Single-Image Dehazing via Compositional Adversarial Network

Hongyuan Zhu, Yi Cheng, Xi Peng, Joey Tianyi Zhou, Zhao Kang, Shijian Lu, Zhiwen Fang, Liyuan Li, and Joo-Hwee Lim

***Abstract*—Single-image dehazing has been an important topic given the commonly occurred image degradation caused by adverse atmosphere aerosols. The key to haze removal relies on an accurate estimation of global air-light and the transmission map. Most existing methods estimate these two parameters using separate pipelines which reduces the efficiency and accumulates errors, thus leading to a suboptimal approximation, hurting the model interpretability, and degrading the performance. To address these issues, this article introduces a novel generative adversarial network (GAN) for single-image dehazing. The network consists of a novel compositional generator and a novel deeply supervised discriminator. The compositional generator is a densely connected network, which combines fine-scale and coarse-scale information. Benefiting from the new generator, our method can directly learn the physical parameters from data and recover clean images from hazy ones in an end-to-end manner. The proposed discriminator is deeply supervised, which enforces that the output of the generator to look similar to the clean images from low-level details to high-level structures. To the best of our knowledge, this is the first end-to-end generative adversarial model for image dehazing, which simultaneously outputs clean images, transmission maps, and air-lights. Extensive experiments show that our method remarkably outperforms the state-of-the-art methods. Furthermore, to facilitate future research, we create the HazeCOCO dataset which is currently the largest dataset for single-image dehazing.**

Manuscript received May 19, 2019; revised September 19, 2019 and October 30, 2019; accepted November 6, 2019. This work was supported in part by the Fundamental Research Funds for the Central Universities under Grant YJ201949 and Grant 2018SCUH0070, and in part by the National Natural Science Foundation of China under Grant 61702182, Grant 61806135, Grant 61625204, and Grant 61836006. This article was recommended by Associate Editor X. Li. *(Corresponding author: Zhiwen Fang.)*

H. Zhu is with the Institute for Infocomm Research, A*Star, Singapore, and also with A*AI, A*STAR CHEEM Program, Singapore (e-mail: zhuh@i2r.a-star.edu.sg).

Y. Cheng, L. Li, and J.-H. Lim are with the Institute for Infocomm Research, A*Star, Singapore (e-mail: cheng_yi@i2r.a-star.edu.sg; lyli@i2r.a-star.edu.sg; joohwee@i2r.a-star.edu.sg).

X. Peng is with the College of Computer Science, Sichuan University, Chengdu 610065, China (e-mail: pangsaai@gmail.com).

J. T. Zhou is with the Institute of High Performance Computing, A*Star, Singapore (e-mail: joey.tianyi.zhou@gmail.com).

Z. Kang is with the School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China (e-mail: zkang@uestc.edu.cn).

S. Lu is with the School of Computer Science and Engineering, Nanyang Technological University, Singapore (e-mail: shijian.lu@ntu.edu.sg).

Z. Fang is with the Guangdong Provincial Key Laboratory of Medical Image Processing, School of Biomedical Engineering, Southern Medical University, Guangzhou 510515, China, and also with the School of Energy and Mechanical-Electronic Engineering, Hunan University of Humanities, Science and Technology, Loudi 417000, China (e-mail: fzw310@gmail.com).

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCYB.2019.2955092

***Index Terms*—Image enhancement, image processing.**

I. Introduction

HAZE is a natural phenomenon which is caused by the flowing particles in the atmosphere which significantly influence the performance of the existing computer vision system. There, image-based haze removal is a highly practical problem with both academic and industrial values [1]–[5].

Different from most existing image restoration and enhancement tasks [6], [7], image dehazing aims to detect and remove haze, whose magnitude depends on the unknown scene radiance which changes with positions and materials of the scene objects, as well as the atmospheric light. Based on the above observation, the key to haze removal is estimating the atmospheric light magnitude and the so-called transmission map induced by the depth of the scene.

Existing methods formulate haze removal as an image composition problem and can be roughly grouped into two categories: 1) prior-based and 2) data-driven. Specifically, the prior-based methods estimate the factors (depth and atmospheric light) based on various priors or assumptions. Typical methods include, but are not limited to, dark channel prior (DCP) [8], [9]; glow prior [10]; and other priors [11]–[14], just to name a few. Much progress has been made in the past few years. However, existing methods assume the haze follows certain priors or assumptions, which are easily violated when the scene is complex.

To address the disadvantages of the prior-based methods, recent efforts have been devoted to develop the data-driven methods, especially, motivated by the success of deep learning [15], [16], some methods [17]–[19] have been proposed. The basic idea behind these methods is utilizing the deep neural networks rather than a shallow model [20], [21], such as convolutional neural networks (CNNs) [22], to learn discriminative features from raw data and regress the physical parameters. However, these methods show the suboptimal result given image recovery and parameter estimation are treated as two separate steps, which could accumulate errors.

Recently, the generative adversarial network (GAN) [23] has become popular in learning end-to-end mapping with applications to tasks, such as image generation [24], object detection [25], and semantic segmentation [26]. Therefore, it is interesting to explore bringing GAN to single-image dehazing. However, simply using GAN to generate dehazed

(a) (b) (c)

Fig. 1. Single-image dehazing aims to remove the visible haze from the input image and our method can recover the input image with faithful color and structure. (a) Input. (b) Ground truth. (c) Our result.

outputs without fully considering the nonuniform and signal-dependent nature of haze will lead to unsatisfactory results.

In this article, we propose a novel end-to-end single-image dehazing method called DehazeGAN, which is specifically designed based on the atmospheric scattering model. Our proposed method embraces adversarial learning for physical parameters learning to directly output the recovered image with an illustration of Fig. 1. Unlike conventional methods with separate physical parameter learning as [17] did, the DehazeGAN introduces a composition generator to simultaneously learn these parameters for better efficiency and recovering quality.

The major contributions of this article are summarized as follows.

