

A Cascaded Convolutional Neural Network for Single Image Dehazing

Chongyi Li, *Student Member, IEEE*, Jichang Guo, Fatih Porikli, *Fellow, IEEE*, Huazhu Fu, and Yanwei Pang *Senior Member, IEEE*

Abstract—Images captured under outdoor scenes usually suffer from low contrast and limited visibility due to suspended atmospheric particles, which directly affects the quality of photos. Despite numerous image dehazing methods have been proposed, effective hazy image restoration remains a challenging problem. Existing learning-based methods usually predict the medium transmission by Convolutional Neural Networks (CNNs), but ignore the key global atmospheric light. Different from previous learning-based methods, we propose a flexible cascaded CNN for single hazy image restoration, which considers the medium transmission and global atmospheric light jointly by two task-driven subnetworks. Specifically, the medium transmission estimation subnetwork is inspired by the densely connected CNN while the global atmospheric light estimation subnetwork is a lightweight CNN. Besides, these two subnetworks are cascaded by sharing the common features. Finally, with the estimated model parameters, the haze-free image is obtained by the atmospheric scattering model inversion, which achieves more accurate and effective restoration performance. Qualitatively and quantitatively experimental results on the synthetic and real-world hazy images demonstrate that the proposed method effectively removes haze from such images, and outperforms several state-of-the-art dehazing methods.

Index Terms—Image dehazing, image degradation, image restoration, convolutional neural networks.

I. INTRODUCTION

DURING recent years, we have witnessed a rapid development of wireless network technologies and mobile devices equipped with various cameras which have revolutionized the way people take and share multimedia content [1], [2]. However, outdoor images (*e.g.*, Figure 1) often suffer from low contrast, obscured clarity, and faded colors due to the floating particles in the atmosphere, such as haze, fog, or dust, that absorb and scatter light. These degraded outdoor images not only affect the quality of photos [3] but also limit

This work was supported in part by the National Key Basic Research Program of China (2014CB340403), the Natural Science Foundation of Tianjin of China (15JCYBJC15500), the National Natural Science Foundation of China (61771334), and the program of China Scholarships Council (CSC) under the Grant CSC No. 201606250063.

Chongyi Li is with the School of Electrical and Information Engineering, Tianjin University, Tianjin, China and Research School of Engineering, College of Engineering and Computer Science, Australian National University, Canberra, ACT 0200, Australia (e-mail: lichongyi@tju.edu.cn).

Fatih Porikli is with the Australian National University, Canberra, ACT 0200, Australia (e-mail: fatih.porikli@gmail.com).

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

Huazhu Fu is with Institute for Infocomm Research, at Agency for Science, Technology and Research, Singapore (e-mail: huazhufu@gmail.com).

(Corresponding author: Jichang Guo.)



Fig. 1. Several examples of images taken under hazy or foggy scenes.

the applications in urban transportation [4], video analysis [5], visual surveillance [6], and driving assistance [7]. Therefore, image dehazing or image defogging has become a promising research area. Additionally, image dehazing methods also provide reference values for the underwater image enhancement and restoration research field [8], [9]. However, it is still a challenging task since the haze concentration is difficult to estimate from the unknown depth in the single image.

Single hazy image restoration methods usually need to estimate two key components in the hazy image formation model (*i.e.*, medium transmission and global atmospheric light). To achieve these two components, traditional prior-based methods either try to find new kinds of haze related priors or propose new ways to use them. However, haze related priors do not always hold, especially for the varying scenes. By contrast, to obtain more robust and accurate estimation, the learning-based methods explore the relations between the hazy images and the corresponding medium transmission in data-driven manner. However, most of the learning-based methods estimate the medium transmission and global atmospheric light separately, and do not consider the joint relations of them. In addition, separate estimation for the medium transmission and global atmospheric light limits the flexibility of previous methods. Thus, it inspires us to explore the joint relations between the medium transmission and the global atmospheric light, and how to directly map an input hazy image to its medium transmission and global atmospheric light simultaneously in pure data-driven manner.

Our contributions In this paper, we propose a cascaded CNN deep model for single image dehazing. Different from previous prior-based methods, we explore the relations between the input hazy images and the corresponding medium transmission in data-driven manner, which achieves more accurate and robust medium transmission. Compared to previous learning-based methods, to our best knowledge, we are first that estimate the medium transmission and global atmospheric light jointly in a cascaded CNN deep model, which advances in dehazing performance and flexibility. Additionally, com-

pared with the existing single image dehazing methods, the proposed method has superior dehazing performance both on perceptually and quantitatively.

The rest of this paper is organized as follows: Section II presents the related work. Section III describes the proposed method in detail. Section IV presents the experimental settings, investigates the network parameter settings, and gives the experimental results. Lastly, Section V concludes and discusses this paper.

II. RELATED WORK

Numerous image dehazing methods have been proposed in the recent decade [10]. These methods can be roughly classified into four categories: extra information-based methods [11]–[14], contrast enhancement-based methods [15]–[18], prior-based methods [19]–[28], and learning-based methods [29]–[33]. Though extra information-based methods can achieve impressive dehazing performance, they show limitations in real-life applications. In general, contrast enhancement-based methods produce under or over enhanced regions, color distortion, and artifacts due to failing to consider the formation principle of the hazy image and image degradation mechanism. As follows, we mainly introduce the prior-based and learning-based methods and summarize the existing problems.

Prior-based methods formulate some restrictions on the visual characteristics of hazy images to solve an ill-posed problem, which has made significant progress recently. Dark channel prior (DCP) method proposed by He *et al.* [20] is one of classical prior-based methods, which is based on statistics that at least one channel has some pixels with very low intensities in most of non-haze patches. Based on the DCP, the medium transmission and global atmospheric light are roughly estimated. Finally, the dehazed image is achieved by the estimated medium transmission refined by soft matting [35] or guided filter [36] as well as the estimated global atmospheric light according to an atmospheric scattering model. Although, the DCP method can obtain outstanding dehazing results in most cases, it tends to over-estimate the thickness of haze, which leads to color casts, especially for the sky regions. Subsequently, many strategies are applied to enhance the performance of the original DCP method. Zhu *et al.* [24] proposed a simple yet effective prior (*i.e.*, CAP) for image dehazing. The scene depth from the camera to the object of a hazy image is modeled in a linear model based on the CAP where unknown model parameters are estimated by a supervised learning strategy. Even though prior-based methods have achieved remarkable progress, they still have some limitations and need to be further improved. For instance, their performance is highly contingent on the accuracy of the estimated medium transmission and global atmospheric light, which is difficult to achieve when the priors are invalid. In addition, they also may entail high computation cost, which makes it infeasible for real-time applications.

