DehazeGAN: When Image Dehazing Meets Differential Programming

Hongyuan Zhu[1], **Xi Peng**[2*], **Vijay Chandrasekhar**[1], **Liyuan Li**[1], **Joo-Hwee Lim**[1]
[1] Institute for Infocomm Research, A*STAR, Singapore
[2] College of Computer Science, Sichuan University, China
{zhuh, vijay, lyli, joohwee}@i2r.a-star.edu.sg, pangsaai@gmail.com

Abstract

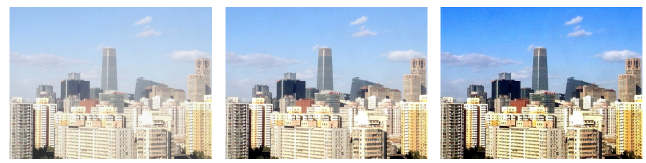
Single image dehazing has been a classic topic in computer vision for years. Motivated by the atmospheric scattering model, the key to satisfactory single image dehazing relies on an estimation of two physical parameters, i.e., the global atmospheric light and the transmission coefficient. Most existing methods employ a two-step pipeline to estimate these two parameters with heuristics which accumulate errors and compromise dehazing quality.

Inspired by differentiable programming, we reformulate the atmospheric scattering model into a novel generative adversarial network (DehazeGAN). Such a reformulation and adversarial learning allow the two parameters to be learned simultaneously and automatically from data by optimizing the final dehazing performance so that clean images with faithful color and structures are directly produced. Moreover, our reformulation also greatly improves the GAN's interpretability and quality for single image dehazing. To the best of our knowledge, our method is one of the first works to explore the connection among generative adversarial models, image dehazing, and differentiable programming, which advance the theories and application of these areas. Extensive experiments on synthetic and realistic data show that our method outperforms state-of-the-art methods in terms of PSNR, SSIM, and subjective visual quality.

1 Introduction

Haze is a typical atmospheric phenomenon in which dust, smoke, and other particles which greatly reduces the quality and visibility of captured images, thus making difficulty in further perception and understanding. Therefore, haze removal, especially, single image dehazing is highly practical and realistic with wide academic and industry value [Li *et al.*, 2017b; Zhang *et al.*, 2017b; Kang *et al.*, 2017a; 2017b; Qin *et al.*, 2016; Zhu *et al.*, 2018; 2016].

Almost all existing methods typically adopt a well-received physical model (see Section III for details) which is parametrized by the global atmospheric light and the pixel-wise transmission coefficient. To recover the transmission map, various prior-based methods have been proposed, e.g. constant albedo prior [Fattal, 2008], dark channel prior (DCP) [He *et al.*, 2011; Tang *et al.*, 2014], color-line prior [Fattal, 2014], boundary constraint [Meng *et al.*, 2013], statistically independent assumption [Nishino *et al.*, 2012], and color attenuation prior [Zhu *et al.*, 2015]. Despite the remarkable performance achieved by these methods, the adopted priors or assumptions are easily violated in practice, especially when the scene contains complex or irregular illumination or corruption.



(a) Input (b) Ground truth (c) Our result

Figure 1: A visual illustration of single image dehazing. The target is to recover a clean image from the input haze image. Our method produces a recovered image with rich details and vivid color information.

To overcome the disadvantages of these prior-based methods, recent focus has shifted to developing data-driven methods which are based on deep learning [Cai *et al.*, 2016; Ren *et al.*, 2016; Li *et al.*, 2017a; Yang *et al.*, 2017]. The basic idea of these methods is utilizing convolutional neural networks (CNNs) to explicitly learn discriminative features from raw data and regress some or all of physical parameters which are further used to recover the clean images. One major disadvantage of these methods is that they employ a two-step rather than end-to-end optimization to produce clean images. Hence the errors within these steps will accumulate, which further results in performance degradation.

On the other side, generative adversarial networks (GAN) have achieved remarkable progress in recent image-to-image translation tasks [Goodfellow *et al.*, 2014; Isola *et al.*, 2017] using a convolutional neural network as image generator and discriminator. Therefore, it is tempting to bridge GAN and singe image dehazing. However, image dehazing is different from other image restoration tasks as haze is a kind of nonuniform and signal-dependent noise. To be specific, the magnitude of haze depends on the depth between a surface

*Corresponding author: X. Peng

and camera, as well as the atmospheric light and the material of objects in the scene. Neglecting such compositional factors will lead to unsatisfactory performance if simply using GAN to generate dehazed outputs. Furthermore, it is difficult to get the ground-truth to learn these parameters as expensive external sensors are typically required. Recently, differentiable programming (DP) [Joey Tianyi Zhou and Goh, 2018] have become popular to formulate the optimization process as a recurrent neural network so that all parameters can be automatically learnt from data without using ground-truths, which have been widely applied in various tasks.

