

 Open access • Journal Article • DOI:10.1109/TIP.2016.2598681

## DehazeNet: An End-to-End System for Single Image Haze Removal — [Source link](#)

[Bolun Cai](#), [Xiangmin Xu](#), [Kui Jia](#), [Chunmei Qing](#) ...+1 more authors

**Institutions:** [South China University of Technology](#), [University of Technology, Sydney](#)

**Published on:** 28 Jan 2016 - [arXiv: Computer Vision and Pattern Recognition](#)

**Topics:** [Feature extraction](#) and [Convolutional neural network](#)

Related papers:

- [DehazeNet: An End-to-End System for Single Image Haze Removal](#)
- [Single Image Haze Removal via Joint Estimation of Detail and Transmission](#)
- [Deep fully convolutional regression networks for single image haze removal](#)
- [Single image dehazing via combining the prior knowledge and CNNs](#)
- [A Light Dual-Task Neural Network for Haze Removal](#)

Share this paper:    

View more about this paper here: <https://typeset.io/papers/dehazenet-an-end-to-end-system-for-single-image-haze-removal-1p2p3nxwmw>

© 2016 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

# DehazeNet: An End-to-End System for Single Image Haze Removal

Bolun Cai, Xiangmin Xu, *Member, IEEE*, Kui Jia, *Member, IEEE*, Chunmei Qing, *Member, IEEE*,  
and Dacheng Tao, *Fellow, IEEE*

**Abstract**—Single image haze removal is a challenging ill-posed problem. Existing methods use various constraints/priors to get plausible dehazing solutions. The key to achieve haze removal is to estimate a medium transmission map for an input hazy image. In this paper, we propose a trainable end-to-end system called DehazeNet, for medium transmission estimation. DehazeNet takes a hazy image as input, and outputs its medium transmission map that is subsequently used to recover a haze-free image via atmospheric scattering model. DehazeNet adopts Convolutional Neural Networks (CNN) based deep architecture, whose layers are specially designed to embody the established assumptions/priors in image dehazing. Specifically, layers of Maxout units are used for feature extraction, which can generate almost all haze-relevant features. We also propose a novel nonlinear activation function in DehazeNet, called Bilateral Rectified Linear Unit (BReLU), which is able to improve the quality of recovered haze-free image. We establish connections between components of the proposed DehazeNet and those used in existing methods. Experiments on benchmark images show that DehazeNet achieves superior performance over existing methods, yet keeps efficient and easy to use.

**Keywords**—Dehaze, image restoration, deep CNN, BReLU.

## I. INTRODUCTION

**H**AZE is a traditional atmospheric phenomenon where dust, smoke and other dry particles obscure the clarity of the atmosphere. Haze causes issues in the area of terrestrial photography, where the light penetration of dense atmosphere may be necessary to image distant subjects. This results in the visual effect of a loss of contrast in the subject, due to the effect of light scattering through the haze particles. For these reasons, haze removal is desired in both consumer photography and computer vision applications.

Haze removal is a challenging problem because the haze transmission depends on the unknown depth which varies at different positions. Various techniques of image enhancement have been applied to the problem of removing haze from a single image, including histogram-based [1], contrast-based [2]

and saturation-based [3]. In addition, methods using multiple images or depth information have also been proposed. For example, polarization based methods [4] remove the haze effect through multiple images taken with different degrees of polarization. In [5], multi-constraint based methods are applied to multiple images capturing the same scene under different weather conditions. Depth-based methods [6] require some depth information from user inputs or known 3D models. In practice, depth information or multiple hazy images are not always available.

Single image haze removal has made significant progresses recently, due to the use of better assumptions and priors. Specifically, under the assumption that the local contrast of the haze-free image is much higher than that of hazy image, a local contrast maximizing method [7] based on Markov Random Field (MRF) is proposed for haze removal. Although contrast maximizing approach is able to achieve impressive results, it tends to produce over-saturated images. In [8], Independent Component Analysis (ICA) based on minimal input is proposed to remove the haze from color images, but the approach is time-consuming and cannot be used to deal with dense-haze images. Inspired by dark-object subtraction technique, Dark Channel Prior (DCP) [9] is discovered based on empirical statistics of experiments on haze-free images, which shows at least one color channel has some pixels with very low intensities in most of non-haze patches. With dark channel prior, the thickness of haze is estimated and removed by the atmospheric scattering model. However, DCP loses dehazing quality in the sky images and is computationally intensive. Some improved algorithms are proposed to overcome these limitations. To improve dehazing quality, Kratz and Nishino *et al.* [10] model the image with a factorial MRF to estimate the scene radiance more accurately; Meng *et al.* [11] propose an effective regularization dehazing method to restore the haze-free image by exploring the inherent boundary constraint. To improve computational efficiency, standard median filtering [12], median of median filter [13], guided joint bilateral filtering [14] and guided image filter [15] are used to replace the time-consuming soft matting [16]. In recent years, haze-relevant priors are investigated in machine learning framework. Tang *et al.* [17] combine four types of haze-relevant features with Random Forests to estimate the transmission. Zhu *et al.* [18] create a linear model for estimating the scene depth of the hazy image under color attenuation prior and learns the parameters of the model with a supervised method. Despite the remarkable progress, these state-of-the-art methods are limited by the very same haze-relevant priors or heuristic cues - they

B. Cai, X. Xu (✉) and C. Qing are with School of Electronic and Information Engineering, South China University of Technology, Wushan RD., Tianhe District, Guangzhou, P.R.China. E-mail: {caibolun@gmail.com, xmxu@scut.edu.cn, qchm@scut.edu.cn}.

K. Jia is with the Department of Electrical and Computer Engineering, Faculty of Science and Technology, University of Macau, Macau 999078, China. E-mail: {kuijia@gmail.com}.

D. Tao is with Centre for Quantum Computation & Intelligent Systems, Faculty of Engineering & Information Technology, University of Technology Sydney, 235 Jones Street, Ultimo, NSW 2007, Australia. E-mail: {dacheng.tao@uts.edu.au}.

are often less effective for some images.

Haze removal from a single image is a difficult vision task. In contrast, the human brain can quickly identify the hazy area from the natural scenery without any additional information. One might be tempted to propose biologically inspired models for image dehazing, by following the success of bio-inspired CNNs for high-level vision tasks such as image classification [19], face recognition [20] and object detection [21]. In fact, there have been a few (convolutional) neural network based deep learning methods that are recently proposed for low-level vision tasks of image restoration/reconstruction [22], [23], [24]. However, these methods cannot be directly applied to single image haze removal.

Note that apart from estimation of a global atmospheric light magnitude, the key to achieve haze removal is to recover an accurate medium transmission map. To this end, we propose DehazeNet, a trainable CNN based end-to-end system for medium transmission estimation. DehazeNet takes a hazy image as input, and outputs its medium transmission map that is subsequently used to recover the haze-free image by a simple pixel-wise operation. Design of DehazeNet borrows ideas from established assumptions/principles in image dehazing, while parameters of all its layers can be automatically learned from training hazy images. Experiments on benchmark images show that DehazeNet gives superior performance over existing methods, yet keeps efficient and easy to use. Our main contributions are summarized as follows.

- 1) DehazeNet is an end-to-end system. It directly learns and estimates the mapping relations between hazy image patches and their medium transmissions. This is achieved by special design of its deep architecture to embody established image dehazing principles.
- 2) We propose a novel nonlinear activation function in DehazeNet, called Bilateral Rectified Linear Unit<sup>1</sup> (BReLU). BReLU extends Rectified Linear Unit (ReLU) and demonstrates its significance in obtaining accurate image restoration. Technically, BReLU uses the bilateral restraint to reduce search space and improve convergence.
- 3) We establish connections between components of DehazeNet and those assumptions/priors used in existing dehazing methods, and explain that DehazeNet improves over these methods by automatically learning all these components from end to end.

The remainder of this paper is organized as follows. In Section II, we review the atmospheric scattering model and haze-relevant features, which provides background knowledge to understand the design of DehazeNet. In Section III, we present details of the proposed DehazeNet, and discuss how it relates to existing methods. Experiments are presented in

<sup>1</sup>During the preparation of this manuscript (in December, 2015), we find that a nonlinear activation function called adjustable bounded rectifier is proposed in [25] (arXived in November, 2015), which is almost identical to BReLU. Adjustable bounded rectifier is motivated to achieve the objective of image recognition. In contrast, BReLU is proposed here to improve image restoration accuracy. It is interesting that we come to the same activation function from completely different initial objectives. This may also suggest the general usefulness of the proposed BReLU.