1) The proposed method is a novel GAN which is specifically designed by incorporating the atmospheric scattering model, which is one of the first works to marry GAN and single-image dehazing.
2) We propose a compositional generator for explicitly learning the transmission map and global atmospheric light coefficient using the multiscale networks which combine coarse-scale and fine-scale information to produce recovered images with faithful color and structures.
3) We propose a novel deeply supervised discriminator to regularize the generated image as realistic as possible from low-level details to high-level structures.
4) We create and will release a synthesized dataset called HazeCOCO, the largest dataset for single-image dehazing to date, to facilitate future research in single-image dehazing.

II. Related Works

A. Single-Image Dehazing

Recently, the data-driven methods have become a major trend of single-image dehazing given that they can learn discriminative features for physical parameters (e.g., atmospheric light and transmission map). These methods could be categorized into two approaches: 1) sequential approach and 2) approximation approach.

The sequential approach (e.g., [17]) estimates atmospheric light and transmission map in separate optimization and learning steps; hence, the whole pipeline is heuristically optimized which can produce unsatisfactory results.

The approximation approach [18] performs image dehazing by using the surrogate parameter to replace the two aforementioned physical parameters to relax the learning difficulty; however, the approximation quality could be hard to estimate which results in suboptimal performance.

Different from the aforementioned approach, our method can simultaneously learn these parameters to optimize the recovering qualities.

B. Generative Adversarial Network

Recently, GAN can produce high-quality synthesis/transformed images by embracing a two-network architecture with a generator and a discriminator. The generator tries to produce results to fool the discriminator, while the discriminator learns to penalize the structural difference between the samples from generator and the samples from the reference group. Since the first work of GAN [23], various applications have been proposed, such as image generation [24], [27]; object detection [25]; etc.

This article is also related to GAN. Actually, it is one of the first works [28]–[30] that introduces adversarial learning into single-image dehazing. Specifically, Li *et al.* [28] proposed using a conditional GAN with a skip-connected generator for single-image dehazing, and Yang *et al.* [30] proposed using a variant of CycleGAN for dehazing without paired annotation. Pan *et al.* [29] proposed a dual network which can learn structure and details for image enhancement.

Different from these works and our conference version [31], we propose a novel densely connected composition generator which automatically learns transmission map and global air-light from data to composite with input image to produce clean recovered images. In addition, we also introduce a novel deeply supervised discriminator to regularize the output of the composition generator to look as realistic as the ground-truth clean images.

III. End-to-End Adversarial Dehazing

The proposed DehazeGAN is based on the following atmospheric scattering model [32]:

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

where the input hazy image $I(x)$ is a linear combination of global airlight A and clean image $J(x)$. The two coefficients are weighted by the transmission map $t(x)$, which is induced by the scene depth $d(x)$ via

$$t(x) = e^{-\beta d(x)} \tag{2}$$

and β is a scatter coefficient. To recover $J(x)$ from $I(x)$, the atmospheric light A and the transmission map $t(x)$ should be estimated. Therefore, an accurate estimation/prediction of these two parameters is the key for successful recovery of clean images.

Estimating A and $t(x)$ from a single image is an ill-posed problem, most existing methods estimate them sequentially with different handcraft priors, which accumulate errors with unsatisfactory performance. Recently, Li *et al.* [18] proposed the AOD-Net to avoid directly modeling A and t by using an intermediate variable K

$$J(x) = K(x)I(x) - K(x) + b. \tag{3}$$

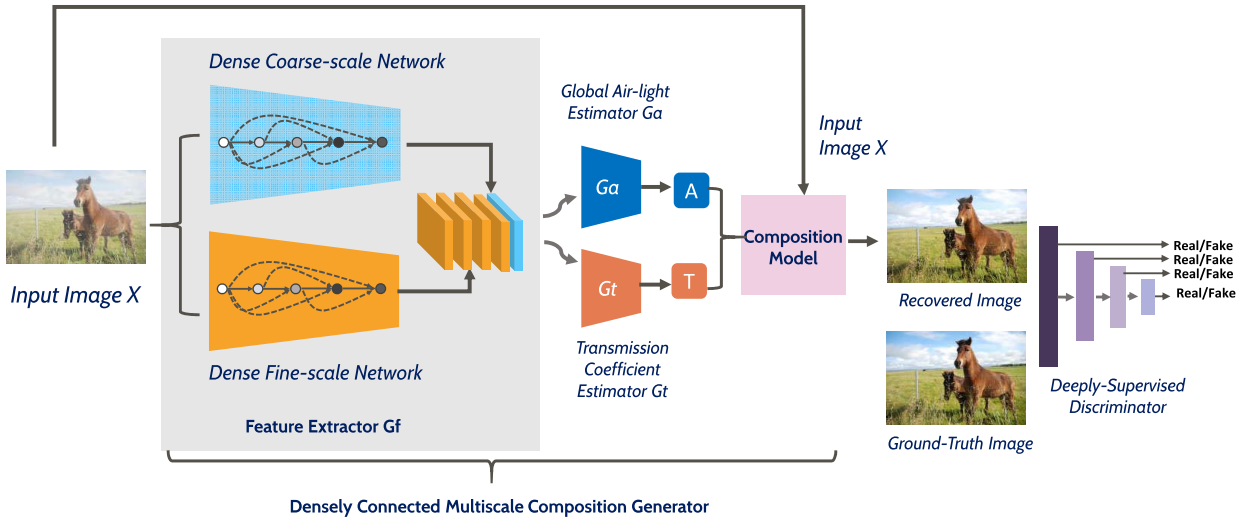


Fig. 2. Pipeline of our method. Our DehazeGAN consists of a multiscale composition generator and a deeply supervised discriminator. The composition generator is specifically designed to learn discriminative features for transmission matrix T and global atmospheric light coefficient A prediction, and then these two factors are further composited to generate the dehazed image. The deeply supervised discriminator helps regularize the generated recovered image as close as possible to the ground-truth image.

However, it is unclear whether $K(x)$ could approximate $[((1/t(x))(I(x)-A)+(A-b))/I(x)-1]$ so that (1) could be recovered.

In this article, we propose DehazeGAN, which consists of a multiscale composition generator and a deeply supervised discriminator, with an illustration in Fig. 2. The composition generator is specifically designed to learn discriminative features, which explicitly estimates the transmission coefficient $T=[t(x)]$ and atmospheric light coefficient A simultaneously. With these two factors, the dehazed image is recovered by

$$J(x)=\frac{I(x)-A}{t(x)}+A. \tag{4}$$

The deeply supervised generator consists of side-outputs at multiple scales, which helps to regularize the generated clean image such that it appears similar to the ground truth from low-level details to high-level structures.