Based on the above observations, we propose a novel single image dehazing method by elaborately reformulating the atmospheric scattering model into a novel generative adversarial network (a.k.a DehazeGAN), inspired by and beyond existing differential programming. The proposed DehazeGAN works in an end-to-end manner, which not only learns the transmission map and the atmospheric light magnitude by embracing adversarial learning, but also explicitly outputs the recovered image. Instead of modeling the physical parameters in two steps as [Cai *et al.*, 2016; Ren *et al.*, 2016; Yang *et al.*, 2017] did, the DehazedGAN introduces a composition generator using convolutional neural network to simultaneously learn these parameters from raw data and further composite them together with hazy images to generate clean ones. The discriminator of DehazeGAN regularize the recovered image to have faithful color and structures.

The contributions of this work is given in following aspects. On the one hand, we specifically design a novel GAN for single image dehazing, which significantly improves the interpretability of GAN because the intermediate variables directly model two physical parameters in a data-driven way. **To the best of our knowledge, this is one of the first works to marry image dehazing and GAN.** On the other hand, this work remarkably advances the boundary of differentiable programming in theory and applications. To be specific, almost all existing differentiable programming studies recast an existing **optimization process** (e.g. L1-optimization) as a **recurrent neural network**, whereas this work directly models the **physical variables** as a **GAN**. Clearly, our idea is more close to the essence of differentiable programming, namely, treating the neural network as a language instead of a machine learning method and describing the physical world in it. Extensive experiments on synthesized and real hazy image datasets prove that our method can learn accurate intermediate parameters from data to recover clean images with faithful color and structures and achieved state-of-the-art dehazing performance.

2 Related Works

Our work mainly relates to three topics, i.e., single image dehazing, generative adversarial networks and differentiable programming which are briefly discussed in this section.

2.1 Single Image Dehazing

In very recent, interests in single image dehazing have shifted to data-driven methods which estimate the atmospheric light and the transmission map induced by depth from raw data without the help of priors. These approaches could be further divided into sequential method and approximation method. Sequential method [Cai *et al.*, 2016; Ren *et al.*, 2016; Yang *et al.*, 2017] first learns a mapping from hazy images to the transmission map and then estimates the atmospheric light using a heuristic approach. As the whole pipeline is not optimized for dehazing, the error in these two separate steps will accumulate and potentially amplify each other, thus resulting in undesirable performance. In recent, [Li *et al.*, 2017a] proposed an approximation method which absorbs the transmission map and the global atmospheric light coefficient into an intermediate parameter and adopts a neural network to learn it. As the approximation quality is not theoretically guaranteed, sub-optimal performance would be given.

Different from these existing works, we propose a holistic approach that can simultaneously learn these parameters including the recovered images by optimizing the final dehazing quality and preserving the perceptual details.

2.2 Generative Adversarial Networks

Recent developments have witnessed the promising performance of generative adversarial networks [Goodfellow *et al.*, 2014; Arjovsky *et al.*, 2017; Zhao *et al.*, 2017] in unsupervised learning [Peng *et al.*, 2017; 2016]. GAN implicitly learns rich distributions over various data such as images and text, whose basic idea is transforming the white noise (or other specified prior) through a parametric model to generate candidate samples with the help of a discriminator and a generator. By optimizing a minimax two-player game, the generator aims to learn the training data distribution, and the discriminator aims to judge that a sample comes from the training data or the generator. Inspired by the huge success of GANs, various works have been proposed, such as image super resolution [Ledig *et al.*, 2017], text2image [Zhang *et al.*, 2017a], image2image [Yi *et al.*, 2017] and etc.

Different from these works, this is one of the first works to introduce adversarial learning into single image dehazing. Our work is remarkably distinct from GAN and its variants [Isola *et al.*, 2017]. Specifically, [Isola *et al.*, 2017] proposes a generator using the U-Net architecture by directly mapping input images to output ones. As discussed in Introduction, such an architecture is useful for recovering signal independent noise, whereas the haze is dependent on the underlying scene depth and other physical factors. Hence, a generator without explicit modeling the signal-dependent parameters probably lead to unsatisfactory dehazing results.

