




# Single Image Dehazing via Multi-scale Convolutional Neural Networks with Holistic Edges

Wenqi Ren<sup>1,6</sup> · Jinshan Pan<sup>4</sup> · Hua Zhang<sup>1</sup> · Xiaochun Cao<sup>1,2,3</sup> · Ming-Hsuan Yang<sup>5</sup> 

Received: 24 December 2017 / Accepted: 12 September 2019 / Published online: 24 September 2019  
© Springer Science+Business Media, LLC, part of Springer Nature 2019

## Abstract

Single image dehazing has been a challenging problem which aims to recover clear images from hazy ones. The performance of existing image dehazing methods is limited by hand-designed features and priors. In this paper, we propose a multi-scale deep neural network for single image dehazing by learning the mapping between hazy images and their transmission maps. The proposed algorithm consists of a coarse-scale net which predicts a holistic transmission map based on the entire image, and a fine-scale net which refines dehazed results locally. To train the multi-scale deep network, we synthesize a dataset comprised of hazy images and corresponding transmission maps based on the NYU Depth dataset. In addition, we propose a holistic edge guided network to refine edges of the estimated transmission map. Extensive experiments demonstrate that the proposed algorithm performs favorably against the state-of-the-art methods on both synthetic and real-world images in terms of quality and speed.

**Keywords** Image dehazing · Image defogging · Convolutional neural network · Transmission map

---

Communicated by Srinivasa Narasimhan.

---

Wenqi Ren and Jinshan Pan have contribute equally to this work.

---

✉ Xiaochun Cao  
caoxiaochun@iie.ac.cn

Wenqi Ren  
renwenqi@iie.ac.cn

Jinshan Pan  
sdluran@gmail.com

Hua Zhang  
zhanghua@iie.ac.cn

Ming-Hsuan Yang  
mhyang@ucmerced.edu

<sup>1</sup> Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100093, China

<sup>2</sup> School of Cyber Security, University of Chinese Academy of Sciences, Beijing 100049, China

<sup>3</sup> The Cyberspace Security Research Center, Peng Cheng Laboratory, Shenzhen 518055, China

<sup>4</sup> School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China

<sup>5</sup> School of Engineering, University of California at Merced, Merced, CA 95343, USA

<sup>6</sup> Tianjin University, Tianjin 300072, China

## 1 Introduction

The recent years have witnessed significant progress in image dehazing (Liu et al. 2018; Narasimhan and Nayar 2000, 2003; Ren et al. 2018; Schechner et al. 2001; Shwartz et al. 2006; Treibitz and Schechner 2009; Zhang et al. 2015). Images acquired in the hazy weather conditions usually contain significant haze as shown in Fig. 1a. The hazy image formation model proposed by Koschmieder (1924) has been widely used (Berman et al. 2016; Fattal 2008; He et al. 2009; Zhang et al. 2018):

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

where  $I(x)$  and  $J(x)$  are the observed hazy image and the clear scene radiance, respectively; the atmospheric light  $A$ , which usually satisfies the uniform assumption, describes the intensity of the scattered light in the scene at each color channel of an image; and the scene transmission  $t(x)$  describes the attenuation in intensity as a function of distance due to scattering.

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

where  $\beta$  is the medium extinction coefficient caused by the turbid medium such as particles and water droplets, and  $d(x)$

is the scene depth. The goal of image dehazing is to recover the clear scene radiance  $J(x)$  from  $I(x)$ . If we know the atmospheric light  $A$  and the transmission  $t(x)$ , the clear scene radiance  $J(x)$  can be recovered based on (1). Since only the input image  $I(x)$  is known, single image dehazing is an ill-posed problem.

Numerous haze removal methods have been proposed (Ancuti et al. 2010; Cai et al. 2016; Caraffa and Tarel 2012, 2013; Gibson and Nguyen 2013; Li et al. 2015; Tan et al. 2007; Schaul et al. 2009; Wang and Fan 2014) in recent years with significant advancements. Most dehazing methods use a variety of visual cues to capture deterministic and statistical properties of hazy images (Ancuti and Ancuti 2013; He et al. 2009, 2011; Ren and Cao 2017; Tan 2008; Tang et al. 2014; Zhu et al. 2015). The extracted features model chromatic (He et al. 2011), textural and contrast (Tan 2008) properties of hazy images to determine the transmission in the scenes. Although these feature representations are useful, the assumptions in these aforementioned methods do not hold in all cases. For example, the prior in He et al. (2011) is based on the assumption that the values of dark channel in haze-free images are close to zero. However, this assumption does not always hold when the haze-free images do not contain zero-intensity pixels, especially when the object colors in the hazy image are similar to the atmospheric light (He et al. 2009). Furthermore, these methods involve a considerable amount of effort to design hand-crafted features for scene transmission estimation [e.g., the use of ensemble features to learn a mapping between hazy images and transmission maps (Tang et al. 2014)]. More importantly, methods based on hand-crafted features are often sensitive to image variations such as changes in illumination, viewpoints, and scenes.

As the main goal of image dehazing is to estimate the transmission map from an input image, we propose a multi-scale convolutional neural network (CNN) to learn effective feature representations for this task based on the depth estimation network (Eigen et al. 2014). The features learned by the proposed algorithm do not depend heavily on statistical priors of the scene images or haze-relevant properties. Since the learned features are based on a data-driven approach, they are able to describe the intrinsic properties of hazy images and help estimate transmission maps. To learn these features, we directly regress on the transmission maps using a neural network with three modules. The first module is the coarse-scale network which estimates the holistic structure of the scene transmission, and then a fine-scale network refines it using local information and the output from the coarse-scale module. Finally, we use a network based on holistic edges to refine transmission maps. The holistic edge guided network transfers the structure of the holistic edges to the filtering output. This removes isolated and spurious pixel transmission estimates, meanwhile, it encourages neighboring pixels

to have the same labels. We evaluate the proposed algorithm against the state-of-the-art methods on numerous datasets comprised of synthetic and real-world hazy images.

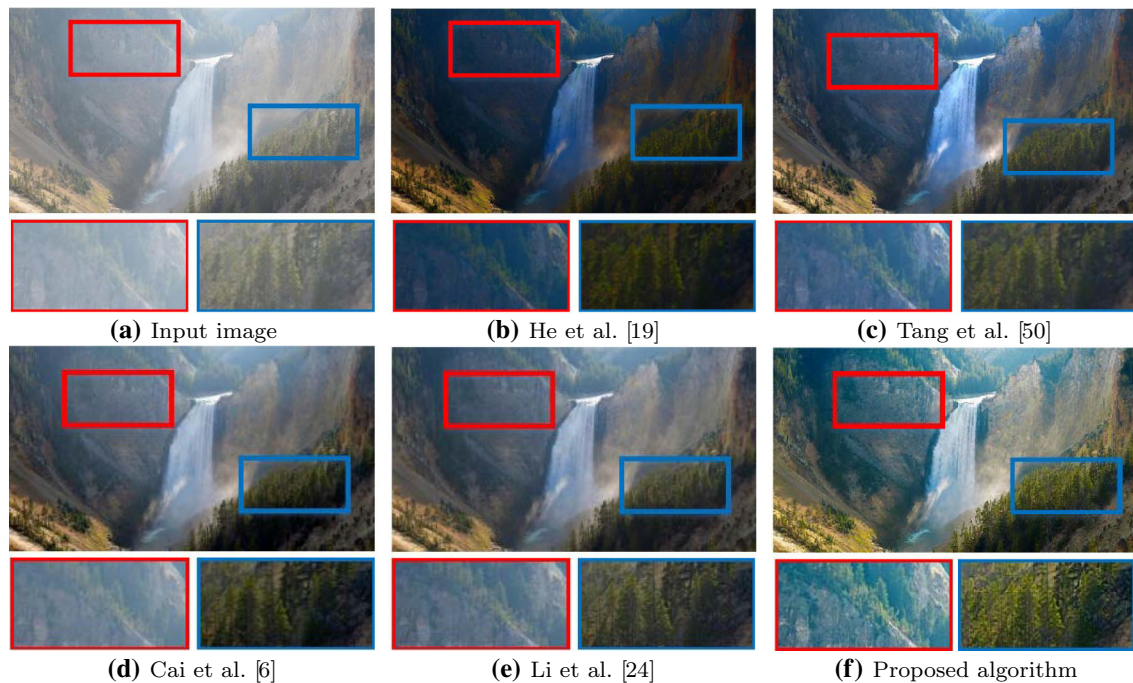
The contributions of this work are summarized as follows:

- We propose a multi-scale CNN to learn effective features from hazy images for the estimation of scene transmission map. The scene transmission map is first estimated by a coarse-scale network and then refined by a fine-scale network.
- We present a novel holistic edge guided network to refine transmission maps based on the holistic edge information of hazy images.
- We develop a benchmark dataset consisting of hazy images and their transmission maps by synthesizing clean images and ground truth depth maps from the NYU Depth database (Silberman et al. 2012). Although trained with the synthetic dataset, we show the learned multi-scale CNN is able to dehaze real-world hazy images well.
- We analyze the differences between hand-crafted and learned features for single image dehazing, and show that the proposed algorithm performs favorably against the state-of-the-art methods.

In this paper, we extend our preliminary work (Ren et al. 2016) in three aspects. First, we simplify the multi-scale network by removing pooling and up-sampling layers (Sect. 3.1) while the performance is still preserved. Second, we develop a novel holistic edge guided network for edge refinement (Sect. 3.3). Third, we present more technical details, performance evaluation and quantitative analysis of the proposed algorithm.

## 2 Related Work

As the image dehazing problem is ill-posed, early approaches often require multiple frames to deal with this problem (Treibitz and Schechner 2009; Narasimhan and Nayar 2003, 2000; Schechner et al. 2001; Shwartz et al. 2006). In Narasimhan and Nayar (2003), a method is proposed to solve the image dehazing problem by processing several images, which are taken in different atmospheric conditions. Narasimhan et al. (Nayar and Narasimhan 1999; Narasimhan and Nayar 2000) propose haze removal approaches with multiple images of the same scene under different weather conditions. Treibitz and Schechner (2009) use different angles of polarized filters to capture multiple images of the same scene, and then analyze different degrees of polarization of images for haze removal. In Kopf et al. (2008) an approximated 3D geometrical model of the scene is assumed to be available, from which a data-driven dehazing method



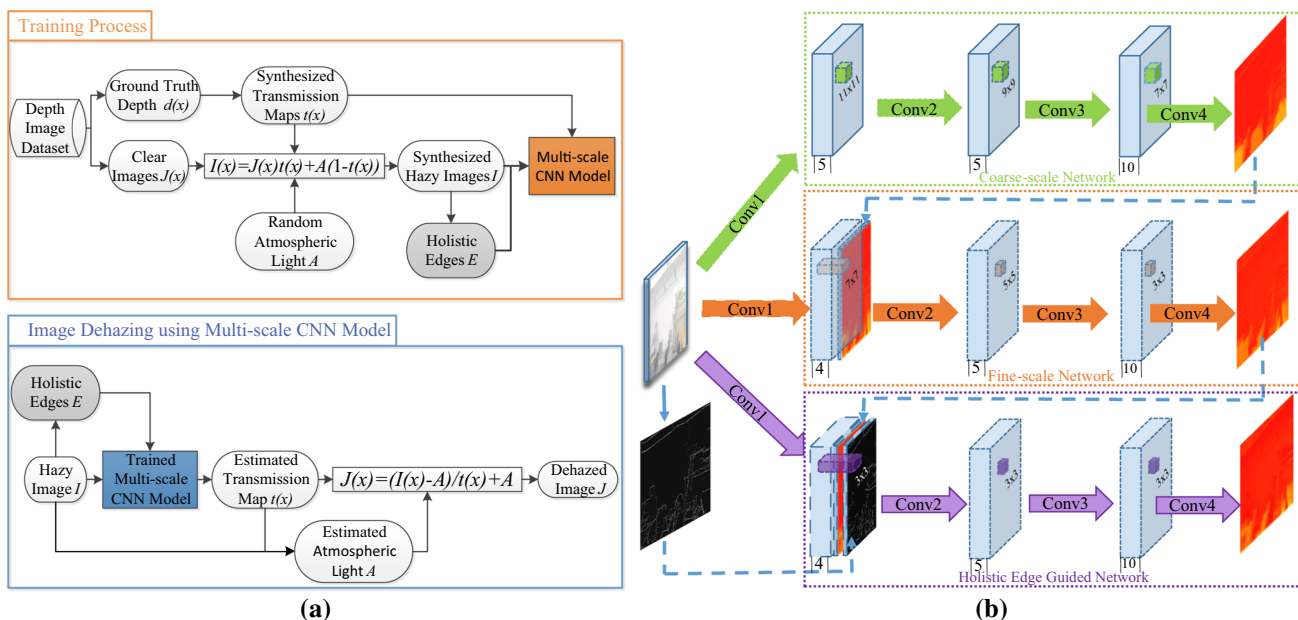
**Fig. 1** Dehazed results of real images by the state-of-the-art methods and proposed algorithm. The recovered image in **f** contains rich details and vivid color information (Color figure online)

is developed for image dehazing. These methods assume that multiple images from the same scene can be obtained under different conditions. However, there may exist only one image for a scene at our disposal.

