

Hierarchical Density-Aware Dehazing Network

Jingang Zhang^{ID}, Wenqi Ren^{ID}, Member, IEEE, Shengdong Zhang^{ID}, He Zhang^{ID}, Yunfeng Nie^{ID}, Member, IEEE, Zhe Xue^{ID}, and Xiaochun Cao^{ID}, Senior Member, IEEE

Abstract—The commonly used atmospheric model in image dehazing cannot hold in real cases. Although deep end-to-end networks were presented to solve this problem by disregarding the physical model, the transmission map in the atmospheric model contains significant haze density information, which cannot simply be ignored. In this article, we propose a novel hierarchical density-aware dehazing network, which consists of a densely connected pyramid encoder, a density generator, and a Laplacian pyramid decoder. The proposed network incorporates density estimation but alleviates the constraint of the atmospheric model. The predicted haze density then guides the Laplacian pyramid decoder to generate a haze-free image in a coarse-to-fine fashion. In addition, we introduce a multiscale discriminator to preserve global and local consistency for dehazing. We conduct extensive experiments on natural and synthetic hazy images, which prove that the proposed model performs favorably against the state-of-the-art dehazing approaches.

Manuscript received June 22, 2020; revised October 13, 2020 and January 28, 2021; accepted March 25, 2021. This work was supported in part by the National Key Research and Development Program of China under Grant 2019YFB1406500; in part by the National Natural Science Foundation of China under Grant 61775219, Grant 61640422, Grant 61802403, Grant 62025604, and Grant U1803264; in part by the Equipment Research Program of the Chinese Academy of Sciences under Grant YJKYYQ20180039 and Grant Y70X25A1HY; in part by the Beijing Nova Program under Grant Z201100006820074; in part by the Elite Scientist Sponsorship Program by the Beijing Association for Science and Technology; in part by the Youth Innovation Promotion Association CAS; and in part by the Peng Cheng Laboratory Project of Guangdong Province under Grant PCL2018KP004. This article was recommended by Associate Editor S. Chen. (Corresponding author: Shengdong Zhang.)

Jingang Zhang is with Intelligent Imaging Center, University of Chinese Academy of Sciences, Beijing 100049, China (e-mail: zhangjg@ucas.ac.cn).

Wenqi Ren is with the State Key Laboratory of Information Security, Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100093, China (e-mail: rwq.renwenqi@gmail.com).

Shengdong Zhang is with the State Key Laboratory of Information Security, Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100093, China, and also with ITS Lab, Huawei Technologies Company Ltd., Hangzhou 315600, China (e-mail: shengdong1986@gmail.com).

He Zhang is with the Adobe Research, Adobe Inc., San Jose, CA 95110 USA (e-mail: zhanghe920312@gmail.com).

Yunfeng Nie is with the Brussel Photonics, Department of Applied Physics and Photonics, Vrije Universiteit Brussel, 1050 Brussel, Belgium (e-mail: ynie@b-photon.org).

Zhe Xue is with the Beijing Key Laboratory of Intelligent Telecommunication Software and Multimedia, School of Computer Science, Beijing University of Posts and Telecommunications, Beijing 100876, China (e-mail: xuezhe@bupt.edu.cn).

Xiaochun Cao is with the State Key Laboratory of Information Security, Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100093, China, also with the School of Cyber Security, University of Chinese Academy of Sciences, Beijing 100049, China, and also with the Cyberspace Security Research Center, Peng Cheng Laboratory, Shenzhen 518055, China (e-mail: caoxiaochun@iee.ac.cn).

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/TCYB.2021.3070310>.

Digital Object Identifier 10.1109/TCYB.2021.3070310

Index Terms—Dehazing, density aware, hierarchical network.

I. INTRODUCTION

THE TURBID medium in the atmosphere often degrades outdoor images captured in bad weather, which makes many computer vision tasks more difficult, such as object detection and tracking. Therefore, image dehazing becomes an essential component in the multimedia and computer vision community [1]–[4]. The physical haze model [5], [6] can be formulated as

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

where $I(x)$ represents the observed hazy image, $J(x)$ denotes the corresponding haze-free radiance to be recovered, A is the global atmospheric light, and $t(x)$ is the transmission map, which describes the portion of the light that is not scattered and reaches the camera. As there only exists one single hazy image $I(x)$, restoring the corresponding clean image $J(x)$ from (1) is severely ill posed.

Conventional single image dehazing methods often remove haze based on the atmospheric model (1). By estimating atmospheric light and transmission maps using various image priors [9], [10] or convolutional neural networks (CNNs) [11], these methods can recover final dehazed images via inverting the physical model (1). However, the inaccurate result of the transmission map or atmospheric light will affect the quality of the recovered image directly, resulting in an undesired visual dehazing result.

To address this problem, some end-to-end dehazing methods are designed [12]–[14] to avoid the process of computing transmission map and atmospheric light. Li *et al.* [12] proposed an AOD-Net by fusing the transmission map and atmospheric light into a new variable. Nevertheless, the reformulation of AOD-Net is still a simplified atmospheric scattering model. Recently, a gated fusion network (GFN) [15] and a conditional generative adversarial network (cGAN) [16] are introduced to break the constraint of the atmospheric model. However, these networks tend to show undesirable artifacts in the restored results since the trained models completely disregard the guidance information of transmission maps.

Although the simplified atmospheric model (1) cannot hold for real hazy images since the complexity of the imaging process [17], the transmission map in the atmospheric model contains significant haze density information since it is inversely proportional to the haze concentration. Therefore, the transmission map cannot be ignored as the end-to-end



Fig. 1. Visual comparisons on a challenging real-world hazy example against the state-of-the-art end-to-end dehazing methods. As shown, E蒲DN [7] generates some halo artifacts around the person contour in (b), while the result by GridDehazeNet [8] has some haze residual in (c). In contrast, our method yields a clearer result by utilizing the haze density but alleviating the atmospheric constraint (1). (a) Hazy input. (b) E蒲DN [7]. (c) GridDehazeNet [8]. (d) Our result.

dehazing methods [15], [16]. However, it is still a difficult task to recover accurate transmission maps on the full resolution directly. In contrast, Karras *et al.* [18] proved that a low-resolution image is easier to recover than high resolution since there is less class information and fewer modes in low-resolution images. This observation is also applied to the transmission map estimation and the dehazed result reconstruction.

To this end, we propose a hierarchical density-aware dehazing network (HDDNet) that effectively utilizes the density information. Specifically, we first propose a densely connected pyramid encoder to extract multiscale haze-relevant features. Different from the conventional encoder with stacked convolutional layers, our proposed densely connected pyramid encoder concatenates downsampled features from prior layers to all the subsequent layers. This allows us to better exploit multiscale features and capture global information of the hazy input. Then, the extracted multiscale features are fed to the proposed density generator for haze concentration estimation. In contrast to previous methods that learn the transmission map on full resolution, we derive a density map by estimating the transmission map on a low-resolution size of the original size. Estimating on a low resolution has several advantages: 1) a low-resolution transmission map contains sufficient haze density information of the hazy input and 2) transmission maps are locally smooth, so it is important to capture the global structure on a low-resolution size in contrast to estimate details on a high resolution. After obtaining the haze density information, we propose a Laplacian pyramid decoder to reconstruct dehazed images progressively. We employ a multiscale generative adversarial network (GAN) to recover perceptual haze-free images, starting from easier low-resolution images and then shifting attention to increasingly finer-scale details, instead of having to learn the finest scale directly. As shown in Fig. 1, the proposed method is able to remove haze and recover more image details.

The contributions of this article are as follows.

- 1) We propose a HDDNet for high-quality single image dehazing. The employed densely connected pyramid encoder and the Laplacian pyramid decoder make full use of all the global structures and local details.
- 2) We present a density generator in HDDNet to estimate haze density maps, and the estimated density is then employed to guide the haze-free image reconstruction. The proposed HDDNet incorporates density estimation but alleviates the atmospheric model constraint to overcome the limitations of the traditional conventional atmospheric model.
- 3) We evaluate the proposed dehazing method through extensive experiments on both synthetic and real image datasets. In addition, ablation studies are conducted to demonstrate the function of different modules in the proposed HDDNet.

II. RELATED WORK

Single image dehazing methods can be mainly divided into two groups: 1) physical model-based recovering methods and 2) image enhancement approaches.

Dehazing via Recovering Methods: Physical model-based methods [19]–[21] assume that hazy images can be formulated by (1), which model hazy images as the sum of scene reflection and atmospheric light [5]. Scene reflection means the scene information that is not affected by medium particles. Based on this model, most existing algorithms focus on recovering the transmission map $t(x)$ and atmospheric light [22] for each hazy image.

He *et al.* [23] found a dark channel prior (DCP) according to the statistics of natural clean images. This prior is that at least one color channel has some pixels, whose values are close to zeros. Zhu *et al.* [24] observed that the difference between brightness and saturation should be very small in a clear image patch, and proposed a color attenuation prior (CAP) to compute transmission maps. Berman *et al.* [25] introduced a nonlocal dehazing (NLD) method for single image dehazing according to the observation that an image can be represented with a few hundred distinct colors. However, these hand-crafted features are obtained from statistical properties over a huge number of images and cannot always hold in practical scenarios [23].