With rapid development of learning technology in computer vision tasks [37], [38], the learning-based methods have been adopted in image dehazing. For example, Tang *et al.* [29] extracted multi-scale handcrafted haze-relevant features, and then

employed random forests regressor [39] to learn the correlation between the handcrafted features and the medium transmission. However, these handcrafted features are less effective and insufficient for some challenging scenes, which limits its performance. Generally, for the handcrafted features-based methods, inappropriate feature extraction often leads to poor dehazing results. Different from the handcrafted features, Cai *et al.* [30] proposed a CNN-based image dehazing method, named DehazeNet, which trained a regressor to predict the medium transmission. The DehazeNet includes four sequential operations, *i.e.*, feature extraction, multi-scale mapping, local extremum, and non-linear regression. The training dataset is generated by haze-free patches collected from Internet, random medium transmission value, and fixed global atmospheric light value (*i.e.*, 1) based on an atmospheric scattering model. With the optimized network weights, the medium transmission of an input hazy image can be estimated by network forward propagation. After that, the guided filtering [36] as post-processing is used to remove the blocking artifacts of the estimated medium transmission caused by the patch based estimation. Additionally, the authors applied an empirical method to estimate the global atmospheric light. Similar with DehazeNet [30], Ren *et al.* [31] designed a multi-scale CNN for single image dehazing. Recently, Li *et al.* [33] proposed an all-in-one deep model for single image dehazing, which directly generated the clean image using CNN. Additionally, such all-in-one network architecture has been extended to the video dehazing [34], which fills in the blank of video dehazing by deep learning strategies. For CNN-based methods, the accuracy of the estimated medium transmission and the dehazing performance need to be further improved, especially for varying scenes. Moreover, most of CNN-based methods estimate the global atmospheric light by the empirical methods, which limits the flexibility of network and the accuracy of restoration.

III. PROPOSED DEHAZING METHOD

To have a better understanding of our work, we first briefly review the atmospheric scattering model and then a detailed introduction of our cascaded CNN framework and the loss functions used in the optimization is presented. Lastly, we illustrate how to use the estimated medium transmission and global atmospheric light to achieve the haze-free image. More details are introduced as follows.

A. Atmospheric Scattering Model

Haze results from air pollution such as dust, smoke, and other dry particles that obscure the clarity of sky. Image captured under hazy or foggy day, only a part of the scene reflected light reaches the imaging equipment due to the effects of atmosphere absorption and scattering caused by haze, which decreases the visibility of scene, introduces the faded colors, and reduces the visual quality. Figure 2 briefly shows the formation of a hazy image, which mainly consists of two parts: direct attenuation and airlight.

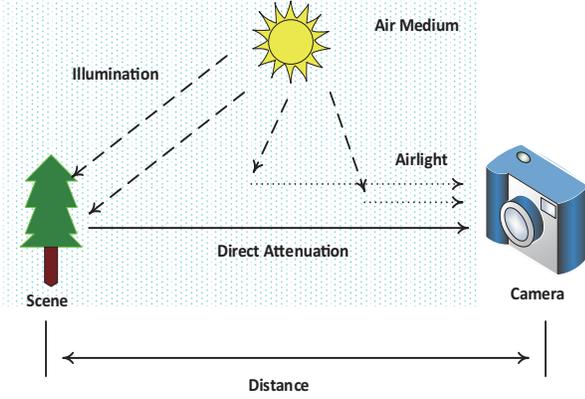


Fig. 2. Hazy image formation based on an atmospheric scattering model.

According to the atmospheric scattering model [40], a hazy image formation can be described as

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

where x denotes a pixel, $I(x)$ is the observed image, $J(x)$ is the haze-free image, $B(x)$ is the global atmospheric light, and $t(x) \in [0, 1]$ is the medium transmission which represents the percentage of the scene radiance reaching the camera. In Equation (1), the first term $J(x)t(x)$ is the direct attenuation, which represents how much the scene reflected light reaches the camera without scattering in the medium. The second term $B(x)(1-t(x))$ is airlight, which results from the scattered light caused by the ambient illumination and leads to the shift of the scene colors. The medium transmission $t(x)$ can be further expressed in an exponential decay term as

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

where β is the attenuation coefficient of the atmosphere and $d(x)$ is the distance from the scene to the camera. The purpose of single image dehazing is to restore $J(x)$, $B(x)$, and $t(x)$ from $I(x)$, which is an ill-posed problem.

B. The Proposed Cascaded CNN Framework

We aim to learn a cascaded CNN model that discerns the statistical relations between the hazy image and the corresponding medium transmission and global atmospheric light. The specific design of our cascaded CNN is presented in Figure 3 for a clear explanation.

In Figure 3, the cascaded CNN includes three parts that one is the shared hidden layers part, which extracts common features for subsequent subnetworks; one is the global atmospheric light estimation subnetwork, which takes the outputs of the shared hidden layers part as the inputs to map the global atmospheric light; one is medium transmission estimation subnetwork, which takes the outputs of the shared hidden layers part as the inputs to map the medium transmission. By such network architecture, our cascaded CNN can predict the global atmospheric light and medium transmission simultaneously.

The shared hidden layers part includes 4 convolutional layers with filter size of $f_i \times f_i \times n_i = 3 \times 3 \times 16$ followed by ReLU nonlinearity function [41]. Here, f_i is the spatial support

of a filter and n_i is the number of filters. Since we found that the task of the global atmospheric light estimation is easy for CNN, we employ a light-weight CNN architecture for the global atmospheric light estimation subnetwork. Specifically, the global atmospheric light estimation subnetwork includes 4 convolutional layers with filter size of $3 \times 3 \times 8$ followed by ReLU nonlinearity function [41], except for the last one. The medium transmission estimation subnetwork architecture is inspired by the densely connected network [42] which stacks early layers at the end of each block, which strengthens feature propagation and alleviates the vanishing-gradient problem. Specifically, the medium transmission estimation subnetwork includes 7 convolutional layers with filter size of $3 \times 3 \times 16$ followed by ReLU nonlinearity function [41], except for the last one. The network parameter settings will be discussed in Section IV. Next, we describe loss functions used in the cascaded CNN optimization.

C. Loss Functions

For image dehazing problem, most of learning-based methods employ Mean Squared Error (MSE) loss function for network optimization. Following previous methods, we also use MSE loss function for our medium transmission estimation subnetwork. For the convenience of training, we first assume that the format of the global atmospheric light is a map with dimension of M . Moreover, every pixel in the global atmospheric light map has the same value. Such assumption is reasonable because previous methods usually assume that every pixel in the input hazy image has the same global atmospheric light value. Then, for the global atmospheric light estimation subnetwork, we first tried MSE loss function, however, we found that the predicted global atmospheric light map is inconsistent with our assumption that every pixel in the input hazy image has the same global atmospheric light value. Thus, to avoid this problem, we use Structural Similarity Index (SSIM) loss function [43] for our global atmospheric light estimation subnetwork, which makes the values in the predicted global atmospheric light map same.