Different from the aforementioned methods, another line of research work is based on physical properties of hazy images. For example, Fattal (2008) proposes a refined image formation model for surface shading and scene transmission. Based on this model, a hazy image can be separated into regions of constant albedo, and then the scene transmission can be inferred. However, this approach requires time-consuming operations and focuses on images that contain a slight amount of haze. Based on a similar model, Tan (2008) proposes to enhance the visibility of hazy images by maximizing their local contrast, but the restored images often contain distorted colors and significant halos.

He et al. (2009) propose a dark channel prior (DCP) based on the statistics of haze-free images. This method assumes that at least one color channel has some pixels whose intensities are close to zeros. The dark channel prior has been shown to be effective for image dehazing when soft-matting operations (He et al. 2011) are used. However, it is computationally expensive (Tarel and Hautiere 2009; Zhang et al. 2018; Gibson et al. 2012; He et al. 2013) and less effective for sky images and scenes where the color of objects are inherently similar to the atmospheric light. Since the dark channel prior (He et al. 2011) is introduced, numerous DCP based dehazing methods (Kratz and Nishino 2009;

Tarel et al. 2012; Nishino et al. 2012; Meng et al. 2013) have been developed for improvements. Gibson et al. (2012) replace the time-consuming soft matting (He et al. 2011) with standard median filtering to improve computing efficiency. Kratz and Nishino (2009) model an image as a factorial Markov Random Field and use a canonical expectation maximization algorithm to analyze images. A haze-free image with fine edge details can be recovered, but the result often tends to be over-enhanced. Meng et al. (2013) propose an effective regularization dehazing method to restore the haze-free image by exploring the inherent boundary constraint. A variety of multi-scale haze-relevant features are analyzed by Tang et al. (2014) in a regression framework based on random forests (Breiman et al. 2001). Nevertheless, this feature fusion approach relies largely on the dark channel features. In Zhu et al. (2014, 2015) find that the difference between brightness and saturation in a clear image patch should be very small, and propose a color attenuation prior for haze removal. Nevertheless, since this method is based on depth estimation rather than transmission estimation, it needs to tune the parameter for the scattering coefficient. Berman et al. (2016) introduce a non-local method for single image dehazing, based on the assumption that an image can be faithfully represented with just a few hundreds of distinct colors. Despite significant advances in this field, the state-of-the-art dehazing methods (Tang et al. 2014; Zhu et al. 2015) are developed based on hand-crafted features.



**Fig. 2** **a** Main steps of the proposed single image dehazing algorithm. For training the multi-scale network, we synthesize hazy images and the corresponding transmission maps based on a depth image dataset. In the test stage, we estimate the transmission map of the input hazy image based on the trained model, and then generate the dehazed image using the estimated atmospheric light and computed transmission map. **b** Proposed multi-scale convolutional network. Given a hazy image, the

coarse-scale network (in the green dashed rectangle) predicts a holistic transmission map and feeds it to the fine-scale network (in the orange dashed rectangle) in order to generate a refined transmission map. We then use holistic edges to refine transmission maps to be smooth inside the same object. The blue dashed lines denote concatenate operation (Color figure online)

Data-driven dehazing models recently become popular due to the success of machine learning in various vision applications (Ren et al. 2018; Li et al. 2019). To avoid designing hand-crafted features, several algorithms use deep CNN for image dehazing (Zhang et al. 2018). In Cai et al. (2016), use a deep neural network for transmission estimation (DehazeNet) and then follow the conventional method to estimate atmospheric light. However, Cai et al. synthesize hazy images based on the assumption that the context of an image patch is independent of transmission map, which does not hold in practice. In addition, this network is trained on the patch-level and does not fully utilize the high-level information from a larger region. Instead of estimating the transmission map and atmospheric light separately, Li et al. (2017) propose the atmosphere scattering model where the atmosphere light and transmission map are formulated in a matrix form and propose an AOD-Net to estimate clear images directly. Although the AOD-Net algorithm does not explicitly require estimations of the transmission map and atmosphere light, it needs to estimate the parameters of the matrix. As the matrix prediction does not use the information of transmission maps, these final restored images still contain some haze residue.

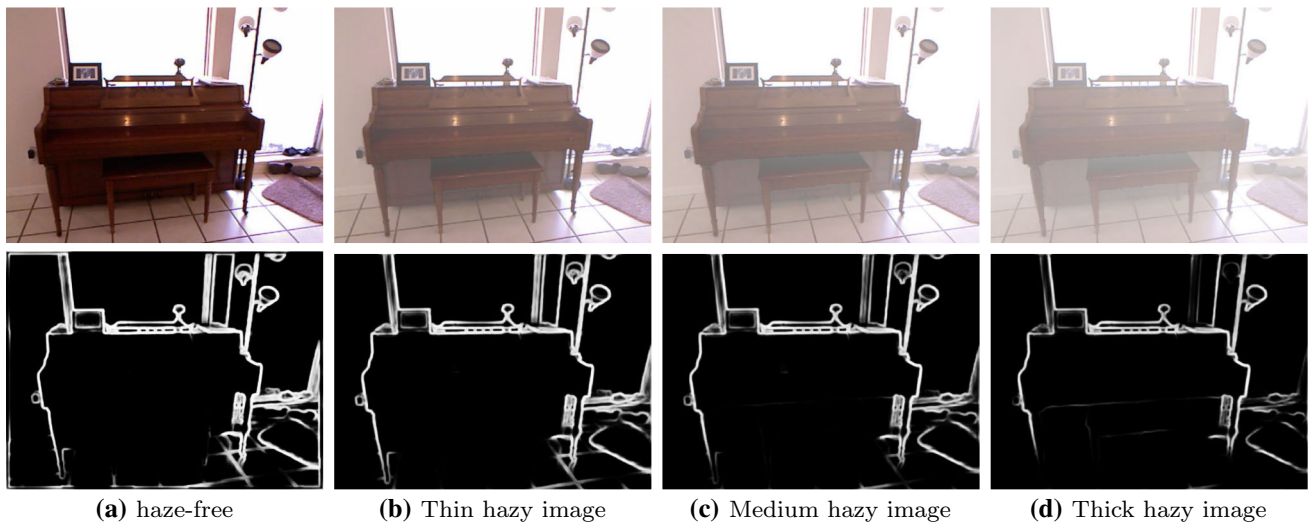
Different from these learning-based methods, our algorithm directly estimates transmission maps from hazy images, where the proposed network is constrained by the ground

truth transmission maps in the training processing. As such, it is able to keep the correlation between hazy images and transmission maps, which leads to more realistic images. In addition, we propose a new multi-scale CNN with a holistic edge guided network to automatically learn the mapping between hazy images and transmission maps.

### 3 Multi-scale Network for Transmission Estimation

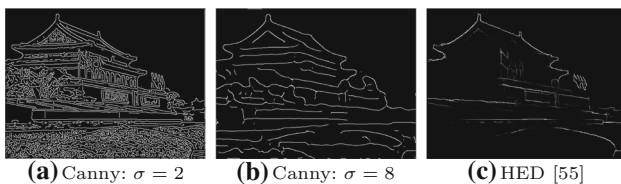
Given a single hazy input, we aim to recover the latent clean image by estimating the scene transmission map. The main steps of the proposed algorithm are shown in Fig. 2a. We first describe how to estimate the scene transmission map  $t(x)$  and present the method to compute atmospheric light  $A$  in Sect. 4.

For each scene, we estimate the scene transmission map  $t(x)$  based on a multi-scale CNN with a holistic edge guided network. The coarse structure of the scene transmission map for each image is obtained from the coarse-scale convolutional net, and then refined by the fine-scale network. Both networks are applied to the original input hazy image, but in addition, the output of the coarse network is passed to the fine network as additional information. Thus, the fine-



**Fig. 3** Edges extracted by the holistic edge detector (Xie and Tu 2015). Top: Clear image and synthetic hazy images with different medium extinction coefficient  $\beta$ . Bottom: Holistic edges detected by the HED

(Xie and Tu 2015). All of these holistic edges detected by the HED (Xie and Tu 2015) are similar for different haze concentrations



**Fig. 4** Effectiveness of the holistic edge detector. **a, b** Edge responses from the Canny detector (Canny et al. 1986) at the scale  $\sigma = 2$  and  $\sigma = 8$ , for the image in Fig. 5a. **c** Holistic edges extracted by the HED method (Xie and Tu 2015) which is able to capture the main boundaries of the object

scale network can refine the coarse prediction with details. Furthermore, we use the holistically-nested edge detection (HED) method (Xie and Tu 2015) to predict the holistic edges of the input hazy image which are used to refine the transmission map. The proposed multi-scale CNN for learning haze-relevant features is shown in Fig. 2b.

### 3.1 Coarse-Scale Network

The task of the coarse-scale network is to predict a holistic transmission map of the scene. As illustrated in the green dashed rectangle in Fig. 2b, the coarse-scale network contains four feature extraction layers. Each convolution layer is followed by the rectified linear unit (ReLU) (Nair and Hinton 2010) except the last layer. Rather than adding max-pooling and up-sampling layers to limit the feature maps and the output transmission map size to be the same as the input hazy image in Ren et al. (2016), we remove these layers and only use convolution layers with zero padding to maintain the size of feature and output maps.

**Convolution layers** This network takes an RGB image as an input. The convolution layers consist of filter banks which are convolved with the input feature maps. The response of each convolution layer is given by

$$f_n^{l+1} = \sigma \left( \sum_m (f_m^l * k_{m,n}^{l+1}) + b_n^{l+1} \right), \quad (3)$$

where  $f_m^l$  and  $f_n^{l+1}$  are the feature maps of the  $l$ th layer and the next  $(l+1)$ th layer, respectively. The feature maps for the first convolution layer ( $l = 1$ ) are based on input hazy image. In addition,  $k$  is the convolution kernel, indices  $(m, n)$  show the mapping from the current layer  $m$ th feature map to the next layer  $n$ th, and  $*$  denotes the convolution operator. The function  $\sigma(\cdot)$  denotes the ReLU (Nair and Hinton 2010) on the filter responses and  $b_n^{l+1}$  is the bias.

**Transmission reconstruction** We produce the coarse transmission map  $t_c$  prediction by

$$t_c = \sum_m (f_m^l * k_{m,n}^{l+1}) + b_n^{l+1}, \quad (4)$$

where  $l = 4$  and  $n = 1$  denote the output from the fourth convolution layer is a one-dimensional image.

### 3.2 Fine-Scale Network

After the coarse scene transmission map is estimated, this is refined by the fine-scale network. The receptive field in this network is smaller than the ones in the coarse-scale network. The architecture of the fine-scale network stack is similar to the coarse-scale network except for the first and second

convolution layers. The structure of our fine-scale network is shown in Fig. 2b (orange dashed rectangle) where the coarse output transmission map is used as an additional low-level feature map. We concatenate these two together in the fine-scale network to refine the scene transmission map. In addition, we maintain the size of the features maps in subsequent layers using zero-padded convolutions.