2.3 Differentiable Programming

Our work belongs to the family of differentiable programming which treats the neural network as a language such that the physical phenomenon could be modeled and parametrized by a neural network, and the model is further optimized in a data-driven way. The first well-known work of differentiable programming may be the Learned ISTA (LISTA) [Gregor and LeCun, 2010] which unfolds a popular ℓ_1-solver (i.e. ISTA) as a simple RNN such that the number of layers corresponds to the iteration number and the weight corresponds to dictionary. LISTA-like paradigms have been explored and applied in a wide range of tasks, e.g. hashing [Wang *et al.*, 2016a],

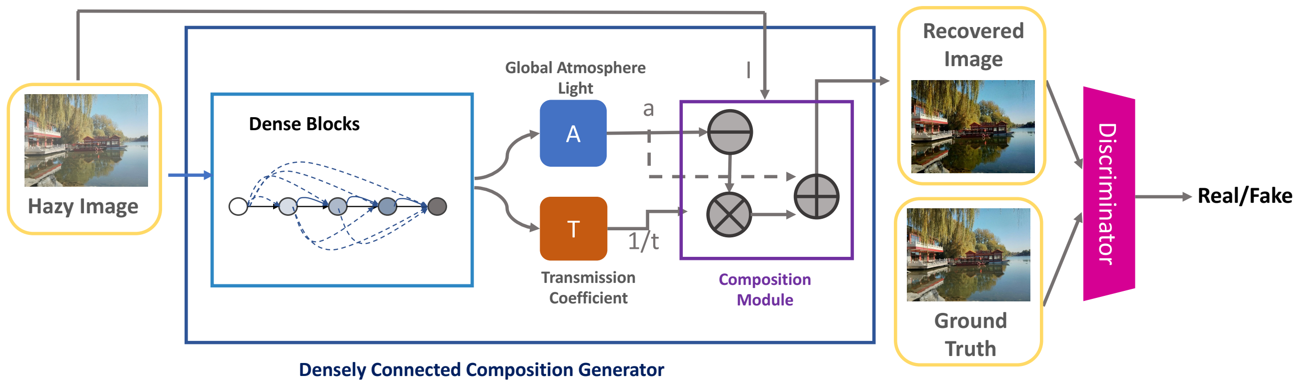


Figure 2: Pipeline of Our DehazeGAN.

classification [Wang *et al.*, 2016b], image restoration [Zuo *et al.*, 2015], data reconstruction [Joey Tianyi Zhou and Goh, 2018], etc.

Different from existing differentiable programming, our method reformulate the atmospheric scattering physical model instead of existing statistical inference models as a feed-forward convolutional neural network with prior knowledge rather than a recurrent neural network.

3 End-to-End Adversarial Dehazing

We first introduce the atmospheric scattering model based on which our DehazeGAN is specifically designed, and then further explain the architecture of our network.

3.1 Physical Model for DehazeGAN

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

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

where $I(x)$ denotes the observed hazy image, $J(x)$ denotes the corresponding clean image, A is the atmospheric light and the transmission map $t(x)$ is induced by the scene depth via

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

More specifically, $t(x)$ follows an exponential decay of the distance to the camera (i.e., $d(x)$) and β denotes the scatter coefficient. The formulation shows that if the atmospheric light A and the transmission map $t(x)$ are known, one could easily recover $J(x)$ for $I(x)$. In other words, the key of image dehazing is estimating A and $t(x)$ given $J(x)$.

To estimate A and $t(x)$, most existing methods employ an alternative optimization framework which first estimates $t(x)$ from $J(x)$ with various priors (e.g. dark channel), and then computes A by solving a regression problem. Such a two-step optimization may lead to accumulation of errors, thus resulting in undesirable recovery.

To overcome these drawbacks, we propose DehazeGAN (see Fig. 2) which proposes a novel composition generator. The generator is specifically designed to explicitly estimate the transmission matrix T and global atmospheric light coefficient A which are further composited to generate the dehazed image via:

$$J(x) = \frac{I(x) - A}{t(x)} + A \tag{3}$$

3.2 Network Architecture

Our novel generator consists of four modules, namely, a feature extractor G_f, a transmission map estimator G_t, a global atmospheric light estimator G_a and a compositional module. In details, G_f extracts rich features to support accurate estimation of T (i.e., t(x)) and A from G_t and G_a. The composition module absorbs the input image I and the obtained A and T to generate the dehazed image.

To embrace state-of-the-art neural network, our feature extractor G_f consists of four densely connected convolutional blocks [Huang *et al.*, 2017] is in the form of C(1)-C(3)-C(5)-C(7), where $C(k)$ denotes the convolution with a Relu function, a filter of size $k \times k$ and a stride of 1. Fig. 2 illustrates how the dense connections connect from lower layers to higher ones.