For the global atmospheric light estimation subnetwork, we minimize the SSIM loss function between the estimated global atmospheric light and the global atmospheric light ground truth. Firstly, the SSIM value for every pixel between the predicted global atmospheric light $F_{gal}(I_i)$ and the corresponding ground truth of the global atmospheric light B_i is calculated as follows:

$$SSIM(p) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \cdot \frac{2\sigma_{xy} + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}, \quad (3)$$

where x and y are the corresponding image patches with size 13×13 (default in the SSIM loss function [43]) in the predicted global atmospheric light and the corresponding ground truth, respectively. Above, p is the center pixel of image patch, μ_x is the mean of x , σ_x is the standard deviations of x , μ_y is the mean of y , σ_y is the standard deviations of y , σ_{xy} is the covariance between x and y . Using the defaults in the SSIM loss function [43], we set the values of C_1 and C_2 to 0.02 and 0.03. In fact, our network is insensitive to those parameters. Besides, F_{gal} is the learned global atmospheric light mapping

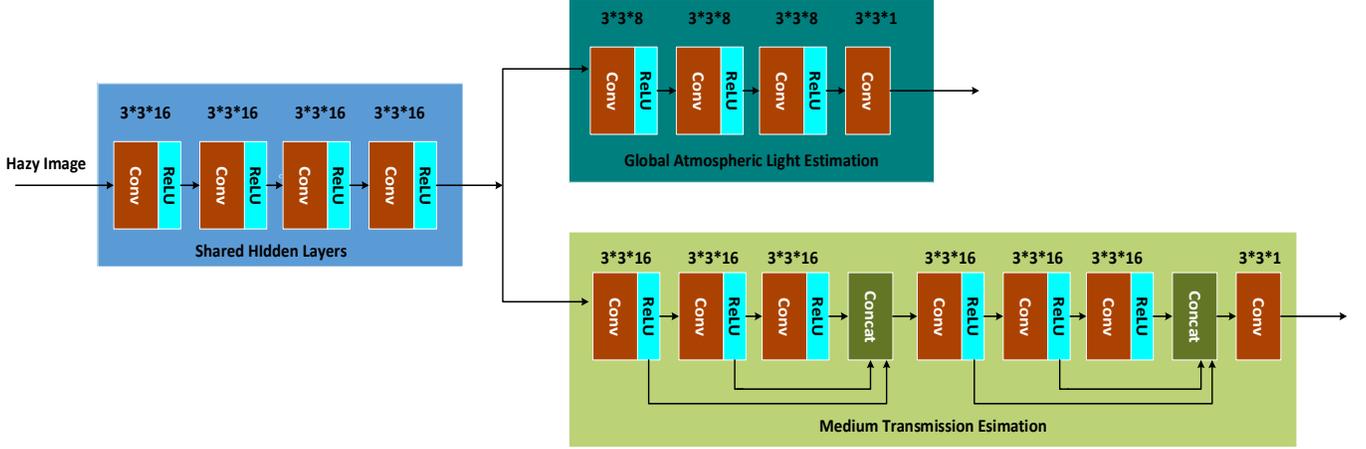


Fig. 3. The diagram of the proposed cascaded CNN structure. The cascaded CNN includes three parts: the shared hidden layers part, the global atmospheric light estimation subnetwork, and the medium transmission estimation subnetwork. In the network diagram, different color blocks represent the different operations. Red block Conv: convolution; light blue block ReLU: ReLU nonlinearity function; dark green block Concat: concatenation.

function. I_i is the input hazy image. Using Equation (3), the SSIM loss between the predicted global atmospheric light $F_{gal}(I_i)$ and the corresponding ground truth B_i is expressed as

$$L_{SSIM} = \frac{1}{N} \sum_{i=1}^N \left(1 - \frac{1}{M} \sum_{p=1}^M (SSIM(p))\right), \quad (4)$$

where N is the number of each batch, $M = H \times W$ is the dimension of the predicted global atmospheric light.

For the medium transmission estimation subnetwork, we minimize the MSE loss function between the predicted medium transmission $F_{mt}(I_i)$ and the corresponding ground truth of the medium transmission t_i , and is expressed as

$$L_{MSE} = \frac{1}{NHW} \sum_{i=1}^N \|F_{mt}(I_i) - t_i\|^2, \quad (5)$$

where N is the number of each batch, F_{mt} is the learned medium transmission mapping function, $H \times W$ is the dimension of the predicted medium transmission.

The final loss function for the cascaded CNN is the linear combination of the above-introduced losses with the following weights:

$$L_{total} = L_{SSIM} + L_{MSE}. \quad (6)$$

The blending weights are picked empirically based on preliminary experiments on the training data, which makes the contributions of SSIM loss and MSE loss same. In addition, these two subnetworks share the weights of the shared hidden layers part and are optimized jointly.

D. Haze Removal

Finally, with the achieved medium transmission and global atmospheric light, the haze-free image can be obtained by

$$J(x) = \frac{I(x) - B(x)}{t(x)} + B(x). \quad (7)$$

where $J(x)$ is the haze-free image, $I(x)$ is the input hazy image, $B(x)$ is the estimated atmospheric light, and $t(x)$ is the

estimated medium transmission refined by the guided image filtering [36]. For our results shown in this paper, the filter size of the guided image filtering is 33×33 .

In most of patch-based image dehazing methods, after estimating the coarse medium transmission, soft matting [35] or guided image filtering [36] is used to suppress the blocking artifacts. Different from these methods, we observed that our results also look pleasing even though we do not use refinement post-processing (please see Figure 4(c)). This might be because we optimize the proposed cascaded CNN using full-size images, which reduces the effects of blocking artifacts.

In contrast to the coarse medium transmission (please see Figure 4(b)), the refined medium transmission by the guided image filtering [36] is more smooth and unveils more structure information (please see Figure 4(d)). Compared with the results in Figure 4(c), the results in Figure 4(e) have better details and are without artifacts (please see the leaves and parterre). The guided image filtering refinement post-processing is beneficial to our final dehazing performance. Thus, our results shown in this paper are achieved using the refined medium transmission by the guided image filtering. Besides, we do not present the estimated global atmospheric light because it is hard to distinguish the estimated global atmospheric light in figure format. Generally, the accuracy of our global atmospheric light estimation reaches around 90% in spite of using a light-weight CNN architecture, which also indicates that the task of the global atmospheric light estimation for CNN is easy.

IV. EXPERIMENTS

In this section, we first describe the experimental settings. Then, the effects of network parameter settings are investigated. Finally, we compare the proposed method with several the state-of-the-art single image dehazing methods, such as regularization-based method (Meng *et al.* [21]), color attenuation prior method (Zhu *et al.* [24]), and recent CNN-based methods (Cai *et al.* [30] and Ren *et al.* [31]), on the synthetic and real-world hazy images. The results presented