### 3.3 Holistic Edge Guided Network

The fine-scale network can estimate fine edges and remove halo artifacts (see Sect. 6.3). However, if the image contains strong textures as shown in Fig. 5a, the fine-scale net is likely to transfer extraneous edges to the transmission map, thereby including unnecessary details as shown in Fig. 5b. In such cases, the assumption of the model (2) does not hold since the amount of haze at each pixel does not depend only on depth but also its texture or color. Ideally, the transmission map should be smooth in the regions of the same object and discontinuous across the boundaries of different objects. Therefore, we expect the refined transmission map to be smooth inside the same object, and discontinuous only along depth edges. As such, we propose a new holistic edge guided network by using the holistic edge detector (Xie and Tu 2015). Since the holistic edge detector (Xie and Tu 2015) learns rich hierarchical representations for an input image, it is important to resolve the ambiguity in edge and object boundary detection.

Given a haze-free image in Fig. 3a, we use different medium extinction coefficient  $\beta$  to synthesize images with different haze concentrations. As shown in the second row of Fig. 3b–d, all edges detected by the HED (Xie and Tu 2015) are similar for different haze concentration images. Thus, we first use the HED (Xie and Tu 2015) to extract the holistic edges, and then concatenate the extracted edges with the first convolution layer in the holistic edge guided network which can further refine edges in transmission maps. The architecture of the holistic edge guided network is the same as the fine-scale net. In addition, we also use the output from the fine-scale network as the additional feature map in the holistic edge guided network. The structure of the holistic edge guided network is shown at the bottom of Fig. 2b.

Figure 4 shows the edges by the Canny et al. (1986) and HED (Xie and Tu 2015) detectors for the image with rich textures in Fig. 5a. The Canny detector (Canny et al. 1986) extracts unnecessary fine edges. These complex structures are likely to be transferred to the transmission map. In contrast, the edges extracted by the HED detector (Xie and Tu 2015) contain the main structures of the scene without extraneous details, which also demonstrates the effectiveness of HED (Xie and Tu 2015) in image dehazing task.

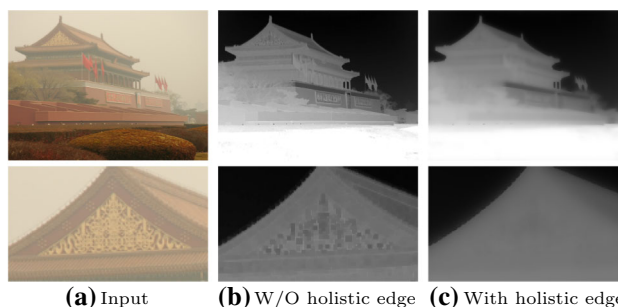


Fig. 5 The network without holistic edge tends to transfer incorrect textures to the transmission map

### 3.4 Training

Learning the mapping function between hazy images and corresponding transmission maps is achieved by minimizing the loss between the reconstructed transmission  $t_i(x)$  and the corresponding ground truth map  $t_i^*(x)$  at every scale  $s$ :

$$L(\theta) = \frac{1}{q} \sum_{s=1}^3 \sum_{i=1}^q \|t_i(x) - t_i^*(x)\|^2, \quad (5)$$

where  $\theta$  are the model parameters,  $q$  is the number of hazy images in the training set and  $s$  denotes the scale index. Here, we have three scales as we use coarse and fine-scale nets as well as a holistic edge guided network. The training loss (5) is used in all of these three scale networks.

We minimize the loss (5) using the stochastic gradient descent method with the back propagation learning rule. The implementation details are included in Sect. 5.

## 4 Dehazing with the Multi-scale Network

### 4.1 Atmospheric Light Estimation

After obtaining the scene transmission map  $t(x)$ , we can use existing algorithms, e.g., (Berman et al. 2017; Sulami et al. 2014; He et al. 2009; Zhu et al. 2015) to estimate the atmospheric light. According to Fattal (2008), a constant atmospheric light is a proper approximation when the aerosol reflectance properties and dominant scene illumination are approximately uniform across the scene. Therefore, we treat  $A$  as a constant across the image and use the method (He et al. 2009; Zhu et al. 2015) for estimation. We compute  $A$  from the estimated transmission map directly. From the hazy image formation model (1), we derive  $A$  when  $t(x) = 0$ , i.e.,  $I(x) = A$  when  $t(x) \rightarrow 0$ . Thus we estimate the atmosphere light  $A$  by giving a threshold  $t_h$ ,

$$I(x) = A, \quad t(x) < t_h. \quad (6)$$



**Fig. 6** Atmospheric light estimation on synthetic images. The method (Sulami et al. 2014) tends to under-estimate atmospheric light and results in darker dehazed results, while the approach by Berman et al.

(2017) over-enhances the dehazed results. In contrast, our estimated atmospheric light is close to the ground truth in **a**. The haze-free image is shown in the third row of Fig. 9g

According to (6), we select 0.1% darkest pixels in a transmission map  $t(x)$  (He et al. 2009; Zhu et al. 2015). These pixels have the most haze concentration. Among these pixels,

the one with the highest intensity value in the corresponding hazy image  $I$  is selected as the atmospheric light. Figure 6 shows three examples of synthetic hazy images. We compare

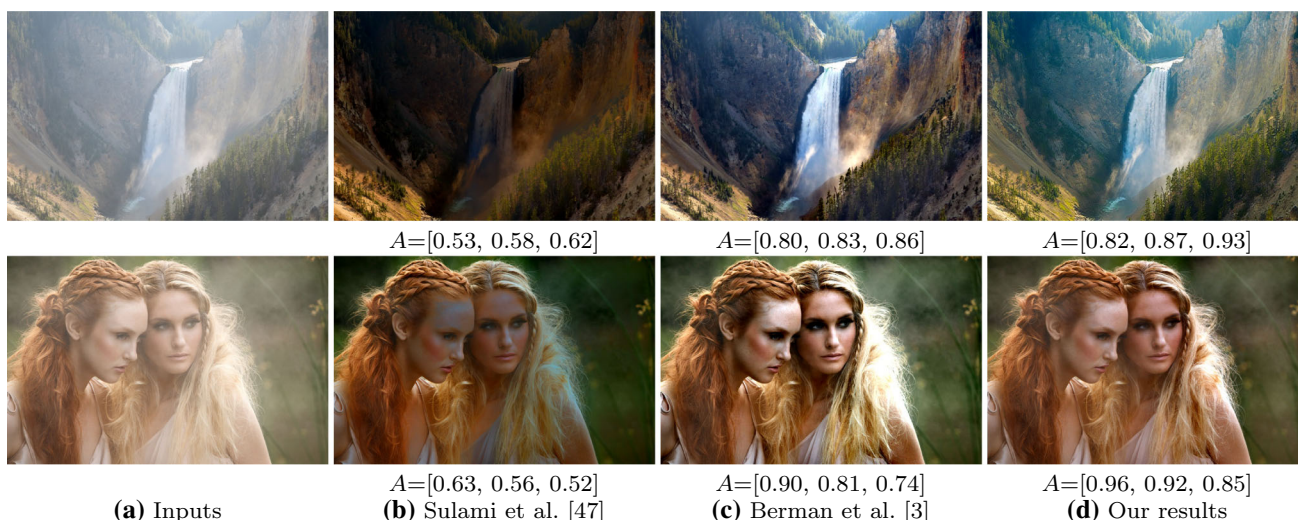


Fig. 7 Atmospheric light estimation on real-world images in which the sky is not present in the inputs

the proposed atmospheric light estimation algorithm against two state-of-the-art methods by Berman et al. (2017) and Sulami et al. (2014). The method (Sulami et al. 2014) underestimates the atmospheric light which leads to the color distortion in the dehazed images (Fig. 6b). The algorithm by Berman et al. (2017) estimates better atmospheric light than Sulami et al. (2014). However, this method over-enhances the dehazed results due to less inaccurate atmospheric light (Fig. 6c). In contrast, the atmospheric light by our algorithm is close to the ground truth, which accordingly lead to visually better results. In addition to the achromatic atmospheric light, we evaluate our method on the non-achromatic light (i.e.,  $A = [0.7, 0.8, 0.9]$ ) as shown in the fourth row of Fig. 6. In addition, the proposed algorithm performs favorably against the state-of-the-art methods on non-achromatic light estimation.

For the real-world examples in Fig. 7, the algorithms by Sulami et al. (2014) and Berman et al. (2017) do not generate clear images due to inaccurate atmospheric light estimation. The dehazed images in Fig. 7d show that our method is able to estimate atmospheric light for real-world image dehazing even the scene does not contain sky regions.

### 4.2 Haze Removal

Once the atmospheric light  $A$  and the scene transmission map  $t(x)$  are estimated, the haze-free image can be estimated by  $J(x) = (I(x) - A)/t(x) + A$ . However, directly estimating  $J(x)$  by this model is prone to noise when  $t(x)$  is close to 0. Therefore, we estimate the scene radiance  $J(x)$  by

$$J(x) = \frac{I(x) - A}{\max\{0.1, t(x)\}} + A. \tag{7}$$

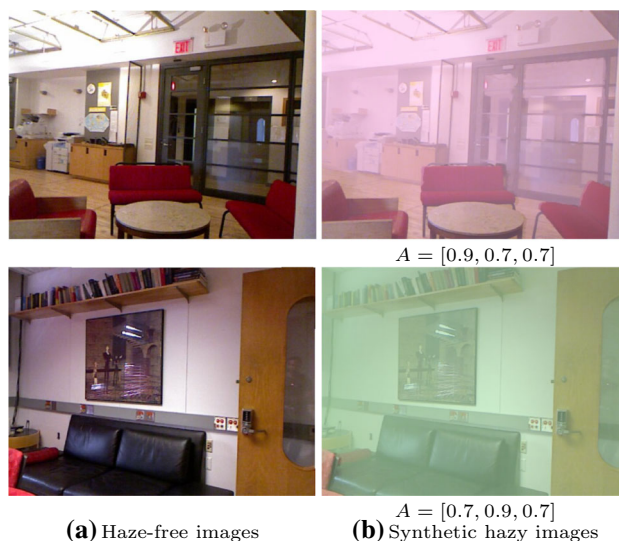


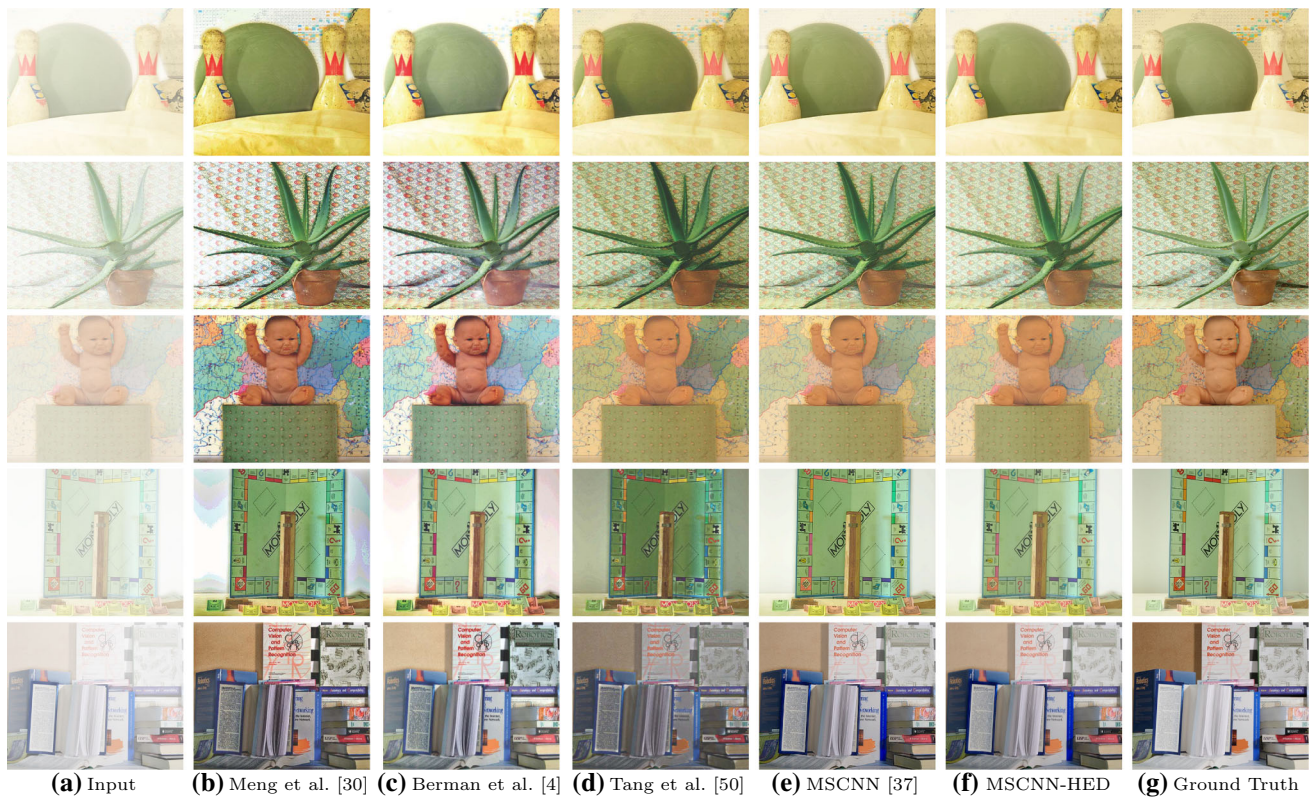
Fig. 8 Using the non-achromatic atmospheric light may generate some unrealistic rufescent or green-bluish images (Color figure online)