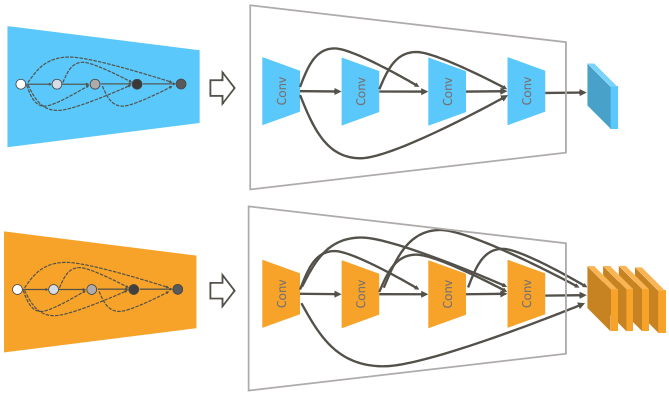


Fig. 3. Close-up of the coarse-scale branch (light blue) and fine-scale branch (light orange) of the feature extractor G_f.

A. Densely Connected Multiscale Composition Generator

GAN has become popular in recent image-to-image translation tasks. However, designing an effective generator which can end-to-end map a hazy image to a clean image is challenging due to two reasons. On the one hand, the traditional generator (e.g., the U-Net [33] commonly used in recent state-of-the-art GAN works [27], [34]) consists of several layers of feature extractions, and the color information may not be well preserved for producing faithful recovered images. On the other hand, haze is a kind of signal-dependent noise, which introduces further difficulties in learning the mapping if without modeling of the scene-dependent physical parameters.

In practice, a desirable generator should possess the following properties: 1) it should be able to learn discriminative features for dehazing; 2) it should involve the scene-dependent physical parameters; and 3) it should produce clean images with faithful color and structures. Based on these desiderata and the physical model of (1), we propose a new generator which consists of four modules, namely, a G_f for feature extraction G_f, a prediction branch G_t for transmission map prediction, another prediction branch G_a for global air-light estimation, and a compositional module use physical formulation (1) to recover the clean image.

To learn the discriminative features for dehazing, our feature extractor G_f embraces recently popular dense connection [35]. Specifically, G_f consists of two branches, namely, fine-scale branch and coarse-scale branch (see Fig. 3). The fine-scale branch applies densely connected modular with small kernels to capture fine-scale structures for recovering pleasant detail structures. The coarse-scale branch, on the other hand, uses larger kernels to capture longer range context to complement the fine-scale branch, thus leading to better estimation of transmission and air-light coefficients. The feature maps from these two branches are finally concatenated to provide the discriminative features for the transmission coefficient estimator G_t and global air-light estimator G_a.

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

Fig. 4. Visual comparison on the effectiveness of different generators. (a) Input. (b) U-Net generator (commonly used in recent GAN structures). (c) Our generator without composition components. (d) Our compositional generator (i.e., DehazeGAN).

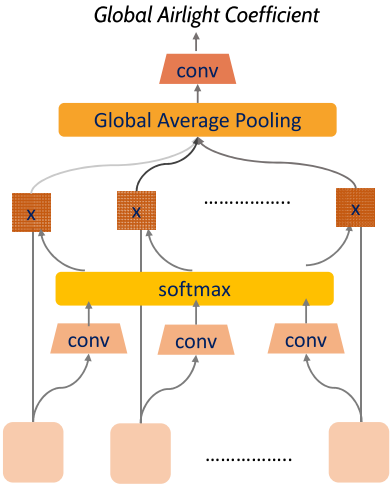


Fig. 5. 1-D close-up illustration of the global air-light estimator G_a. The round squares at the bottom represent the features from G_f, these features are weighted by attention modules (conv + softmax + element − wise product) and are further aggregated by global average pooling for air-light coefficient estimation.

The transmission map prediction branch G_t consists of a convolutional layer with the sigmoid function and outputs pixel wise transmission map t. The global atmosphere light estimator G_a is used to estimate the air-light coefficient A based on the global image features. Since not all regions are necessary for parameters estimation, we hence utilize a recent popular attention mechanism used in the nature language processing [36] so that the coefficient can be estimated according to the features in relevant regions. An illustration of G_a is shown in Fig. 5. After obtaining T and A, the dehazed image is produced by using the composition module in (1).

To show the effectiveness of our generator, Fig. 4 gives an illustration example. The only one difference among Fig. 4(b)–(d) demonstrates the used generator. Specifically, Fig. 4(b) adopts the commonly used U-Net generator in some popular GAN structures [27], [34], Fig. 4(c) is the result of using our generator without composition components (i.e., G_t and G_a are not considered), and Fig. 4(d) is the result of using our composition generator (i.e., DehazeGAN). One can observe that Fig. 4(b) and (c) gives the brownish and blurry outputs since they are not specifically designed for dehazing. In contrast, our composition generator explicitly estimates A and T with clean image composited at the last layer which could generate outputs with faithful color and structures.

B. Deeply Supervised Discriminator

GAN plays a min–max game so that the generator tries to generate examples as close as possible to the real examples for the purpose of fooling the discriminator. Recently, Isola *et al.* [27] proposed a patch-based discriminator, which achieves good performance in image translation tasks. Isola *et al.*'s discriminator is a shallowly supervised discriminator with supervision at the last layer. In other words, it prefers to penalize the differences in high-level structures.