To avoid artifacts caused by hand-crafted priors, several algorithms employ deep neural networks for single image dehazing [26]–[31]. Instead of estimating the transmission and the atmospheric light separately, AOD-Net [12] combines the transmission $t(x)$ and the airlight A in an integrated variable $K(x)$. However, we note that the variable $K(x)$ is a reformulation of the physical model (1). Even the recent work of densely connected pyramid dehazing network (DCPDN) [32] needs to estimate transmission maps and atmospheric lights first, and then compute final dehazed images by inverting the atmospheric model (1). Yang *et al.* [33] designed a disentanglement dehazing network (DDN) based on the physics model and adversarial training to restore realistic clean images

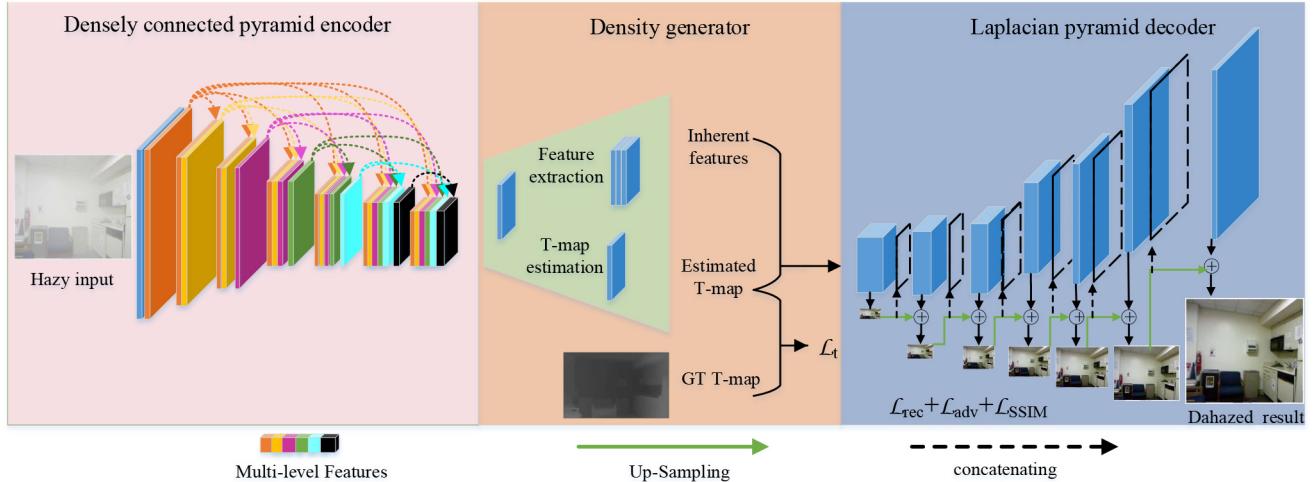


Fig. 2. Architecture of the proposed HDDNet. The proposed network consists of three main components: 1) a densely connected pyramid encoder; 2) a density-aware module; and 3) a Laplacian pyramid decoder. The densely connected pyramid encoder aims to extract multiscale image features, while the Laplacian pyramid decoder progressively refines the haze-free image. The proposed density-aware module provides haze concentration information and guides haze removal. The encoder features are pooled to different sizes to construct a feature pyramid, which helps the model capture multiscale information of the hazy input. The multiscale features are then fed into the density generator for low-resolution transmission estimation (we use 1/64 size of the original input). Finally, we progressively reconstruct the dehazed result by the proposed Laplacian pyramid decoder through six scales of reconstruction and adversarial losses.

using only unpaired supervision. Li *et al.* [34] proposed a cascaded dehazed model, which solves the transmission map and global atmospheric light simultaneously by two subnetworks. The dehazed results are then obtained by the atmospheric scattering model inversion. PMS-Net [35] is a deep learning method to select the patch size corresponding to each pixel and improve the accuracy of the transmission map. Golts *et al.* [36] proposed an unsupervised dehazing method, which uses real hazy images to train a dehazing model. This method can obtain a more satisfying dehazed result than fully supervised learning-based methods and prior-based methods. Hong *et al.* [37] proposed to employ haze-free images to guide the training of the dehazing network and introduced a spatial-aware block to learn more effective features for dense haze area. Kar *et al.* [38] proposed a progressive dehazing method, which refines the transmission map and atmospheric light in each iterative step. A learning-based filter method [39] can also be adapted to solve the dehazing problem. Although promising results have been obtained, the assumption that hazy images are the sum of scene reflection and airlight does not hold in complex scenes.

Dehazing via Enhancement Approaches: In general, humans are treated as the ultimate judge of the quality of images [40]. Therefore, another line of research attempts to remove haze based on image enhancement algorithms without considering the physical model (1). For example, the methods based on the famous Retinex theory [41], which restores the spectral properties of object surfaces by the ratio of the reflected light in this area to others, can enhance hazy images with or without prior. In [42], a multiscale fusion method is used for single image dehazing by blending several images into a single one. Based on this pipeline, an end-to-end GFN [15] is proposed to predict weight maps and combine three derived inputs into one by fusing the most significant features. However, GFN also needs handcrafted inputs for fusion, and required to compute

intermediate confidence maps. The enhanced pix2pix dehazing network (EPDN) [7] adopts image-to-image translation to remove haze and avoids relying on the physical scattering model. However, this method cannot deal heavily hazy scene well. To boost the dehazing performance on natural hazy images, Li *et al.* [43] designed a perception-inspired haze removal with a refinement module called PDR-Net. The dehazed results by PDR-Net are more colorful. These enhancement-based methods only consider the visual effect. Therefore, the dehazed results may not be the real scenario in the images.

Different from these approaches, our method takes the haze density map into the proposed network rather than direct dehazing by inverting the physical model (1). Therefore, the proposed HDDNet efficiently utilizes density information but alleviates the atmospheric constraint of (1).

III. PROPOSED METHOD

In this section, we present the details of the proposed HDDNet. The input of the network is an original hazy image, and its output is the corresponding dehazed result. We refer to this network as HDDNet.

A. Network Architecture

An overview framework of the proposed network is shown in Fig. 2. Our HDDNet mainly consists of three parts: 1) a densely connected pyramid encoder; 2) a density generator; and 3) a density-aware Laplacian pyramid decoder.

Densely Connected Pyramid Encoder: The encoder network aims to map hazy inputs to feature representations. Conventional encoders extract features by stacking some convolutional layers. However, such an encoder cannot capture the global and local information well. Since contextual information can be used to improve the performance of CNNs and the size of the receptive

TABLE I

CONFIGURATIONS OF THE NETWORK. THE FEATURE MAPS ARE DOWNSAMPLED BY MAX-POOLING AND UPSAMPLED BY BILINEAR INTERPOLATION. SIX SKIP LINKS ARE ADDED BETWEEN THE ENCODER AND THE DECODER

	Densely connected pyramid encoder												Density generator				
layer	conv1	conv2	pool1	conv3	pool2	conv4	pool3	conv5	pool4	conv6	pool5	conv7	pool6	conv8	density	conv9	clean1
size	3	3	2	3	2	3	2	3	2	3	2	3	2	3	3	3	3
channel	24	24		32		40		48		56		64		64	1	64	3
stride	1	1	↓2	1	↓2	1	↓2	1	↓2	1	↓2	1	↓2	1	1	1	1
pad	1	1		1		1		1		1		1		1	1	1	1
concatenate			pool1		pool2		pool3		pool4		pool5				conv8		
	Laplacian pyramid decoder																
layer	resize1	conv10	residual	clean2	resize2	conv11	residual	clean3	3	3	3	3	3	3	3	3	3
size	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
channel	24	3	3	3	24	3	3	24	3	3	24	3	3	24	3	3	3
stride	↑2	1	1	1	↑2	1	1	1	↑2	1	1	1	1	↑2	1	1	1
pad	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
concatenate	conv7		conv10	conv6		conv11	conv5		conv12	conv4		conv13	conv3		conv14	conv2	

field can roughly indicate how much contextual information is used. To increase the receptive field, we employ downsampling in the encoder to construct pyramid features of the hazy input, and concatenate downsampled features to all the subsequent layers, which make the proposed encoder effectively capture high-level and low-level features simultaneously. For example, the feature maps in the first layer are pooled to 1/2, 1/4, 1/8, 1/16, 1/32, and 1/64 of the original image size, and then concatenated to the following layers.

Density Generator: We propose a density generator to estimate the haze density. Since the transmission map directly depicts concentration-related information of the hazy input, and an accurate pixelwise transmission map is hard to estimate [32]. Therefore, we propose to estimate a low-resolution transmission by the density generator, which takes features from the last layer of the encoder as the input and outputs a haze density map and inherent features by two separate convolution layers (e.g., conv8 and conv9 in Table I). The estimated density is then fed into the decoder to guide haze-free image reconstruction. Specifically, we use a 1/64 resolution transmission to approximate the density information as shown in Fig. 2. The effectiveness of using the low-resolution transmission to approximate the density can be found in Section IV-A. By incorporating the transmission map into the network, our HDDNet is able to understand the distribution of haze well and then know where and how much haze needs to be removed.

Laplacian Pyramid Decoder: The Laplacian pyramid decoder has been widely applied in several computer vision tasks (e.g., semantic segmentation [44], image generation [45], and super-resolution [46]), which progressively reconstruct high-resolution feature maps and refine the output in a coarse-to-fine fashion. We construct the multiscale structure of images by building a series of reconstruction models, where each removes haze and captures image global structure at a particular scale of a Laplacian pyramid. We apply transposed convolutional layers for upsampling the feature maps to the next finer level. Then, we apply convolutional layers to predict the subband residuals (i.e., the differences between the ground-truth image and the upsampled result at the corresponding pyramid level). Different from the existing end-to-end dehazing networks [7], [12], [15], our HDDNet reconstructs haze-free images capitalizing on the haze density information, at the same time, without the constraint of the atmospheric model (1).

The Laplacian pyramid decoder consists of six blocks, for example, conv10–conv16 in Table I. Each of them predicts

a residual image. Then, we add the residual and the upsampled dehazed result from the previous layer to generate the desired output. Subsequent levels repeat this process, always basing on the result from the previous scale until the final dehazed result is obtained. Therefore, the dehazed result is reconstructed in a coarse-to-fine fashion, commencing with the estimated transmission map from the density generator. In other words, we first reconstruct a coarse dehazed result with the help of the density map, and then refine the details by estimating the residual at each level. The endpoint of the proposed model is the residual between the upsampled dehazed result and fine-scale ground truth. We add the residual to the upsampled dehazed result to obtain the final dehazed result. To avoid gradient vanishing and to accelerate training, skip connections are added between the encoder and decoder. This strategy allows the proposed Laplacian pyramid decoder to make effective use of features from both the encoder and the decoder, estimated haze density information, and dehazed result of each scale to correct any potential errors introduced in the decoder.