## 5 Experimental Results

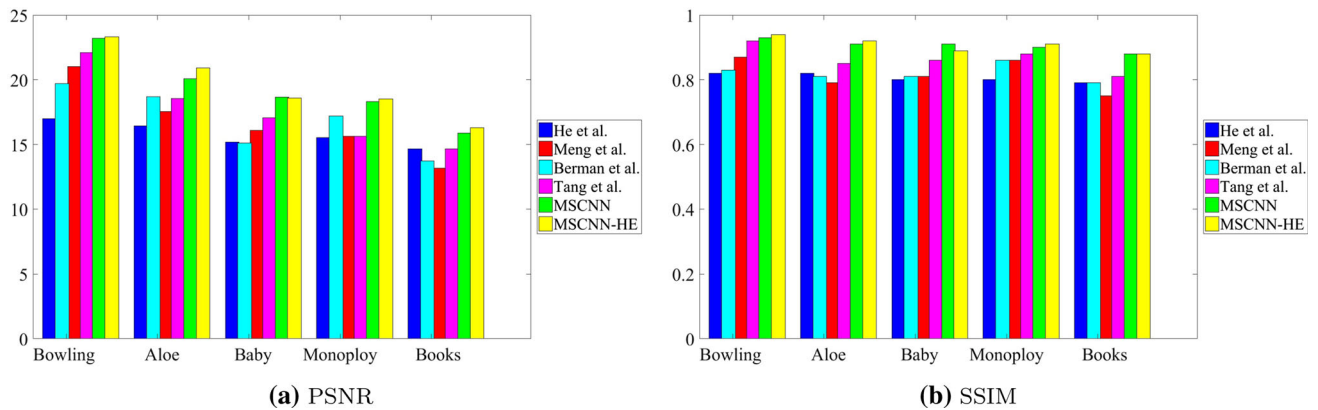
We quantitatively evaluate the proposed algorithm on two synthetic datasets and real-world hazy images, with comparisons to the state-of-the-art methods in terms of accuracy and run time. The implementation code will be made available to the public. The Multi-Scale CNN in our previous work (Ren et al. 2016) is referred to as MSCNN, and the proposed Multi-Scale CNN with Holistic Edge guided network as MSCNN-HE.

*Experimental Settings* The proposed network is trained using the stochastic gradient descent method. The momentum value, weight decay parameter, and batch size are set to be 0.9,  $5 \times 10^{-4}$ , and 10, respectively. Each batch is a whole





**Fig. 9** Dehazed results on synthetic hazy images using stereo images: *Bowling*, *Aloe*, *Baby*, *Monopoly*, and *Books*



**Fig. 10** Quantitative comparisons of the dehazed images shown in Fig. 9

image, whose size is  $320 \times 240$  pixels. The initial learning rate is 0.001 and decreased by 0.1 after every 20 epochs and the number of epoch is set to be 70. The training time is approximately ten hours on a desktop computer with a 2.8 GHz CPU and an NVIDIA K40 GPU.

**Training Data** To train the multi-scale network, we generate a dataset with synthesized hazy images and their corresponding transmission maps. Although there exist some outdoor datasets such as the Make3D (Saxena et al. 2009) and KITTI (Geiger et al. 2012), the depth maps are less precise and incomplete compared to the existing indoor datasets (Sil-

berman et al. 2012). Therefore, we randomly sample 6000 clean images and the corresponding depth maps from the NYU Depth dataset (Silberman et al. 2012) to construct the training set. In addition, we generate a validation set of 50 synthesized hazy images using the Middlebury stereo dataset (Scharstein and Szeliski 2002, 2003).

Given a clear image  $J(x)$  and the ground truth depth  $d(x)$ , we synthesize a hazy image using the physical model described by (1) and (2). We generate the random atmospheric light  $A = [k, k, k]$ , where  $k \in [0.7, 1.0]$ , and sample three random  $\beta \in [0.5, 1.5]$  for every image. We use

achromatic atmospheric light since non-achromatic atmospheric light tend to generate some unnatural rufescent or green-bluish images as shown in Fig. 8b. For example, the atmospheric light of  $A = [0.9, 0.7, 0.7]$  would result in a rufescent output as shown in the first row of Fig. 8. In addition, we do not use small  $\beta \in (0, 0.5)$  because it would lead to thin haze and boost noise (He et al. 2011). On the other hand, we do not use large  $\beta \in (1.5, \infty)$  as the resulting transmission maps are close to zero. Therefore, we have 18, 000 hazy images and transmission maps (6000 images  $\times$  3 medium extinction coefficients  $\beta$ ) in the training set.

## 5.1 Quantitative Evaluation on Benchmark Dataset

We compare the proposed algorithm with the state-of-the-art dehazing methods (He et al. 2011; Meng et al. 2013; Berman et al. 2016; Tang et al. 2014; Cai et al. 2016) using the Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity (SSIM) metrics.

We use five examples: *Bowling*, *Aloe*, *Baby*, *Monopoly*, and *Books* for illustration. Figure 9a shows the input hazy images which are synthesized from the haze-free images with known depth maps (Scharstein and Szeliski 2002). As the method by Meng et al. (He et al. 2011) is designed based on the DCP which assumes that dark channel values of clear images are zeros, it tends to overestimate the haze thickness and results in darker results as shown in Fig. 9b. We note that the dehazed images generated by Berman et al. (Meng et al. 2013) tends to have some color distortions. For example, the colors of the *Books* dehazed image become darker as shown in Fig. 9c, d. Although the dehazed results in Fig. 9d by Tang et al. (2014) are better than those by Meng et al. (2013), Berman et al. (2016) in the first three images, the colors are still darker than the ground truth in the last two images in Fig. 9d. In contrast, the dehazed results by the MSCNN method (Ren et al. 2016) and the proposed MSCNN-HE are close to the ground truth images, which indicates that better transmission maps are estimated. Figure 10 shows that the proposed algorithm performs well on each image in Fig. 9 against the state-of-the-art dehazing methods (He et al. 2011; Tarel and Hautiere 2009; Meng et al. 2013; Tang et al. 2014) in terms of PSNR and SSIM. Although the visual effect in Fig. 9f, g are similar, the average PSNR and SSIM values by the proposed method are higher than the MSCNN method (Ren et al. 2016).

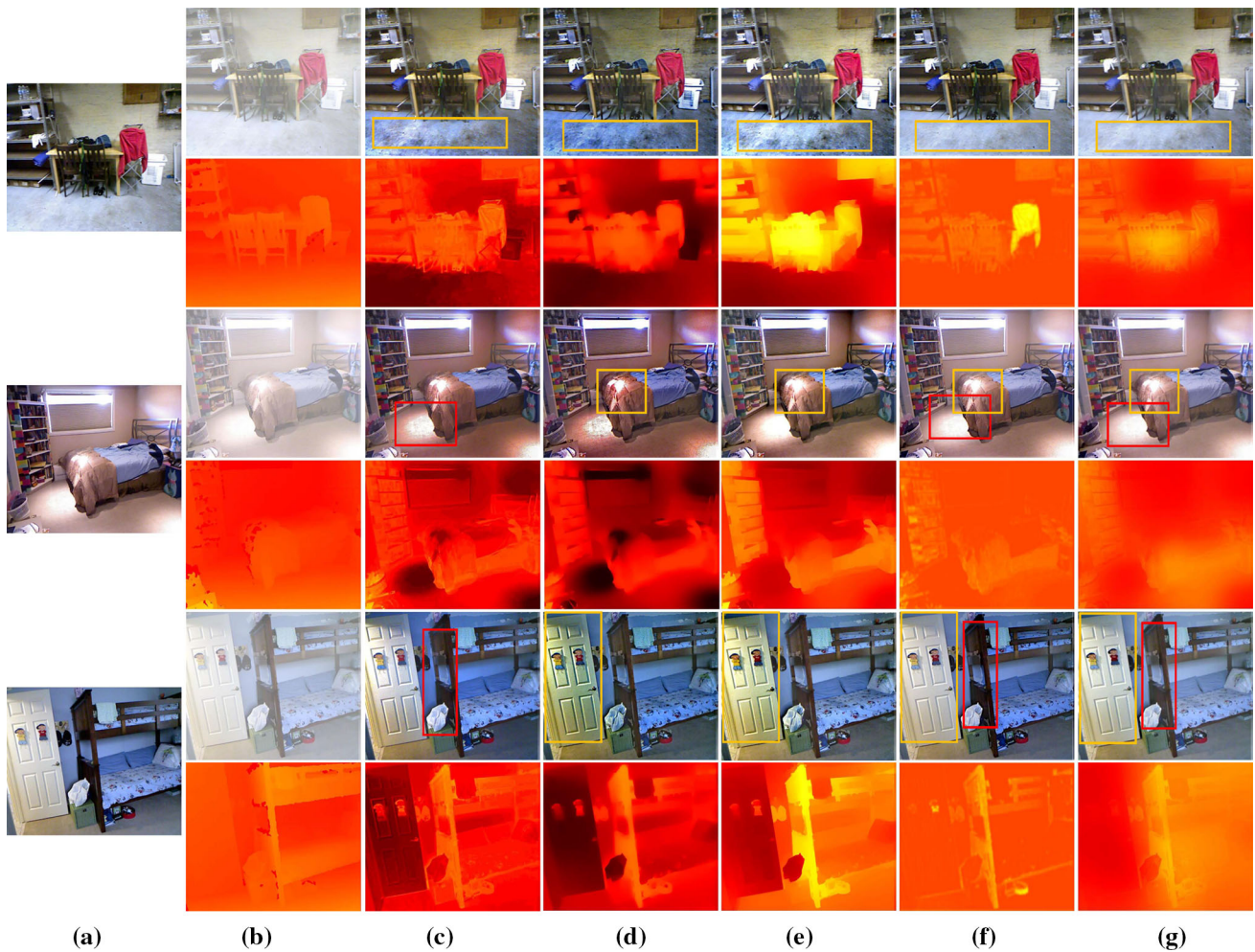
*New synthetic dataset* For quantitative performance evaluation, we construct a new dataset of synthesized hazy images and compare the proposed algorithm with the state-of-the-art dehazing methods (He et al. 2011; Meng et al. 2013; Berman et al. 2016; Cai et al. 2016). In addition, we also compare with the method without using holistic edge guided network (Ren et al. 2016). We randomly select 40 images and their depth maps from the NYU Depth dataset (Silberman et al. 2012)

(different from those used for training) to synthesize 40 transmission maps and hazy images using (1). In (1), we assume pure white atmospheric air light, i.e.,  $A = [1, 1, 1]$ , and then use medium extinction coefficient of  $\beta = 1$  in the experiments.