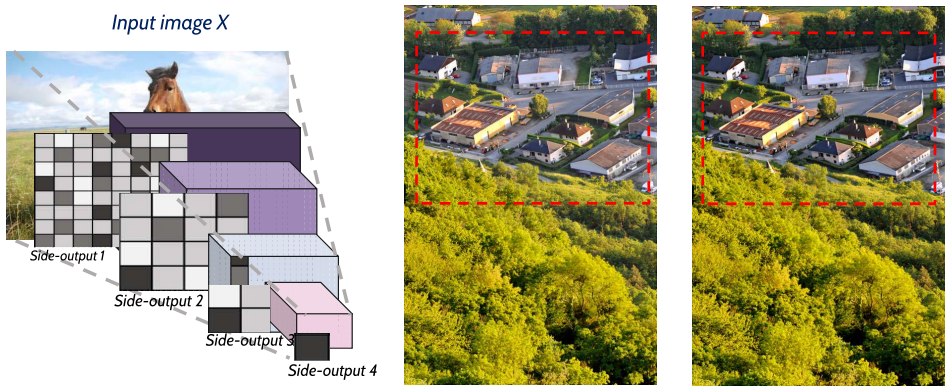
Different from image generation, the image enhancement tasks require recovering both details and structures. Hence, we propose a novel deeply supervised discriminator to fulfill this motivation. Our discriminator consists of four convolutional layers each with a stride of two, followed by batch normalization and leakly relu, which judges whether each patch in an image comes from the ground truth or the generator. Moreover, each convolutional feature map will also be passed into a sigmoid function to generate the classification prediction, hence, we will have four predictions which provide multiple level supervision to optimize our model. Fig. 6 shows the effectiveness of our deeply supervised discriminator, that is, it could yield a sharper image than the original discriminator with the supervision at only the last convolutional layer (i.e., shallowly supervised discriminator).

C. Objective Function

Our objective function consists of three terms: 1) the dehazing loss L_r; 2) the perceptual loss L_p; and 3) the adversarial learning loss L_g. These three terms are designed for the purpose of minimizing the reconstruction error, improving the perceptual quality and enhancing the details, respectively. In mathematics

$$\mathcal{L} = \mathcal{L}_r + \lambda \mathcal{L}_p + \gamma \mathcal{L}_g \tag{5}$$

where λ and γ are two tradeoff factors.



(a) (b) (c)

Fig. 6. (a) Illustration of our deeply supervised discriminator. Each convolutional layer is with a stride of two. Moreover, it has a probability side-output to indicate the similarities with the ground-truth. (b) Result by a shallowly supervised discriminator (only the last layer with supervision). (c) Result by our deeply supervised discriminator.

Dehazing Loss: We apply a mean-square-error loss to make output image I_h as close to ground-truth I_l

$$\mathcal{L}_r = \frac{1}{C \times W \times H} \sum_{i=1}^{W} \sum_{j=1}^{H} \sum_{c=1}^{C} \left\| I_h^{i,j,c} - I_l^{i,j,c} \right\|_2. \tag{6}$$

The W, H, and C are the width, height, and channel number of the input image I_h as in [31].

Perceptual Loss: To achieve better perceptual quality, we employ that the perceptual loss is modeled with a 19-layer VGG network [37], [38] pretrained on ImageNet. With the neural network Ψ, the perceptual loss actually measures the difference between the recovery and the ground truth in the feature space. In mathematics

$$\mathcal{L}_p = \frac{1}{C_f \times W_f \times H_f} \sum_{i=1}^{W_f} \sum_{j=1}^{H_f} \sum_{c=1}^{C_f} \left\| \Psi\left(I_h^{i,j,c}\right) - \Psi\left(I_l^{i,j,c}\right) \right\|_2 \tag{7}$$

where Ψ denotes a nonlinear transformation, and W_f, H_f, and C_f are the width, height, and channel number of the feature maps.

Multiscale Adversarial Loss: In addition to the content losses described above, we also consider the loss of adversarial learning at multiple layers to encourage the generator G to recover high-quality clean image $G(x)$ to fool the discriminator D. To improve the stability of the training phase, we employ the least-square loss [39] as the side-outputs of our discriminator's layers, that is

$$\mathcal{L}_g(G, D) = \sum_{i=1}^{K} E_{x,y}\left[(D_i(x, y))^2 \right] + E_x\left[(1 - D_i(x, G(x)))^2 \right] \tag{8}$$

where K is the layer number of discriminator, $D(x, y)$ is the conditional probability of the ground-truth image y and hazy image x, and $D(x, G(x))$ is the probability that the recovered image $G(x)$ is conditioned on x.

D. Implementation Details

Our model is implemented using Pytorch and trained on a Nvidia Titan X with the Adam optimizer [40] using a learning rate of 0.0002 and a batch size of 8. In experiments, we set $\lambda = 10^{-2}$ and $\gamma = 10^{-4}$ through the cross-validation.

The coarse-scale branch of feature extractor G_f is with the network structure of C(7, 3)–C(9, 3)–C(11, 3)–C(1, 3), where $C(k, l)$ denotes the convolution with a filter of size $k \times k$, an output channel number of l and a stride of 1 with ReLu output. The fine-scale branch of feature extractor G_f is with the structure of C(1, 3)–C(3, 3)–C(5, 3)–C(7, 3). The global air-light estimator G_a and the transmission map estimator G_t are with a convolution layer with 3-kernels to output air-light coefficient and transmission map, respectively. The deeply supervised discriminator consists of four convolutional layers, which is with the structure of C(3, 64)–C(3, 128)–C(3, 256)–C(3, 512), where each convolutional layer has a stride of two. Moreover, each side-output follows by an additional convolution layer of C(1, 1) with sigmoid to output the probability of each patch assigning the ground truth.

IV. Experiments

We quantitatively and qualitatively evaluate our method and compared with the other state-of-the-arts on our HazeCOCO dataset and natural hazy images.

A. HazeCOCO Dataset

It is very difficult to collect same visual scene with and without haze due to the changing environmental factors, such as lighting, wind, and temperature. Therefore, recent data-driven methods [17], [18] choose to synthesize haze images from clean ones using the indoor dataset with depth images, which is limited in terms of visual patterns in the real-world.

To learn more discriminative features for dehazing, we introduce the HazeCOCO dataset (see Fig. 7) which consists of 200 000 indoor images synthesized from the SUN-RGBD dataset [41] and 1 400 000 images synthesized from the COCO

Fig. 7. Image samples from the HazeCOCO dataset.