Section IV, before conclusion is drawn in Section V.

## II. RELATED WORKS

Many image dehazing methods have been proposed in the literature. In this section, we briefly review some important ones, paying attention to those proposing the atmospheric scattering model, which is the basic underlying model of image dehazing, and those proposing useful assumptions for computing haze-relevant features.

### A. Atmospheric Scattering Model

To describe the formation of a hazy image, the atmospheric scattering model is first proposed by McCartney [26], which is further developed by Narasimhan and Nayar [27], [28]. The atmospheric scattering model can be formally written as

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

where  $I(x)$  is the observed hazy image,  $J(x)$  is the real scene to be recovered,  $t(x)$  is the medium transmission,  $\alpha$  is the global atmospheric light, and  $x$  indexes pixels in the observed hazy image  $I$ . Fig. 1 gives an illustration. There are three unknowns in equation (1), and the real scene  $J(x)$  can be recovered after  $\alpha$  and  $t(x)$  are estimated.

The medium transmission map  $t(x)$  describes the light portion that is not scattered and reaches the camera.  $t(x)$  is defined as

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

where  $d(x)$  is the distance from the scene point to the camera, and  $\beta$  is the scattering coefficient of the atmosphere. Equation (2) suggests that when  $d(x)$  goes to infinity,  $t(x)$  approaches zero. Together with equation (1) we have

$$\alpha = I(x), d(x) \rightarrow \inf \quad (3)$$

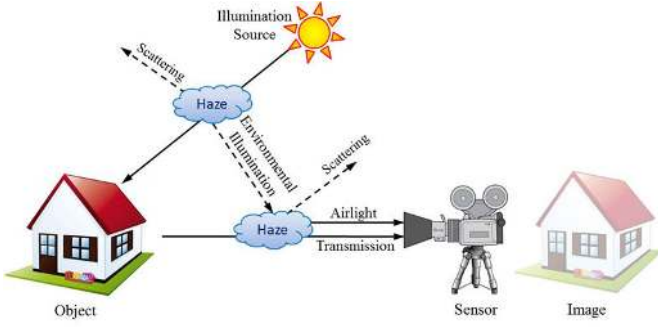
In practical imaging of a distance view,  $d(x)$  cannot be infinity, but rather be a long distance that gives a very low transmission  $t_0$ . Instead of relying on equation (3) to get the global atmospheric light  $\alpha$ , it is more stably estimated based on the following rule

$$\alpha = \max_{y \in \{x | t(x) \leq t_0\}} I(y) \quad (4)$$

The discussion above suggests that *to recover a clean scene (i.e., to achieve haze removal), it is the key to estimate an accurate medium transmission map.*

### B. Haze-relevant features

Image dehazing is an inherently ill-posed problem. Based on empirical observations, existing methods propose various assumptions or prior knowledge that are utilized to compute intermediate haze-relevant features. Final haze removal can be achieved based on these haze-relevant features.



(a) The process of imaging in hazy weather. The transmission attenuation  $J(x)t(x)$  caused by the reduction in reflected energy, leads to low brightness intensity. The airlight  $\alpha(1-t(x))$  formed by the scattering of the environmental illumination, enhances the brightness and reduces the saturation.

Fig. 1. Imaging in hazy weather and atmospheric scattering model

1) *Dark Channel*: The dark channel prior is based on the wide observation on outdoor haze-free images. In most of the haze-free patches, at least one color channel has some pixels whose intensity values are very low and even close to zero. The dark channel [9] is defined as the minimum of all pixel colors in a local patch:

$$D(x) = \min_{y \in \Omega_r(x)} \left( \min_{c \in \{r, g, b\}} I^c(y) \right), \quad (5)$$

where  $I^c$  is a RGB color channel of  $I$  and  $\Omega_r(x)$  is a local patch centered at  $x$  with the size of  $r \times r$ . The dark channel feature has a high correlation to the amount of haze in the image, and is used to estimate the medium transmission  $t(x) \propto 1 - D(x)$  directly.

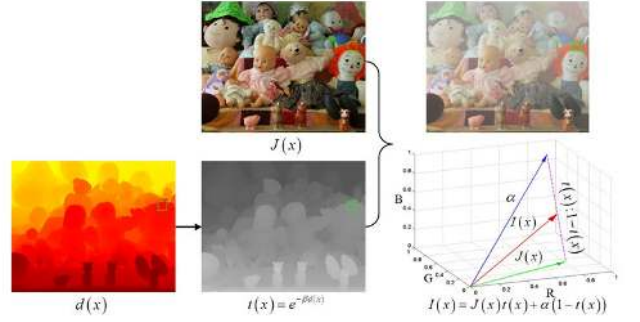
2) *Maximum Contrast*: According to the atmospheric scattering, the contrast of the image is reduced by the haze transmission as  $\sum_x \|\nabla I(x)\| = t \sum_x \|\nabla J(x)\| \leq \sum_x \|\nabla J(x)\|$ . Based on this observation, the local contrast [7] as the variance of pixel intensities in a  $s \times s$  local patch  $\Omega_s$  with respect to the center pixel, and the local maximum of local contrast values in a  $r \times r$  region  $\Omega_r$  is defined as:

$$C(x) = \max_{y \in \Omega_r(x)} \sqrt{\frac{1}{|\Omega_s(y)|} \sum_{z \in \Omega_s(y)} \|I(z) - I(y)\|^2}, \quad (6)$$

where  $|\Omega_s(y)|$  is the cardinality of the local neighborhood. The correlation between the contrast feature and the medium transmission  $t$  is visually obvious, so the visibility of the image is enhanced by maximizing the local contrast showed as (6).

3) *Color Attenuation*: The saturation  $I^s(x)$  of the patch decreases sharply while the color of the scene fades under the influence of the haze, and the brightness value  $I^v(x)$  increases at the same time producing a high value for the difference. According to the above color attenuation prior [18], the difference between the brightness and the saturation is utilized to estimate the concentration of the haze:

$$A(x) = I^v(x) - I^s(x), \quad (7)$$



(b) Atmospheric scattering model. The observed hazy image  $I(x)$  is generated by the real scene  $J(x)$ , the medium transmission  $t(x)$  and the global atmospheric light  $\alpha$ .

where  $I^v(x)$  and  $I^h(x)$  can be expressed in the HSV color space as  $I^v(x) = \max_{c \in \{r, b, g\}} I^c(x)$  and  $I^s(x) = (\max_{c \in \{r, b, g\}} I^c(x) - \min_{c \in \{r, b, g\}} I^c(x)) / \max_{c \in \{r, b, g\}} I^c(x)$ . The color attenuation feature is proportional to the scene depth  $d(x) \propto A(x)$ , and is used for transmission estimation easily.

4) *Hue Disparity*: Hue disparity between the original image  $I(x)$  and its semi-inverse image,  $I_{si}(x) = \max[I^c(x), 1 - I^c(x)]$  with  $c \in \{r, g, b\}$ , has been used to detect the haze. For haze-free images, pixel values in the three channels of their semi-inverse images will not all flip, resulting in large hue changes between  $I_{si}(x)$  and  $I(x)$ . In [29], the hue disparity feature is defined:

$$H(x) = |I_{si}^h(x) - I^h(x)|, \quad (8)$$

where the superscript "h" denotes the hue channel of the image in HSV color space. According to (8), the medium transmission  $t(x)$  is in inverse propagation to  $H(x)$ .

### III. THE PROPOSED DEHAZENET

The atmospheric scattering model in Section II-A suggests that estimation of the medium transmission map is the most important step to recover a haze-free image. To this end, we propose DehazeNet, a trainable end-to-end system that explicitly learns the mapping relations between raw hazy images and their associated medium transmission maps. In this section, we present layer designs of DehazeNet, and discuss how these designs are related to ideas in existing image dehazing methods. The final pixel-wise operation to get a recovered haze-free image from the estimated medium transmission map will be presented in Section IV.