Multiscale GANs: Our decoder has constructed an image pyramid of the dehazed result, which captures the global structure and local details well. However, only using the image pyramid cannot generate both perceptually pleasing and haze-free images. Therefore, we introduce multiscale GANs to improve the dehazing quality, that is, a separate generative adversarial model is trained at each pyramid level. We note that the multiscale adversarial training procedure has been used to generate both perceptually pleasing and haze-free images in [33]. However, unlike [33] only trains a multiscale discriminator to judge an image is *real* or *fake*. We train six discriminators to judge an image is *real* or *fake* on each scale to obtain realistic image textures at different levels. In addition, in contrast to conventional discriminators used to judge the entire image, we design discriminators to differentiate an image patch that is *real* or *fake*. The patch discriminators combine with the multiscale reconstruction images that help the proposed network recover more photorealistic dehazed results. The details of the proposed discriminator structure and parameter settings are reported in Table II. We apply the same architecture at each scale of the discriminator.

B. Loss Function

To train the proposed density-aware dehazing model, we first define a density loss function to generate a fine

TABLE II
DETAILED PARAMETERS AND STRUCTURE OF THE DISCRIMINATOR

	conv1	conv2	pool1	conv3	pool2	conv4	conv5
size	3	3		3		3	3
channel	24	24		24		24	1
stride	1	1	↓2	1	↓2	1	1
pad	1	1		1		1	1
concatenate		conv1		pool1		pool2	

transmission map. Our density loss \mathcal{L}_t is imposed on both the intensity and gradient of estimated low-resolution transmission maps

$$\mathcal{L}_t = \frac{1}{N} \sum_{i=1}^N (\|\mathcal{F}_d(I_i) - (t_i)_{\downarrow}\|_1 + \lambda_s \|\nabla(\mathcal{F}_d(I_i)) - \nabla(t_i)_{\downarrow}\|_1) \quad (2)$$

where $\mathcal{F}_d(I)$ denotes the predicted density from the density-aware module, t is the ground-truth transmission map, $(\cdot)_{\downarrow}$ represents the downsampling operation, $\nabla(I_i, t_i)$ denotes the gradient extractor, and N is the number of training samples. Since L_2 loss tends to blur results in image restoration tasks, we use the L_1 norm to constrain the loss function. In addition, λ_s is used to control the importance of gradient loss of density map, which is set to 0.01.

For the dehazed results, we also define a loss between the restored images and the ground truths based on L_1 norm of multiscale reconstruction

$$\mathcal{L}_{\text{rec}} = \frac{1}{N} \sum_{i=1}^N \sum_{k=1}^6 (\|\mathcal{F}_k(I_i, \Theta) - J_{i,k}\|_1 + \lambda_s \|\nabla(\mathcal{F}_k(I_i, \Theta)) - \nabla(J_{i,k})\|_1) \quad (3)$$

where $\mathcal{F}_k(I_i, \Theta)$ denotes the dehazed output of I_i at the k th scale of the proposed Laplacian pyramid decoder, J is the ground-truth haze-free image (downsampled for each corresponding scale), and Θ keeps the weights of the learned filters. Since we use multiscale GANs to capture the global structure and local detail information, we also employ multiscale adversarial loss at each level of the pyramid to optimize the network, which can be expressed as

$$\mathcal{L}_{\text{adv}} = \frac{1}{N} \sum_{i=1}^N \sum_{k=1}^6 \left(\mathbb{E}_{J \sim p_{\text{clear}}(J)} [\log(D_k(J))] + \mathbb{E}_{I \sim p_{\text{hazy}}(I)} [\log(1 - D_k(\mathcal{F}_k(I_i, \Theta)))] \right) \quad (4)$$

where D_k denotes the discriminator at the k th scale.

In addition, since structural similarity index metric (SSIM) indicates the recovered structures of dehazed results, we exploit an SSIM loss at the finest level to further improve the dehazed quality by

$$\mathcal{L}_{\text{SSIM}} = \frac{1}{N} \sum_{i=1}^N \|1 - \text{SSIM}(\mathcal{F}_k(I_i, \Theta), J_i)\|_1, \quad k = 1 \quad (5)$$

where SSIM factors the local structure and contrast of the images by [57]

$$\text{SSIM}(x, y) = \frac{(2\mu_{x(p)}\mu_{y(p)} + c_1)(2\sigma_{x(p)y(p)} + c_2)}{(\mu_{x(p)}^2 + \mu_{y(p)}^2 + c_1)(\sigma_{x(p)}^2 + \sigma_{y(p)}^2 + c_1)} \quad (6)$$

where $\mu_{x(p)}$ and $\mu_{y(p)}$ are the mean value of patch centered around $x(p)$ and $y(p)$, $\sigma_{x(p)}^2$ and $\sigma_{y(p)}^2$ are the variances of x and y , respectively, and $\sigma_{x(p)y(p)}$ is the covariance of patches centered at $x(p)$ and $y(p)$. In addition, c_1 and c_2 are small positive values added on the numerator and denominator to avoid numerical instability. We use the default $c_1 = 10^{-4}$ and $c_2 = 9 \times 10^{-4}$ in this article. We demonstrate the effectiveness of the SSIM loss in Section V-A.

Finally, by combining the density and multiscale content losses, our final loss function is

$$\mathcal{L}_{\text{total}} = \mathcal{L}_t + \lambda_1 \mathcal{L}_{\text{rec}} + \lambda_2 \mathcal{L}_{\text{adv}} + \lambda_3 \mathcal{L}_{\text{SSIM}} \quad (7)$$

where λ_1 , λ_2 , and λ_3 are the positive weights, which are used to control the importance of the corresponding loss.

C. Implementation Details

The detailed architecture and parameter settings of the proposed network are presented in Table I. Our model contains six scales. The feature maps in the first scale are pooled to 1/2, 1/4, 1/8, 1/16, 1/32, and 1/64 of the original image size, and concatenated to the subsequent scales, that is, the feature maps at the last scale are 1/64 of original image size.

We initialize the weights using Gaussian random variables. Each convolution layer is followed by a nonlinear activation function. In this article, we adopt the leaky rectified linear units (LReLU) with a negative slope of 0.2 as the nonlinear activation function. The patch size is set as 256×256 , and the batch size is set to be 1. We use Adam optimization to train our network. We set β_1 and β_2 to 0.9 and 0.999, respectively. Our model is trained for a total of 100 epochs in a fully end-to-end fashion. The initial learning rate is 0.0005 and we decrease the learning rate by 0.5 every 30 epochs.

In the proposed HDDNet, each level k has its own loss function and the corresponding ground-truth haze-free image J_k as demonstrated in (3) and (4). We implement our model with the TensorFlow framework and run on a workstation with a 3.5 GHz CPU, 32-GB RAM, and Nvidia K80 GPU. It takes less than two days to converge.

For a fair comparison of the existing CNN-based methods, we train the proposed network by using the public training data from the RESIDE dataset [54], which contains both indoor and outdoor training images. The RESIDE dataset contains five subsets: 1) indoor training set (ITS); 2) outdoor training set (OTS); 3) synthetic objective testing set (SOTS); 4) real-world task-driven testing set (RTTS); and 5) hybrid subjective testing set (HSTS). The ITS contains 110 500 synthetic hazy indoor images and OTS contains 313 950 synthetic hazy outdoor images. The different atmospheric lights A are between [0.7, 1.0], and β is between [0.6, 1.8]. For a given haze-free image, we choose $A \in [0.7, 1.0]$ and $\beta \in [0.6, 1.8]$, then generate a transmission map using $t(x) = e^{-\beta d(x)}$, and finally, the hazy image can be obtained via model (1). We randomly select 2000 indoor and 2000 outdoor training images from the RESIDE dataset.

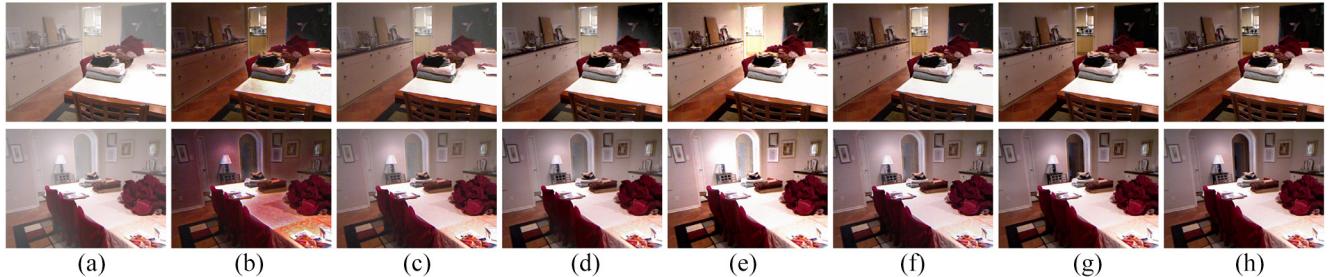


Fig. 3. Visual comparison of simulated hazy images using several state-of-the-art dehazing methods. (a) Simulated hazy images, (h) Ground truths. Notice the color distortions in the results by DCP [23] in (b), the under-dehazing effect of AOD-Net [12] (c) and PDNet [47] (d), and over-enhancement by DCPDN [32] in (e). The color distortion in the area of the door in the first image and second image of EPDN [7] (f).