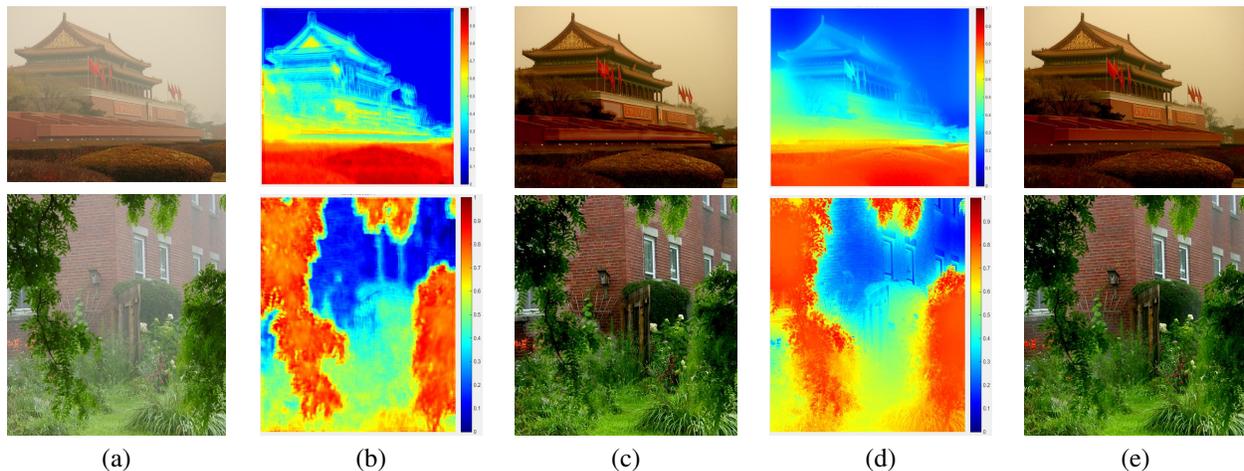


Fig. 4. Examples of our results. (a) Raw hazy images. (b) The medium transmission estimated by our cascaded network. (c) The dehazed results achieved by our method. (d) The medium transmission estimated by our cascaded network and refined by the guided image filtering [36]. (e) The dehazed results achieved by our method using the refined medium transmission. In the medium transmission, different color represents different values (red is close to 1 and blue is close to 0). (Best viewed on high-resolution display with zoom-in.)

in this paper are achieved by the source code provided by authors.

A. Experimental Settings

Dataset Deep learning strategy usually benefits from big data training. However, there is no easy way to have a amount of the labelled data for our network training. In order to train our cascaded CNN, we generate synthetic hazy images using an indoor RGB-D dataset based on Equation (1) and Equation (2).

Specifically, we assume that (i) the random global atmospheric light $B(x) \in [0.7, 1]$; (ii) the atmospheric attenuation coefficient β ranging from 0.6 to 2.8 (including haze thickness from light to heavy); (iii) the RGB channels of a hazy image have the same medium transmission and global atmospheric light values. Then, we divide NYU2 Depth dataset [44] into two parts: one part with 1300 RGB-D images for training data synthesis and another part with 101 RGB-D images for validation data synthesis. For each RGB-D image, we randomly select 5 global atmospheric light and atmospheric attenuation coefficient values to synthesize 5 hazy images. In this way, we synthesize a training set including 1300×5 training samples and a validation set including 101×5 validation samples. Those synthetic samples include hazy images with different haze concentration and light intensities as well as the corresponding medium transmission maps and global atmospheric light maps. We resize these samples to size 207×154 based on our limited memory. The depth images in the NYU2 Depth dataset have been normalized to $[0,1]$ by us. Figure 5 presents several synthetic hazy images, the corresponding medium transmission maps, and the haze-free images.

Implementation In the stage of training our network, the filter weights of each layer are initialized randomly from a Gaussian distribution, and the biases are set to 0. The learning rate is 0.001. The momentum parameter is set to 0.9. A batch-mode learning method with a batch size of 32 is applied. Our



Fig. 5. Synthetic samples. (a) Haze-free images from NYU2 Depth dataset [44]. (b) The synthetic medium transmission maps using the depth images from NYU2 Depth dataset [44] and the random atmospheric attenuation coefficient based on Equation (2). (c) Synthetic hazy images using (a), (b), and the random global atmospheric light based on Equation (1).

cascaded network is implemented in TensorFlow framework. Adam [45] is used to optimize our network. The network training with the basic parameter settings shown in Figure 3 is done on a PC with a Intel(R) i7-6700 CPU @3.40GHz and a Nvidia GTX 1080 Ti GPU.

B. Investigation of Network Parameter Settings

We mainly investigate the effects of the parameter settings of the shared hidden layers part and the medium transmission estimation subnetwork. Thus, we fix the network parameter settings of the global atmospheric light estimation subnetwork. We do not discuss the parameter settings of the global atmospheric light estimation subnetwork since it already has light enough network weights and reaches high accuracy of the

global atmospheric light estimation. The basic filter number of our network is 16, denoted as fn_{16} , except for the last layer of the medium transmission estimation subnetwork (*i.e.*, 1). The basic filter size used in our network is 3×3 , denoted as fs_3 . The network depth for the shared hidden layer part and the medium transmission estimation subnetwork is 4 and 7, respectively.

First, we fix other network parameter settings and then only modify one of them. Next, we denote another 3 filter numbers as fn_8 , fn_{32} , and fn_{64} and another 2 filter sizes as fs_5 and fs_7 . Besides, we denote another 3 network depth for the shared hidden layer part as ds_3 , ds_7 and ds_9 . After that, we denote another 2 network depth for the medium transmission estimation subnetwork as dm_3 and dm_5 . Here, dm_3 means 3 Concat blocks in the medium transmission estimation subnetwork. Our basic network architecture includes 2 Concat blocks. The final loss values on the validation dataset for different network parameter settings are summarized in Table I. The final loss value for our basic network parameter settings is marked in bold.

TABLE I
THE FINAL LOSS FOR DIFFERENT NETWORK PARAMETER SETTINGS

Network Depth	Filter Number	Filter Size	Loss
Basic	Basic	Basic	0.043
Basic	fn_8	Basic	0.059
Basic	fn_{32}	Basic	0.033
Basic	fn_{64}	Basic	0.018
Basic	Basic	fs_5	0.030
Basic	Basic	fs_7	0.028
ds_3	Basic	Basic	0.055
ds_7	Basic	Basic	0.031
ds_9	Basic	Basic	0.023
dm_3	Basic	Basic	0.029
dm_5	Basic	Basic	0.017

Finally, we use the above-mentioned basic parameter settings for our cascaded CNN based on its simplicity and efficiency. The cascaded CNN can have varied settings for accuracy and computation trade-off. With more filters, larger kernel sizes and more layers, we observed that the final loss is decreased. Nevertheless, the complexity, difficulty of training and running time are increased.

C. Comparisons on Synthetic Images

In this part, we compare our method with the state-of-the-art methods on synthetic hazy images. Firstly, we synthesize a hazy image testing dataset using the same approach with our training data generation. Several dehazed results and the estimated medium transmission on this testing dataset are shown in Figure 6 and Figure 7, respectively.

In Figure 6, it is obvious that our method can remove the haze on the input hazy images and restore the color and appearance. Moreover, our results are most close to the ground truth images. It is almost difficult to distinguish our results from the ground truth images. Meng *et al.* [21]’s method produces over-enhanced and over-saturated results, while Zhu *et al.* [24] and Cai *et al.* [30]’s methods have same dehazing performance which has less effect on the input

hazy images, especially for heavy haze. Ren *et al.* [31]’s method can remove the haze but still remains haze on several regions. Besides, we also present the corresponding medium transmission estimated by different methods in Figure 7. We do not show the medium transmission of Ren *et al.* [31] method because the code for medium transmission output is unavailable. In addition, to demonstrate the effectiveness of the refinement post-processing, we also show the coarse medium transmission directly estimated by our medium transmission estimation subnetwork.

As shown in Figure 7, all of the estimated medium transmission can indicate the concentration of haze in the hazy images. However, observing Figure 6, the dehazed results are different, which indicates that the accuracy of the global atmospheric light estimation is also significant for image dehazing. In addition, compared with the our coarse medium transmission, the refined medium transmission is more smooth, which leads to superior details and textures of our final dehazed results.