#### A. Layer Designs of DehazeNet

The proposed DehazeNet consists of cascaded convolutional and pooling layers, with appropriate nonlinear activation functions employed after some of these layers. Fig. 2 shows the architecture of DehazeNet. Layers and nonlinear activations

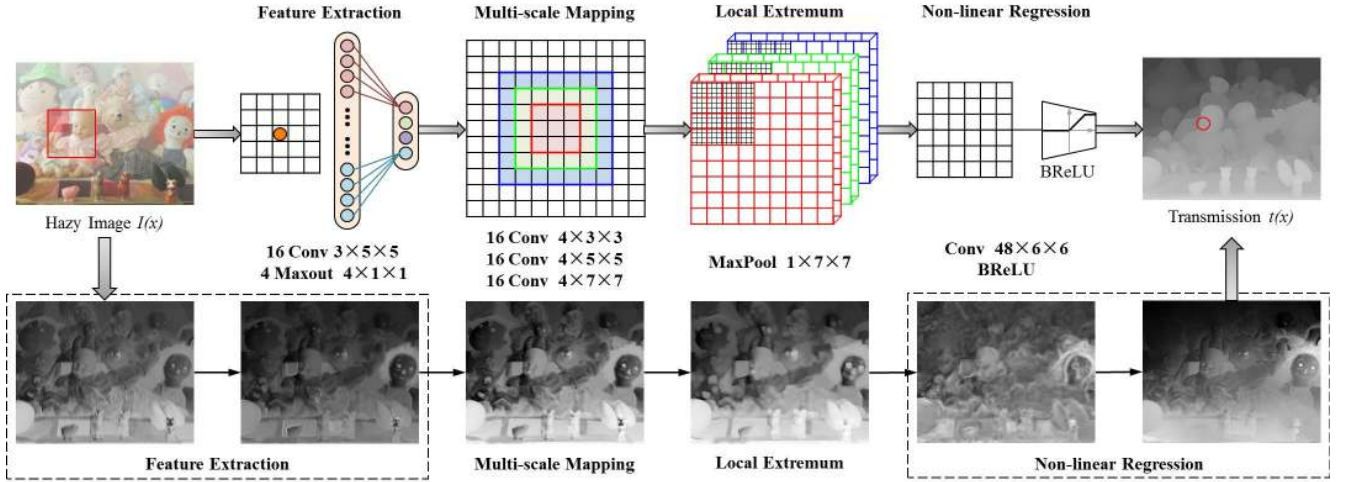


Fig. 2. The architecture of DehazeNet. DehazeNet conceptually consists of four sequential operations (feature extraction, multi-scale mapping, local extremum and non-linear regression), which is constructed by 3 convolution layers, a max-pooling, a Maxout unit and a BReLU activation function.

of DehazeNet are designed to implement four sequential operations for medium transmission estimation, namely, feature extraction, multi-scale mapping, local extremum, and nonlinear regression. We detail these designs as follows.

1) *Feature Extraction*: To address the ill-posed nature of image dehazing problem, existing methods propose various assumptions and based on these assumptions, they are able to extract haze-relevant features (e.g., dark channel, hue disparity, and color attenuation) densely over the image domain. Note that densely extracting these haze-relevant features is equivalent to convolving an input hazy image with appropriate filters, followed by nonlinear mappings. Inspired by extremum processing in color channels of those haze-relevant features, an unusual activation function called Maxout unit [30] is selected as the non-linear mapping for dimension reduction. Maxout unit is a simple feed-forward nonlinear activation function used in multi-layer perceptron or CNNs. When used in CNNs, it generates a new feature map by taking a pixel-wise maximization operation over  $k$  affine feature maps. Based on Maxout unit, we design the first layer of DehazeNet as follows

$$F_1^i(x) = \max_{j \in [1, k]} f_1^{i, j}(x), f_1^{i, j} = W_1^{i, j} * I + B_1^{i, j}, \quad (9)$$

where  $W_1 = \{W_1^{i, j}\}_{(i, j)=(1, 1)}^{(n_1, k)}$  and  $B_1 = \{B_1^{i, j}\}_{(i, j)=(1, 1)}^{(n_1, k)}$  represent the filters and biases respectively, and  $*$  denotes the convolution operation. Here, there are  $n_1$  output feature maps in the first layer.  $W_1^{i, j} \in \mathbb{R}^{3 \times f_1 \times f_1}$  is one of the total  $k \times n_1$  convolution filters, where 3 is the number of channels in the input image  $I(x)$ , and  $f_1$  is the spatial size of a filter (detailed in Table I). Maxout unit maps each of  $kn_1$ -dimensional vectors into an  $n_1$ -dimensional one, and extracts the haze-relevant features by automatic learning rather than heuristic ways in existing methods.

2) *Multi-scale Mapping*: In [17], multi-scale features have been proven effective for haze removal, which densely compute features of an input image at multiple spatial scales. Multi-scale feature extraction is also effective to achieve

scale invariance. For example, the inception architecture in GoogLeNet [31] uses parallel convolutions with varying filter sizes, and better addresses the issue of aligning objects in input images, resulting in state-of-the-art performance in ILSVRC14 [32]. Motivated by these successes of multi-scale feature extraction, we choose to use parallel convolutional operations in the second layer of DehazeNet, where size of any convolution filter is among  $3 \times 3$ ,  $5 \times 5$  and  $7 \times 7$ , and we use the same number of filters for these three scales. Formally, the output of the second layer is written as

$$F_2^i = W_2^{[i/3], (i \setminus 3)} * F_1 + B_2^{[i/3], (i \setminus 3)}, \quad (10)$$

where  $W_2 = \{W_2^{p, q}\}_{(p, q)=(1, 1)}^{(3, n_2/3)}$  and  $B_2 = \{B_2^{p, q}\}_{(p, q)=(1, 1)}^{(3, n_2/3)}$  contain  $n_2$  pairs of parameters that is break up into 3 groups.  $n_2$  is the output dimension of the second layer, and  $i \in [1, n_2]$  indexes the output feature maps.  $\lceil \cdot \rceil$  takes the integer upwardly and  $\setminus$  denotes the remainder operation.

3) *Local Extremum*: To achieve spatial invariance, the cortical complex cells in the visual cortex receive responses from the simple cells for linear feature integration. Ilan *et al.* [33] proposed that spatial integration properties of complex cells can be described by a series of pooling operations. According to the classical architecture of CNNs [34], the neighborhood maximum is considered under each pixel to overcome local sensitivity. In addition, the local extremum is in accordance with the assumption that the medium transmission is locally constant, and it is commonly to overcome the noise of transmission estimation. Therefore, we use a local extremum operation in the third layer of DehazeNet.

$$F_3^i(x) = \max_{y \in \Omega(x)} F_2^i(y), \quad (11)$$

where  $\Omega(x)$  is an  $f_3 \times f_3$  neighborhood centered at  $x$ , and the output dimension of the third layer  $n_3 = n_2$ . In contrast to max-pooling in CNNs, which usually reduce resolutions of feature maps, the local extremum operation here is densely

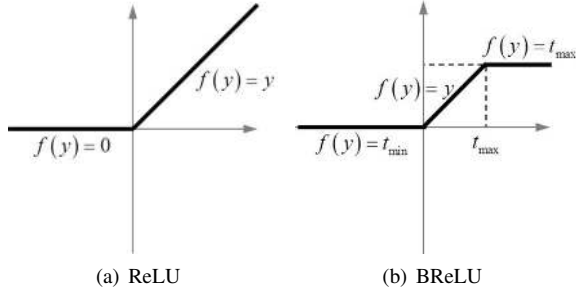


Fig. 3. Rectified Linear Unit (ReLU) and Bilateral Rectified Linear Unit (BReLU)

applied to every feature map pixel, and is able to preserve resolution for use of image restoration.