The transmission map estimator G_t is a fully convolutional network (FCN). To be specific, to estimate the pixel-wise transmission map t, a convolutional layer with the sigmoid function is added on the output of G_f. The global atmosphere light estimator is modeled as a classification network which first performs global average pooling on the output of G_f and then passes the obtained result into a fully connected layer with three neurons and the sigmoid function.

Our discriminator is similar to [Isola *et al.*, 2017], which consists of four convolutional layers with a stride of two, which classifies if each $N \times N$ patch in an image comes from the ground truth or the generator. Each convolutional feature map will be passed through a batch normalization layer and a leaky-relu to feed into the next convolutional layer. The motivation is that performing regularization to make the generated image as realistic as the ground-truth image in terms of low-level details and high-level structures. Fig. 3 shows the effectiveness of using discriminator in helping yield a sharper image.

Figure 3: Our method with and without adversarial learning.

3.3 Objective Function

The objective function of our method consists of two terms, i.e., the dehazing loss L_r and the adversarial learning loss L_g which are used to minimize the reconstruction error and enhance the details, respectively. In mathematical,

$$\mathcal{L} = \mathcal{L}_r + \gamma \mathcal{L}_g \tag{4}$$

where γ is a trade-off factor.

Dehazing Loss: To encourage the network recover the image as close as possible to the ground-truth, we minimize the discrepancy between the recovered image I_h and the ground-truth I_l via

$$\mathcal{L}_r = \frac{1}{C \times W \times H} \sum_{i=1}^{W} \sum_{j=1}^{H} \sum_{c=1}^{C} \| I_h^{i,j,c} - I_l^{i,j,c} \|_2 \tag{5}$$

The W, H, and C are the width, height, and channel number of the input image I_h.

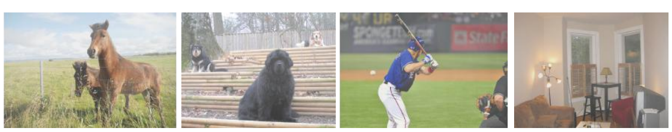


Figure 4: Image samples from the synthesized dataset.

Adversarial Loss In addition to the content losses described so far, we also consider the loss of adversarial learning. More specifically, it encourages the generator G to recover image $G(x)$ as realistic as the ground-truth image y so that the discriminator D is fooled. To the end, the loss is defined based on the probabilities of the discriminator overall training samples as:

$$\mathcal{L}_g(G, D) = E_{x,y}[\log D(x, y)] + E_x[\log[1 - D(x, G(x))]] \tag{6}$$

3.4 Implementation Details

The entire network is trained on a Nvidia Titan X GPU in PyTorch. For training, we employ the ADAM [Kingma and Ba, 2015] optimizer with a learning rate of 0.002 and a batch size of eight. We set $\gamma = 10^{-4}$ through the cross-validation.

4 Experiments

To demonstrate the effectiveness of our approach, we conduct experiments on both synthetic and natural hazy image datasets which are with a variety of haze conditions. On synthetic datasets, we quantitatively compare the proposed DehazeGAN and seven state of the arts on the indoor and outdoor image subsets of our synthesized dataset in terms of PSNR and SSIM. We also report the running time of our method and the baselines. On natural hazy image dataset, we provide qualitative results to illustrate our superior performance on generating perceptually pleasant recovered images.

4.1 Synthesized Dataset

Existing data-driven methods [Ren *et al.*, 2016; Li *et al.*, 2017a] are usually trained on synthesized images that are converted from RGB-D indoor images. However, the colors and texture patterns appeared in indoor images only take a small portion of the natural visual world, which may be insufficient to learning discriminative features for dehazing. Moreover, the depths of indoor images are relatively shallower than that of the natural scene.

To facilitate further research and benchmarking, we create the HazeCOCO dataset (see Fig. 4) which consists of 0.7 million synthetic indoor and outdoor images. The dataset is synthesized using the indoor images from the SUN-RGBD dataset [Song *et al.*, 2015], NYU-Depth dataset [Silberman *et al.*, 2012] and natural images from the COCO dataset [Lin *et al.*, 2014].

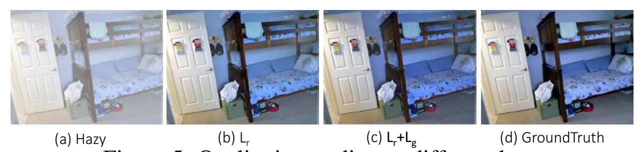


Figure 5: Qualitative studies on different loss.

Metrics	$\mathcal{L}_r$	$\mathcal{L}_r + \mathcal{L}_g$
PSNR	24.56	24.94
SSIM	0.8098	0.9169

Table 1: Quantitative studies on different losses.