Furthermore, we apply the metrics of MSE, Peak Signal-to-Noise Ratio (PSNR) and SSIM [46] to quantitatively evaluate different methods. Besides, we also compare the running time (RT) for different methods. The compared methods are implemented in MATLAB and evaluated on the same machine with our model training. Our method is implemented in Python and our RT is calculated on the same machine with the compared methods but with GPU acceleration. Quantitative comparison is conducted on 50 synthetic hazy images which are generated by the same approach with our training data generation. Some of hazy images and the restored results have been shown in Figure 6. Table II summarizes the average values of the MSE, PSNR, SSIM, and RT for the image with average size of 640×480 . The values in bold represent the best results.

TABLE II
QUANTITATIVE RESULTS ON SYNTHETIC HAZY IMAGES IN TERMS OF MSE, PSNR, SSIM, AND RT.

Method	MSE	PSNR (dB)	SSIM	RT(s)
Meng <i>et al.</i> [21]’s	7.5742×10^3	9.2841	0.7942	2.5460
Zhu <i>et al.</i> [24]’s	6.5163×10^3	9.9908	0.8355	0.9486
Cai <i>et al.</i> [30]’s	5.1967×10^3	10.9635	0.8474	1.6905
Ren <i>et al.</i> [31]’s	1.2533×10^3	17.2971	0.8191	1.8785
Ours without refinement	1.0991×10^3	17.7249	0.8607	0.0936
Ours	958.1711	18.3298	0.8857	0.1029

As shown in Table II, our method outperforms the compared methods in terms of the average values of MSE, PSNR, SSIM and RT. Moreover, our method without refinement post-processing ranks second, which indicates that refinement post-processing is beneficial to final dehazing performance. The lowest MSE (highest PSNR) indicates that our results are most closed to the corresponding haze-free images in term of image content. The highest SSIM indicates that our results are most close to the corresponding haze-free images in term of image structure and texture. Besides, the speed of our method is faster than other methods because of GPU acceleration and our light-weight network parameter settings. Our method is fast enough for real-time applications.

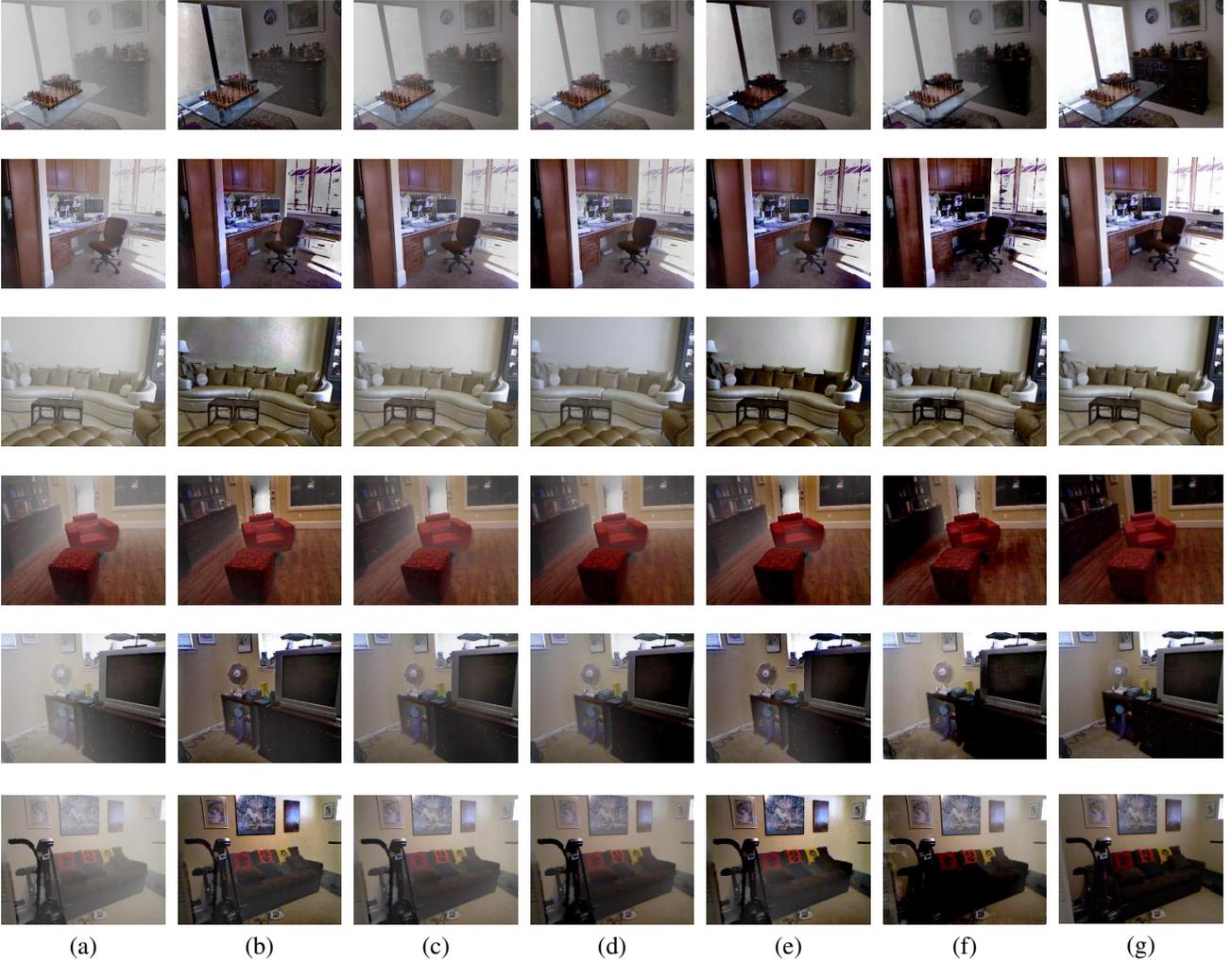


Fig. 6. Qualitative comparison on the synthetic hazy images generated by the same approach with our training data generation. (a) The synthetic hazy images. (b) The results of Meng *et al.* [21]. (c) The results of Zhu *et al.* [24]. (d) The results of Cai *et al.* [30]. (e) The results of Ren *et al.* [31]. (f) Our results. (g) The corresponding haze-free images.

D. Comparisons on Real-World Images

We conduct several comparisons on the real-world images to verify the performance of the proposed method. We first select several real-world hazy images which are usually used to qualitatively compare and hard to be handled. We compare the proposed method with the above-mentioned state-of-the-art methods in Figure 8. Additionally, the corresponding medium transmission is shown in Figure 9.

In Figure 8(b) and Figure 8(e), the results of Meng *et al.* [21] and Ren *et al.* [31] have over-enhanced regions and even introduce color deviation, (*e.g.*, the regions of sky) since these two methods tend to over-estimate the thickness of the haze and are sensitive to sky regions. In Figure 8(c) and Figure 8(d), the results of Zhu *et al.* [24] and Cai *et al.* [30] have significant improvement on the sky regions, but still have some remaining haze on the dense haze regions. Observing Figure 8(f), our method produces good dehazing performance in the challenging sky regions and our results have good contrast, vivid color and visually pleasing visibility, which benefits from data-driven non-linear regression.