4) *Non-linear Regression*: Standard choices of nonlinear activation functions in deep networks include Sigmoid [35] and Rectified Linear Unit (ReLU). The former one is easier to suffer from vanishing gradient, which may lead to slow convergence or poor local optima in networks training. To overcome the problem of vanishing gradient, ReLU is proposed [36] which offers sparse representations. However, ReLU is designed for classification problems and not perfectly suitable for the regression problems such as image restoration. In particular, ReLU inhibits values only when they are less than zero. It might lead to response overflow especially in the last layer, because for image restoration, the output values of the last layer are supposed to be both lower and upper bounded in a small range. To this end, we propose a Bilateral Rectified Linear Unit (BReLU) activation function, shown in Fig. 3, to overcome this limitation. Inspired by Sigmoid and ReLU, BReLU as a novel linear unit keeps bilateral restraint and local linearity. Based on the proposed BReLU, the feature map of the fourth layer is defined as

$$F_4 = \min(t_{\max}, \max(t_{\min}, W_4 * F_3 + B_4)) \quad (12)$$

Here  $\mathcal{W}_4 = \{W_4\}$  contains a filter with the size of  $n_3 \times f_4 \times f_4$ ,  $\mathcal{B}_4 = \{B_4\}$  contains a bias, and  $t_{\min, \max}$  is the marginal value of BReLU ( $t_{\min} = 0$  and  $t_{\max} = 1$  in this paper). According to (12), the gradient of this activation function can be shown as

$$\frac{\partial F_4(x)}{\partial F_3} = \begin{cases} \frac{\partial F_4(x)}{\partial F_3}, & t_{\min} \leq F_4(x) < t_{\max} \\ 0, & \text{otherwise} \end{cases} \quad (13)$$

The above four layers are cascaded together to form a CNN based trainable end-to-end system, where filters and biases associated with convolutional layers are network parameters to be learned. We note that designs of these layers can be connected with expertise in existing image dehazing methods, which we specify in the subsequent section.

### B. Connections with Traditional Dehazing Methods

The first layer feature  $F_1$  in DehazeNet is designed for haze-relevant feature extraction. Take dark channel feature [9] as an

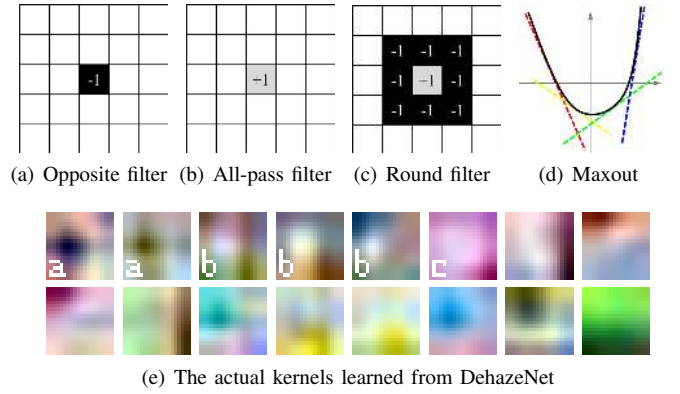


Fig. 4. Filter weight and Maxout unit in the first layer operation  $F_1$

example. If the weight  $W_1$  is an opposite filter (sparse matrices with the value of -1 at the center of one channel, as in Fig. 4(a)) and  $B_1$  is a unit bias, then the maximum output of the feature map is equivalent to the minimum of color channels, which is similar to dark channel [9] (see Equation (5)). In the same way, when the weight is a round filter as Fig. 4(c),  $F_1$  is similar to the maximum contrast [7] (see Equation (6)); when  $\mathcal{W}_1$  includes all-pass filters and opposite filters,  $F_1$  is similar to the maximum and minimum feature maps, which are atomic operations of the color space transformation from RGB to HSV, then the color attenuation [18] (see Equation (7)) and hue disparity [29] (see Equation (8)) features are extracted. In conclusion, upon success of filter learning shown in Fig. 4(e), almost all haze-relevant features can be potentially extracted from the first layer of DehazeNet. On the other hand, Maxout activation functions can be considered as piece-wise linear approximations to arbitrary convex functions. In this paper, we choose the maximum across four feature maps ( $k = 4$ ) to approximate an arbitrary convex function, as shown in Fig. 4(d).

White-colored objects in an image are similar to heavy haze scenes that are usually with high values of brightness and low values of saturation. Therefore, almost all the haze estimation models tend to consider the white-colored scene objects as being distant, resulting in inaccurate estimation of the medium transmission. Based on the assumption that the scene depth is locally constant, local extremum filter is commonly to overcome this problem [9], [18], [7]. In DehazeNet, local maximum filters of the third layer operation remove the local estimation error. Thus the direct attenuation term  $J(x)t(x)$  can be very close to zero when the transmission  $t(x)$  is close to zero. The directly recovered scene radiance  $J(x)$  is prone to noise. In DehazeNet, we propose BReLU to restrict the values of transmission between  $t_{\min}$  and  $t_{\max}$ , thus alleviating the noise problem. Note that BReLU is equivalent to the boundary constraints used in traditional methods [9], [18].

### C. Training of DehazeNet

1) *Training Data*: It is in general costly to collect a vast amount of labelled data for training deep models [19]. For



Fig. 5. Example haze-free training images collected from the Internet

training of DehazeNet, it is even more difficult as the pairs of hazy and haze-free images of natural scenes (or the pairs of hazy images and their associated medium transmission maps) are not massively available. Instead, we resort to synthesized training data based on the physical haze formation model [17].

More specifically, we synthesize training pairs of hazy and haze-free image patches based on two assumptions [17]: first, image content is independent of medium transmission (the same image content can appear at any depths of scenes); second, medium transmission is locally constant (image pixels in a small patch tend to have similar depths). These assumptions suggest that we can assume an arbitrary transmission for an individual image patch. Given a haze-free patch  $J^P(x)$ , the atmospheric light  $\alpha$ , and a random transmission  $t \in (0, 1)$ , a hazy patch is synthesized as  $I^P(x) = J^P(x)t + \alpha(1 - t)$ . To reduce the uncertainty in variable learning, atmospheric light  $\alpha$  is set to 1.

In this work, we collect haze-free images from the Internet, and randomly sample from them patches of size  $16 \times 16$ . Different from [17], these haze-free images include not only those capturing people’s daily life, but also those of natural and city landscapes, since we believe that this variety of training samples can be learned into the filters of DehazeNet. Fig. 5 shows examples of our collected haze-free images.

2) *Training Method*: In the DehazeNet, supervised learning requires the mapping relationship  $\mathcal{F}$  between RGB value and medium transmission. Network parameters  $\Theta = \{\mathcal{W}_1, \mathcal{W}_2, \mathcal{W}_4, \mathcal{B}_1, \mathcal{B}_2, \mathcal{B}_4\}$  are achieved through minimizing the loss function between the training patch  $I^P(x)$  and the corresponding ground truth medium transmission  $t$ . Given a set of hazy image patches and their corresponding medium transmissions, where hazy patches are synthesized from haze-free patches as described above, we use Mean Squared Error (MSE) as the loss function:

$$L(\Theta) = \frac{1}{N} \sum_{i=1}^N \|\mathcal{F}(I_i^P; \Theta) - t_i\|^2 \quad (14)$$

Stochastic gradient descent (SGD) is used to train DehazeNet. We implement our model using the *Caffe* package [37]. Detailed configurations and parameter settings of our proposed DehazeNet (as shown in Fig. 2) are summarized in Table I, which includes 3 convolutional layers and 1 max-pooling layer, with Maxout and BReLU activations used respectively after the first and last convolutional operations.

TABLE I. THE ARCHITECTURES OF THE DEHAZENET MODEL

Formulation	Type	Input Size	Num $n$	Filter $f \times f$	Pad
Feature Extraction	Conv	$3 \times 16 \times 16$	16	$5 \times 5$	0
	Maxout	$16 \times 12 \times 12$	4	–	0
Multi-scale Mapping	Conv	$4 \times 12 \times 12$	16	$3 \times 3$	1
			16	$5 \times 5$	2
			16	$7 \times 7$	3
Local Extremum	Maxpool	$48 \times 12 \times 12$	–	$7 \times 7$	0
Non-linear Regression	Conv	$48 \times 6 \times 6$	1	$6 \times 6$	0
	BReLU	$1 \times 1$	1	–	0

#### IV. EXPERIMENTS

To verify the architecture of DehazeNet, we analyze its convergence and compare it with the state-of-art methods, including FVR [38], DCP [9], BCCR [11], ATM [39], RF [17], BPNN [40], RF [17] and CAP [18].