TABLE I
QUANTITATIVE STUDIES ON DIFFERENT GENERATOR

Metrics	U-Net	Our generator without composition	Our composition generator
PSNR	14.41	20.77	28.04
SSIM	0.47	0.7383	0.9273

TABLE II
QUANTITATIVE STUDIES ON GENERATOR WITH/WITHOUT COARSE-SCALE BRANCH

Metrics	Our generator without coarse-scale branch	Our composition generator
PSNR	27.29	28.04
SSIM	0.9146	0.9273

TABLE III
QUANTITATIVE COMPARISON BETWEEN SHALLOWLY SUPERVISED DISCRIMINATOR AND DEEPLY SUPERVISED DISCRIMINATOR

Metrics	Our discriminator with last layer supervision	Our deeply-supervised discriminator
PSNR	27.45	28.04
SSIM	0.9168	0.9273

TABLE IV
QUANTITATIVE STUDIES ON DIFFERENT LOSSES

Metrics	$\mathcal{L}_r$	$\mathcal{L}_r + \mathcal{L}_p$	$\mathcal{L}_r + \mathcal{L}_p + \mathcal{L}_g$
PSNR	27.72	27.91	28.04
SSIM	0.9143	0.9209	0.9273
# Parameters	Generator (G)	Discrimiantor (D)	VGG-16
	2238	276, 8705	763, 5264

dataset [42]. To the best of our knowledge, HazeCOCO is currently the largest dataset in the community.

As COCO is an RGB dataset, we use the method in [43] to predict the depth for the RGB images. After that, the haze images are synthesized using (1) as [17] did. Specifically, the atmospheric light $A = [k, k, k]$ is randomly sampled with a $k \in [0.6, 1.0]$ with a step-size of 0.1 and the β is sampled from $[0.4, 1.6]$ with a step-size of 0.2. With the haze image generated, we randomly use 5% images for testing, 10% for validation, and the remaining 85% for training.

B. Ablation Study

In this section, we investigate the performance of the alternative generator, discriminator, and loss function.

1) Generator Comparison: The qualitative comparison of different generator architecture has been demonstrated in Fig. 4. We report the quantitative performance of U-Net generator [27], [33], our generators with and without composition components on the HazeCOCO dataset. As illustrated in Table I, one can observe that our generator yields significant better PSNR and SSIM than the other two baselines.

The superior performance of our composition generator could be attributed to two reasons: 1) our method explicitly models the physical parameters to handle nonuniform, signal-dependent noise, which are largely neglected in existing generators and 2) our generator use composition module to recover faithful color and details.

In addition, we compare the network with and without combining the coarse-scale branch in Table II, one can observe that the PSNR and SSIM are improved thanks to the large context information captured by the coarse-scale branch.

2) Discriminator Comparison: We compare the performance of shallowly supervised discriminator and deeply supervised discriminator which utilize the supervision at the last layer and all layers, respectively. Table III shows that our deeply supervised discriminator remarkably improves the quality of dehazed images.

3) Loss Comparison: We also perform ablation study to study the effectiveness of dehazing loss $\mathcal{L}_r$, the perceptual loss $\mathcal{L}_p$, and the adversarial loss $\mathcal{L}_g$. Moreover, we also list the parameters of generator, discriminator and perceptual loss regularizer, one can see that discriminator has more parameters and perceptual loss has more parameters as higher complexity network can capture more regularity in the training data. The ablation study of loss on the HazeCOCO is shown in Table IV with Fig. 8 shows an visual examples from HazeCOCO. We also list the number of parameters used in Generator (G), Discriminator (D), and VGG-16 network used for perceptual loss in Table IV. One can see the VGG network and the discriminator makes up most of the parameters as they need higher capacity to help learn discriminative features for improving visual qualities.

From Fig. 8, one could see that after considering the perceptual loss $\mathcal{L}_p$, the haze located at dark areas is removed. With the additional adversarial loss $\mathcal{L}_g$, more high-frequency details are preserved. When all terms are used, our method yields the best performance as shown in Table IV.

4) Influence of Using Depth Prediction or Real Depth: Although synthesizing haze images from clean images by estimating the depth or using available depth dataset is popular in recent works, there could be some differences between using real depth maps and estimated depth maps to synthesize hazy images. To study the influence of depth estimation, we conduct experiments using images from the Make3D dataset [44] with the ground truth of outdoor depth like [43]. In particular, we synthesize two hazy images datasets, where one is using the predicted depth from [43] and the other is using the corresponding Make3Ds real depth. Using these two new datasets, we retrain the network and obtain corresponding

Fig. 8. Qualitative studies on different lossy functions. (a) Hazy. (b) L_r. (c) $L_r + L_p$. (d) $L_r + L_p + L_g$. (e) GroundTruth.

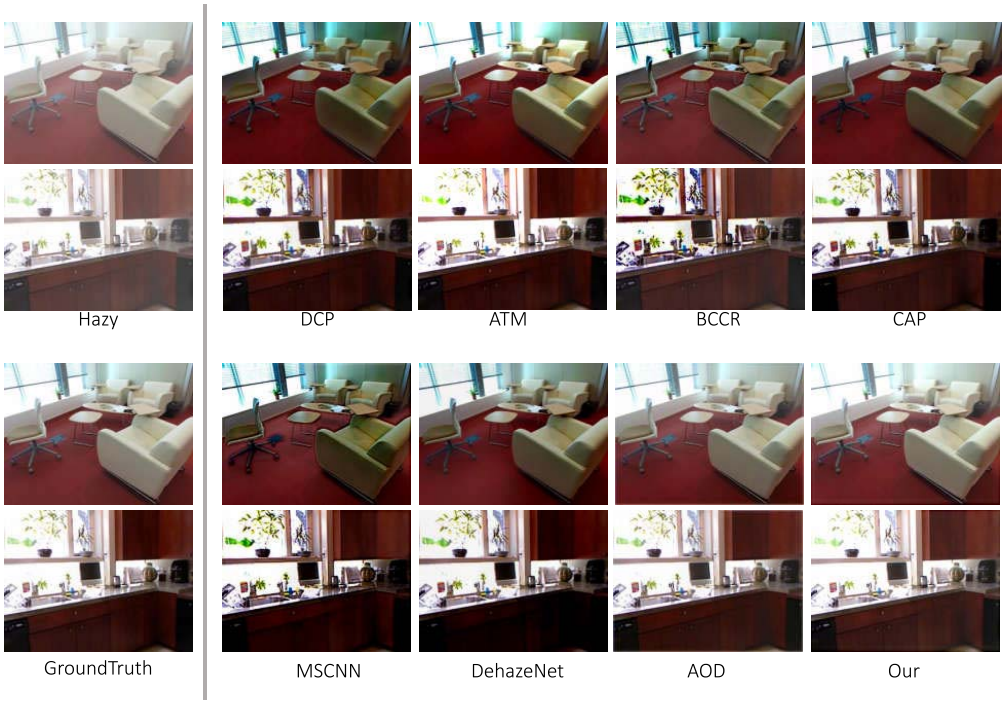


Fig. 9. Qualitative results on some randomly sampled HazeCOCO indoor testing images.