This comparison results are in accordance with those of the synthetic hazy images. Although our cascaded CNN is trained on synthetic hazy images, the experimental results show that our method can be applied for real-world hazy images as well. To further illustrate the performance of different methods, we also present the corresponding medium transmission estimated by the above-mentioned methods in Figure 9. Observing Figure 9, all of the estimated medium transmission indicates the concentration of haze in the input hazy images. However, the final dehazed results are different, which demonstrates that the global atmospheric light as key component also has significant effects on final results, even for real-world hazy images. Thus, our good dehazing performance benefits from the joint estimation of the global atmospheric light and medium transmission. More results of our method on challenging hazy images are presented in Figure 10.

V. DISCUSSION AND CONCLUSION

In this paper, we have introduced a novel CNN model for single image dehazing. Inspired by the advances on big data

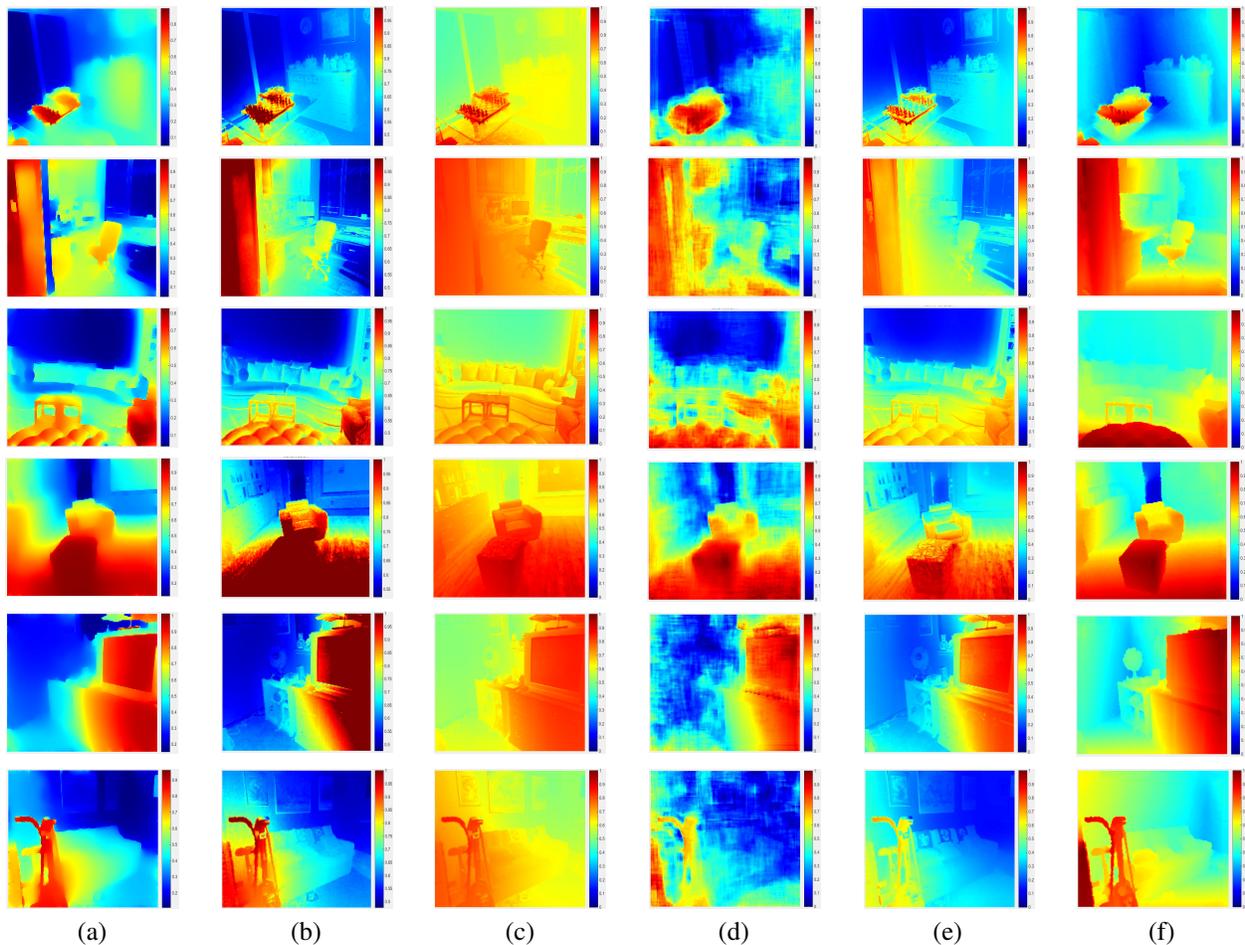


Fig. 7. Qualitative comparison on the estimated medium transmission. (a) The medium transmission estimated by Meng *et al.* [21]. (b) The medium transmission estimated by Zhu *et al.* [24]. (c) The medium transmission estimated by Cai *et al.* [30]. (d) The medium transmission estimated by our network. (e) The refined results using the guided image filtering [36] of (d). (f) The corresponding medium transmission ground truth.

driven low-level vision problems, we formulate a cascaded CNN with special design, which estimates the medium transmission and global atmospheric light jointly. Experimental results show that the proposed method outperforms the state-of-the-art methods both on the synthetic and real-world hazy images.

Regarding our method, the remaining question is that similar with most existing image dehazing methods, our method tends to amplify existing image artifacts because our training dataset is generated based on the atmospheric scattering model which does not take artifacts into account. For future work, we intend to suppress artifacts as an integral part in the proposed dehazing model. Additionally, we will investigate end-to-end networks for image dehazing where the networks directly produce haze-free results.

REFERENCES

- [1] W. Yin, T. Mei, C. Chen, and S. Li, "Socialized mobile photography: learning to photograph with social context via mobile devices," *IEEE Trans. Multimedia*, vol. 11, no. 1, pp. 184-200, 2014.
- [2] R. Cong, J. Lei, H. Fu, Q. Huang, X. Cao, and C. Hou, "Co-saliency detection for RGBD images based on multi-constraint feature matching and cross label propagation," *IEEE Trans. Image Process.*, vol. 27, no. 2, pp. 568-579, 2018.
- [3] X. Tian, Z. Dong, K. Yang, and T. Mei, "Query-dependent aesthetic model with deep learning for photo quality assessment," *IEEE Trans. Multimedia*, vol. 17, no. 11, pp. 2035-2048, 2015.
- [4] S. Huang, B. Chen, and Y. Cheng, "An efficient visibility enhancement algorithm for road scenes captured by intelligent transportation systems," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 5, pp. 2321-2332, 2014.
- [5] Z. Zhang and D. Tao, "Slow feature analysis for human action recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 3, pp. 436-450, 2012.
- [6] B. Tian, Y. Li, B. Li, and D. Wen, "Rear-view vehicle detection and tracking by combining multiple parts for complex urban surveillance," *IEEE Trans. Intell. Transport. Syst.*, vol. 15, no. 2, pp. 597-606, 2014.
- [7] M. Negru, S. Nedevschi, and R. Peter, "Exponential contrast restoration in fog conditions for driving assistance," *IEEE Trans. Intell. Transport. Syst.*, vol. 16, no. 4, pp. 2257-2268, 2015.
- [8] C. Li, J. Guo, R. Cong, Y. Pang, and B. Wang, "Underwater image enhancement by dehazing with minimum information loss and histogram distribution prior," *IEEE Trans. Image Process.*, vol. 25, no. 12, pp. 5664-5677, 2016.
- [9] C. Li, J. Guo, C. Guo, R. Cong, and J. Gong, "A hybrid method for underwater image correction," *Pattern Recognition Letters*, vol. 94, no. 15, pp. 62-67, 2017.
- [10] Y. Li, S. You, M. Brown, and R. Tan, "Haze visibility enhancement: a survey and quantitative benchmarking," *arXiv preprint arXiv:1607.06235*, 2016.
- [11] Y. Schechner, S. Narasimhan, and S. Nayar, "Instant dehazing of images using polarization," in *Proc. of IEEE Int. Conf. Comput. Vis. Pattern Rec. (CVPR)*, 2001, pp. 325-332.
- [12] S. Narasimhan and S. Nayar, "Contrast restoration of weather degraded