Regarding the training data, 10,000 haze-free patches are sampled randomly from the images collected from the Internet. For each patch, we uniformly sample 10 random transmissions  $t \in (0, 1)$  to generate 10 hazy patches. Therefore, a total of 100,000 synthetic patches are generated for DehazeNet training. In DehazeNet, the filter weights of each layer are initialized by drawing randomly from a Gaussian distribution (with mean value  $\mu = 0$  and standard deviation  $\sigma = 0.001$ ), and the biases are set to 0. The learning rate decreases by half from 0.005 to  $3.125e-4$  every 100,000 iterations. Based on the parameters above, DehazeNet is trained (in 500,000 iterations with a batch-size of 128) on a PC with Nvidia GeForce GTX 780 GPU.

Based on the transmission estimated by DehazeNet and the atmospheric scattering model, haze-free images are restored as traditional methods. Because of the local extremum in the third layer, the blocking artifacts appear in the transmission map obtained from DehazeNet. To refine the transmission map, guided image filtering [15] is used to smooth the image. Referring to Equation (4), the boundary value of 0.1 percent intensity is chosen as  $t_0$  in the transmission map, and we select the highest intensity pixel in the corresponding hazy image  $I(x)$  among  $x \in \{y | t(y) \leq t_0\}$  as the atmospheric light  $\alpha$ . Given the medium transmission  $t(x)$  and the atmospheric light  $\alpha$ , the haze-free image  $J(x)$  is recovered easily. For convenience, Equation (1) is rewritten as follows:

$$J(x) = \frac{I(x) - \alpha(1 - t(x))}{t(x)} \quad (15)$$

Although DehazeNet is based on CNNs, the lightened network can effectively guarantee the realtime performance, and runs without GPUs. The entire dehazing framework is tested in MATLAB 2014A only with a CPU (Intel i7 3770, 3.4GHz), and it processes a  $640 \times 480$  image with approximately 1.5 seconds.

##### A. Model and performance

In DehazeNet, there are two important layers with special design for transmission estimation, feature extraction  $F_1$  and non-linear regression  $F_4$ . To proof the effectiveness of DehazeNet, two traditional CNNs (SRCNN [41] and CNN-L [23])



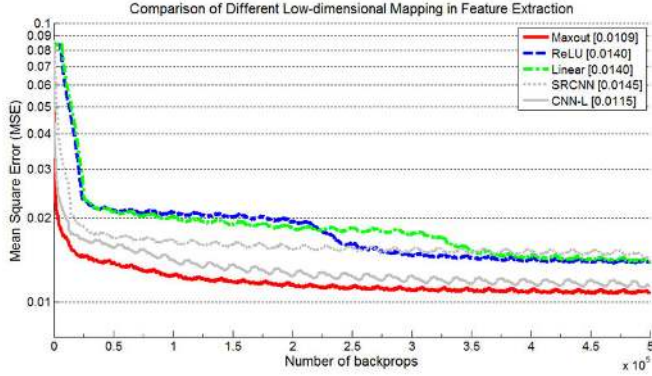


Fig. 6. The training process with different low-dimensional mapping in  $F_1$

with the same number of 3 layers are regarded as baseline models. The number of parameters of DehazeNet, SRCNN, and CNN-L is 8,240, 18,400, and 67,552 respectively.

1) *Maxout unit in feature extraction  $F_1$* : The activation unit in  $F_1$  is a non-linear dimension reduction to approximate traditional haze-relevant features extraction. In the field of image processing, low-dimensional mapping is a core procedure for discovering principal attributes and for reducing pattern noise. For example, PCA [42] and LDA [43] as classical linear low-dimensional mappings are widely used in computer vision and data mining. In [23], a non-linear sparse low-dimensional mapping with ReLU is used for high-resolution reconstruction. As an unusual low-dimensional mapping, Maxout unit maximizes feature maps to discover the prior knowledge of the hazy images. Therefore, the following experiment is designed to confirm the validity of Maxout unit. According to [23], linear unit maps a 16-dimensional vector into a 4-dimensional vector, which is equivalent to applying 4 filters with a size of  $16 \times 1 \times 1$ . In addition, the sparse low-dimensional mapping is connecting ReLU to the linear unit.

Fig. 6 presents the training process of DehazeNet with Maxout unit, compared with ReLU and linear unit. We observe in Fig. 6 that the speed of convergence for Maxout network is faster than that for ReLU and linear unit. In addition, the values in the bracket present the convergent result, and the performance of Maxout is improved by approximately  $0.30e-2$  compared with ReLU and linear unit. The reason is that Maxout unit provides the equivalent function of almost all of haze-relevant features, and alleviates the disadvantage of the simple piecewise functions such as ReLU.

2) *BReLU in non-linear regression  $F_4$* : BReLU is a novel activation function that is useful for image restoration and reconstruction. Inspired by ReLU and Sigmoid, BReLU is designed with bilateral restraint and local linearity. The bilateral restraint applies a priori constraint to reduce the solution space scale; the local linearity overcomes the gradient vanishing to gain better precision. In the contrast experiment, ReLU and Sigmoid are used to take the place of BReLU in the non-linear regression layer. For ReLU,  $F_4$  can be rewritten as  $F_4 = \max(0, W_4 * F_3 + B_4)$ , and for Sigmoid, it can be rewritten as  $F_4 = 1/(1 + \exp(-W_4 * F_3 - B_4))$ .

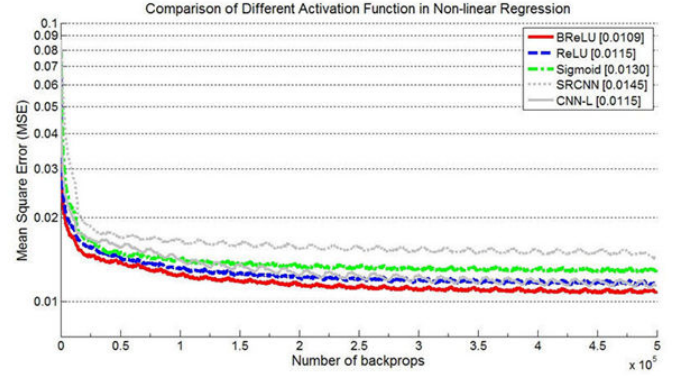


Fig. 7. The training process with different activation function in  $F_4$

TABLE II. THE RESULTS OF USING DIFFERENT FILTER NUMBER OR SIZE IN DEHAZENET ( $\times 10^{-2}$ )

Filter	Architecture	Train MSE	Test MSE	#Param
Number ( $n_1-n_2$ )	<b>4-(16<math>\times</math>3)</b>	<b>1.090</b>	<b>1.190</b>	<b>8,240</b>
	8-(32 $\times$ 3)	0.972	1.138	27,104
	16-(64 $\times$ 3)	0.902	1.112	96,704
$F_2$ Size ( $f_1-f_2-f_3-f_4$ )	5-3-7-6	1.184	1.219	4,656
	5-5-7-6	1.133	1.225	7,728
	5-7-7-6	1.021	1.184	12,336
	<b>5-M-7-6</b>	<b>1.090</b>	<b>1.190</b>	<b>8,240</b>
$F_4$ Size ( $f_1-f_2-f_3-f_4$ )	5-M-6-7	1.077	1.192	8,864
	<b>5-M-7-6</b>	<b>1.090</b>	<b>1.190</b>	<b>8,240</b>
	5-M-8-5	1.103	1.201	7,712

Fig. 7 shows the training process using different activation functions in  $F_4$ . BReLU has a better convergence rate than ReLU and Sigmoid, especially during the first 50,000 iterations. The convergent precisions show that the performance of BReLU is improved approximately  $0.05e-2$  compared with ReLU and by  $0.20e-2$  compared with Sigmoid. Fig. 8 plots the predicted transmission versus the ground truth transmission on the test patches. Clearly, the predicted transmission centers around the 45 degree line in BReLU result. However, the predicted transmission of ReLU is always higher than the true transmission, and there are some predicted transmissions over the limit value  $t_{\max} = 1$ . Due to the curvature of Sigmoid function, the predicted transmission is far away from the true transmission, closing to 0 and 1. The MSE on the test set of BReLU is  $1.19e-2$ , and that of ReLU and Sigmoid are  $1.28e-2$  and  $1.46e-2$ , respectively.