TABLE V
QUANTITATIVE STUDIES ON GENERATOR ON USING MAKE3D'S REAL DEPTH MAP AND PREDICTED DEPTH OF [43]

Metrics	Using real depth of Make3d	Using the predicted depth of [43]
PSNR	23.57	22.38
SSIM	0.8901	0.8628

results. Since the Make3D is much smaller than HazeCOCO, the performance of model degrades. The PSNR and SSIM of using the predicted depth from [43] and the outdoor depth of Make3D are as shown in Table V.

Table V demonstrates that there are some differences in the performance using predicted depth and ground-truth depth, but the results are comparable. Actually, our method is independent of the existing single image depth estimation and, thus, with the development of depth estimators, the performance of our method performance could be further improved.

C. Comparisons With State-of-the-Art

We compare our method with seven state-of-the-art methods on HazeCOCO, which can be divided into two groups: the first group consists of prior-based methods, for example, DCP [8], BCCR [13], ATM [45], and CAP [14]. The second group consists of data-driven approaches, for example, DehazeNet [19], MSCNN [17], and AOD-Net [18].

1) On Synthetic HazeCOCO Dataset: We first report all methods' average PSNR, SSIM, and running time on HazeCOCO in Table VI. Our method achieves superior performance in all metrics thanks to the physical modeling and adversarial learning.

Our method outperforms the second best methods by at least 1.66% and 2.0% in terms of PSNR and SSIM, respectively. On the outdoor dataset, our method again outperforms the AOD-Net by 2.06% and 1.74% in terms of PSNR and SSIM, respectively. Our method takes about 0.73 s to handle one image, which is the second fastest method. As the entire pipeline of our method and AOD-Net can be parallelized by GPU, hence these end-to-end methods are remarkably faster than the methods (DCP, BCCR, ATM, MSCNN, and DehazeNet) with separate estimation.

Some qualitative examples are provided in Figs. 9 and 10 with the following observations.

1) Prior-based ATM, BCCR, CAP, and DCP show color aberration which could attribute to the overestimation or underestimation of the transmission map.

Fig. 10. Qualitative results on some randomly sampled HazeCOCO outdoor testing images.

TABLE VI
AVERAGE PSNR, SSIM, AND RUNNING TIME ON HAZECOCO DATASET. THE RED INDICATES THE BEST RESULT AND THE BLUE INDICATES THE SECOND BEST RESULT

	Metrics	ATM	BCCR	CAP	DCP	MSCNN	DehazeNet	AOD-Net	Ours
Running Time	Seconds	35.19	1.77	0.81	18.38	1.70	1.81	**0.65**	**0.73**
Outdoor	PSNR	18.09	19.06	**23.90**	20.71	21.96	22.67	23.26	**25.32**
	SSIM	0.7751	0.7963	0.8822	0.8330	0.7725	0.8645	**0.8954**	**0.9197**
Indoor	PSNR	19.45	18.30	22.01	18.81	20.71	23.33	**26.38**	**28.04**
	SSIM	0.7651	0.7568	0.7863	0.7515	0.6428	0.7079	**0.9089**	**0.9273**

TABLE VII
MSE-ERROR ON TRANSMISSION MAP AND GLOBAL-AIR-LIGHT. THE **BOLD** FACE INDICATES THE BEST RESULT

Metrics	DCP	BCCR	ATM	CAP	DehazeNet	MSCNN	Ours
Average MSE Error on Transmission Map	0.0758	0.0521	0.1084	0.0431	0.0405	0.0446	**0.02356**
Average MSE Error on Air-light	0.0436	0.0931	0.3703	0.0389	0.0530	0.0521	**0.0608**

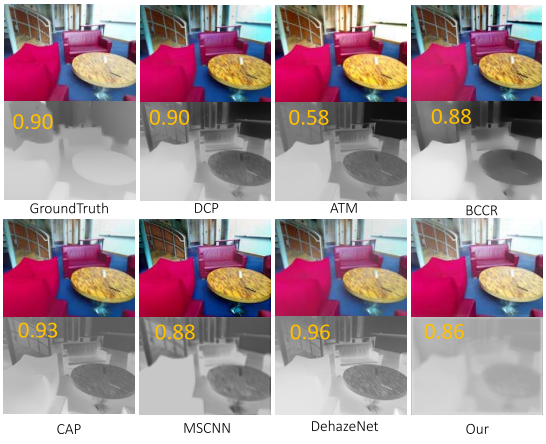


Fig. 11. Qualitative results with the generated transmission maps where the yellow number on the left-hand corner is the predicted global air-light coefficient.

2) DehazeNet, MSCNN, and AOD-Net show better results than the prior-based methods in some examples, some shadow areas still have haze as haze level of underestimation of haze level.

3) DehazeGAN yields the most nature recovery result, which suggests that our method can learn accurate physical parameters for superior recovery performance.

2) Quality of Predicted Transmission and Air-Light Coefficient: We also compare the estimated transmission and global air-light coefficient from different methods using average MSE error on the HazeCOCO dataset with the best performance highlighted in bold.

From Table VII, one can observe that our method achieves the best transmission map and a good air-light estimation, though these factors are intermediate result which are learned through data without using the ground truth. Note that DehazeNet and MSCNN have access to the pixel-level transmission map as ground truth.

In addition, we also attach the transmission maps estimated by different methods in Fig. 11. One could observe that our physical parameters learned from data are quite accurate to help produce results close to the ground truth.

