4Blocks**
- the haze removal subnetwork including four blocks is trained by the blending data of indoor and outdoor training data mentioned in Section III-D, denoted as **PD-Net-4Blocks-Blend**

The purpose of the first experiment is to demonstrate the importance of outdoor training data used in our refinement subnetwork. The second experiment is to demonstrate the importance of loss function used in our refinement subnetwork. The third experiment is to study the effect of the number of network parameters on the haze removal subnetwork. The fourth experiment is to demonstrate the importance of two-stage training. In these experiments, we use the same parameters settings, loss function, and training data mentioned in Section III unless special explanations. Note that, the input of haze removal subnetwork is the hazy image while the input of refinement subnetwork is the dehazed result by our original haze removal subnetwork. The proposed refinement subnetwork is to further improve the visual quality of the results of our haze removal subnetwork. Thus, we mainly present the subjective results in Figure 13 and also report the quantitative results on **Test set 2** in Table VII.

In comparison, the visual quality of the results of PD-Net-RefineLoss, PD-Net-4Blocks, and PD-Net falls behind the results of the PDR-Net-DehazeLoss and our PDR-Net as shown in Figure 13(f) and (g). This is because of the indoor training data with low contrast and monotonous color. Such a result implies the importance of outdoor training data. In addition, the results of PDR-Net have better color and brightness than the results of PDR-Net-DehazeLoss as shown in Figure 13(f), which indicates the effectiveness of the multi-term loss used in our refinement subnetwork. As shown in Figure 13(c) and (e), the results of PD-Net-4Blocks and PD-Net have similar visual quality, which indicates that enlarging the number of haze removal subnetwork parameters cannot significantly improve the dehazing performance. At last, as shown in Figure 13(d), one-stage training by enlarging network parameters and using the blending training data cannot produce comparable visual quality to the two-stage training (*i.e.*, the results in Figure 13(g)), which indicates the indispensability of two-stage training and refinement subnetwork.

In Table VII, our PDR-Net achieves the best quantitative results on **Test Set 2**. The results of the PDR-Net-DehazeLoss are a little lower than the results of our PDR-Net, which demonstrates the effectiveness of the multi-term loss used

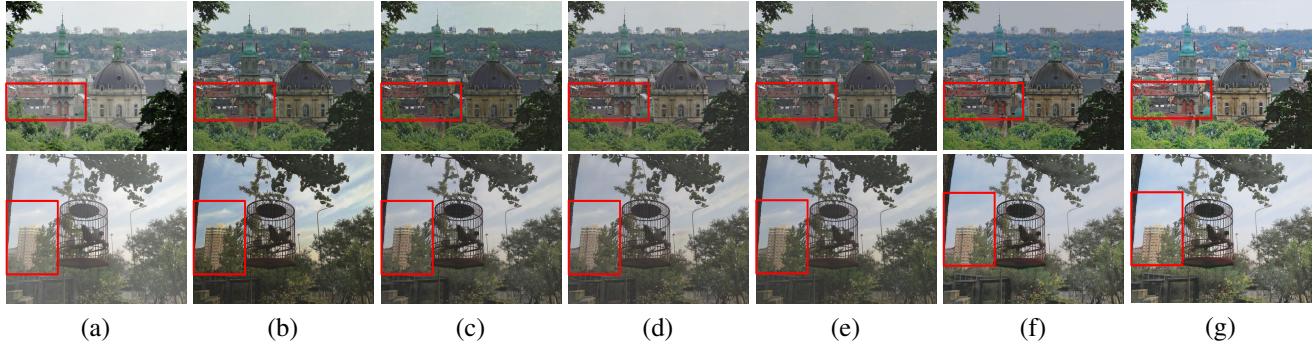


Fig. 13. Subjective results of the ablation study on the effectiveness of the refinement subnetwork. From (a) to (g) are the hazy images, PD-Net-RefineLoss, PD-Net-4Blocks, PD-Net-4Blocks-Blend, PD-Net, PDR-Net-DehazeLoss, and PDR-Net.