### B. Filter number and size

To investigate the best trade-off between performance and parameter size, we progressively modify the parameters of DehazeNet. Based on the default settings of DehazeNet, two experiments are conducted: (1) one is with a larger filter number, and (2) the other is with a different filter size. Similar to Sec. III-C2, these models are trained on the same dataset and Table II shows the training/testing MSEs with the corresponding parameter settings.

In general, the performance would improve when increasing the network width. It is clear that superior performance could be achieved by increasing the number of filter. However, if a

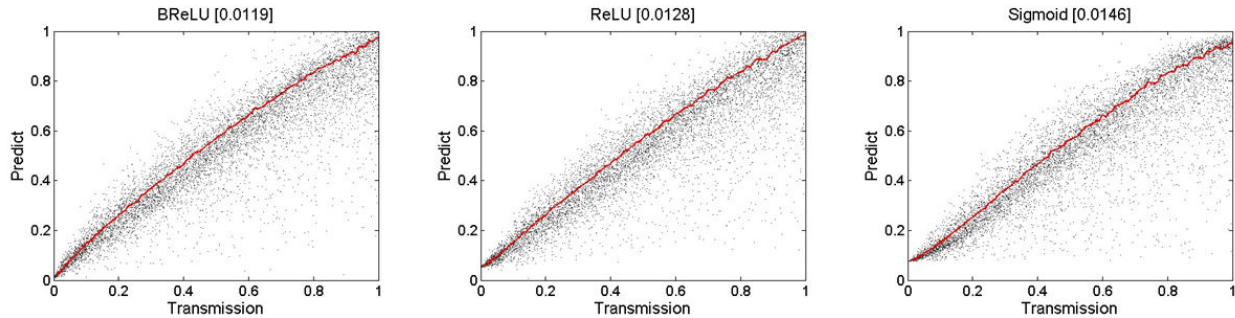


Fig. 8. The plots between predicted and truth transmission on different activation function in the non-linear regression  $F_4$

fast dehazing speed is desired, a small network is preferred, which could still achieve better performance than other popular methods. In this paper, the lightened network is adopted in the following experiments.

In addition, we examine the network sensitivity to different filter sizes. The default network setting, whose specifics are shown in Table I, is denoted as 5-M-7-6. We first analyze the effects of varying filter sizes in the second layer  $F_2$ . Table II indicates that a reasonably larger filter size in  $F_2$  could grasp richer structural information, which in turn leads to better results. The multi-scale feature mapping with the filter sizes of 3/5/7 is also adopted in  $F_2$  of DehazeNet, which achieves similar testing MSE to that of the single-scale case of  $7 \times 7$  filter. Moreover, we demonstrate in Section IV-D that the multi-scale mapping is able to improve the scale robustness.

We further examine networks with different filter sizes in the third and fourth layer. Keeping the same receptive field of network, the filter sizes in  $F_3$  and  $F_4$  are adjusted simultaneously. It is showed that the larger filter size of non-linear regression  $F_4$  enhances the fitting effect, but may lead to over-fitting. The local extremum in  $F_3$  could improve the robustness on testing dataset. Therefore, we find the best filter setting of  $F_3$  and  $F_4$  as 5-M-7-6 in DehazeNet.

### C. Quantitative results on synthetic patches

In recent years, there are three methods based on learning framework for haze removal. In [18], dehazing parameters are learned by a linear model, to estimate the scene depth under Color Attenuation Prior (CAP). A back propagation neural network (BPNN) [40] is used to mine the internal link between color and depth from the training samples. In [17], Random Forests (RF) are used to investigate haze-relevant features for haze-free image restoration. All of the above methods and DehazeNet are trained with the same method as RF. According to the testing measure of RF, 2000 image patches are randomly sampled from haze-free images with 10 random transmission  $t \in (0, 1)$  to generate 20,000 hazy patches for testing. We run DehazeNet and CAP on the same testing dataset to measure the mean squared error (MSE) between the predicted transmission and true transmission. DCP [9] is a classical dehazing method, which is used as a comparison baselines.

Table III shows the MSE between predicted transmissions and truth transmissions on the testing patches. DehazeNet

TABLE III. MSE BETWEEN PREDICTED TRANSMISSION AND GROUND TRUTH ON SYNTHETIC PATCHES

Methods	DCP [9]	BPNN [40]	CAP [18]	RF [17]	DehazeNet
MSE ( $\times 10^{-2}$ )	3.18	4.37	3.32	1.26	<b>1.19</b>

achieves the best state-of-the-art score, which is 1.19e-2; the MSE between our method and the next state-of-art result (RF [17]) in the literature is 0.07e-2. Because in [17], the feature values of patches are sorted to break the correlation between the haze-relevant features and the image content. However, the content information concerned by DehazeNet is useful for the transmission estimation of the sky region and white objects. Moreover, CAP [18] achieves satisfactory results in follow-on experiments but poor performance in this experiment, due to some outlier values (greater than 1 or less than 0) in the linear regression model.

### D. Quantitative results on synthetic images

To verify the effectiveness on complete images, DehazeNet is tested on synthesized hazy images from stereo images with a known depth map  $d(x)$ , and it is compared with DCP [9], FVR [38], BCCR [11], ATM [39], CAP<sup>2</sup>[18], and RF [17]. There are 12 pairs of stereo images collected in Middlebury Stereo Datasets (2001-2006) [44], [45], [46]. In Fig. 9, the hazy images are synthesized from the haze-free stereo images based on (1), and they are restored to haze-free images by DehazeNet.

To quantitatively assess these methods, we use a series of evaluation criteria in terms of the difference between each pair of haze-free image and dehazing result. Apart from the widely used mean square error (MSE) and the structural similarity (SSIM) [47] indices, we used additional evaluation matrices, namely peak signal-to-noise ratio (PSNR) and weighted peak signal-to-noise ratio (WPSNR) [48]. We define one-pass evaluation (OPE) as the conventional method, which we run with standard parameters and report the average measure for the performance assessment. In Table IV, DehazeNet is compared with six state-of-the-art methods on all of the hazy images by OPE (hazy images are synthesized with the single scattering

<sup>2</sup>The results outside the parenthesis are run with the code implemented by authors [18], and the results in the parenthesis are re-implemented by us.

TABLE IV. THE AVERAGE RESULTS OF MSE, SSIM, PSNR AND WSNR ON THE SYNTHETIC IMAGES ( $\beta = 1$  AND  $\alpha = 1$ )

Metric	ATM [39]	BCCR [11]	FVR [38]	DCP [9]	CAP <sup>2</sup> [18]	RF [17]	DehazeNet
MSE	0.0689	0.0243	0.0155	0.0172	0.0075 (0.0068)	<u>0.0070</u>	<b>0.0062</b>
SSIM	0.9890	0.9963	0.9973	0.9981	<u>0.9991</u> (0.9990)	0.9989	<b>0.9993</b>
PSNR	60.8612	65.2794	66.5450	66.7392	70.0029 (70.6581)	<u>70.0099</u>	<b>70.9767</b>
WSNR	7.8492	12.6230	13.7236	13.8508	16.9873 (17.7839)	<u>17.1180</u>	<b>18.0996</b>



Fig. 9. Synthetic images based on Middlebury Stereo Datasets and DehazeNet results

coefficient  $\beta = 1$  and the pure-white atmospheric airlight  $\alpha = 1$ ). It is exciting that, although DehazeNet is optimized by the MSE loss function, it also achieves the best performance on the other types of evaluation matrices.

The dehazing effectiveness is sensitive to the haze density, and the performance with a different scattering coefficient  $\beta$  could become much worse or better. Therefore, we propose an evaluation to analyze dehazing robustness to scattering coefficient  $\beta \in \{0.75, 1.0, 1.25, 1.5\}$ , which is called as coefficient robustness evaluation (CRE). As shown in Table V, CAP [18] achieve better performances on the mist ( $\beta = 0.75$ ), but the dehazing performance reduces gradually when the amount of haze increases. The reason is that CAP estimates the medium transmission based on predicted scene depth and a assumed scattering coefficient ( $\beta = 1$ ). In [17], 200 trees are used to build random forests for non-linear regression and shows greater coefficient robustness. However, the high-computation of random forests in every pixel constraints to its practicality. For DehazeNet, the medium transmission is estimated directly by a non-linear activation function (Maxout) in  $F_1$ , resulting



Fig. 10. Image enhancement for anti-halation by DehazeNet

in excellent robustness to the scattering coefficient.