TABLE III
QUANTITATIVE COMPARISON ON INDOOR HAZY IMAGES OF SOTS TEST IMAGES FROM THE RESIDE DATASET [54]

		DCP	FVR	BCCR	CAP	NLD	GRM	MSCNN	DehazeNet	AOD-Net	DCPDN	GFN	PDNet	HDDNet
w/o noise	PSNR	16.62	15.72	16.88	19.05	17.29	18.86	17.57	21.14	19.06	15.86	22.30	22.8282	24.79
	SSIM	0.82	0.75	0.79	0.84	0.86	0.81	0.75	0.85	0.85	0.82	0.88	0.91	0.94
1% noise	PSNR	17.13	15.68	17.89	18.69	16.44	19.00	17.16	20.75	18.96	15.82	22.21	21.80	22.37
	SSIM	0.65	0.53	0.67	0.71	0.6211	0.79	0.71	0.70	0.73	0.80	0.82	0.78	0.83
3% noise	PSNR	17.03	15.53	16.78	17.39	16.14	18.33	17.06	19.94	18.06	15.82	20.97	20.80	21.07
	SSIM	0.64	0.51	0.65	0.70	0.60	0.75	0.69	0.68	0.70	0.75	0.76	0.75	0.79

IV. EXPERIMENTAL RESULTS

In this section, we quantitatively and qualitatively evaluate our method against 11 state-of-the-art dehazing methods, including DCP [23], BCCR [49], CAP [24], GRM [58], NLD [25], DehazeNet [28], MSCNN [29], AOD-Net [12], GFN [15], DCPDN [32], PDNet [47], and EPDN [7] on synthetic datasets and real-world hazy images. We set $\lambda_2 = 0.01$ and $\lambda_1 = \lambda_3 = 1$ in all our experiments. The source code, trained model, as well as the dataset will be publicly available on our project website.

A. Evaluation on Benchmarks

We quantitatively evaluate the proposed network on three public dehazing test datasets of RESIDE [54], HazeRD [55], and O-HAZE dataset [48].

Evaluation on the RESIDE Dataset: We first evaluate the proposed HDDNet on SOTS from the RESIDE dataset. This test data contain 500 indoor images synthesized from the NYU2 depth dataset [59]. Fig. 3 shows some dehazed results generated by the state-of-the-art methods. The DCP algorithm tends to overenhance the dehazed image. The results of AOD-Net tend to have some remaining haze. PDNet improves the dehazing ability to some extent, but this method still remains some haze as shown in the fourth row of Fig. 3(d). DCPDN tends to overexposure the result as demonstrated in the second and fourth images of Fig. 3(e). EPDN can obtain a much better dehazed result. However, the dehazed images still contain haze residual as shown in the door region of the second image in Fig. 3(f).

To further show the robustness of the proposed method, we carry out experiments on hazy and noisy images using the SOTS dataset [54]. We add 1% and 3% Gaussian noise to each hazy image in the dataset and report average PSNR and SSIM in Table III. As shown, although the proposed HDDNet degrades when the noise variance increases, our

method achieves the best dehazing performance for different noise levels on the SOTS dataset. These results prove that our proposed algorithm is robust for different noise levels.

In addition to the synthesized indoor hazy images, we also compare the proposed HDDNet against the state-of-the-art dehazing approaches on outdoor hazy images from the RESIDE dataset. As shown in Table IV, the proposed model obtains favorable dehazing results against the state-of-the-art methods on outdoor hazy images.

Evaluation on the HazeRD Dataset: The HazeRD dataset provides natural outdoor images with high-quality depth maps, therefore, simulating more realistic haze. Note that all the CNN-based methods do not include images in HazeRD as training data, and it is fair to compare them on this dataset. As shown in Table IV, the proposed HDDNet achieves highest PNSR and SSIM on the HazeRD testing data, and exceeds the suboptimal method of PDNet by up to 1.24 dB and 0.117 in terms of PSNR and SSIM, respectively.

Evaluation on the O-HAZE Dataset: To evaluate the dehazing ability on real-captured hazy images, we compare our algorithm against the state-of-the-art dehazing methods on the O-HAZE dataset [48], which generates haze using a professional haze machine that simulates with high fidelity real hazy conditions. As reported in Table IV, we can observe that our model achieves the highest performance in terms of PSNR and SSIM. The visual comparisons in Fig. 4 also show that our method performs favorably against other state-of-the-art approaches.

Evaluations on the Dense-Haze Dataset [53]: The Dense-Haze dataset contains 55 pairs of real-world dense hazy images and the corresponding haze-free counterpart of various outdoor scenes, in which the training set contains 50 pairs and the test set includes 5 pairs. To show the performance of the proposed HDDNet, we compare our method with the state-of-the-art methods on these dense hazy images. In addition to the popular dehazing methods, we also compare our algorithm

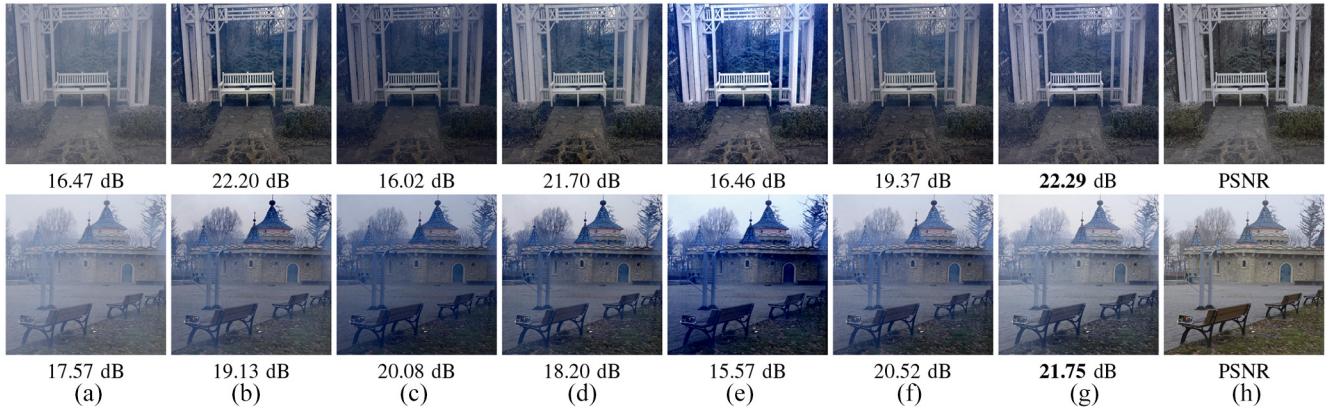


Fig. 4. Visual results of the O-HAZE dataset [48]. PSNR values are reported for each dehazed result. (a) Hazy inputs. (b) MSCNN [29]. (c) AOD-Net [12]. (d) PDNet [47]. (e) DCPDN [32]. (f) GFN [15]. (g) HDDNet. (h) Ground truths.

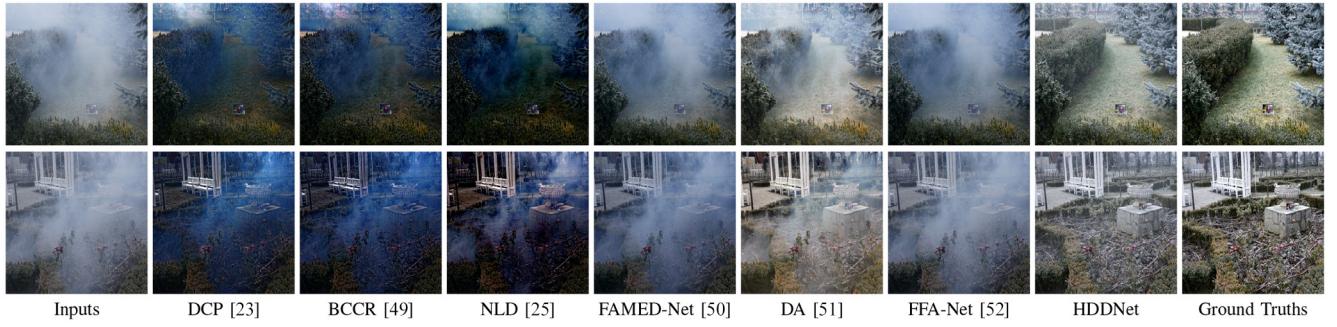


Fig. 5. Qualitative evaluations on the dense nonhomogeneous haze images [53]. The proposed method obtains much clearer images with clearer structures and characters.

TABLE IV
QUANTITATIVE COMPARISON ON OUTDOOR HAZY IMAGES OF SOTS TEST IMAGES FROM THE RESIDE DATASET [54], THE HAZERD DATASET [55], AND THE O-HAZE DATASET [48]

		DCP	FVR	BCCR	CAP	NLD	GRM	MSCNN	DehazeNet	AOD-Net	DCPDN	GFN	PDNet	HDDNet
RESIDE	PSNR	19.13	15.39	15.08	22.27	16.85	17.35	18.64	22.46	19.06	19.93	21.55	20.89	22.52
	SSIM	0.81	0.72	0.74	0.90	0.78	0.82	0.78	0.85	0.88	0.84	0.84	0.85	0.91
HazeRD	PSNR	17.66	16.17	16.31	18.56	18.82	17.47	19.10	19.53	18.13	18.82	19.18	20.14	21.38
	SSIM	0.84	0.85	0.83	0.83	0.84	0.79	0.85	0.85	0.83	0.89	0.86	0.89	0.91
O-HAZE	PSNR	16.59	14.92	17.44	17.62	16.61	15.62	19.07	16.21	16.72	15.62	18.30	18.52	19.28
	SSIM	0.74	0.70	0.75	0.71	0.75	0.58	0.77	0.67	0.68	0.62	0.72	0.75	0.77

TABLE V
QUANTITATIVE COMPARISON ON THE DENSE-HAZE DATASET [53]

	DCP	BCCR	NLD	DM ² F-Net	LDP	LAP-Net	FAMED-Net	HR-Dehazer	RI-GAN	Big FastNet	Dense 123	HDDNet
PSNR	14.56	14.62	13.18	17.62	18.65	16.62	16.06	16.47	16.47	16.37	19.92	20.19
SSIM	0.40	0.35	0.36	0.57	0.62	0.55	0.51	0.52	0.54	0.57	0.65	0.67

against eight state-of-the-art dense hazy removal approaches, including the one-to-three color enhancement dehazing (123-CEDH) network [60], RI-GAN [13], HR-Dehazer [61], DM²F-Net [11], LDP [26], FAMED-Net [50], GridDehazeNet [8], and LAP-Net [27]. The quantitative results in terms of PSNR and SSIM are reported in Table V. As shown, the proposed method achieves the highest scores on the Dense-Haze dataset.