Figure 11 shows some dehazed images by different methods. The methods of He et al. (2011) and Meng et al. (2013) tend to overestimate the haze thickness as shown in the estimated transmission maps in Fig. 11c, d. This indicates that the dehazed results tend to be darker than the ground truth images in some regions. For example, the floor color is changed from gray to blue in the first image and the door color is changed from white to yellow in the third image. The regions that contain color distortions in the dehazed images correspond to the darker areas in the estimated transmission maps. The estimated transmission maps and the final dehazed results by Berman et al. (2016) are similar to the results by Meng et al. (2013). As shown in Fig. 11e, the dehazed images still contain some color distortions. Figure 11f, g show the estimated transmission maps and the final recovered images by the proposed algorithm without and with the holistic edge guided network, respectively. The results in Fig. 11f, g are very similar in most cases. However, the holistic edge guided network still improves the estimated transmission map. For example, the estimated transmission maps in (g) are smoother and closer to the ground truth in (b) than the results in (f), which demonstrates the effectiveness of the proposed holistic edge guided network. Overall, the dehazed results by the proposed algorithm have higher visual quality and fewer color distortions. In addition, the qualitative results in Table 1 demonstrate that the proposed algorithm performs favorably against state-of-the-art methods in terms of PSNR and SSIM metrics. We also compare the MSE of the estimated transmission maps by He et al. (2011), Meng et al. (2013), Berman et al. (2016), Cai et al. (2016) and our algorithm in Table 1. *RESIDE dataset* We evaluate the proposed algorithm on the SOTS data from the RESIDE dataset (Li et al. 2018) against the state-of-the-art methods (He et al. 2011; Meng et al. 2013; Berman et al. 2016; Cai et al. 2016; Li et al. 2017; Ren et al. 2016). Table 3 shows that our method performs competitively against state-of-the-art algorithms in terms of PSNR and SSIM. In addition, compared to the MSCNN (Ren et al. 2016), the proposed algorithm achieves performance gain of 1.7 dB in terms of PSNR on the SOTS dataset, which demonstrates the proposed holistic edge guided network is able to better estimate transmission maps and restore clear images.

## 5.2 Real Images

Although our multi-scale network is trained on synthetic indoor images, it can be applied to real-world outdoor images as well. We evaluate the proposed algorithm against the state-of-the-art image dehazing methods (He et al. 2009; Tang et al.



**Fig. 11** Dehazed results (odd rows) and estimated transmission maps (even rows) on our synthetic images. **a** Haze-free images. **b** Hazy images and ground truth transmission maps. **c** He et al. (2011). **d** Meng et al. (2013). **e** Berman et al. (2016). **f** MSCNN (Ren et al. 2016). **g** Ours with

the holistic edge guided network. The red rectangles are for comparison of our methods with He et al. (2011), and the yellow rectangles are for comparison of our methods with Meng et al. (2013) and Berman et al. (2016) (Color figure online)

**Table 1** Average PSNR and SSIM of dehazed results, and MSE of the estimated transmission maps, on the new synthetic dataset

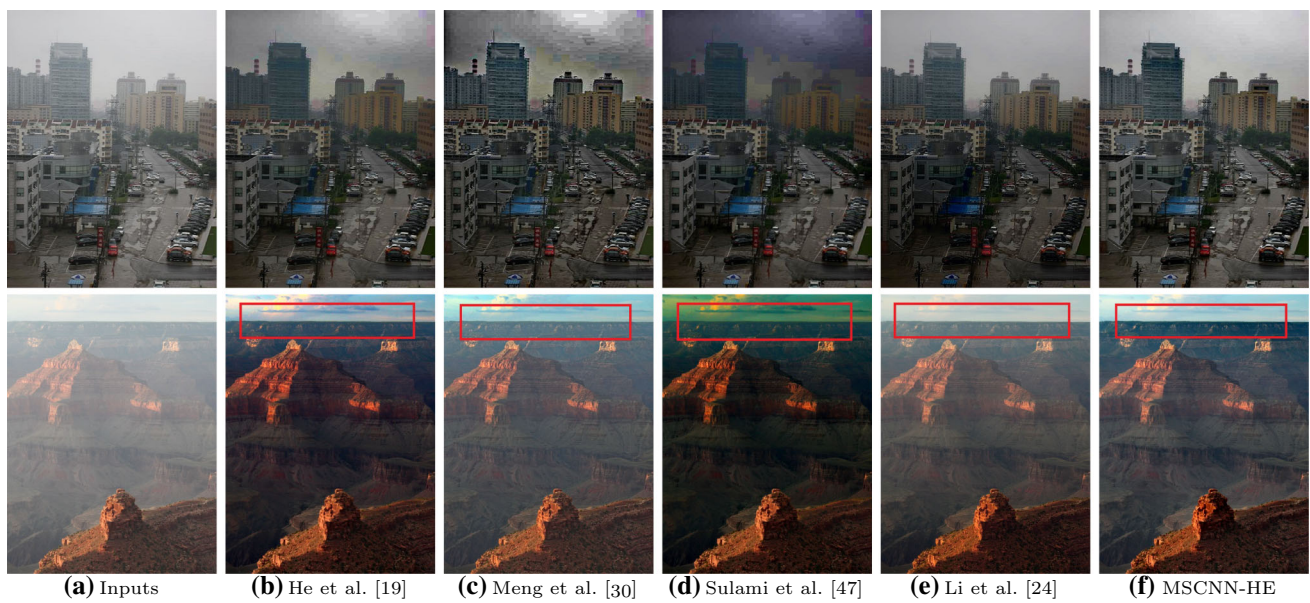
	He et al. (2011)	Meng et al. (2013)	Berman et al. (2016)	Cai et al. (2016)	MSCNN (Ren et al. 2016)	MSCNN-HE
PSNR ( $I$ )	20.28	16.79	19.26	<u>21.29</u>	21.27	<b>21.32</b>
SSIM ( $I$ )	0.80	0.41	0.73	<u>0.84</u>	<b>0.85</b>	<b>0.85</b>
MSE ( $t$ )	0.0660	0.0675	0.0549	<u>0.0357</u>	0.0338	<b>0.0332</b>

The bold and underline values indicate the best and second best results

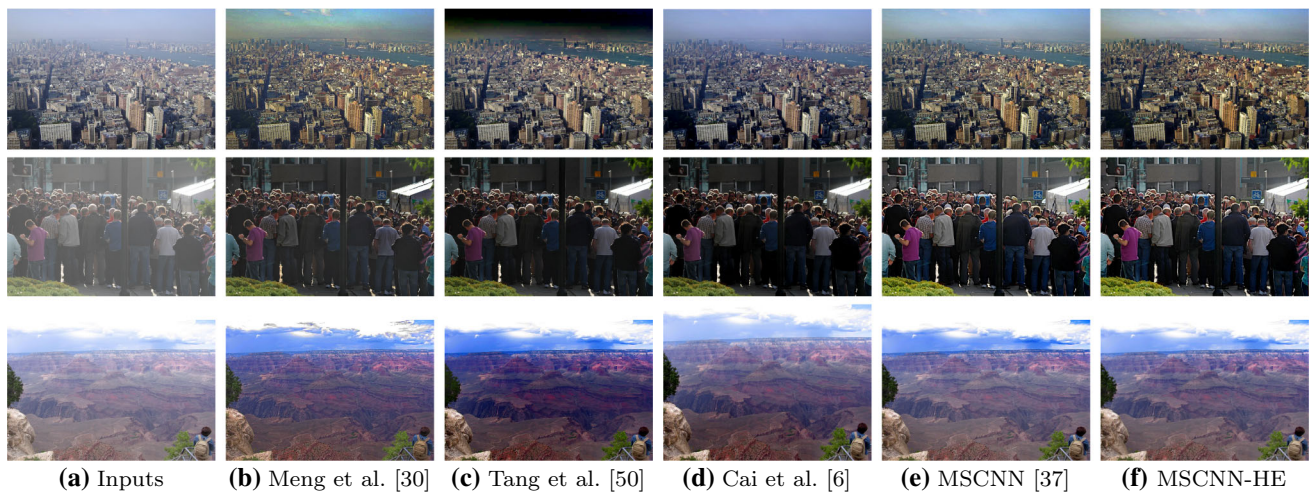
**Table 2** Average run time (in seconds) on test images

Image size	Fattal (2008)	He et al. (2011)	Tarel et al. (2012)	Meng et al. (2013)	Berman et al. (2016)	Cai et al. (2016)	Ours
$427 \times 370$	25.68	13.15	2.02	2.29	4.17	<b>0.63</b>	0.83
$640 \times 480$	63.09	26.90	7.02	3.23	6.23	1.34	<b>1.25</b>

The bold values indicate the best results



**Fig. 12** Visual comparisons for real image dehazing



**Fig. 13** Visual comparisons for real image dehazing

**Table 3** Average PSNR/SSIM of dehazed results on the SOTS dataset (Li et al. 2018)

He et al. (2009)	Meng et al. (2013)	Berman et al. (2016)	Cai et al. (2016)	Li et al. (2017)	MSCNN (Ren et al. 2016)	MSCNN-HE
16.62/0.82	16.88/0.79	17.29/0.75	<u>21.14/0.85</u>	19.06/0.85	17.57/0.81	<b>21.56/0.86</b>

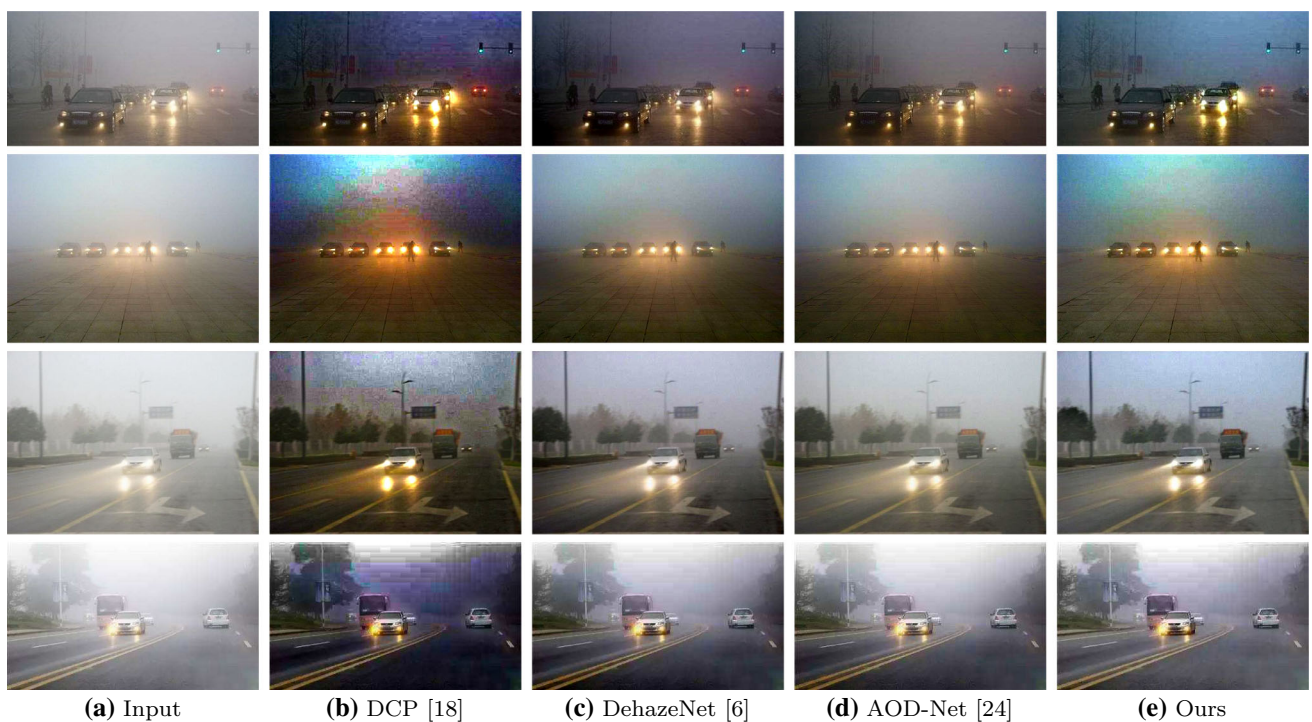
2014; Sulami et al. 2014; Meng et al. 2013; Cai et al. 2016; Li et al. 2017) using seven challenging real-world images in Figs. 12 and 13.