Fig. 8. Qualitative comparison on the real-world hazy images. (a) Real-world hazy images. (b) The results of Meng *et al.* [21]. (c) The results of Zhu *et al.* [24]. (d) The results of Cai *et al.* [30]. (e) The results of Ren *et al.* [31]. (f) Our results. (Best viewed on high-resolution display with zoom-in.)

- images,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 6, pp. 713-724, 2003.
- [13] L. Caraffa and J. P. Tarel, “Stereo reconstruction and contrast restoration in daytime fog,” in *Proc. of Asian Conf. Comput. Vis. (ACCV)*, 2012, pp. 12-25.
- [14] Z. Li, P. Tan., R. Tan, S. Zhou, and L. Cheong “Simultaneous video defogging and stereo reconstruction,” in *Proc. of IEEE Int. Conf. Comput. Vis. Pattern Rec. (CVPR)*, 2015. pp. 4988-4997.
- [15] J. Stark, “Adaptive image contrast enhancement using generalizations of histogram equalization,” *IEEE Trans. Image Process.*, vol. 9, no. 5, pp. 889-896, 2000.
- [16] J. Kim, L. Kim, and S. Hwang, “An advanced contrast enhancement using partially overlapped sub-block histogram equalization,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 4, pp. 475-484, 2001.
- [17] R. Tan, “Visibility in bad weather from a single image,” in *Proc. of IEEE Int. Conf. Comput. Vis. Pattern Rec. (CVPR)*, 2008, pp. 1-8.
- [18] C. O. Ancuti and C. Ancuti, “Single image dehazing by multi-scale fusion,” *IEEE Trans. Image Process.*, vol. 22, no. 8, pp. 3271-3282, 2013.
- [19] R. Fattal, “Single image dehazing,” *ACM Trans Graph.*, vol. 27, no. 3, pp. 1-9, 2008.
- [20] K. He, J. Sun, and X. Tang, “Single image haze removal using dark channel prior,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 12, pp. 2341-2353, 2011.
- [21] G. Meng, Y. Wang, J. Duan, S. Xiang, and C. Pan, “Efficient image dehazing with boundary constraint and contextual regularization,” in *Proc. of IEEE Int. Conf. Comput. Vis. (ICCV)*, 2013, pp. 617-624.
- [22] R. Fattal, “Dehazing using color-lines,” *ACM Trans Graph.*, vol. 34, no. 1, pp. 13:1-13:14, 2014.
- [23] Y. Lai, Y. Chen, C. Chiou, and C. Hsu, “Single-image dehazing via optimal transmission map under scene priors,” *IEEE Trans. Intell. Transport. Syst.*, vol. 25, no. 1, pp. 1-14, 2015.
- [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, 2015.
- [25] N. Baig, M. Riaz, A. Ghafoor, and A. Siddiqui, “Image dehazing using quadtree decomposition and entropy-based contextual regularization,” *IEEE Sig. Process. Letters*, vol. 23, no. 6, pp. 853-857, 2016
- [26] D. Berman and S. Avidan, “Non-local image dehazing,” in *Proc. of IEEE Int. Conf. Comput. Vis. Pattern Rec. (CVPR)*, 2016, pp. 1674-1682.
- [27] C. Chen, N. Do, and J. Wang, “Robust image and video dehazing with visual artifact suppression via gradient residual minimization,” in *Proc.*