Evaluations on the NH-Haze Dataset [56]: The NH-Haze dataset contains 55 pairs of real-world dense hazy images and corresponding haze-free counterpart of various outdoor scenes, in which the training set contains 50 pairs and the test set includes 5 pairs. We compare the proposed model with the state-of-the-art methods, such as HardGAN [62],

FFA-Net [52], Dense 123, and DA [51]. Following the protocol in [56], the quantitative results in terms of PSNR and SSIM are reported in Table VI. As shown, the proposed method achieves the highest scores on the NH-Haze dataset. In addition, we display two examples from the Dense-Haze dataset in Fig. 5, which shows that our method generates visually pleasant results and performs favorably against the state-of-the-art dehazing approaches.

B. Evaluation on Real-World Images

We further qualitatively evaluate the proposed method on the real hazy images from [21]. We choose four challenging

TABLE VI
QUANTITATIVE COMPARISON ON THE NH-HAZE DATASET [56]

	DCP	BCCR	NLD	MSCNN	DA	FAMED-Net	EDPN	FFA-Net	Dense 123	Trident	HardGAN	HDDNet
PSNR	12.35	12.15	12.01	19.72	19.65	19.62	19.06	19.87	18.47	21.41	21.70	21.89
SSIM	0.40	0.38	0.38	0.67	0.64	0.63	0.65	0.59	0.67	0.71	0.70	0.71

TABLE VII

QUANTITATIVE PSNR AND SSIM RESULTS ON THE SYNTHETIC DATASET USING DIFFERENT NETWORK CONFIGURATIONS AND LOSS FUNCTIONS

	PED	w/o HE	w/o GAN	w/o MSGAN	w/o density	PED+density	w/o DC	\mathcal{L}_{rec}	$\mathcal{L}_{rec} + \mathcal{L}_t$	$\mathcal{L}_{rec} + \mathcal{L}_{adv}$	$\mathcal{L}_{rec} + \mathcal{L}_{SSIM}$	HDDNet
PSNR	19.07	23.27	21.90	23.5	24.51	20.07	23.27	23.00	23.85	24.47	23.20	24.79
SSIM	0.85	0.87	0.86	0.93	0.93	0.85	0.92	0.83	0.91	0.86	0.90	0.94

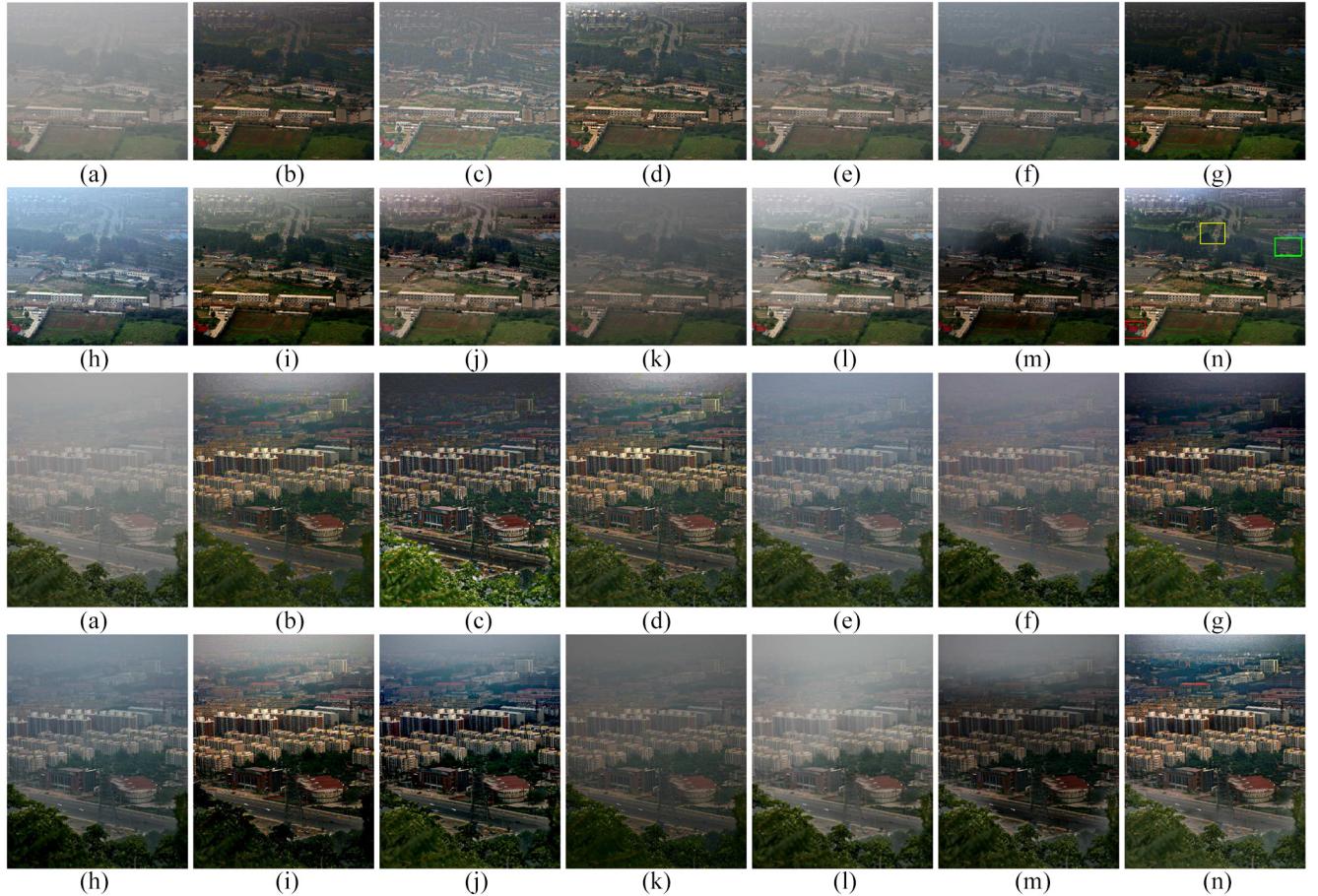


Fig. 6. Qualitative evaluations on the real hazy images. The proposed model generates much clearer images with finer structures and characters. The results of our model are more colorful than others. (a) Hazy input. (b) DCP [23]. (c) FVR [63]. (d) BCCR [49]. (e) RF [21]. (f) CAP [24]. (g) ATM [22]. (h) DehazeNet [28]. (i) MSCNN [29]. (j) NLD [25]. (k) AOD-Net [12]. (l) DCPDN [32]. (m) GFN [15]. (n) HDDNet.

hazy images, which contain dense haze and a large area of white objects or sky. Fig. 6 shows several real-world hazy images and the dehazed results generated by the proposed approach and 12 state-of-the-art dehazing methods [12], [15], [21]–[25], [28], [29], [32], [49], [63]. The traditional dehazing methods of FVR [63] and ATM [22] fail to generate clear images and tend to generate color distortion as shown in Fig. 6(c) and (g). For the first result of FVR, we can note that this method cannot remove haze completely. For the first result of ATM, we can note that it loses some detail in the final dehazed result. Image priors-based approaches of DCP, BCCR, and NLD overestimate the haze concentration and obtain darker results than others. We can note that

the overenhancement of the sky area in the result of DCP and BCCR. For the third result of NLD, we also note the overenhancement artifacts of human regions. We note that Cai *et al.* [28], Ren *et al.* [29], and Zhang and Patel [32] developed CNN-based methods for transmission map estimation and then used the conventional atmospheric model (1) to restore clean images. However, the final recovered images contain some artifacts because of imperfect estimated transmission maps as shown in Fig. 6(h), (i), and (l). The end-to-end deep learning networks by Li *et al.* [12] and Ren *et al.* [15] used a CNN to directly recover clean images from hazy images. However, these methods fail to obtain clean images and retain some haze as shown in Fig. 6(k) and (m).

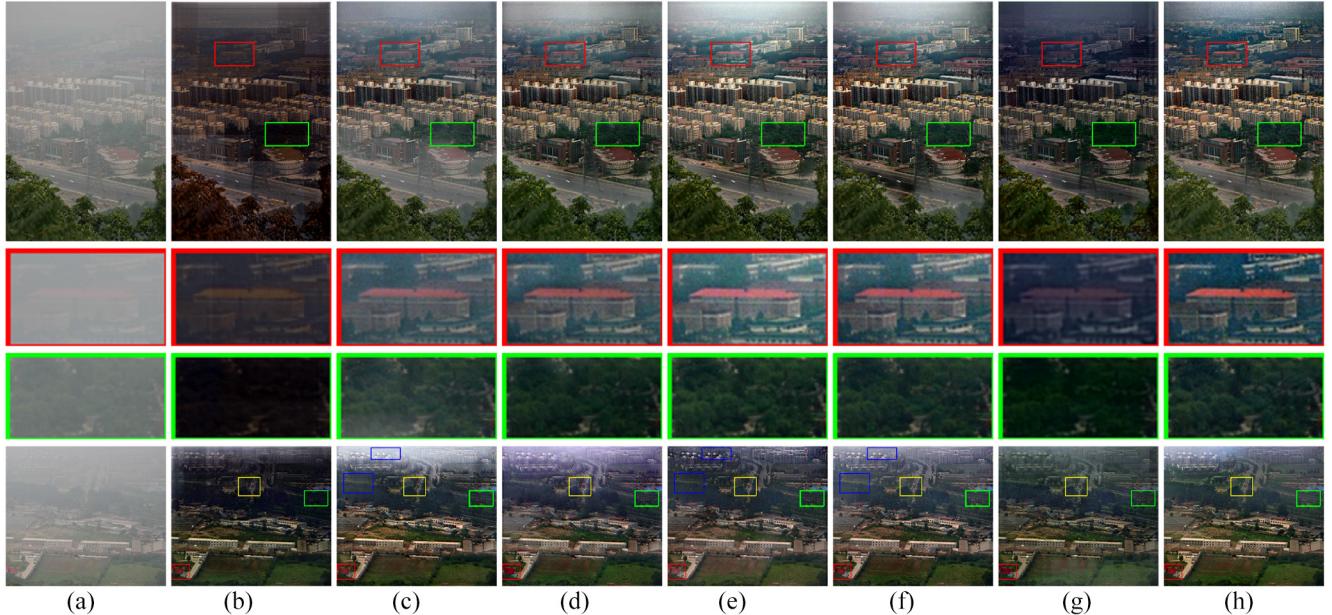


Fig. 7. Visual results of the proposed network with different modules. (a) Input. (b) PED. (c) w/o HE. (d) w/o GAN. (e) w/o MSGAN. (f) w/o density. (g) PED+density. (h) Our results.