TABLE VIII
THE MAP VALUES OF ALL DETECTION RESULTS ON RTTS (IN %).

	Haze	DCP [18]	CAP [23]	DehazeNet [17]	MSCNN [24]	All-in-One [5]	ProximalNet [32]	cGAN-based [10]	PDR-Net
mAP	44.72	45.00	45.40	44.48	46.60	44.44	46.72	45.52	46.67

TABLE VII

QUANTITATIVE EVALUATIONS ON TEST SET 2 FOR ABLATION STUDY ON THE EFFECTIVENESS OF THE REFINEMENT SUBNETWORK.

Method	PSNR (dB)	SSIM
PD-Net-RefineLoss	21.47	0.8538
PD-Net-4Blocks	21.58	0.8503
PD-Net-4Blocks-Blend	21.62	0.8511
PD-Net	21.50	0.8491
PDR-Net-DehazeLoss	21.66	0.8559
PDR-Net	21.70	0.8847

in our refinement subnetwork. The quantitative comparisons among the PD-Net-RefineLoss, PD-Net-4Blocks, PD-Net-4Blocks-Blend, and PDR-Net indicate the importance of the outdoor training data and the two-stage training strategy.

In summary, 1) the haze removal subnetwork cannot significantly improve the color and contrast of the dehazed results in spite of using the multi-term loss optimization and enlarging the number of network parameters; 2) the multi-term loss designed for outdoor training data is important for the refinement subnetwork; 3) the two-stage training strategy can better improve the visual quality of final result than the one-stage blending data training, and 4) the good performance of our method benefits from the combination of the advantages of indoor and outdoor training data and well-designed losses.

H. Experiments on Application

Hazy images affect the performance of computer vision applications since the deep networks usually are trained by high-quality images or the algorithms assume the inputs are clear images. To illustrate the potential usage of our PDR-Net in practical applications, we employ it on object detection and recognition. Specifically, we use the fast R-CNN [54] trained on PASCAL VOC 2007 dataset [55] as the detector (model: VGG16, batch size: 4, max_epoch: 9)¹ and feed the dehazed results of different methods to it. Here, we directly

use the codes and models of the compared methods to produce dehazed results. We compare the mean average precision (mAP) values of different methods on the RTTS subset in RESIDE dataset [52] which consists of 4,322 real-world hazy images with annotated bounding boxes for five categories (*i.e.*, person, bicycle, car, bus, and motorbike). The mAP values of all detection results on RTTS are reported in Table VIII.

In Table VIII, the methods such as DCP [18], CAP [23], MSCNN [24], ProximalNet [32], cGAN-based [10], and our PDR-Net increase the mAP values to different extent. In comparison, the DehazeNet [17] and All-in-One [5] methods decrease the mAP values when compared with the result of the input hazy images. Furthermore, the results of the ProximalNet [32] method achieve the best mAP value (*i.e.*, 46.72%) while our results achieve the second best mAP value (*i.e.*, 46.67%) which is a little lower than the best one. Additionally, our network is light-weight and runs faster than the ProximalNet [32]. In short, experiments on application provide additional evidence for the potential usage of our PDR-Net.

V. CONCLUSION

We have presented a deep model for single image dehazing, named PDR-Net. For haze removal, we propose a perception-driven image dehazing subnetwork, which is designed to reconstruct a clear image. To improve the contrast and color of the dehazed result, we propose a novel refinement subnetwork. Experimental results have demonstrated the effectiveness of the proposed method. At last, an application experiment demonstrates the potential usage of the proposed method in high-level vision problems. In the future, we plan to add a transmission estimation subnetwork to our current model to improve the dehazing performance and accelerate network convergence. In addition, we will try to design new discriminators to measure the quality of the dehazed results.