Due to the color offset of haze particles and light sources, the atmosphere airlight is not a proper pure-white. An airlight robustness evaluation (ARE) is proposed to analyze the dehazing methods for different atmosphere airlight  $\alpha$ . Although DehazeNet is trained from the samples generated by setting  $\alpha = 1$ , it also achieves the greater robustness on the other values of atmosphere airlight. In particular, DehazeNet performs better than the other methods when sunlight haze is [1.0, 1.0, 0.9]. Therefore, DehazeNet could also be applied to remove halation, which is a bright ring surrounding a source of light as shown in Fig. 10.

The view field transformation and image zoom occur often in real-world applications. The scale robustness evaluation (SRE) is used to analyze the influence from the scale variation. Compared with the same state-of-the-art methods in OPE, there are 4 scale coefficients  $s$  selected from 0.4 to 1.0 to generate different scale images for SER. In Table V, DehazeNet shows excellent robustness to the scale variation due to the multi-scale mapping in  $F_2$ . The single scale used in CAP [18], DCP [9] and ATM [39] results in a different prediction accuracy on a different scale. When an image shrinks, an excessively large-scale processing neighborhood will lose the image's details. Therefore, the multi-scale mapping in DehazeNet provides a variety of filters to merge multi-scale features, and it achieves the best scores under all of different scales.

In most situations, noise is random produced by the sensor or camera circuitry, which will bring in estimation error. We also discuss the influences of varying degrees of image noise to our method. As a basic noise model, additive white Gaussian (AWG) noise with standard deviation  $\sigma \in \{10, 15, 20, 25\}$  is used for noise robustness evaluation (NRE). Benefiting from the Maxout suppression in  $F_1$  and the local extremum in  $F_3$ , DehazeNet performs more robustly in NRE than the others do. RF [17] has a good performance in most of the evaluations but fails in NRE, because the feature values of patches are sorted

TABLE V. THE MSE ON THE SYNTHETIC IMAGES BY DIFFERENT SCATTERING COEFFICIENT, IMAGE SCALE AND ATMOSPHERIC AIRLIGHT

Evaluation		ATM [39]	BCCR [11]	FVR [38]	DCP [9]	CAP <sup>2</sup> [18]	RF [17]	DehazeNet
CRE ( $\beta =$ )	0.75	0.0581	0.0269	0.0122	0.0199	<b>0.0043</b> (0.0042)	<u>0.0046</u>	0.0063
	1.00	0.0689	0.0243	0.0155	0.0172	0.0077 (0.0068)	<u>0.0070</u>	<b>0.0062</b>
	1.25	0.0703	0.0230	0.0219	0.0147	0.0141 (0.0121)	<u>0.0109</u>	<b>0.0084</b>
	1.50	0.0683	0.0219	0.0305	0.0134	0.0231 (0.0201)	<u>0.0152</u>	<b>0.0127</b>
CRE Average		0.0653	0.0254	0.0187	0.0177	0.0105 (0.0095)	<u>0.0094</u>	<b>0.0084</b>
ARE ( $\alpha =$ )	[1.0, 1.0, 1.0]	0.0689	0.0243	0.0155	0.0172	0.0075 (0.0068)	<u>0.0070</u>	<b>0.0062</b>
	[0.9, 1.0, 1.0]	0.0660	0.0266	0.0170	0.0210	0.0073 (0.0069)	<b>0.0071</b>	<u>0.0072</u>
	[1.0, 0.9, 1.0]	0.0870	0.0270	0.0159	0.0200	<b>0.0070</b> (0.0067)	<u>0.0073</u>	0.0074
	[1.0, 1.0, 0.9]	0.0689	0.0239	0.0152	0.0186	<u>0.0081</u> (0.0069)	0.0083	<b>0.0062</b>
ARE Average		0.0727	0.0255	0.0159	0.0192	0.0075 (0.0068)	<u>0.0074</u>	<b>0.0067</b>
SRE ( $s =$ )	0.40	0.0450	0.0238	0.0155	0.0102	0.0137 (0.0084)	<u>0.0089</u>	<b>0.0066</b>
	0.60	0.0564	0.0223	0.0154	0.0137	0.0092 (0.0071)	<u>0.0076</u>	<b>0.0060</b>
	0.80	0.0619	0.0236	0.0155	0.0166	0.0086 (0.0066)	<u>0.0074</u>	<b>0.0062</b>
	1.00	0.0689	0.0243	0.0155	0.0172	0.0077 (0.0068)	<u>0.0070</u>	<b>0.0062</b>
SRE Average		0.0581	0.0235	0.0155	0.0144	0.0098 (0.0072)	<u>0.0077</u>	<b>0.0062</b>
NRE ( $\sigma =$ )	10	0.0541	0.0138	0.0150	0.0133	<u>0.0065</u> (0.0070)	0.0086	<b>0.0059</b>
	15	0.0439	0.0144	0.0148	0.0104	<u>0.0072</u> (0.0074)	0.0112	<b>0.0061</b>
	20	-	0.0181	0.0151	0.0093	<u>0.0083</u> (0.0085)	0.0143	<b>0.0058</b>
	25	-	0.0224	0.0150	-	0.0100 (0.0092)	0.0155	<b>0.0051</b>
	30	-	0.0192	0.0151	-	<u>0.0085</u> (0.0119 (0.0112))	0.0191	<b>0.0049</b>
NRE Average		-	0.0255	0.0150	0.0100	<u>0.0088</u> (0.0087)	0.0137	<b>0.0055</b>

to break the correlation between the medium transmission and the image content, which will also magnify the effect of outlier.

### E. Qualitative results on real-world images

Fig. 11 shows the dehazing results and depth maps restored by DehazeNet, and more results and comparisons can be found at <http://caibolun.github.io/DehazeNet/>. Because all of the dehazing algorithms can obtain truly good results on general outdoor images, it is difficult to rank them visually. To compare them, this paper focuses on 5 identified challenging images in related studies [9], [17], [18]. These images have large white or gray regions that are hard to handle, because most existing dehazing algorithms are sensitive to the white color. Fig. 12 shows a qualitative comparison with six state-of-the-art dehazing algorithms on the challenging images. Fig. 12 (a) depicts the hazy images to be dehazed, and Fig. 12 (b-g) shows the results of ATM [39], BCCR [11], FVR [38], DCP [9], CAP [18] and RF [17], respectively. The results of DehazeNet are given in Fig. 12 (h).

The sky region in a hazy image is a challenge of dehazing, because clouds and haze are similar natural phenomena with the same atmospheric scattering model. As shown in the first three figures, most of the haze is removed in the (b-d) results, and the details of the scenes and objects are well restored. However, the results significantly suffer from over-enhancement in the sky region. Overall, the sky region of these images is much darker than it should be or is oversaturated and distorted. Haze generally exists only in the atmospheric surface layer, and thus the sky region almost does not require handling. Based on the learning framework, CAP and RF avoid color distortion in the sky, but non-sky regions are enhanced poorly because of the non-content regression model (for example, the rock-soil of the first image and the green flatlands in the third image). DehazeNet appears to be capable of finding the sky region to keep the color, and assures a good dehazing effect in other regions. The reason is that the patch attribute can be learned in the hidden layer of DehazeNet, and it contributes to the dehazing effects in the sky.

Because transmission estimation based on priors are a type of statistics, which might not work for certain images. The fourth and fifth figures are determined to be failure cases in [9]. When the scene objects are inherently similar to the atmospheric light (such as the fair-skinned complexion in the fourth figure and the white marble in the fifth figure), the estimated transmission based on priors (DCP, BCCR, FVR) is not reliable. Because the dark channel has bright values near such objects, and FVR and BCCR are based on DCP which has an inherent problem of overestimating the transmission. CAP and RF learned from a regression model is free from oversaturation, but underestimates the haze degree in the distance (see the brown hair in the fourth image and the red pillar in the fifth image). Compared with the six algorithms, our results avoid image oversaturation and retain the dehazing validity due to the non-linear regression of DehazeNet.