Different from these methods, the proposed model utilizes the haze density information but alleviates the traditional atmospheric constraint in (1), thus, facilitating haze removal and avoiding artifacts. Our method first extracts representing features using the densely connected pyramid encoder, which can leverage high-level and low-level features well. Then, a density generator is designed to estimate haze density from multiscale features. Based on the estimated haze density and multiscale features, our model estimates a high quality coarse dehazed result. Then, a multiscale refinement decoder is employed to refine the dehazed result from coarse to fine. Based on this carefully designed framework, the images recovered by the proposed HDDNet look much clearer than those of other algorithms as shown in Fig. 6(n).

V. ANALYSIS AND DISCUSSION

A. Ablation Study

Ablation Study of Different Configurations: The first ablation study is designed to show the effectiveness of each module proposed in the HDDNet. We conduct the following seven experiments.

- 1) *PED*: The plain encoder-decoder network without the density generator, GAN, and Gaussian and Laplacian pyramids.
- 2) *w/o HE*: Our full model without the hierarchical refinement in the Laplacian pyramid decoder, that is, only reconstruct haze-free images at the last layer.
- 3) *w/o GAN*: Our full model without the adversarial loss in the last layer of the Laplacian pyramid decoder.
- 4) *w/o MSGAN*: Our full model without multiscale adversarial losses at all the scales in the Laplacian pyramid decoder.

- 5) *w/o Density*: Our full model without the proposed density generator.
- 6) *w/o Dense Connection*: Our full model without the dense connection in the encoder.
- 7) *PED+Density*: The plain encoder-decoder with the proposed density generator but without the multiscale reconstruction and multiscale adversarial losses.

The corresponding results can be found in Fig. 7. As shown, the PED tends to generate dark results and many visual block artifacts in Fig. 7(b). While without using the hierarchical refinement in the decoder, it cannot capture multiscale contextual information well and thus lost some details in the dehazed results as shown in Fig. 9(c). Without using GAN and MSGAN tends to overenhance the results as shown in Fig. 7(d) and (e). In addition, without using the proposed density generator would remain haze in some regions [e.g., the *road* area in Fig. 7(f)], which demonstrates that the proposed density-aware module could help the network understand the distribution of haze and how much haze should be removed. Compared with Fig. 7(b), we also notice that adding density can alleviate the block artifacts. In contrast, the result in Fig. 7(h) generated by the full model is able to preserve clearer and sharper structures and details. The quantitative results on the SOTS dataset in Table VII also demonstrate the effectiveness of each component in the proposed HDDNet.

We further show that the density map is critical for dehazing. The proposed model decomposes the dehazing problem into two subtasks. The first is to recover a coarse dehazed result, and the second is to refine the missing details of the coarse dehazed result. If a high-quality dehazed image is obtained, the decoder can focus on refining the missing details by residual learning. We use the haze density map to make the model know the haze distribution and avoid the overdehazing or underdehazing problem. With the help of the density map,

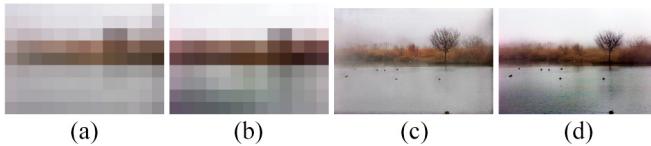


Fig. 8. Visual comparisons on a challenging real-world hazy example. (a) Global structure estimated without density map. (b) Global structure estimated with density amp. (c) Final result recovered from (a). (d) Final result recovered from (b). As shown, the density map can help remove haze in the low-resolution result, which can help the decoder recover a much clear and sharp result. The resolutions of coarse dehazed results are 1/64 size of the hazy inputs.

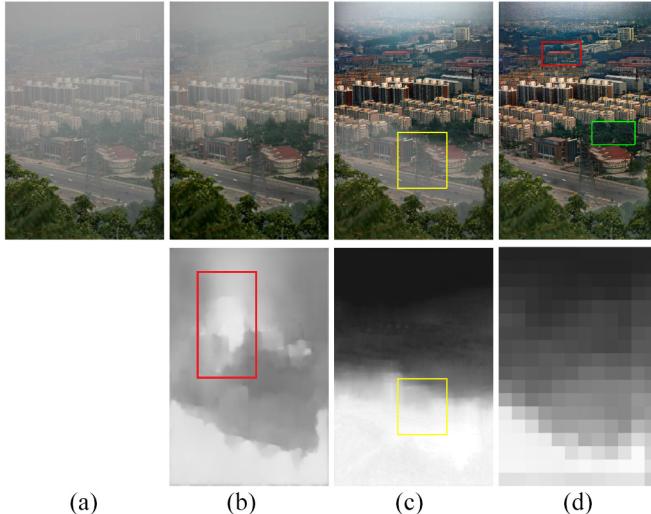


Fig. 9. Qualitative evaluation on the transmission maps. Both DCPDN and the proposed HDDNet-full fails to generate accurate pixelwise transmission maps in (b) and (c). In contrast, HDDNet effectively preserves the scene structure well in the estimated density map in (d). The resolutions of our estimated density maps are 1/64 size of the hazy inputs.

the proposed decoder can restore a high-quality coarse dehazed result. Then, the image details are recovered from coarse to fine with the help of multiscale GANs. As shown in Fig. 8, since the coarse result [Fig. 8(b)] contains less haze, the final result becomes much clear and sharp as shown in Fig. 8(d).

Ablation Study of Different Losses: In this section, we illustrate the effectiveness of each loss function in the proposed network. Table VII shows the quantitative evaluations on the SOTS dataset using different losses. Only using the reconstruction loss \mathcal{L}_{rec} generates the lowest score in terms of PSNR and SSIM. While adding the multiscale adversarial loss \mathcal{L}_{adv} on the reconstruction loss \mathcal{L}_{rec} boosts the dehazing performance significantly. In addition, the SSIM loss $\mathcal{L}_{\text{SSIM}}$ could preserve structures of images and yield a higher SSIM value. Overall, Table VII indicates the effectiveness of each loss function for image dehazing.

Ablation Study on Different Resolutions of Transmission Maps: In the proposed network, we use the coarse transmission map as a haze density map, which can be used to improve the dehazing performance. To show the effectiveness of the proposed strategy, we evaluate the performance of dehazing with different resolution transmission maps. We

TABLE VIII
AVERAGE PSNR OF DEHAZING WITH DIFFERENT RESOLUTIONS OF ESTIMATED TRANSMISSION MAPS

1/1	1/4	1/8	1/16	1/32	1/64
24.54	24.59	24.60	24.65	24.71	<u>24.79</u>

design another five networks with a density generator at different decoder levels to obtain transmission maps with different resolutions. Table VIII shows that the recovered image with the lowest resolution transmission map achieves the highest dehazing performance. We can draw two reasons for why the low-resolution transmission map achieves the highest dehazing performance. First, the proposed model decomposes the dehazing problem into two phases: 1) restoring a coarse dehazed result and 2) refining the image details. With the help of a low-resolution transmission map, the proposed model can restore a high-quality coarse dehazed result, and then the decoder can refine the image details with multiscale GANs. Second, the high-resolution transmission map contains more details than the low-resolution transmission map. The details are hard to be predicted accurately. Suppose the transmission map is not correct, which will mislead the dehazing processing. As shown in Fig. 9, the dehazed results of DCPDN and HDDNet-full contain residual haze due to the incorrect transmission map.

These ablation studies demonstrate that the proposed pyramid encoder and decoder, density estimator, and different loss functions are effective for image dehazing.

B. Running Time

We select 100 hazy images from the RESIDE dataset to evaluate the proposed network and the state-of-the-art dehazing methods on the same machine. The average running time of all the methods is shown in Table IX. The results show the high efficiency of the proposed model. A possibly answer is that the proposed model requires fewer parameters than traditional dehazing networks since the proposed model reuse the features map. The traditional dehazing networks have many layers that contribute little to remove haze, and the features are redundant. The dehazing models without features reusing need to learn new features and pass on information that needs to be preserved, which requires more parameters and memory. Furthermore, the proposed model employs a multiscale network architecture to reduce computation time and memory.