¹<https://github.com/jwyang/faster-rcnn.pytorch>

REFERENCES

- [1] W. Yin, T. Nei, C. Chen, and S. Li, "Socialized mobile photography: learning to photograph with social context via mobile devices," *IEEE Trans. Multimedia*, vol. 11, no. 1, pp. 184-200, 2014. [1](#)
- [2] F. Li, F. Shuang, Z. Liu, and X. Qian, "A cost-constrained video quality satisfaction study on mobile devices," *IEEE Trans. Multimedia*, vol. 20, no. 5, pp. 1154-1168, 2018. [1](#)
- [3] D. Berman and S. Avidan, "Non-local image dehazing," in *Proc. of IEEE Int. Conf. Comput. Vis. Pattern Rec. (CVPR)*, 2016, pp. 1674-1682. [1](#), [2](#)
- [4] Y. Li, S. You, S. Brown, and R. Tan, "Haze visibility enhancement: a survey and quantitative benchmarking," *Comput. Vis. Image Understand.*, vol. 165, pp. 1-16, 2017. [1](#)
- [5] B. Li, X. Peng, Z. Wang, J. Xu, and D. Feng, "Aod-net: All-in-one dehazing network," in *Proc. of IEEE Int. Conf. Comput. Vis. (ICCV)*, 2017, pp. 4780-4788. [1](#), [2](#), [5](#), [6](#), [7](#), [8](#), [9](#), [10](#), [11](#)
- [6] T. Zhang, W. Zheng, Z. Cui, Y. Zong, J. Yan, and K. Yan, "A deep neural network-driven feature learning method for multi-view facial expression recognition," *IEEE Trans. Multimedia*, vol. 18, no. 12, pp. 2528-2536, 2016. [1](#)
- [7] C. Li, R. Cong, J. Hou, S. Zhang, Y. Qian, and S. Kwong, "Nested network with two-stream pyramid for salient object detection in optical remote sensing images," *arXiv preprint arXiv:1906.08462*, 2019. [1](#)
- [8] C. Guo, C. Li, J. Guo, R. Cong, H. Fu, and P. Han, "Hierarchical features driven residual learning for depth map super-resolution," *IEEE Trans. Image Process.*, vol. 28, no. 5, pp. 2545-2557, 2019. [1](#)
- [9] Y. Song, J. Li, X. Wang, and X. Chen, "Single image dehazing using ranking convolutional neural network," *IEEE Trans. Multimedia*, vol. 20, no. 6, pp. 1548-1560, 2018. [1](#), [2](#)
- [10] R. Li, J. Pan, Z. Li, and J. Tang, "Single image dehazing via conditional generative adversarial network," in *Proc. of IEEE Int. Conf. Comput. Vis. Pattern Rec. (CVPR)*, 2018, pp. 8202-8211. [1](#), [2](#), [5](#), [6](#), [7](#), [8](#), [9](#), [10](#), [11](#)
- [11] J. Kopf, B. Neubert, B. Chen, M. Cohen, and D. Cohen-Or, "Deep photo: model-based photograph enhancement and viewing," *ACM Trans. Graph.*, vol. 27, no. 5, pp. 1-20, 2008. [2](#)
- [12] K. Tan and J. Oakley, "Enhancement of color images in poor visibility conditions," in *Proc. of IEEE Int. Conf. Image Process. (ICIP)*, 2000, pp. 788-791. [2](#)
- [13] S. Narasimhan and S. Nayar, "Contrast restoration of weather degraded images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 6, pp. 713-724, 2003. [2](#)
- [14] Y. Schechner, S. Narasimhan, and S. Nayar, "Instant dehazing of images using polarization," in *Proc. of IEEE Int. Conf. Comput. Vis. Pattern Rec. (CVPR)*, 2001, pp. 325-332. [2](#)
- [15] R. Fattal, "Single image dehazing," *ACM Trans. Graph.*, vol. 27, no. 3, pp. 1-9, 2008. [2](#)
- [16] R. Tan, "Visibility in bad weather from a single image," in *Proc. of IEEE Int. Conf. Comput. Vis. Pattern Rec. (CVPR)*, 2008, pp. 1-8. [2](#)