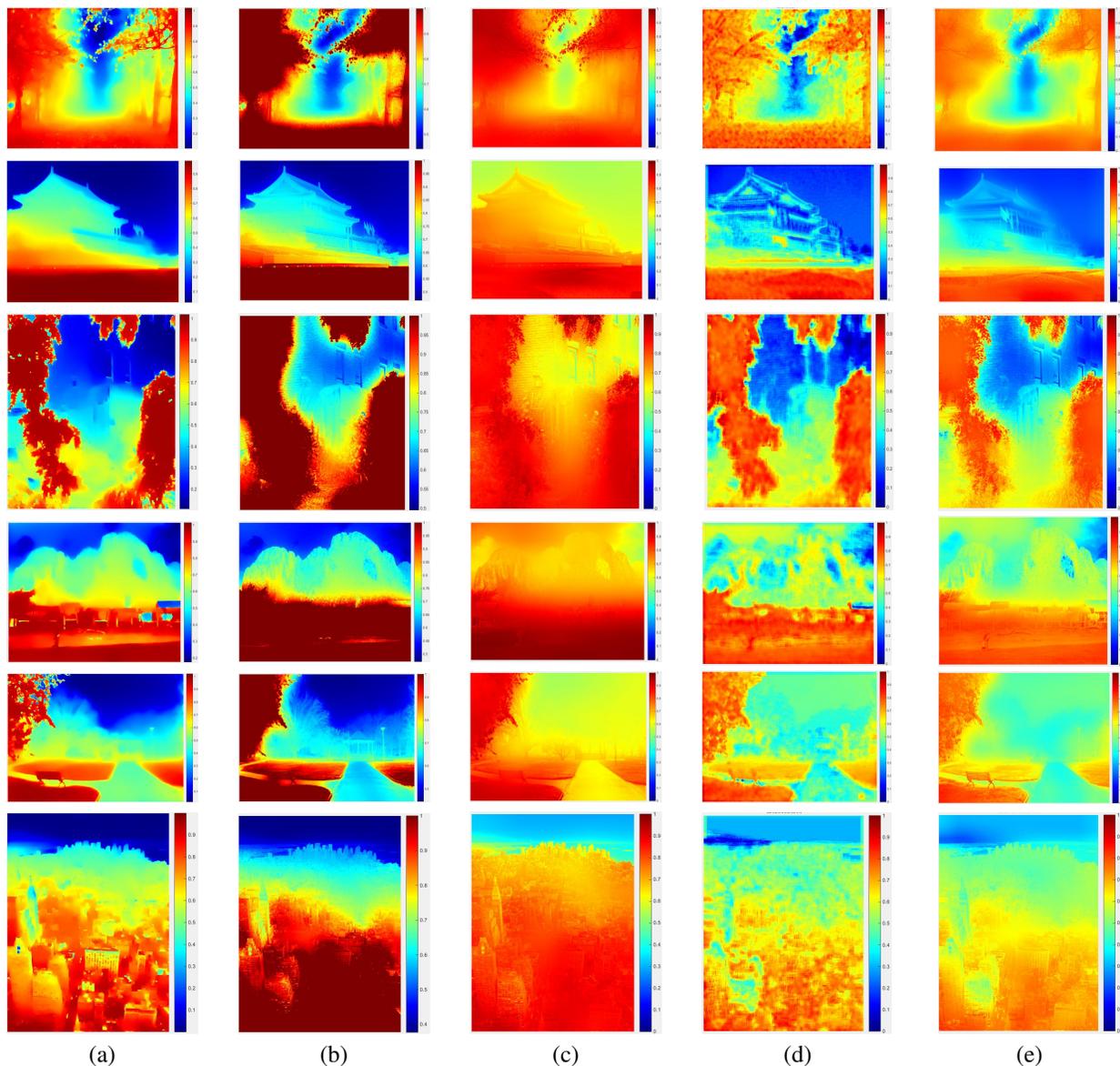


Fig. 9. The corresponding medium transmission maps of Figure 8. (a) The medium transmission estimated by Meng *et al.* [21]. (b) The medium transmission estimated by Zhu *et al.* [24]. (c) The medium transmission estimated by Cai *et al.* [30]. (d) The medium transmission estimated by our network. (e) The refined results using the guided image filtering [36] of (d) . (Best viewed on high-resolution display with zoom-in.)

- of Eur. Conf. Comput. Vis. (ECCV)*, 2016, pp. 576-591.
- [28] W. Wang, X. Yuan, X. Wu, and Y. Liu, "Fast image dehazing method based on linear transformation," *IEEE Trans. Multimedia*, vol. 19, no. 6, pp. 1142-1155, 2017.
- [29] K. Tang, J. Yang, and J. Wang, "Investigating haze-relevant features in a learning framework for image dehazing," in *Proc. of IEEE Int. Conf. Comput. Vis. Pattern Rec. (CVPR)*, 2014, pp. 2995-3002.
- [30] B. Cai, X. Xu, K. Jia, C. Qing, and D. Tao, "DehazeNet: an end-to-end system for single image haze removal," *IEEE Trans. Image Process.*, vol. 25, no. 11, pp. 5187-5198, 2016.
- [31] W. Ren, S. Liu, H. Zhang, J. Pan, and X. Cao, "Single image dehazing via multi-scale convolutional neural networks," in *Proc. of Eur. Conf. Comput. Vis. (ECCV)*, 2016, pp. 154-169.
- [32] X. Fan, Y. Wang, X. Tang, and R. Gao, "Two-Layer Gaussian process regression with example selection for image dehazing," *IEEE Trans. Circuits Syst. Video Technol.*, 2016.
- [33] B. Li, X. Peng, Z. Wang, J. Xu, and D. Feng, "Aod-net: All-in-one dehazing network," in *Proc. of Inter. Conf. Comput. Vis. (ICCV)*, 2017, pp. 4770-4778.
- [34] B. Li, X. Peng, Z. Wang, J. Xu, and D. Feng, "End-to-End United Video Dehazing and Detection," in *arXiv:1709.03919*, 2017.
- [35] A. Levin, D. Lischinski, and Y. Weiss, "A closed form solution to natural image matting," in *Proc. of IEEE Int. Conf. Comput. Vis. Pattern Rec. (CVPR)*, 2006, pp. 61-68.
- [36] K. He, J. Sun, and X. Tang, "Guided image filtering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 6, pp. 1397-1409, 2013.
- [37] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436-444, 2015.
- [38] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *arXiv preprint arXiv:1512.03385*, 2015.
- [39] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 532, 2001.
- [40] H. Koschmieder, "Theorie der horizontalen sichtweite," in *Beitrage zur Physik der freien Atmosphere*, 1924.
- [41] A. Krizhevsky, I. Sutskever, and G. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. of Advances in Neural Information Processing Systems (NIPS)*, 2012, pp. 1097-1105.
- [42] G. Huang, Z. Liu, L. van der Matten, and K. Weinberger, "Densely connected convolutional networks," in *Proc. of IEEE Int. Conf. Comput. Vis. Pattern Rec. (CVPR)*, 2017, pp. 2261-2269.

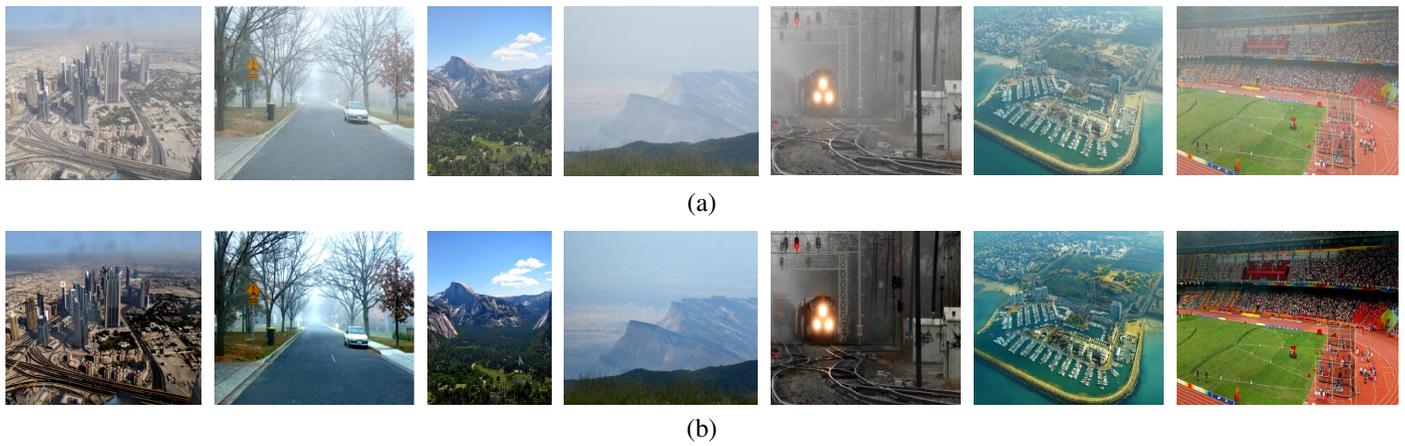


Fig. 10. Our results on varying scenes. (a) Hazy images. (b) Our results. (Best viewed on high-resolution display with zoom-in.)

- [43] H. Zhao, O. Gallo, I. Frosio, and J. Kautz, "Loss functions for image restoration with neural networks," *IEEE Trans. Computational Imaging*, vol. 3, no. 1, pp. 47-57, 2017.
- [44] N. Silberman, D. Hoiem, P. Kohli, and R. Fergus, "Indoor segmentation and support inference from rgb-d images," in *Proc. of Eur. Conf. Comput. Vis. (ECCV)*, 2012, pp. 746-760.
- [45] D. Kingma and J. Ba, "Adam: a method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.
- [46] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600-612, 2004.