C. Discussions

Difference From Traditional Laplacian Pyramid Works: The LAPGAN [45] applies convolution on upsampled input images at each level, while the LapSRN [46] obtains features directly from the LR space at the end of each level. In contrast, the proposed model receives the features from the end of each level and the same level of the encoder. The Laplacian pyramid reconstruction [44] employs the segmentation result to obtain a mask and extracts a residual from the result of the product of the mask and feature, while our model refines the image details by features from multiscale features in the encoder,

TABLE IX
AVERAGE RUNNING TIME ON THE RESIDE DATASET

Method	DCP	BCCR	NLD	MSCNN	DehazeNet	GFN	FAMED-Net	DCPDN	EPDN	PDNet	Our	
Language	Matlab						Python					
Platform												
CPU (s)	25.08	3.52	8.41	2.45	3.09	19.03	1.43	3.24	3.51	4.02	1.32	
Parameters	-	-	-	8.01×10^3	8.02×10^3	5.14×10^5	1.80×10^4	6.69×10^7	1.74×10^7	4.10×10^4	4.58×10^5	

TABLE X
AVERAGE MSE OF ESTIMATED TRANSMISSION MAPS

DCP [23]	BCCR	CAP [24]	NLD [25]	GRM [58]	DehazeNet [28]	MSCNN [29]	PDNet [47]	DCPDN [32]	HDDNet
0.0282	0.0405	0.0562	0.0469	0.0246	0.0312	0.0479	0.0146	0.0046	0.0099

upsampling features from the decoder, and the restored coarse dehazed result.

Difference From Traditional Hazy Models: The hazy model (1) has been used to describe the formation of hazy images. However, we note that it cannot describe complex images well [27], [62]. The scene light should go through more particles to reach the camera. The particles scatter the light many times rather than one time of scattering before reaching the camera [27]. Furthermore, the density of haze depends on various factors, such as temperature, altitude, and humidity, making the distribution of haze at each spatial scene point space-variant and nonhomogeneous [62]. The atmospheric light also cannot be treated as constant in real hazy images [21]. The existing methods based on the hazy model may fail to estimate accurate transmission map and global atmospheric light. Although DCPDN incorporates the hazy model into a learning framework and achieves state-of-the-art performance, we note that learning an accurate pixelwise transmission map is still hard. As shown in Table X, the transmission map generated by DCPDN achieves the lowest MSE value. However, the dehazd results by DCPDN are dependent on the hazy model, which limits the performance. To overcome the weakness of the hazy model (1), we design a density map, which can provide additional information for dehazing and significantly improve the dehazing performance.

D. Limitations

Although our model alleviates the constraints of the atmospheric model and has shown effective dehazing performance on real-world hazy images, our method has some limitations since synthetic images are used. Our training images are selected from the RESIDE dataset, which is synthesized based on the atmospheric model of (1). Therefore, we will address this issue by semisupervised and unsupervised training strategies in future work, which can further enhance the generalization of the proposed method on real hazy images. The capability of the model is critical for dehazing. To increase the capability of the proposed model, we will adopt the technologies used in image super-resolution [64], [65] and segmentation [66].

VI. CONCLUSION

In this article, we proposed an HDDNet to progressively reconstruct the clean image from a hazy input. Our model

consists of a densely connected pyramid encoder, a density generator, and a Laplacian pyramid decoder with multiscale GAN. The proposed HDDNet utilizes haze density information to guide clean image reconstruction and alleviates the constraint of the atmospheric model. In addition, both the pyramid encoder and decoder, as well as the multiscale GAN are employed to capture global structures and fine details of the scene. Extensive experiments and ablation studies prove that the proposed method obtain high-quality dehazing result against the state-of-the-art dehazing methods on both synthetic datasets and real-world hazy images.

REFERENCES

- [1] S. G. Narasimhan and S. K. Nayar, "Contrast restoration of weather degraded images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 6, pp. 713–724, Jun. 2003.
- [2] H. Zhu *et al.*, "Single-image dehazing via compositional adversarial network," *IEEE Trans. Cybern.*, vol. 51, no. 2, pp. 829–838, Feb. 2021.
- [3] Y. Zhang, J. Zhang, B. Huang, and Z. Fang, "Single-image deraining via a recurrent memory unit network," *Knowl. Based Syst.*, vol. 218, Apr. 2021, Art. no. 106832.
- [4] S. Li *et al.*, "A comprehensive benchmark analysis of single image deraining: Current challenges and future perspectives," *Int. J. Comput. Vis.*, vol. 129, pp. 1301–1322, Jan. 2021.
- [5] H. Koschmieder, "Theorie der horizontalen sichtweite," *Beitrage zur Physik der freien Atmosphare*, vol. 12, pp. 33–53, Jan. 1924.
- [6] M. Han, Z. Lyu, T. Qiu, and M. Xu, "A review on intelligence dehazing and color restoration for underwater images," *IEEE Trans. Syst., Man, Cybern.*, Syst., vol. 50, no. 5, pp. 1820–1832, May 2020.
- [7] Y. Qu, Y. Chen, J. Huang, and Y. Xie, "Enhanced Pix2Pix dehazing network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 8160–8168.
- [8] X. Liu, Y. Ma, Z. Shi, and J. Chen, "GridDehazeNet: Attention-based multi-scale network for image dehazing," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2019, pp. 7313–7322.
- [9] R. Fattal, "Dehazing using color-lines," *ACM Trans. Graph.*, vol. 34, no. 1, p. 13, 2014.
- [10] H.-M. Hu, H. Zhang, Z. Zhao, B. Li, and J. Zheng, "Adaptive single image dehazing using joint local-global illumination adjustment," *IEEE Trans. Multimedia*, vol. 22, no. 6, pp. 1485–1495, Jun. 2020.
- [11] Z. Deng *et al.*, "Deep multi-model fusion for single-image dehazing," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2019, pp. 2453–2462.
- [12] B. Li, X. Peng, Z. Wang, J. Xu, and D. Feng, "AOD-Net: All-in-one dehazing network," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 4780–4788.
- [13] A. Dudhane, H. S. Aulakh, and S. Murala, "RI-GAN: An end-to-end network for single image haze removal," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2019, pp. 2014–2023.
- [14] S. Zhang, F. He, and W. Ren, "NLNDN: Non-local dehazing network for dense haze removal," *Neurocomputing*, vol. 410, pp. 363–373, Oct. 2020.
- [15] W. Ren *et al.*, "Gated fusion network for single image dehazing," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 3253–3261.