As COCO does not contain the depth information, we generate the depth for the COCO images by using the method presented in [Liu *et al.*, 2016]. With the images with depth and clean images, we synthesize hazy images using the physical model of Eq. 1 as [Ren *et al.*, 2016] did. Specifically, we generate the random atmospheric light $A = [k, k, k]$ with $k \in [0.6, 1.0]$ and determine the value of β from $\{0.4, 0.6, 0.8, 1.0, 1.2, 1.4, 1.6\}$. After generating the hazy images, we randomly choose 85% data for training, 10% data for validation, and a small number of test images to form the indoor and outdoor subsets.

4.2 Baselines

Seven state-of-the-art methods are used as baselines in our experiments, which are divided into two groups: prior-based approach and data-driven approach. The first group consists of **DCP** [He *et al.*, 2011], **BCCR** [Meng *et al.*, 2013], **ATM** [Fattal, 2014], and **CAP** [Zhu *et al.*, 2015]. For the recent proposed data-driven approach, we investigate the performance of **DehazeNet** [Cai *et al.*, 2016], **MSCNN** [Ren *et al.*, 2016] and **AOD-Net** [Li *et al.*, 2017a].

4.3 Ablation Study

To better demonstrate the effectiveness of our objective function, we conduct an ablation study by considering the combinations of the proposed dehazing loss $\mathcal{L}_r$ and the adversarial loss $\mathcal{L}_g$. Figure 5 and Table 1 demonstrate qualitative and quantitative results on the HazeCOCO indoor testing data, respectively.

From Fig. 5, one could see that by further considering the adversarial loss $\mathcal{L}_g$, our method obtains images which are sharper and preserve more details. Table 1 shows that the performance of DehazeGAN is consistently improved when more terms are adopted.

4.4 Comparisons with State of the Arts

In this section, we conduct comparisons with seven recently proposed methods on synthetic testing data and a natural hazy dataset.

On synthetic dataset: Table 2 reports the average PSNR, SSIM and running time of all methods on the synthesized indoor and outdoor testing sets. From the result, one could see that our method consistently outperforms existing methods by a large margin thanks to our explicit physical modeling and adversarial learning.

On the indoor dataset, our method outperforms the other methods at least 1% in terms of PSNR. The gap between our method and the second best method (DCP) is about 7% in terms of SSIM. On the outdoor dataset, our method again outperforms the second best method by 1.04% and 2.15% in PSNR and SSIM, respectively.

In terms of running time, our method ranks the second best place, which takes about 0.72s for handling one image. Actually, one can observe that end-to-end methods (ours and AOD-Net) are remarkably faster than the off-the-shelf dehazing methods (DCP, BCCR, ATM, MSCNN and DehazeNet).

Fig. 6 provides a qualitative comparison on the synthesized Indoor and Outdoor testing dataset. One can observe that

- Prior-based methods such as DCP, ATM, and BCCR shows a strong color distortion. The potential reason for such a result may attribute to the inaccurate estimation of the transmission map.
- Although CAP, MSCNN, DehazeNet, and AOD-Net perform better than prior-based methods in quantitative comparisons, the output still contains haze in some scenarios. This could be attributed to their under-estimation of haze level.
- The proposed DehazeGAN shows the best look compared with the ground-truth, which suggests the physical parameters learned by our method is accurate to help recover the clean image. The perceptual loss and adversarial loss regularize the recovered image to have a faithful color with subtle details, as verified in Sec.4.3.

Comparisons on real dataset: To demonstrate the generalization ability of the proposed method, we evaluate DehazeGAN and other methods on three real-world hazy images used by previous works [Ren *et al.*, 2016; Li *et al.*, 2017a].

From Fig. 7, one could observe that DCP, ATM, and BCCR show color distortions in foreground regions and background sky. Moreover, the sky of recovered images given by DehazeNet and AOD-Net still contains haze, which could be blamed by their under-estimation of sky region's haze level. Overall, the proposed method avoids these issues and achieve the best visual result.

5 Conclusion

This paper proposed a novel method for end-to-end single image dehazing. The proposed DehazeGAN automatically learns the mappings between hazy images and clean images using a novel adversarial composition network. More interestingly, the atmospheric light and the transmission map are explicitly learned during optimizing our generator. Extensive experiments show the promising performance of our method in terms of PSNR, SSIM, running time and visual quality.

Acknowledgements

Xi Peng is supported by National Nature Science Foundation of China under grant No.61432012 and No.U1435213, and the Fundamental Research Funds for the Central Universities under grant No.YJ201748.