- [17] B. Cai, X. Xu, K. Jia, C. Qing, and D. Tao, "DehazeNet: an end-to-end system for single image haze removal," *IEEE Trans. Image Process.*, vol. 25, no. 11, pp. 5187-5198, 2016. [2](#), [5](#), [6](#), [7](#), [8](#), [9](#), [10](#), [11](#)
- [18] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 12, pp. 2341-2353, 2011. [2](#), [5](#), [6](#), [7](#), [8](#), [9](#), [10](#), [11](#)
- [19] G. Meng, Y. Wang, J. Duan, S. Xiang, and C. Pan, "Efficient image dehazing with boundary constraint and contextual regularization," in *Proc. of IEEE Int. Conf. Comput. Vis. (ICCV)*, 2013, pp. 617-624. [2](#), [5](#), [6](#), [7](#), [9](#)
- [20] C. O. Ancuti and C. Ancuti, "Single image dehazing by multi-scale fusion," *IEEE Trans. Image Process.*, vol. 22, no. 8, pp. 3271-3282, 2013. [2](#)
- [21] R. Fattal, "Dehazing using color-lines," *ACM Trans. Graph.*, vol. 34, no. 1, pp. 13:1-13:14, 2014. [2](#)
- [22] K. Tang, J. Yang, and J. Wang, "Investigating haze-relevant features in a learning framework for image dehazing," in *Proc. of IEEE Int. Conf. Comput. Vis. Pattern Rec. (CVPR)*, 2014, pp. 2995-3002. [2](#)
- [23] Q. Zhu, J. Mai, and L. Shao, "A fast single image haze removal algorithm using color attenuation prior," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3522-3533, 2015. [2](#), [5](#), [6](#), [7](#), [8](#), [9](#), [10](#), [11](#)
- [24] W. Ren, S. Liu, H. Zhang, J. Pan, and X. Cao, "Single image dehazing via multi-scale convolutional neural networks," in *Proc. of Eur. Conf. Comput. Vis. (ECCV)*, 2016, pp. 154-169. [2](#), [5](#), [6](#), [7](#), [8](#), [9](#), [10](#), [11](#)
- [25] W. Wang, X. Yuan, X. Wu, and Y. Liu, "Fast image dehazing method based on linear transformation," *IEEE Trans. Multimedia*, vol. 19, no. 6, pp. 1142-1155, 2017. [2](#)
- [26] T. Bui and W. Kim, "Single image dehazing using color ellipsoid prior," *IEEE Trans. Image Process.*, vol. 27, no. 2, pp. 999-1009, 2018. [2](#)
- [27] Q. Liu, X. Gao, L. He, and W. Lu, "Single image dehazing with depth-aware non-local total variation regularization," *IEEE Trans. Image Process.*, 2018. [2](#)
- [28] C. Li, J. Guo, F. Porikli, H. Fu, and Y. Pang, "A cascaded convolutional neural network for single image dehazing," *IEEE Access*, vol. 6, pp. 24877-24887, 2018. [2](#)
- [29] S. Santra, R. Mondal, B. Chanda, "Learning a patch quality comparator for single image dehazing," *IEEE Trans. Image Process.*, vol. 27, no. 9, pp. 4598-4607, 2018. [2](#)
- [30] F. Yuan and H. Huang, "Image haze removal via reference retrieval and scene prior," *IEEE Trans. Image Process.*, vol. 27, no. 9, pp. 4395-4409, 2018. [2](#)
- [31] L. Shi, B. Chen, S. Huang, A. Larin, O. Seredin, A. Kopylov, and S. Kuo, "Removing haze particles from single image via exponential inference with support vector data descriprion," *IEEE Trans. Multimedia*, 2018, [Early Access]. [2](#)
- [32] D. Yang and J. Sun, "Proximal dehaze-net: A prior learning-based deep network for single image dehazing," in *Proc. of Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 1-15. [2](#), [5](#), [6](#), [7](#), [8](#), [9](#), [10](#), [11](#)
- [33] H. Zhang and V. Patel, "Densely connected pyramid dehazing network," in *Proc. of IEEE Int. Conf. Comput. Vis. Pattern Rec. (CVPR)*, 2018, pp. 3194-3203. [2](#)