In Fig. 12, the dehazed *City* image by He et al. (2011) and Sulami et al. (2014) tend to overestimate the thickness of the haze and produces dark results as shown in Fig. 12b, d. In addition, the results by Meng et al. (2013) and Sulami et al. (2014) have some color distortions as shown in (c) and (d), especially at the sky regions. The method by Meng

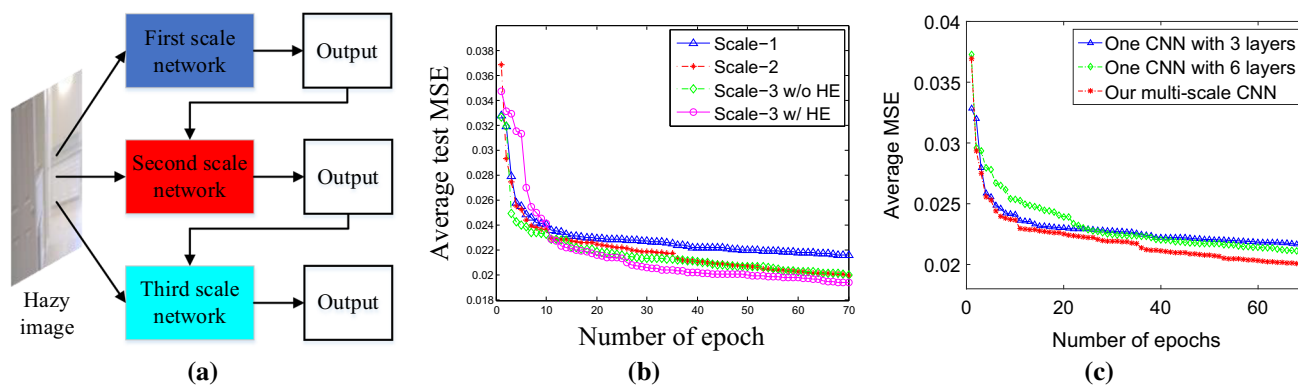
et al. (2013) can augment the image details and enhance the image visibility. However, the recovered images still have some color distortions. For example, the sky color is changed from gray to black in the *City* image in (c). In Fig. 13, due to the methods of Meng et al. (2013) and Tang et al. (2014) still depend on DCP, these approaches also tend to overestimate the thickness of the haze and generates darker images. The results by Cai et al. (2016) have some remaining haze as shown in the third row in Fig. 13d. In contrast, the dehazed



**Fig. 14** Comparison on real-world thick hazy images with deep learning based methods. The methods of DehazeNet and AOD-Net tend to generate dark results, while the dehazed results by DCPDN have some non-uniform dehazing artifacts. In contrast, our results are more pleasant and have vivid scene details



**Fig. 15** Dehazed results on the hazy images where cars' headlights are turned on. The results by DCP and DehazeNet suffer from over-enhancement. Although the details of the scenes and objects are well restored by AOD-Net, the results still have some remaining haze



**Fig. 16** **a** A multi-scale network with three scales. The output of each scale serves as an additional feature in next scale. **b** Comparisons among the first, second and third scale networks with and without holistic edges. As observed, the network with three scales does not lead to better results

results by the proposed algorithm with (MSCNN-HE) and without (MSCNN) using the holistic edge guided network are visually more pleasing in dense haze regions without color distortions or artifacts. Since all the dehazing algorithms are able to get good results by dehazing the general outdoor image as shown in Figs. 12 and 13, we further conduct some experiments on thick hazy images as shown in Fig. 14. In this section, we mainly compare our method against the existing deep learning based approaches since these are the most relevant algorithms to ours. Figure 14a depicts the thick hazy images to be dehazed. Figure 14b–d show the results of DehazeNet (Cai et al. 2016), AOD-Net (Li et al. 2017), and DCPDN (Zhang and Patel 2018), respectively. As shown, both DehazedNet and AOD-Net employ single-scale network and fail to remove dense haze. DCPND exploits the multi-scale strategy but the dehazed results significantly suffer from non-uniform dehazing artifacts (e.g., the nearby region is well dehazed while the distant region still has significant remaining haze for the second image). In contrast, the dehazed results by our algorithm in Fig. 14e are clear and the details are enhanced moderately.

In addition, we collect some hazy images where cars' headlights are turned on from the internet since these scenes are relatively common in hazy days, and compare the proposed algorithm with some state-of-the-art single image dehazing methods. As shown in Fig. 15, the results generated by DCP, DehazeNet, and AOD-Net tend to be dark. In contrast, the dehazed results by the proposed algorithm are visually more pleasing.

## 6 Analysis and Discussions

### 6.1 Generalization Capability

As shown in Sect. 5.2, the proposed multi-scale network performs favorably against the state-of-the-art image dehazing

than the one with two scales. But with the holistic edge guided information, the network with three scales could improve the transmission estimation performance. **c** Comparisons of one CNN with more layers and the proposed multi-scale CNN (Color figure online)

methods for outdoor scenes. In the following, we explain why the proposed network, which is trained on indoor scenes, can handle outdoor images.

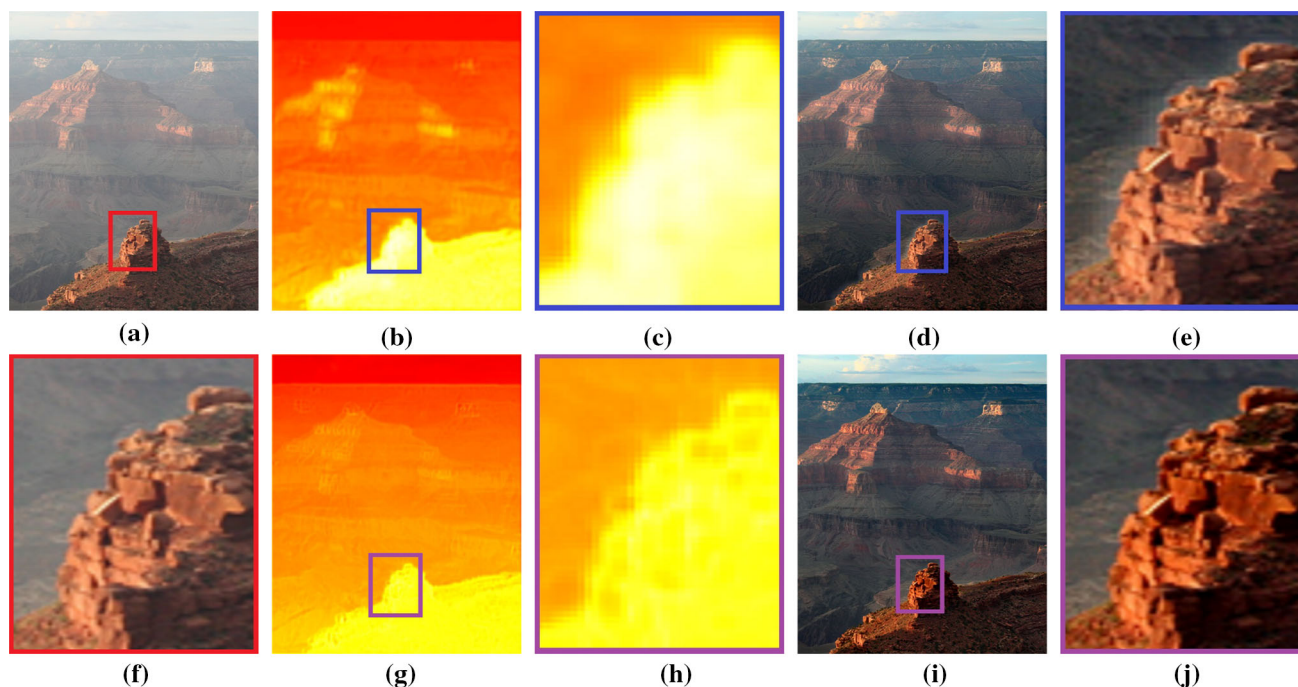
The key observation is that image content is independent of scene depth and medium transmission (Tang et al. 2014), i.e., the same image (or patch) content can appear at different depths in different images. Therefore, although the training images have relatively shallow depths, we could increase the haze concentration by adjusting the value of the medium extinction coefficient  $\beta$ . Based on this premise, the synthetic transmission maps are independent of depth  $d(x)$  and cover a broad range of values in real transmission maps.

### 6.2 Run Time

The proposed algorithm is more efficient than the state-of-the-art image dehazing methods (Fattal 2008; He et al. 2011; Tarel et al. 2012; Meng et al. 2013; Zhu et al. 2015) in terms of run time. We use the five images ( $427 \times 370$  pixels) in Fig. 9 and the 40 images ( $640 \times 480$  pixels) in the new synthetic dataset for evaluation. All the methods are implemented in MATLAB, and we evaluate them on the same machine without GPU acceleration (Intel CPU 3.40 GHz and 16 GB memory). The average run time using two image resolutions is shown in Table 2. Since we add an additional scale in Ren et al. (2016), our algorithm is a little slower than Ren et al. (2016). Nevertheless, the proposed method is still faster than most other methods (Fattal 2008; He et al. 2011; Tarel et al. 2012; Meng et al. 2013; Berman et al. 2016).

### 6.3 Effectiveness of Fine-Scale Network

In this section, we analyze how the fine-scale network helps estimate scene transmission maps. The transmission map from the coarse-scale network serves as additional features in

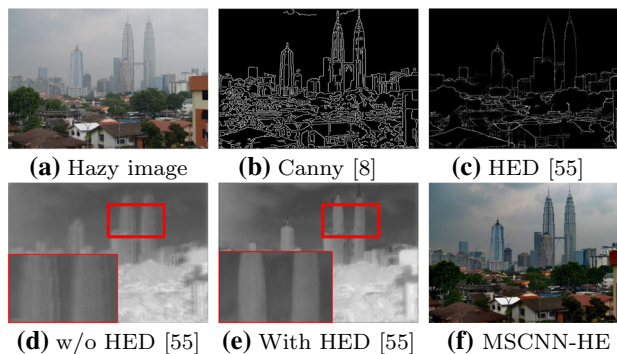


**Fig. 17** Effectiveness of the proposed fine-scale network. **a** Hazy image. **b, d** are the transmission map and dehazed result without the fine-scale network. **g, i** denote transmission map and dehazed result with the fine-scale network. **f, c, e, h, j** are the zoom-in views in **a, b, d, g, i**, respectively

the fine-scale network, which can greatly improve the final estimation of scene transmission map. The validation cost convergence curves (the blue and red lines) in Fig. 16b show that using a fine-scale network could significantly improve the transmission estimation performance. Furthermore, we also train a network with three scales without holistic edge as in Fig. 16a. The output from the second scale also serves as additional features in the third scale network. However, we find that networks with three scales without holistic edge do not help to generate better results as shown in Fig. 16b. The results also show that the proposed network architecture is compact and robust for image dehazing.

To better understand how the fine-scale network affects our method, we conduct a deeper architecture by adding more layers in the single scale network. Figure 16c shows that the CNN with more layers does not perform well compared to the proposed multi-scale CNN. This can be explained by that the output from the coarse-scale network provides sufficiently important features as the input for the fine-scale network. Note that similar observations have been reported in SRCNN (Dong et al. 2014), which indicates that the effectiveness of deeper structures for low-level tasks is not as apparent as that shown in high-level tasks (e.g., image classification).

We show an example of dehazed results with and without the fine-scale network in Fig. 17. Without the fine-scale network, the estimated transmission map in Fig. 17b lacks fine details and the edges of rock do not match with the input hazy image, which accordingly leads to the dehazed results



**Fig. 18** Effectiveness of the proposed holistic edge guided network. With the HED, our estimated transmission map has few trivial edge in the same depth region (Color figure online)

containing halo artifacts around the rock edge as shown in Fig. 17e. In contrast, the transmission map generated with fine-scale network (Fig. 17g) is more informative and thus results in a clearer image (Fig. 17j).

#### 6.4 Effectiveness of Holistic Edge Guided Network

In Sect. 5.1 and Table 1, we have shown that the proposed holistic edge guided network is able to improve the estimated transmission maps and dehazed results. The green and orange lines in Fig. 16b also demonstrate that using the holistic edge information could improve the transmission estimation performance.