- [16] R. Li, J. Pan, Z. Li, and J. Tang, "Single image dehazing via conditional generative adversarial network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 8202–8211.
- [17] Y. Chen, X. Cao, Q. Zhao, D. Meng, and Z. Xu, "Denoising hyperspectral image with non-IID noise structure," *IEEE Trans. Cybern.*, vol. 48, no. 3, pp. 1054–1066, Mar. 2018.
- [18] T. Karras, T. Aila, S. Laine, and J. Lehtinen, "Progressive growing of GANs for improved quality, stability, and variation," in *Proc. Int. Conf. Learn. Rep.*, 2018, pp. 1–9.
- [19] X. Fan, Y. Wang, X. Tang, R. Gao, and Z. Luo, "Two-layer Gaussian process regression with example selection for image dehazing," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 12, pp. 2505–2517, Dec. 2017.
- [20] R. Liu, X. Fan, M. Hou, Z. Jiang, Z. Luo, and L. Zhang, "Learning aggregated transmission propagation networks for haze removal and beyond," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 20, no. 10, pp. 2973–2986, Oct. 2016.
- [21] K. Tang, J. Yang, and J. Wang, "Investigating haze-relevant features in a learning framework for image dehazing," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2014, pp. 2995–3000.
- [22] M. Sulami, I. Glatzer, R. Fattal, and M. Werman, "Automatic recovery of the atmospheric light in hazy images," in *Proc. IEEE Int. Conf. Comput. Photography*, 2014, pp. 1–11.
- [23] 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, Dec. 2011.
- [24] 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, Nov. 2015.
- [25] D. Berman, T. Treibitz, and S. Avidan, "Non-local image dehazing," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 1674–1682.
- [26] Y. Liu, J. Pan, J. Ren, and Z. Su, "Learning deep priors for image dehazing," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2019, pp. 2492–2500.
- [27] Y. Li *et al.*, "LAP-Net: Level-aware progressive network for image dehazing," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2019, pp. 3276–3285.
- [28] 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, Nov. 2016.
- [29] W. Ren, S. Liu, H. Zhang, J. Pan, X. Cao, and M.-H. Yang, "Single image dehazing via multi-scale convolutional neural networks," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 154–169.
- [30] S. Zhang, W. Ren, and J. Yao, "FEED-NET: Fully end-to-end dehazing," in *Proc. IEEE Int. Conf. Multimedia Expo*, 2018, pp. 1–6.
- [31] S. Zhang, F. He, and W. Ren, "Photo-realistic dehazing via contextual generative adversarial networks," *Mach. Vis. Appl.*, vol. 33, no. 31, p. 33, 2020.
- [32] H. Zhang and V. M. Patel, "Densely connected pyramid dehazing network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 3194–3203.
- [33] X. Yang, Z. Xu, and J. Luo, "Towards perceptual image dehazing by physics-based disentanglement and adversarial training," in *Proc. AAAI Conf. Artif. Intell.*, 2018, pp. 7485–7492.
- [34] 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.
- [35] W.-T. Chen, J.-J. Ding, and S.-Y. Kuo, "PMS-Net: Robust haze removal based on patch map for single images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 11681–11689.
- [36] A. Golts, D. Freedman, and M. Elad, "Unsupervised single image dehazing using dark channel prior loss," *IEEE Trans. Image Process.*, vol. 29, pp. 2692–2701, 2020.
- [37] M. Hong, Y. Xie, C. Li, and Y. Qu, "Distilling image dehazing with heterogeneous task imitation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 3459–3468.
- [38] A. Kar, S. K. Dhar, D. Sen, and P. K. Biswas, *Transmission Map and Atmospheric Light Guided Iterative Updater Network for Single Image Dehazing*, 2020. [Online]. Available: arXiv:2008.01701.
- [39] X. Xu, Y. Ma, and W. Sun, "Learning factorized weight matrix for joint image filtering," in *Proc. Int. Conf. Mach. Learn.*, 2020, pp. 1–9.
- [40] A. C. Bovik, "Automatic prediction of perceptual image and video quality," *Proc. IEEE*, vol. 101, no. 9, pp. 2008–2024, Sep. 2013.
- [41] X.-S. Zhang, S.-B. Gao, C.-Y. Li, and Y.-J. Li, "A retina inspired model for enhancing visibility of hazy images," *Front. Comput. Neurosci.*, vol. 9, p. 151, Dec. 2015.
- [42] C. O. Ancuti and C. Ancuti, "Single image dehazing by multi-scale fusion," *IEEE Trans. Image Process.*, vol. 22, no. 8, pp. 3271–3282, Aug. 2013.
- [43] C. Li, C. Guo, J. Guo, P. Han, H. Fu, and R. Cong, "PDR-net: Perception-inspired single image dehazing network with refinement," *IEEE Trans. Multimedia*, vol. 22, no. 3, pp. 704–716, Mar. 2020.
- [44] G. Ghiasi and C. C. Fowlkes, "Laplacian pyramid reconstruction and refinement for semantic segmentation," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 519–534.
- [45] E. L. Denton, S. Chintala, A. Szlam, and R. Fergus, "Deep generative image models using a Laplacian pyramid of adversarial networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 1486–1494.
- [46] W. Lai, J. Huang, N. Ahuja, and M. Yang, "Fast and accurate image super-resolution with deep Laplacian pyramid networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 11, pp. 2599–2613, Nov. 2019.
- [47] D. Yang and J. Sun, "Proximal dehaze-net: A prior learning-based deep network for single image dehazing," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 729–746.
- [48] C. O. Ancuti, C. Ancuti, R. Timofte, and C. De Vleeschouwer, "O-Haze: A dehazing benchmark with real hazy and haze-free outdoor images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2018, pp. 754–762.
- [49] G. Meng, Y. Wang, J. Duan, S. Xiang, and C. Pan, "Efficient image dehazing with boundary constraint and contextual regularization," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2013, pp. 617–624.
- [50] J. Zhang and D. Tao, "FAMED-Net: A fast and accurate multi-scale end-to-end dehazing network," *IEEE Trans. Image Process.*, vol. 29, pp. 72–84, 2020.
- [51] Y. Shao, L. Li, W. Ren, C. Gao, and N. Sang, "Domain adaptation for image dehazing," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2020, pp. 2805–2814.
- [52] X. Qin, Z. Wang, Y. Bai, X. Xie, and H. Xie, "FFA-Net: Feature fusion attention network for single image dehazing," in *Proc. AAAI Conf. Artif. Intell.*, Feb. 2020, pp. 11908–11915.
- [53] C. O. Ancuti, C. Ancuti, M. Sbert, and R. Timofte, "DenseHaze: A benchmark for image dehazing with dense-haze and haze-free images," 2019. [Online]. Available: arXiv:1904.02904.
- [54] B. Li *et al.*, "Benchmarking single image dehazing and beyond," *IEEE Trans. Image Process.*, vol. 28, no. 1, pp. 492–505, Jan. 2019.
- [55] Y. Zhang, L. Ding, and G. Sharma, "HazeRD: An outdoor scene dataset and benchmark for single image dehazing," in *Proc. IEEE Int. Conf. Image Process.*, 2017, pp. 3205–3209.
- [56] C. O. Ancuti, C. Ancuti, and R. Timofte, "NH-HAZE: An image dehazing benchmark with non-homogeneous hazy and haze-free images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2020, pp. 1798–1805.
- [57] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
- [58] C. Chen, M. N. Do, and J. Wang, "Robust image and video dehazing with visual artifact suppression via gradient residual minimization," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 576–591.
- [59] N. Silberman, D. Hoiem, P. Kohli, and R. Fergus, "Indoor segmentation and support inference from RGBD images," in *Proc. Eur. Conf. Comput. Vis.*, 2012, pp. 746–760.
- [60] T. Guo, V. Cherukuri, and V. Monga, "Dense123' color enhancement dehazing network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2019, pp. 2131–2139.
- [61] S. Bianco, L. Celona, F. Piccoli, and R. Schettini, "High-resolution single image dehazing using encoder-decoder architecture," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2019, pp. 1927–1935.
- [62] Q. Deng, Z. Huang, C.-C. Tsai, and C.-W. Lin, "HardGAN: A haze-aware representation distillation gan for single image dehazing," in *Proc. Eur. Conf. Comput. Vis.*, 2020, pp. 722–738.
- [63] J.-P. Tarel and N. Hautiere, "Fast visibility restoration from a single color or gray level image," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2009, pp. 2201–2208.
- [64] K. Zeng, J. Yu, R. Wang, C. Li, and D. Tao, "Coupled deep autoencoder for single image super-resolution," *IEEE Trans. Cybern.*, vol. 47, no. 1, pp. 27–37, Jan. 2017.
- [65] R. Lan, L. Sun, Z. Liu, H. Lu, C. Pang, and X. Luo, "MADNet: A fast and lightweight network for single-image super resolution," *IEEE Trans. Cybern.*, vol. 51, no. 3, pp. 1443–1453, Mar. 2021.
- [66] D. Lin, R. Zhang, Y. Ji, P. Li, and H. Huang, "SCN: Switchable context network for semantic segmentation of RGB-D images," *IEEE Trans. Cybern.*, vol. 50, no. 3, pp. 1120–1131, Mar. 2020.

Jingang Zhang received the B.S. degree from Xidian University, Xi'an, China, in 2005, the M.S. degree in communication engineering from Xidian University in 2008, and the Ph.D. degree in communication and information systems from the Chinese Academy of Sciences University, Beijing, China, in 2017.

He is an Associate Professor with the University of Chinese Academy of Sciences, Beijing, China, where he is the Executive Director of the Intelligent Imaging Center. He has been engaged in the research of computational optical imaging technology for more than ten years, presided over more than ten national and ministerial-level scientific research projects, such as the National Natural Science Foundation of China and the Joint Foundation Program of the Chinese Academy of Sciences for equipment prefeasibility study, more than eight authorized invention patents and four software copyrights, and published over 20 academic papers and coauthored an academic book *Vector-Based Mathematics and Geometry*. His research interests include image denosing, deblurring, and dehazing, image/video analysis and enhancement, and related high-level vision problems.

Wenqi Ren (Member, IEEE) received the Ph.D. degree from Tianjin University, Tianjin, China, in 2017.

He is an Assistant Professor with the Institute of Information Engineering, Chinese Academy of Sciences, Beijing, China. From 2015 to 2016, he was supported by the China Scholarship Council and worked with Prof. M.-H. Yang as a joint-training Ph.D. student with the Electrical Engineering and Computer Science Department, University of California at Merced, Merced, CA, USA. His research interests include image processing and related high-level vision problems.

Dr. Ren received the Tencent Rhino Bird Elite Graduate Program Scholarship in 2017 and the MSRA Star Track Program in 2018.

Shengdong Zhang received the B.S. degree in computer science and technology from Shanxi University, Taiyuan, China, in 2009, the M.S. degree in mechanics from Peking University, Beijing, China, in 2012, and the Ph.D. degree from Wuhan University, Wuhan, China, in 2020.

His research interests include deep learning, computer vision, and image processing.

He Zhang received the bachelor's degree in electronic engineering and its automation from Jiangnan University, Wuxi, China, in 2014, and the Ph.D. degree from the Department of Electrical and Computer Engineering, Rutgers, the State University of New Jersey, New Brunswick, NJ, USA, in 2018.

He is currently a Research Scientist II with Adobe Research, San Jose, CA, USA, where he is the main contributor for the Select Subject Portrait feature shipped in June 2020, and Photoshop Release, Refine Hair feature shipped in 2020 Adobe MAX. His research focuses on deep learning, sparse and low-rank representation, and image processing.

Yunfeng Nie (Member, IEEE) received the B.Sc. degree in mechanical engineering from the University of Science and Technology of China (VUB), Hefei, China, in 2009, the M.Sc. degree in optical engineering from the University of Chinese Academy of Sciences, Beijing, China, in 2012, and the Ph.D. degree in the topic of "advanced freeform optical imaging system" in VUB under EU's FP7 Marie Curie Programme "ADOPSY" with exchanges in LPI Co., Madrid, Spain, and FSU University, Jena, Germany, in 2018.

He has been a Full-Time Researcher with the Faculty of Engineering, Vrije Universiteit Brussel (VUB), Brussels, Belgium, since 2014.

Dr. Nie is a member of OSA and SPIE, and has been very active in computational imaging, opto-digital joint design, freeform optical design algorithms, and advanced optimization algorithms.

Zhe Xue received the B.S. degree in electronic engineering from the Civil Aviation University of China, Tianjin, China, in 2010, and the Ph.D. degree in computer science from the University of Chinese Academy of Sciences, Beijing, China, in 2017.

He is currently an Associate Professor with the School of Computer Science, Beijing University of Posts and Telecommunications, Beijing. His research interests include machine learning, data mining, and multimedia data analysis.

Xiaochun Cao (Senior Member, IEEE) received the B.E. and M.E. degrees in computer science from Beihang University, Beijing, China, in 1999 and 2002, respectively, and the Ph.D. degree in computer science from the University of Central Florida, Orlando, FL, USA, in 2006.

He has been a Professor with the State Key Laboratory of Information Security, Institute of Information Engineering, Chinese Academy of Sciences, Beijing, since 2012. After graduation, he spent about three years with ObjectVideo Inc., Reston, VA, USA, as a Research Scientist. From 2008 to 2012, he was a Professor with Tianjin University, Tianjin, China. He has authored and coauthored more than 120 journal and conference papers.

Prof. Caos dissertation was nominated for the University of Central Florida's university-level Outstanding Dissertation Award. In 2004 and 2010, he was a recipient of the Piero Zamperoni Best Student Paper Award at the International Conference on Pattern Recognition in 2004 and 2010. He is on the Editorial Board of the IEEE TRANSACTIONS ON IMAGE PROCESSING. He is a Fellow of the IET.