3) Qualitative Comparison on Real-World Examples: We also perform comparison on some real-world hazy images

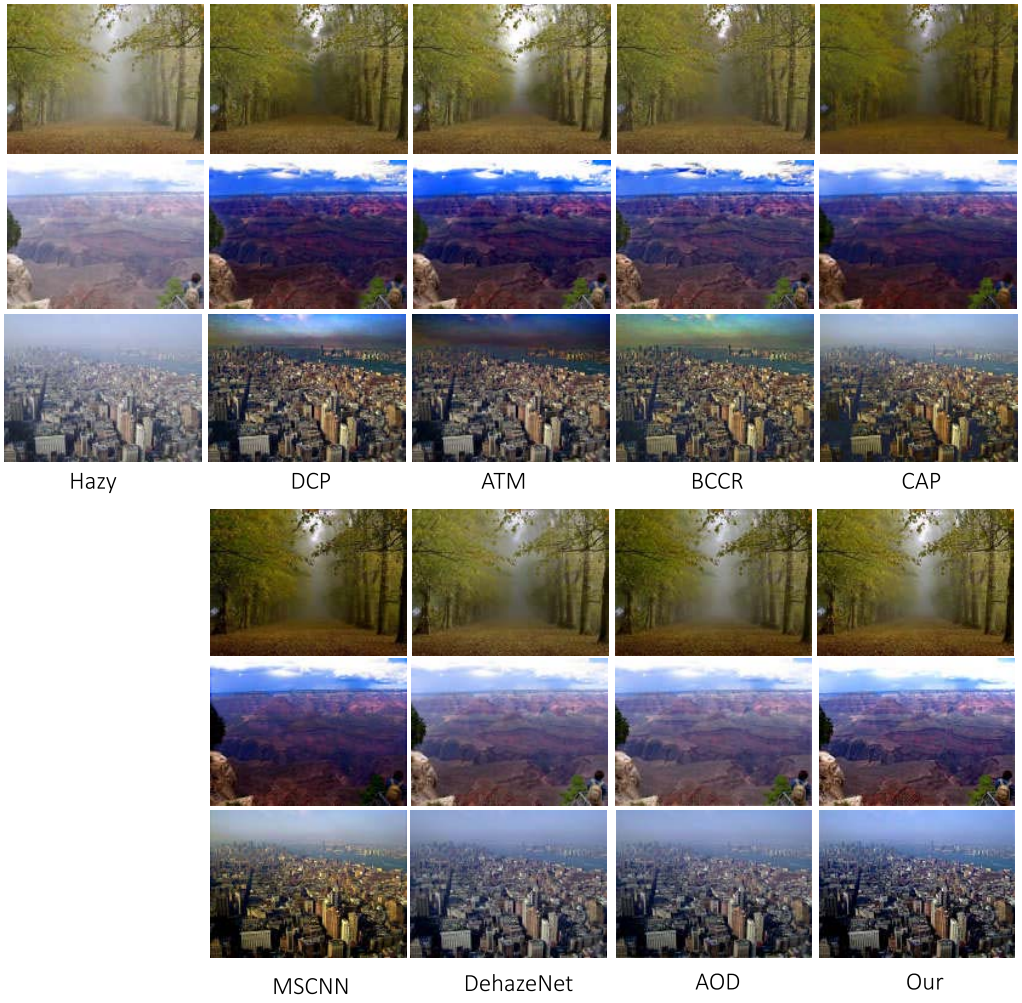


Fig. 12. Qualitative results on real images.

commonly used in [17] and [18] to show the existing methods generalization capability to real cases. From Fig. 12, the results of the prior-based methods (e.g., ATM, DCP, and BCCR) show color distortions. DehazeNet and AOD-Net underestimate the haze level of sky region which results in residual haze in sky area. On the other hand, the proposed method yields the best color and details thanks to the novel generator and discriminator.

V. Conclusion

In his article, we proposed DehazeGAN for single image haze removal. The proposed method uses a novel adversarial composition network to learn discriminative physical parameters for clean image recovery. To facilitate evaluation and comparison, we create HazeCOCO which is currently the largest dataset for single image dehazing currently. Extensive experiments have demonstrated the promising qualitative and quantitative performance of our method in the HazeCOCO dataset.

References

[1] W. Dong *et al.*, "Hyperspectral image super-resolution via non-negative structured sparse representation," *IEEE Trans. Image Process.*, vol. 25, no. 5, pp. 2337–2352, May 2016. [Online]. Available: https://doi.org/10.1109/TIP.2016.2542360

[2] X. Wang *et al.*, "Greedy batch-based minimum-cost flows for tracking multiple objects," *IEEE Trans. Image Process.*, vol. 26, no. 10, pp. 4765–4776, Oct. 2017. [Online]. Available: https://doi.org/10.1109/TIP.2017.2723239

[3] W. Zuo *et al.*, "Distance metric learning via iterated support vector machines," *IEEE Trans. Image Process.*, vol. 26, no. 10, pp. 4937–4950, Oct. 2017. [Online]. Available: https://doi.org/10.1109/TIP.2017.2725578

[4] Z. Wang, Y. Yang, Z. Wang, S. Chang, J. Yang, and T. S. Huang, "Learning super-resolution jointly from external and internal examples," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 4359–4371, Nov. 2015. [Online]. Available: https://doi.org/10.1109/TIP.2015.2462113

[5] Q. Wang, M. Chen, F. Nie, and X. Li, "Detecting coherent groups in crowd scenes by multiview clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, to be published.

[6] R. Liao and Z. Qin, "Image super-resolution using local learnable kernel regression," in *Proc. ACCV*, 2012, pp. 349–360.

[7] W. Zuo, D. Ren, D. Zhang, S. Gu, and L. Zhang, "Learning iteration-wise generalized shrinkage-thresholding operators for blind deconvolution," *IEEE Trans. Image Process.*, vol. 25, no. 4, pp. 1751–1764, Apr. 2016. [Online]. Available: https://doi.org/10.1109/TIP.2016.2531905

[8] 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.

[9] Y. Li, S. You, M. S. Brown, and R. T. Tan, "Haze visibility enhancement: A survey and quantitative benchmarking," *Comput. Vis. Image Understand.*, vol. 165, pp. 1–16, Dec. 2017.