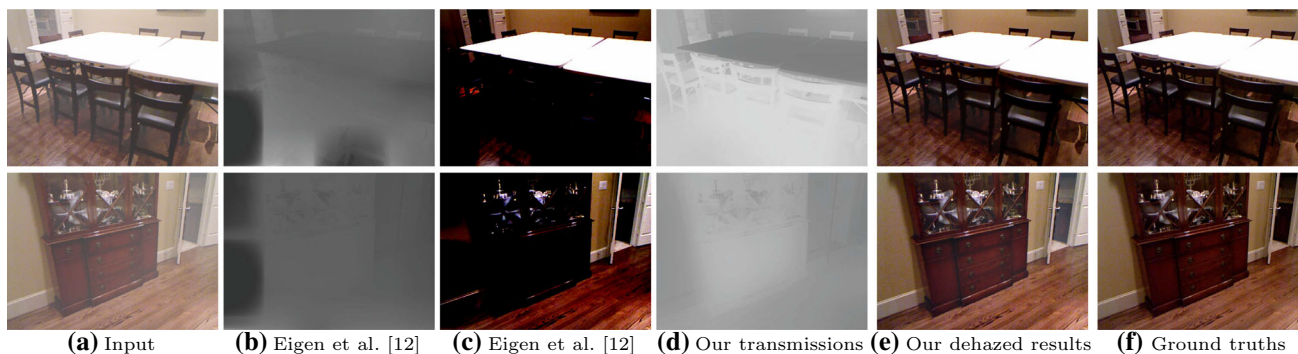


Fig. 19 Dehazed results by the re-trained multi-scale network (Eigen et al. 2014) and our proposed network, respectively

Table 4 Average PSNR and SSIM of dehazed results on the RESIDE dataset (Li et al. 2018) using different configurations

	Depth network (Eigen et al. 2014)	Coarse-scale	Fine-scale	MSCNN-HE
PSNR ( $I$ )	13.17	16.38	<u>17.56</u>	<b>21.56</b>
SSIM ( $I$ )	0.49	0.62	<u>0.81</u>	<b>0.86</b>

The bold and underline values indicate the best and second best results

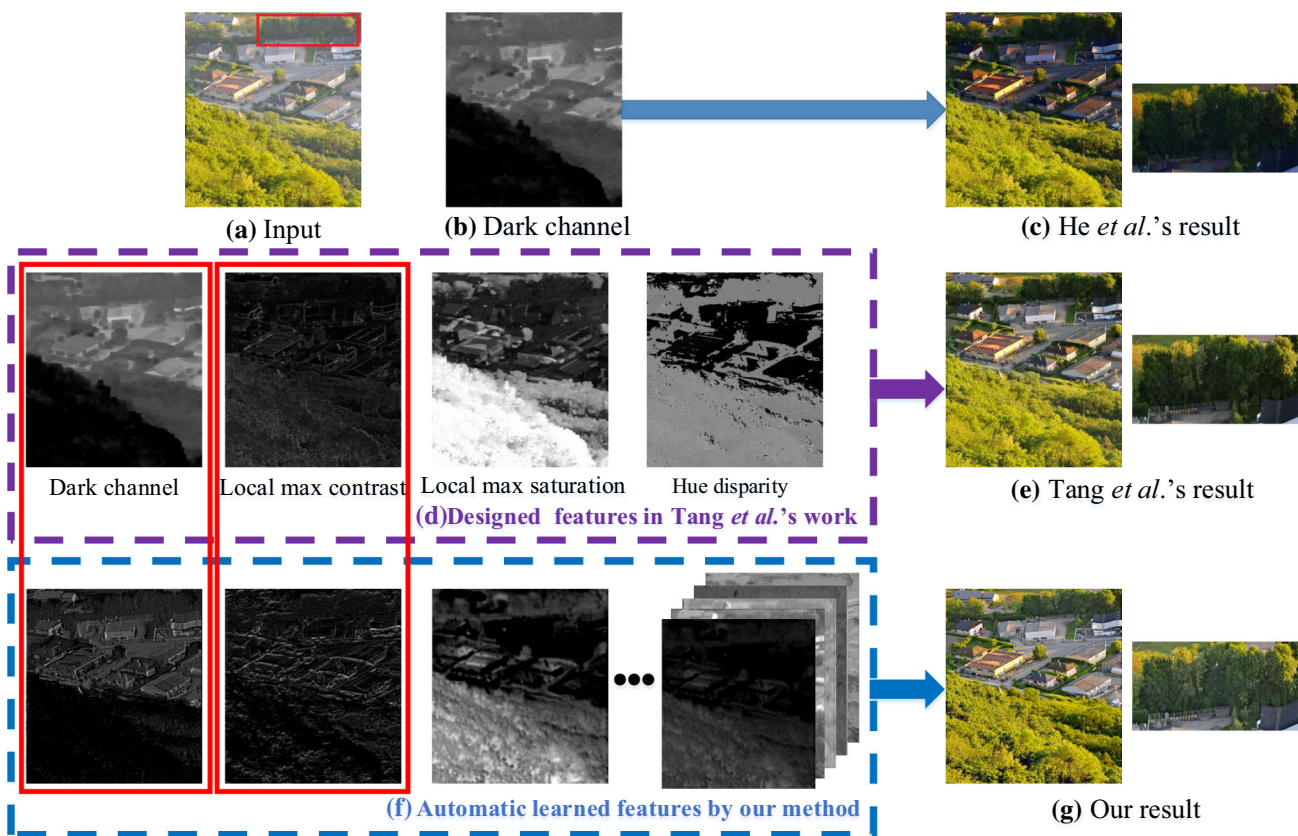
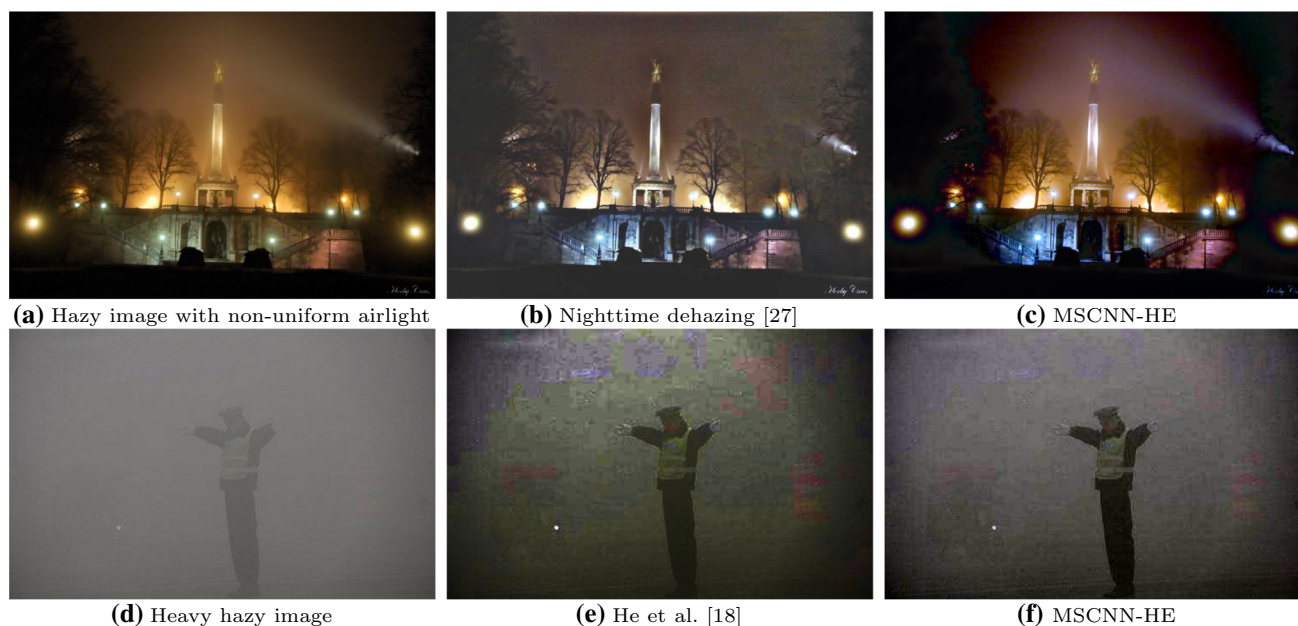


Fig. 20 Effectiveness of learned features. With these diverse features  $f$  automatically learned from the proposed algorithm, our dehazed result is sharper and visually more pleasing than others (Color figure online)

We further demonstrate the effect of the proposed holistic edge guided network. Figure 18 shows the edge detection results and transmission maps with and without the holistic edge guided network. As shown in Fig. 18b, the result

by the Canny detector (Canny et al. 1986) has many trivial edge details. Such details are likely to interfere the transmission map (Fig. 18d). In contrast, the edge detection result in Fig. 18c only contains the main structures





**Fig. 21** Failure cases. Our algorithm does not perform well when the input images contain non-uniform light or heavy haze since the hazy model (1) does not hold for these images

of the hazy images, which are almost consistent with the idea edge of transmission map. Therefore, the buildings belong to same depth have the same transmission as shown in Fig. 18e which results in the clear dehazed image in Fig. 18f.

We note that the proposed method modulates intermediate features based on three-scale networks. We analyze the relationship between transmission estimation accuracy and network configurations in Table 4. The results show that only using the coarse-scale network cannot recover clear images but adding the fine-scale network could improve the performance in terms of PSNR and SSIM. Using the integrated network further helps generate clearer images than those recovered by the state-of-the-art approaches as shown in Table 3.

### 6.5 Connection with Depth Estimation

In Eigen et al. (2014), propose a multi-scale network for depth estimation. The proposed network is also based on the multi-scale CNN (Eigen et al. 2014) but different in several aspects. First, the proposed network does not use a fully-connected layer while the model in Eigen et al. (2014) uses this layer and fixes the output size as one quarter the resolution of the input image. The depth estimation model (Eigen et al. 2014) leads to less accurate transmission maps (e.g., edges in the transmission maps are blurry as shown in Fig. 19b). Second, different from Eigen et al. (2014), we develop an edge information guided approach to better estimate transmission maps. Our analysis and experimental results show

that the proposed network generates clear images and performs favorably against the state-of-the-art methods.

We re-train the network (Eigen et al. 2014) image dehazing and show two examples in Fig. 19b, c. The estimated transmission maps by Eigen et al. (2014) contain blurry artifacts and result in undesired dehazed images. In contrast, the transmission maps estimated by our method contain detailed information and the dehazed images are sharper. Table 4 shows that the re-trained multi-scale network (Eigen et al. 2014) do not dehaze images well on the RESIDE dataset.

### 6.6 Effects of Different Features

In this section, we illustrate the differences between the traditional hand-crafted features and the features learned by the proposed multi-scale CNN model. Traditional methods (He et al. 2011; Tang et al. 2014; Tan 2008) focus on designing hand-crafted features while our method learns the effective haze-relevant features automatically.

Figure 20a shows a hazy input. The dehazed result by only using dark channel feature (b) is shown in (c). As shown, the result has some dark regions when only using the hand-crafted DCP feature. In Tang et al. (2014), propose a learning-based dehazing model. However, this work involves a considerable amount of effort in the design of hand-crafted features including dark channel, local max contrast, local max saturation and hue disparity as shown in (d). By fusing all these features in a regression framework, the dehazed result is shown in (e). In contrast, our network automatically learns the effective features. Figure 20f shows some

features learned by the multi-scale network from the input image. These features are randomly selected from the intermediate layers of the multi-scale CNN model. As shown in Fig. 20f, the learned features include various kinds of information for the input hazy image, including luminance map, intensity map, edge details and amount of haze, and so on. More interestingly, some features learned by the proposed algorithm are similar to the dark channel and local max contrast as shown in the two red rectangles in Fig. 20f, which indicates that the dark channel and local max contrast priors are useful for dehazing as demonstrated by prior studies. With these diverse features learned from the proposed algorithm, the dehazed image shown in Fig. 20g is visually sharper and brighter.

### 6.7 Failure Cases

Our multi-scale CNN model is trained on the synthetic dataset which is created based on the hazy model (1). As the hazy model (1) usually does not hold for the nighttime hazy images (Li et al. 2015; Zhang et al. 2014) or the images with non-uniform atmospheric light since these images often contain other light sources as shown in Fig. 21a, our method is less effective for such images. One failure example is shown in Fig. 21. Our result has some dark region since the inaccuracy of the estimated atmospheric light. Although the proposed algorithm is able to remove thick haze in Fig. 14, our method still cannot handle images with heavy haze where most of the scene information is corrupted by the haze. For example, all the background information is lost in the hazy image in Fig. 21d, the adopted image formation model does not hold for such examples. Figure 21f shows that the proposed method fails to generate a clear image. We aim to address these issues in the future work.