- [34] Y. Du and X. Li, "Perceptually optimized generative adversarial network for single image dehazing," *arXiv preprint arXiv:1805.01084*, 2018. [2](#)
- [35] K. He, J. Sun, and X. Tang, "Guided image filtering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 6, pp. 1397-1409, 2013. [2](#)
- [36] W. Wang, J. Shen, J. Xie, M. Cheng, H. Ling, and B. Ali, "Revisiting video saliency prediction in the deep learning era," *IEEE Trans. Pattern Anal. Mach. Intell.*, 2019 (Early Access). [2](#)
- [37] C. Li, C. Guo, W. Ren, R. Cong, J. Hou, S. Kwong, and D. Tao, "An underwater image enhancemant benchmark and beyond," *arXiv preprint arXiv:1901.05495*, 2019. [2](#)
- [38] W. Wang, J. Shen, and H. Ling, "A deep network solution for attention and aesthetics aware photo cropping," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 7, pp. 1531-1544, 2019. [2](#)
- [39] C. Li, J. Guo, F. Porikli, and Y. Pang, "LightenNet: A convolutional neural network for weakly illuminated image enhancement," *Pattern Rec. Lett.*, vol. 104, pp. 15-22, 2018. [2](#)
- [40] I. Goodfellow, J. Pouget-Adadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proc. of Advances in Neural Information Processing Systems (NeurIPS)*, 2014, pp. 2672-2680. [2](#)
- [41] K. Sohn, H. Lee, and X. Yan, "Learning structured output representation using deep conditional generative models," in *Proc. of Advances in Neural Information Processing Systems (NeurIPS)*, 2015, pp. 3483-3491. [2](#)
- [42] H. Koschmieder, "Theorie der horizontalen sichtweite," *Beitrage zur Physik der freien Atmosphare*, 1924. [3](#)
- [43] G. Huang, Z. Liu, L. van der Matten, and K. Weinberger, "Densely connected convolutional networks," in *Proc. of IEEE Int. Conf. Comput. Vis. Pattern Rec. (CVPR)*, 2017, pp. 2261-2269. [3](#)
- [44] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. of IEEE Int. Conf. Comput. Vis. Pattern Rec. (CVPR)*, 2016, pp. 770-778. [3](#)
- [45] J. Johnson and A. Alahi and L. Fei Fei, "Perceptual losses for real-time style transfer and super-resolution," in *Proc. of Eur. Conf. Comput. Vis. (ECCV)*, 2016, pp. 694-711. [3](#), [4](#)
- [46] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014. [4](#)
- [47] J. Deng, W. Dong, R. Socher, L. Li, K. Li, and L. Fei, "Imagenet: a large-scale hierarchical image database," in *Proc. of IEEE Int. Conf. Comput. Vis. Pattern Rec. (CVPR)*, 2009, pp. 248-255. [4](#)
- [48] N. Silberman, D. Hoiem, P. Kohli, and R. Fergus, "Indoor segmentation and support inference from rgbd images," in *Proc. of Eur. Conf. Comput. Vis. (ECCV)*, 2016, pp. 746-760. [4](#)
- [49] A. Ignatov, N. Kobyshev, K. Vanhoey, R. Timofte, and L. Gool, "Dslr-quality photos on mobile devices with deep convolutional networks," in *Proc. of IEEE Int. Conf. Comput. Vis. (ICCV)*, 2017, pp. 3297-3305. [4](#)
- [50] A. Ignatov, N. Kobyshev, R. Timofte, K. Vanhoey, and L. Gool, "WESPE: Weakly supervised photo enhancer for digital cameras," *arXiv preprint arXiv:1709.01118*, 2017. [4](#)
- [51] C. Ancuti, C. O. Ancuti, and C. D Vleeschouwer, "D-HAZY: A dataset to evaluate quantitatively dehazing algorithms," in *Proc. of IEEE Int. Conf. Image Process. (ICIP)*, 2016, pp. 2226-2230. [4](#)