References

[Arjovsky *et al.*, 2017] Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein GAN. In *International Conference on Machine Learning*, Jan. 2017.

[Cai *et al.*, 2016] B. Cai, X. Xu, K. Jia, C. Qing, and D. Tao. Dehazenet: An end-to-end system for single image haze removal. *IEEE Transactions on Image Processing*, 25(11):5187–5198, Nov. 2016.

[Fattal, 2008] Raanan Fattal. Single image dehazing. *ACM transactions on graphics*, 27(3):72, 2008.

[Fattal, 2014] Raanan Fattal. Dehazing using color-lines. *ACM transactions on graphics*, 34(1):13, 2014.

[Goodfellow *et al.*, 2014] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative Adversarial Nets. In *Advances in Neural Information Processing Systems*, pages 2672–2680, Montreal, Canada, Dec. 2014.

[Gregor and LeCun, 2010] Karol Gregor and Yann LeCun. Learning fast approximations of sparse coding. In *ICML*, pages 399–406, USA, 2010.

[He *et al.*, 2011] Kaiming He, Jian Sun, and Xiaoou Tang. Single image haze removal using dark channel prior. *IEEE transactions on pattern analysis and machine intelligence*, 33(12):2341–2353, 2011.

[Huang *et al.*, 2017] Gao Huang, Zhuang Liu, Laurens van der Maaten, and Kilian Q. Weinberger. Densely connected convolutional networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017*, pages 2261–2269, 2017.

	Metrics	**DCP**	**BCCR**	**ATM**	**CAP**	**MSCNN**	**DehazeNet**	**AOD-Net**	Ours
Indoor	PSNR	19.67	18.00	18.19	20.67	**21.15**	20.48	20.03	**22.15**
	SSIM	**0.8098**	0.7512	0.7335	0.8092	0.8087	0.7739	0.7702	**0.8727**
Outdoor	PSNR	20.71	19.06	18.09	**23.90**	21.96	22.67	23.26	**24.94**
	SSIM	0.8330	0.7963	0.7751	0.8822	0.7725	0.8645	**0.8954**	**0.9169**
Running Time	Seconds	18.38	1.77	35.19	0.81	1.70	1.81	**0.65**	**0.72**

Table 2: Average PSNR, SSIM and Running Time on synthesized Indoor and Outdoor testing data. The red color indicates the best result and the blue color indicates the second best result.

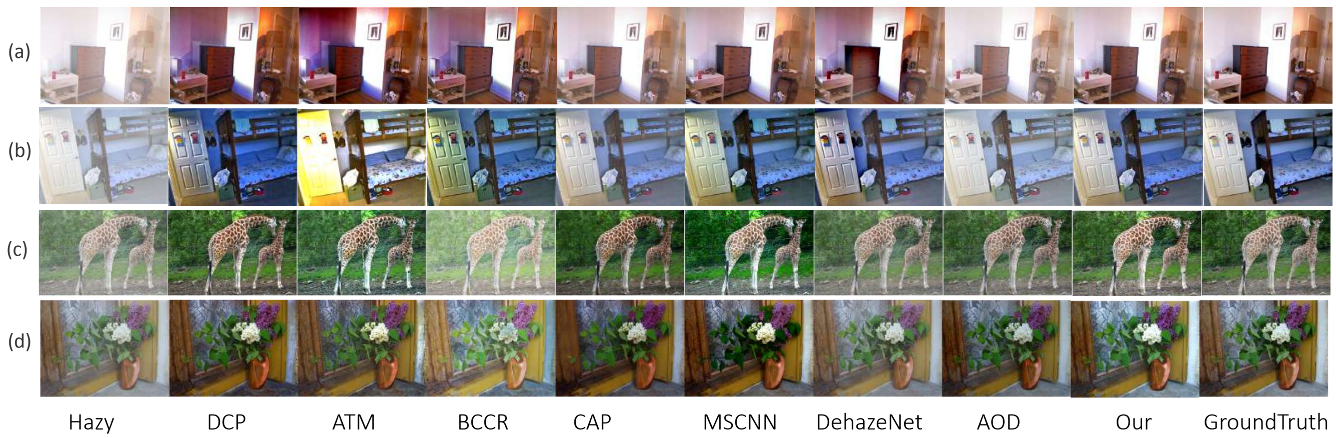


Figure 6: Qualitative Results on the Synthesized Testing Images.