## 7 Conclusions

In this paper, we address the image dehazing problem via a multi-scale deep network which learns effective features to estimate the scene transmission map of a single hazy image. Compared to previous methods which require carefully designed features and combination strategies, the proposed feature learning method is easy to implement and reproduce. In the proposed multi-scale model, we first use a coarse-scale network to learn a holistic estimation of the scene transmission map, and then use a fine-scale network to refine it using local information and the output from the coarse-scale network. In addition, we propose a holistic edge guided network to ensure that the objects with the same depth have the same transmission values. Experimental results on synthetic and real images demonstrate the effectiveness of the proposed algorithm.

**Acknowledgements** This work is supported by the National Key R&D Program of China (Grant No. 2018YFB0803701), Beijing Natural Science Foundation (No. KZ201910005007), National Natural Science Foundation of China (Nos. U1636214, U1803264, U1605252, 61802403, 61602464, 61872421, 61922043), Peng Cheng Laboratory Project of Guangdong Province PCL2018KP004, and the Natural Science Foundation of Jiangsu Province (No. BK20180471). The work of W. Ren is supported in part by the CCF-DiDi GAIA (YF20180101), CCF-Tencent Open Fund, Zhejiang Lab's International Talent Fund for Young Professionals, and the Open Projects Program of the National Laboratory of Pattern Recognition. The work of M.-H. Yang is supported by Directorate for Computer and Information Science and Engineering (CAREER 1149783).

## References

- Ancuti, C. O., & Ancuti, C. (2013). Single image dehazing by multi-scale fusion. *IEEE Transactions on Image Processing*, 22(8), 3271–3282.
- Ancuti, C. O., Ancuti, C., & Bekaert, P. (2010). Effective single image dehazing by fusion. In *IEEE international conference on image processing*, (pp. 3541–3544).
- Berman, D., Treibitz, T., & Avidan, S. (2017). Air-light estimation using haze-lines. In *IEEE international conference on computational photography*, (pp. 1–9).
- Berman, D., Treibitz, T., & Shai, A. (2016). Non-local image dehazing. In *IEEE conference on computer vision and pattern recognition*, (pp. 1674–1682).
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.
- Cai, B., Xu, X., Jia, K., Qing, C., & Tao, D. (2016). Dehazenet: An end-to-end system for single image haze removal. *IEEE Transactions on Image Processing*, 25(11), 5187–5198.
- Cai, B., Xu, X., & Tao, D. (2016). Real-time video dehazing based on spatio-temporal mrf. In *Pacific Rim conference on multimedia*, (pp. 315–325).
- Canny, J. (1986). A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6), 679–98.
- Caraffa, L., & Tarel, J. P. (2012). Stereo reconstruction and contrast restoration in daytime fog. In *Asian conference on computer vision*, (pp. 13–25).
- Caraffa, L., & Tarel, J. P. (2013). Markov random field model for single image defogging. In *IEEE intelligent vehicles symposium*, (pp. 994–999).
- Dong, C., Loy, C. C., He, K., & Tang, X. (2014). Learning a deep convolutional network for image super-resolution. In *European conference on computer vision*, (pp. 184–199).
- Eigen, D., Puhrsch, C., & Fergus, R. (2014). Depth map prediction from a single image using a multi-scale deep network. In *Advances in neural information processing systems*, (pp. 2366–2374).
- Fattal, R. (2008). Single image dehazing. *ACM Transactions on Graphics*, 27(3), 72.
- Fattal, R. (2014). Dehazing using color-lines. *ACM Transactions on Graphics*, 34(1), 13.
- Geiger, A., Lenz, P., & Urtasun, R. (2012). Are we ready for autonomous driving? The kitti vision benchmark suite. In *IEEE conference on computer vision and pattern recognition*, (pp. 3354–3361).
- Gibson, K. B., & Nguyen, T. Q. (2013). Fast single image fog removal using the adaptive wiener filter. In *IEEE international conference on image processing*, (pp. 714–718).
- Gibson, K. B., Vo, D. T., & Nguyen, T. Q. (2012). An investigation of dehazing effects on image and video coding. *IEEE Transactions on Image Processing*, 21(2), 662–673.

- He, K., Sun, J., & Tang, X. (2009). Single image haze removal using dark channel prior. In *IEEE conference on computer vision and pattern recognition*, (pp. 1956–1963).
- He, K., Sun, J., & Tang, X. (2011). Single image haze removal using dark channel prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(12), 2341–2353.
- He, K., Sun, J., & Tang, X. (2013). Guided image filtering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(6), 1397–1409.
- Kopf, J., Neubert, B., Chen, B., Cohen, M., Cohen-Or, D., Deussen, O., et al. (2008). Deep photo: Model-based photograph enhancement and viewing. *ACM Transactions on Graphics*, 27(5), 116.
- Koschmieder, H. (1924). Theorie der horizontalen sichtweite. *Beitrag zur Physik der freien Atmosphäre* (pp. 33–53).
- Kratz, L., & Nishino, K. (2009). Factorizing scene albedo and depth from a single foggy image. In *IEEE international conference on computer vision*, (pp. 1701–1708).
- Li, B., Peng, X., Wang, Z., Xu, J., & Feng, D. (2017). Aod-net: All-in-one dehazing network. In *IEEE international conference on computer vision*.
- Li, B., Ren, W., Fu, D., Tao, D., Feng, D., Zeng, W., et al. (2018). Benchmarking single-image dehazing and beyond. *IEEE Transactions on Image Processing*, 28(1), 492–505.
- Li, S., Araujo, I. B., Ren, W., Wang, Z., Tokuda, E. K., Junior, R. H., Cesar-Junior, R., Zhang, J., Guo, X., & Cao, X. (2019). Single image deraining: A comprehensive benchmark analysis. In *IEEE conference on computer vision and pattern recognition*, (pp. 3838–3847).
- Li, Y., Tan, R. T., & Brown, M. S. (2015). Nighttime haze removal with glow and multiple light colors. In *IEEE international conference on computer vision*, (pp. 226–234).
- Li, Z., Tan, P., Tan, R. T., Zou, D., Zhou, S. Z., & Cheong, L. F. (2015). Simultaneous video defogging and stereo reconstruction. In *IEEE conference on computer vision and pattern recognition*, (pp. 4988–4997).
- Liu, Y., Zhao, G., Gong, B., Li, Y., Raj, R., Goel, N., Kesav, S., Gottimukkala, S., Wang, Z., Ren, W., et al. (2018). Improved techniques for learning to dehaze and beyond: A collective study. *arXiv preprint arXiv:1807.00202*
- Meng, G., Wang, Y., Duan, J., Xiang, S., & Pan, C. (2013). Efficient image dehazing with boundary constraint and contextual regularization. In *IEEE international conference on computer vision*, (pp. 617–624).
- Nair, V., & Hinton, G. E. (2010). Rectified linear units improve restricted Boltzmann machines. In *International conference on machine learning*, (pp. 807–814).
- Narasimhan, S. G., & Nayar, S. K. (2000). Chromatic framework for vision in bad weather. In *IEEE conference on computer vision and pattern recognition*, (pp. 598–605).
- Narasimhan, S. G., & Nayar, S. K. (2003). Contrast restoration of weather degraded images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(6), 713–724.
- Nayar, S. K., & Narasimhan, S. G. (1999). Vision in bad weather. In *IEEE international conference on computer vision*, (pp. 820–827).
- Nishino, K., Kratz, L., & Lombardi, S. (2012). Bayesian defogging. *International Journal of Computer Vision*, 98(3), 263–278.
- Ren, W., & Cao, X. (2017). Deep video dehazing. In *Pacific rim conference on multimedia*, (pp. 14–24).
- Ren, W., Liu, S., Zhang, H., Pan, J., Cao, X., & Yang, M. H. (2016). Single image dehazing via multi-scale convolutional neural networks. In *European conference on computer vision*, (pp. 154–169).
- Ren, W., Ma, L., Zhang, J., Pan, J., Cao, X., Liu, W., & Yang, M. H. (2018). Gated fusion network for single image dehazing. In *IEEE conference on computer vision and pattern recognition*, (pp. 3253–3261).
- Ren, W., Zhang, J., Xu, X., Ma, L., Cao, X., Meng, G., et al. (2018). Deep video dehazing with semantic segmentation. *IEEE Transactions on Image Processing*, 28(4), 1895–1908.
- Saxena, A., Sun, M., & Ng, A. Y. (2009). Make3d: Learning 3d scene structure from a single still image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(5), 824–840.
- Scharstein, D., & Szeliski, R. (2002). A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, 47(1–3), 7–42.
- Scharstein, D., & Szeliski, R. (2003). High-accuracy stereo depth maps using structured light. In *IEEE conference on computer vision and pattern recognition*, (Vol. 1, pp. 1–195).
- Schaul, L., Fredembach, C., & Süsstrunk, S. (2009). Color image dehazing using the near-infrared. In *IEEE international conference on image processing*, (pp. 1629–1632).
- Schechner, Y. Y., Narasimhan, S. G., & Nayar, S. K. (2001). Instant dehazing of images using polarization. In *IEEE conference on computer vision and pattern recognition*, (Vol. 1, pp. 1–325).
- Shwartz, S., Namer, E., & Schechner, Y. Y. (2006). Blind haze separation. In *IEEE conference on computer vision and pattern recognition*, (pp. 1984–1991).
- Silberman, N., Hoiem, D., Kohli, P., & Fergus, R. (2012). Indoor segmentation and support inference from RGBD images. In *European conference on computer vision*, (pp. 746–760).
- Sulami, M., Glatzer, I., Fattal, R., & Werman, M. (2014). Automatic recovery of the atmospheric light in hazy images. In *IEEE international conference on computational photography*.
- Tan, R. T. (2008). Visibility in bad weather from a single image. In *IEEE conference on computer vision and pattern recognition*, (pp. 1–8).
- Tan, R. T., Pettersson, N., & Pettersson, L. (2007). Visibility enhancement for roads with foggy or hazy scenes. In *IEEE intelligent vehicles symposium*, (pp. 19–24).
- Tang, K., Yang, J., & Wang, J. (2014). Investigating haze-relevant features in a learning framework for image dehazing. In *IEEE conference on computer vision and pattern recognition*, (pp. 2995–3002).
- Tarel, J. P., & Hautiere, N. (2009). Fast visibility restoration from a single color or gray level image. In *IEEE international conference on computer vision*, (pp. 2201–2208).
- Tarel, J. P., Hautière, N., Caraffa, L., Cord, A., Halmaoui, H., & Gruyer, D. (2012). Vision enhancement in homogeneous and heterogeneous fog. *Intelligent Transportation Systems Magazine*, 4(2), 6–20.
- Treibitz, T., & Schechner, Y. Y. (2009). Polarization: Beneficial for visibility enhancement? In *IEEE conference on computer vision and pattern recognition*, (pp. 525–532).
- Wang, Y. K., & Fan, C. T. (2014). Single image defogging by multiscale depth fusion. *IEEE Transactions on Image Processing*, 23(11), 4826–4837.
- Xie, S., & Tu, Z. (2015). Holistically-nested edge detection. In *IEEE international conference on computer vision*, (pp. 1395–1403).
- Zhang, H., & Patel, V. M. (2018). Densely connected pyramid dehazing network. In *IEEE conference on computer vision and pattern recognition*.
- Zhang, H., Sindagi, V., & Patel, V. M. (2018). Multi-scale single image dehazing using perceptual pyramid deep network. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, (pp. 902–911).
- Zhang, J., Cao, Y., & Wang, Z. (2014). Nighttime haze removal based on a new imaging model. In *IEEE international conference on image processing*, (pp. 4557–4561).
- Zhang, S., He, F., Ren, W., & Yao, J. (2018). Joint learning of image detail and transmission map for single image dehazing. *The Visual Computer*, 34, 1–12.

- Zhang, S., Ren, W., & Yao, J. (2018). Feed-net: Fully end-to-end dehazing. In *IEEE international conference on multimedia and expo*, (pp. 1–6).
- Zhang, X. S., Gao, S. B., Li, C. Y., & Li, Y. J. (2015). A retina inspired model for enhancing visibility of hazy images. *Frontiers in Computational Neuroscience*, 9, 1–13.
- Zhu, Q., Mai, J., & Shao, L. (2014). Single image dehazing using color attenuation prior. In *British machine vision conference*
- Zhu, Q., Mai, J., & Shao, L. (2015). A fast single image haze removal algorithm using color attenuation prior. *IEEE Transactions on Image Processing*, 24(11), 3522–3533.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.