- [52] B. Li, W. Ren, D. Fu, D. Tao, D. Feng, W. Zeng, and Z. Wang, "RESIDE: A benchmark for single image dehazing," *IEEE Trans. Image Process.*, vol. 28, no. 1, pp. 492-505, 2019. [4](#) [8](#) [11](#)
- [53] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 8, pp. 600-612, 2004. [5](#)
- [54] R. Girshick, "Fast r-cnn," in *Proc. of IEEE Int. Conf. Comput. Vis. (ICCV)*, 2015, pp. 1440-1448. [11](#)
- [55] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The PASCAL Visual Object Classes Challenge 2007 (VOC2007) Results," <http://www.pascal-network.org/challenges/VOC/voc2007/workshop/index.html>. [11](#)



Chongyi Li received his Ph.D. degree from Tianjin University, China, in June 2018. From 2016 to 2017, he took one year study at the Research School of Engineering, Australian National University (ANU) as a visiting Ph.D. student. Now, he is a Postdoc Research Fellow at the Department of Computer Science, City University of Hong Kong (CityU), Hong Kong. He received Excellent Doctoral Degree Dissertation Award from BSIG. His research interests include image processing, computer vision, and deep learning, particularly in the domains of image dehazing, underwater image enhancement, image super-resolution, low-light image enhancement, and salient object detection.



Chunle Guo received his B.S. degree from School of Electronic Information Engineering in Tianjin University. He is currently pursuing his Ph.D. degree with the School of Electrical and Information Engineering, Tianjin University, Tianjin, China. His current research focuses on image processing and computer vision, particularly in the domains of deep learning-based image restoration and enhancement.



Jichang Guo received his M.S. and Ph.D. degrees from the School of Electronic Information Engineering, Tianjin University, Tianjin, China, in 1993 and 2006, respectively. He is a Professor at Tianjin University. His current research interests include image processing, video coding, and computer vision.



Ping Han received her Ph.D. degree from Tianjin University, Tianjin, China in 2004. She is currently a professor with the College of Electronic Information and Automation, Civil Aviation University of China (CAUC), Tianjin, China. Her current research interests include SAR image processing, target detection, and pattern recognition.



Huazhu Fu (SM'18) received the B.S. degree in mathematical sciences from Nankai University in 2006, the M.E. degree in mechatronics engineering from the Tianjin University of Technology in 2010, and the Ph.D. degree in computer science from Tianjin University, China, in 2013.

He was a Research Fellow with Nanyang Technological University, Singapore, for two years. He is currently a Research Scientist with the Institute for Infocomm Research, Agency for Science, Technology and Research, Singapore. His research interests include computer vision, image processing, and medical image analysis. He is the Associate Editor of IEEE Access and BMC Medical Imaging. He is IEEE senior member.



Runmin Cong (M'19) received the Ph.D. degree in information and communication engineering from Tianjin University, Tianjin, China, in June 2019. He was a visiting student at Nanyang Technological University (NTU), Singapore, from Dec. 2016 to Feb. 2017. Since May 2018, he has spent one year as a Research Associate at the Department of Computer Science, City University of Hong Kong (CityU), Hong Kong. Since July 2019, he has been an Associate Professor with the Institute of Information Science, Beijing Jiaotong University, Beijing, China.

He was a recipient of the Best Student Paper Runner-Up at IEEE ICME in 2018. He is a Reviewer for the IEEE TIP, TMM, and TCSVT, etc. His research interests include computer vision, image processing, saliency detection, and 3-D imaging.