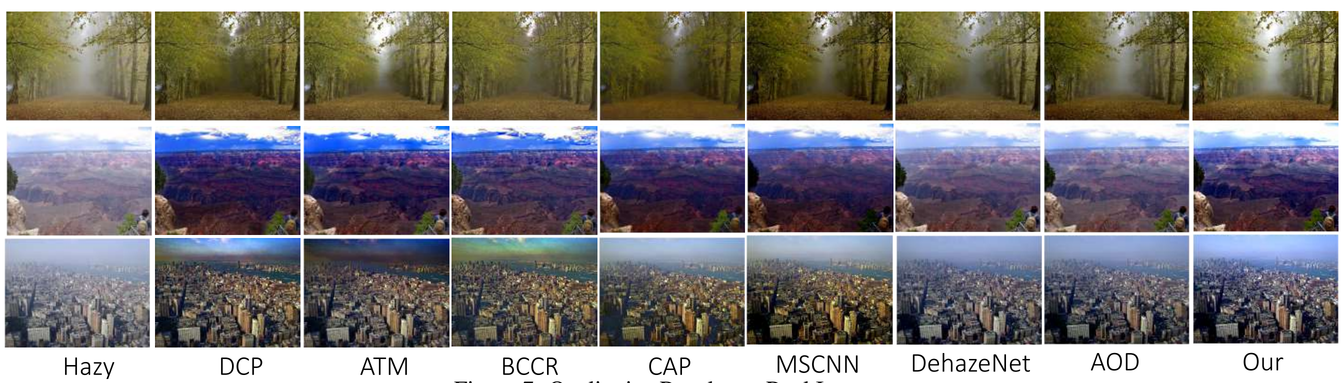


Figure 7: Qualitative Results on Real Images.

[Isola *et al.*, 2017] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. Image-to-image translation with conditional adversarial networks. In *CVPR*, pages 5967–5976, 2017.

[Joey Tianyi Zhou and Goh, 2018] Jiawei Du Xi Peng Hao Yang Sinno Jialin Pan Ivor W. Tsang Yong Liu Zheng Qin Joey Tianyi Zhou, Kai Di and Rick Siow Mong Goh. Sc2net: Sparse lstms for sparse coding. In *Proc. of 31th AAAI Conf. on Artif. Intell.*, pages –, New Orleans, Louisiana, Feb. 2018. AAAI.

[Kang *et al.*, 2017a] Zhao Kang, Chong Peng, and Qiang Cheng. Kernel-driven similarity learning. *Neurocomputing*, 267:210–219, 2017.

[Kang *et al.*, 2017b] Zhao Kang, Chong Peng, and Qiang Cheng. Twin learning for similarity and clustering: A unified kernel approach. In *AAAI*, pages 2080–2086, 2017.

[Kingma and Ba, 2015] Diederik P. Kingma and Jimmy Ba. May 2015.

[Ledig *et al.*, 2017] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi. Photo-realistic single image super-resolution using a generative adversarial network. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 105–114, Honolulu, HI, Jul. 2017.

[Li *et al.*, 2017a] B. Li, X. Peng, Z. Wang, J. Xu, and D. Feng. Aod-net: All-in-one dehazing network. In *IEEE International Conference on Computer Vision (ICCV)*, pages 4780–4788, Venice, Italy, Oct. 2017.

[Li *et al.*, 2017b] Boyi Li, Xiulian Peng, Zhangyang Wang, Jizheng Xu, and Dan Feng. An all-in-one network for dehazing and beyond. *CoRR*, abs/1707.06543, 2017.

[Lin *et al.*, 2014] Tsung-Yi Lin, Michael Maire, Serge J. Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. Microsoft COCO: common objects in context. In *ECCV*, pages 740–755, 2014.

[Liu *et al.*, 2016] Fayao Liu, Chunhua Shen, Guosheng Lin, and Ian D. Reid. Learning depth from single monocular images using deep convolutional neural fields. *IEEE Trans. Pattern Anal. Mach. Intell.*, 38(10):2024–2039, 2016.

[Meng *et al.*, 2013] G. Meng, Y. Wang, J. Duan, S. Xiang, and C. Pan. Efficient image dehazing with boundary constraint and contextual regularization. In *IEEE International Conference on Computer Vision*, pages 617–624, Dec. 2013.

[Nishino *et al.*, 2012] Ko Nishino, Louis Kratz, and Stephen Lombardi. Bayesian defogging. *International Journal of Computer Vision*, 98(3):263–278, Jul. 2012.

[Peng *et al.*, 2016] Xi Peng, Shijie Xiao, Jiashi Feng, Wei-Yun Yau, and Zhang Yi. Deep subspace clustering with sparsity prior. In *Proc. of 25th Int. Joint Conf. Artif. Intell.*, pages 1925–1931, New York, NY, USA, Jul. 2016.