[10] Y. Li, R. T. Tan, and M. S. Brown, "Nighttime haze removal with glow and multiple light colors," in *Proc. ICCV*, 2015, pp. 226–234.

[11] X. Fan, Y. Wang, R. Gao, and Z. Luo, "Haze editing with natural transmission," *Vis. Comput.*, vol. 32, no. 1, pp. 137–147, 2016. [Online]. Available: https://doi.org/10.1007/s00371-015-1083-1

[12] R. Fattal, "Single image dehazing," *ACM Trans. Graph.*, vol. 27, no. 3, p. 72, 2008.

[13] 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.*, Dec. 2013, pp. 617–624.

[14] 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.

[15] X. Peng, J. Feng, S. Xiao, W.-Y. Yau, J. T. Zhou, and S. Yang, "Structured autoencoders for subspace clustering," *IEEE Trans. Image Process.*, vol. 27, no. 10, pp. 5076–5086, Oct. 2018.

[16] Q. Wang, J. Gao, W. Lin, and Y. Yuan, "Learning from synthetic data for crowd counting in the wild," in *Proc. CVPR*, 2019, pp. 8198–8207.

[17] W. Ren, S. Liu, H. Zhang, J.-S. 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.

[18] 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. (ICCV)*, Venice, Italy, Oct. 2017, pp. 4780–4788.

[19] 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.

[20] Z. Huang, H. Zhu, J. T. Zhou, and X. Peng, "Multiple marginal fisher analysis," *IEEE Trans. Ind. Electron.*, vol. 66, no. 12, pp. 9798–9807, Dec. 2019.

[21] X. Peng, C. Lu, Y. Zhang, and H. Tang, "Connections between nuclear norm and Frobenius norm based representation," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 1, pp. 218–224, Jan. 2018.

[22] T. Zhang, X. Wang, X. Xu, and C. L. P. Chen, "GCB-Net: Graph convolutional broad network and its application in emotion recognition," *IEEE Trans. Affect. Comput.*, to be published.

[23] I. Goodfellow *et al.*, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, Montreal, QC, Canada, Dec. 2014, pp. 2672–2680.

[24] J. Yang, A. Kannan, D. Batra, and D. Parikh, "LR-GAN: Layered recursive generative adversarial networks for image generation," in *Proc. 5th Int. Conf. Learn. Represent. (ICLR)*, Toulon, France, Apr. 2017. [Online]. Available: https://openreview.net/forum?id=HJ1kmv9xx

[25] M. Zhang, K. T. Ma, J. H. Lim, Q. Zhao, and J. Feng, "Deep future gaze: Gaze anticipation on egocentric videos using adversarial networks," in *Proc. CVPR*, 2017, pp. 3539–3548.

[26] Q. Wang, J. Gao, and X. Li, "Weakly supervised adversarial domain adaptation for semantic segmentation in urban scenes," *IEEE Trans. Image Process.*, vol. 28, no. 9, pp. 4376–4386, Sep. 2019.

[27] P. Isola, J. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. CVPR*, 2017, pp. 5967–5976. [Online]. Available: https://doi.org/10.1109/CVPR.2017.632

[28] R. Li, J. Pan, Z. Li, and J. Tang, "Single image dehazing via conditional generative adversarial network," in *Proc. CVPR*, 2018, pp. 8202–8211.

[29] J. Pan *et al.*, "Learning dual convolutional neural networks for low-level vision," in *Proc. CVPR*, 2018, pp. 3070–3079.

[30] X. Yang, Z. Xu, and J. Luo, "Towards perceptual image dehazing by physics-based disentanglement and adversarial training," in *Proc. AAAI*, 2018, pp. 7485–7492.

[31] H. Zhu, X. Peng, V. Chandrasekhar, L. Li, and J.-H. Lim, "DehazeGAN: When image dehazing meets differential programming," in *Proc. IJCAI*, 2018, pp. 1234–1240.

[32] S. G. Narasimhan and S. K. Nayar, "Vision and the atmosphere," *Int. J. Comput. Vis.*, vol. 48, no. 3, pp. 233–254, Jul. 2002. [Online]. Available: https://doi.org/10.1023/A:1016328200723

[33] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. MICCAI*, 2015, pp. 234–241.

[34] J. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. ICCV*, 2017, pp. 2242–2251.

[35] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 2261–2269. [Online]. Available: https://doi.org/10.1109/CVPR.2017.243

[36] A. Vaswani *et al.*, "Attention is all you need," in *Proc. NIPS*, 2017, pp. 5998–6008.

[37] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. 3rd Int. Conf. Learn. Represent. (ICLR)*, San Diego, CA, USA, May 2015. [Online]. Available: http://arxiv.org/abs/1409.1556

[38] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *Proc. ECCV*, 2016, pp. 694–711.

[39] X. Mao, Q. Li, H. Xie, R. Y. K. Lau, Z. Wang, and S. P. Smolley, "Least squares generative adversarial networks," in *Proc. ICCV*, 2017, pp. 2813–2821.

[40] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. 3rd Int. Conf. Learn. Represent. (ICLR)*, San Diego, CA, USA, May 2015. [Online]. Available: http://arxiv.org/abs/1412.6980

[41] S. Song, S. P. Lichtenberg, and J. Xiao, "SUN RGB-D: A RGB-D scene understanding benchmark suite," in *Proc. CVPR*, 2015, pp. 567–576.

[42] T. Lin *et al.*, "Microsoft COCO: Common objects in context," in *Proc. ECCV*, 2014, pp. 740–755.

[43] F. Liu, C. Shen, G. Lin, and I. D. Reid, "Learning depth from single monocular images using deep convolutional neural fields," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 10, pp. 2024–2039, Feb. 2016. [Online]. Available: https://doi.org/10.1109/TPAMI.2015.2505283

[44] A. Saxena, S. H. Chung, and A. Y. Ng, "Learning depth from single monocular images," in *Proc. NIPS*, 2005, pp. 1161–1168.

[45] R. Fattal, "Dehazing using color-lines," *ACM Trans. Graph.*, vol. 34, no. 1, p. 13, 2014.