[Peng *et al.*, 2017] Xi Peng, Jiashi Feng, Jiwen Lu, Wei-Yun Yau, and Zhang Yi. Cascade subspace clustering. In *Proc. of 31th AAAI Conf. on Artif. Intell.*, pages 2478–2484, SFO, USA, Feb. 2017. AAAI.

[Qin *et al.*, 2016] Zengchang Qin, Farhan Khawar, and Tao Wan. Collective game behavior learning with probabilistic graphical models. *Neurocomputing*, 194:74–86, 2016.

[Ren *et al.*, 2016] Wenqi Ren, Si Liu, Hua Zhang, Jinshan Pan, Xiaochun Cao, and Ming-Hsuan Yang. Single image dehazing via multi-scale convolutional neural networks. In Bastian Leibe, Jiri Matas, Nicu Sebe, and Max Welling, editors, *European Conference on Computer Vision*, pages 154–169, Cham, 2016. Springer International Publishing.

[Silberman *et al.*, 2012] Nathan Silberman, Derek Hoiem, Pushmeet Kohli, and Rob Fergus. Indoor segmentation and support inference from RGBD images. In *ECCV*, pages 746–760, 2012.

[Song *et al.*, 2015] Shuran Song, Samuel P. Lichtenberg, and Jianxiong Xiao. SUN RGB-D: A RGB-D scene understanding benchmark suite. In *CVPR*, pages 567–576, 2015.

[Tang *et al.*, 2014] K. Tang, J. Yang, and J. Wang. Investigating haze-relevant features in a learning framework for image dehazing. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2995–3002, Jun. 2014.

[Wang *et al.*, 2016a] Zhangyang Wang, Qing Ling, and Thomas S. Huang. Learning deep l_0 encoders. In *AAAI*, pages 2194–2200, 2016.

[Wang *et al.*, 2016b] Zhangyang Wang, Yingzhen Yang, Shiyu Chang, Qing Ling, and Thomas S. Huang. Learning A deep l_∞ encoder for hashing. In *IJCAI*, pages 2174–2180, 2016.

[Yang *et al.*, 2017] Hui Yang, Jinshan Pan, Qiong Yan, Wenxiu Sun, Jimmy Ren, and Yu-Wing Tai. Image dehazing using bilinear composition loss function. *arXiv preprint arXiv:1710.00279*, 2017.

[Yi *et al.*, 2017] Zili Yi, Hao Zhang, and Ping Tan Minglun Gong. DualGAN: Unsupervised Dual Learning for Image-to-Image Translation. In *IEEE International Conference on Computer Vision*, pages 502–510, Venice, Italy, October 2017.

[Zhang *et al.*, 2017a] Han Zhang, Tao Xu, Hongsheng Li, Shaoting Zhang, Xiaolei Huang, Xiaogang Wang, and Dimitris N Metaxas. StackGAN - Text to Photo-realistic Image Synthesis with Stacked Generative Adversarial Networks. In *IEEE International Conference on Computer Vision*, Venice, Italy, Oct. 2017.

[Zhang *et al.*, 2017b] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a gaussian denoiser: Residual learning of deep CNN for image denoising. *IEEE Trans. Image Processing*, 26(7):3142–3155, 2017.

[Zhao *et al.*, 2017] Junbo Zhao, Michael Mathieu, and Yann LeCun. Energy-based Generative Adversarial Network. In *International Conference on Learning Representations*, Toulon, France, Apr 2017.

[Zhu *et al.*, 2015] Q. Zhu, J. Mai, and L. Shao. A fast single image haze removal algorithm using color attenuation prior. *IEEE Transactions on Image Processing*, 24(11):3522–3533, Nov. 2015.

[Zhu *et al.*, 2016] Hongyuan Zhu, Jiangbo Lu, Jianfei Cai, Jianmin Zheng, Shijian Lu, and Nadia Magnenat-Thalmann. Multiple human identification and cosegmentation: A human-oriented CRF approach with poselets. *IEEE Trans. Multimedia*, 18(8):1516–1530, 2016.

[Zhu *et al.*, 2018] H. Zhu, R. Vial, S. Lu, X. Peng, H. Fu, Y. Tian, and X. Cao. Yotube: Searching action proposal via recurrent and static regression networks. *IEEE Transactions on Image Processing*, 27(6):2609–2622, June 2018.

[Zuo *et al.*, 2015] Wangmeng Zuo, Dongwei Ren, Shuhang Gu, Liang Lin, and Lei Zhang. Discriminative learning of iteration-wise priors for blind deconvolution. In *CVPR*, pages 3232–3240, Boston, MA, Jun. 2015.