## V. CONCLUSION

In this paper, we have presented a novel deep learning approach for single image dehazing. Inspired by traditional haze-relevant features and dehazing methods, we show that medium transmission estimation can be reformulated into a trainable end-to-end system with special design, where the feature extraction layer and the non-linear regression layer are distinguished from classical CNNs. In the first layer  $F_1$ , Maxout unit is proved similar to the priori methods, and it is more effective to learn haze-relevant features. In the last layer  $F_4$ , a novel activation function called BReLU is instead of ReLU or Sigmoid to keep bilateral restraint and local linearity for image restoration. With this lightweight architecture, DehazeNet achieves dramatically high efficiency and outstanding dehazing effects than the state-of-the-art methods.

Although we successfully applied a CNN for haze removal, there are still some extensibility researches to be carried out. That is, the atmospheric light  $\alpha$  cannot be regarded as a global constant, which will be learned together with medium transmission in a unified network. Moreover, we think

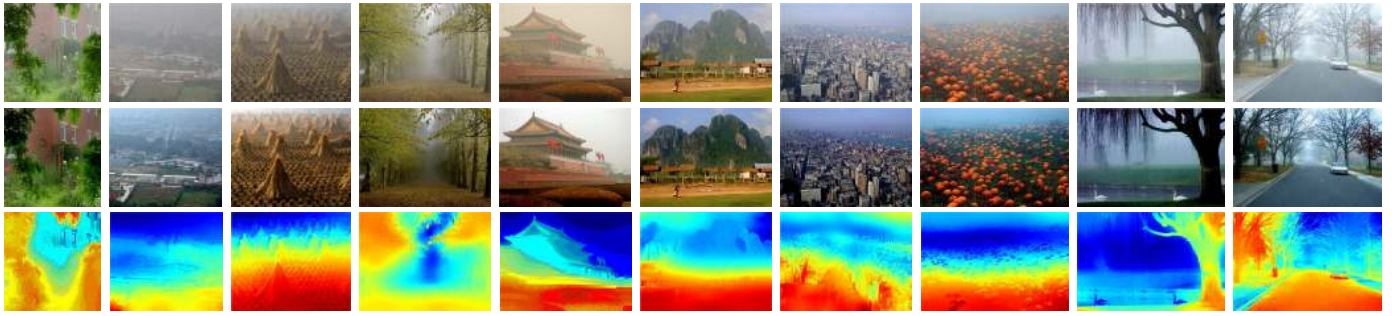


Fig. 11. The haze-free images and depth maps restored by DehazeNet

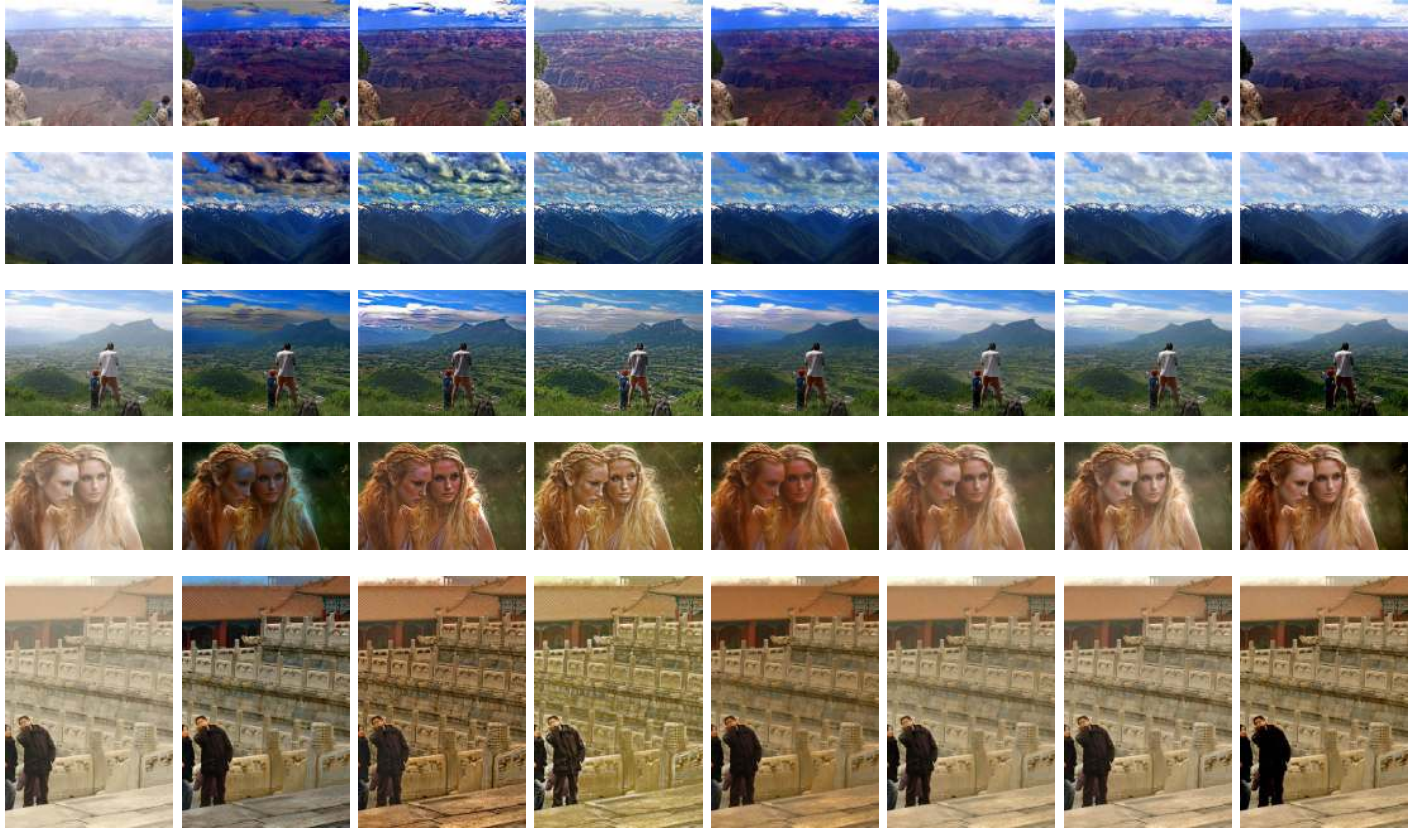


Fig. 12. Qualitative comparison of different methods on real-world images.

atmospheric scattering model can also be learned in a deeper neural network, in which an end-to-end mapping between haze and haze-free images can be optimized directly without the medium transmission estimation. We leave this problem for future research.

## REFERENCES

- [1] T. K. Kim, J. K. Paik, and B. S. Kang, "Contrast enhancement system using spatially adaptive histogram equalization with temporal filtering," *IEEE Transactions on Consumer Electronics*, vol. 44, no. 1, pp. 82–87, 1998.
- [2] J. A. Stark, "Adaptive image contrast enhancement using generalizations of histogram equalization," *IEEE Transactions on Image Processing*, vol. 9, no. 5, pp. 889–896, 2000.
- [3] R. Eschbach and B. W. Kolpatzik, "Image-dependent color saturation correction in a natural scene pictorial image," Sep. 12 1995, uS Patent 5,450,217.
- [4] Y. Y. Schechner, S. G. Narasimhan, and S. K. Nayar, "Instant dehazing of images using polarization," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 1, 2001, pp. 1–325.
- [5] S. G. Narasimhan and S. K. Nayar, "Contrast restoration of weather degraded images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 6, pp. 713–724, 2003.
- [6] J. Kopf, B. Neubert, B. Chen, M. Cohen, D. Cohen-Or, O. Deussen,

- M. Uyttendaele, and D. Lischinski, "Deep photo: Model-based photograph enhancement and viewing," in *ACM Transactions on Graphics (TOG)*, vol. 27, no. 5, 2008, p. 116.
- [7] R. T. Tan, "Visibility in bad weather from a single image," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2008, pp. 1–8.
- [8] R. Fattal, "Single image dehazing," in *ACM Transactions on Graphics (TOG)*, vol. 27, no. 3, 2008, p. 72.
- [9] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 12, pp. 2341–2353, 2011.
- [10] K. Nishino, L. Kratz, and S. Lombardi, "Bayesian defogging," *International journal of computer vision*, vol. 98, no. 3, pp. 263–278, 2012.
- [11] G. Meng, Y. Wang, J. Duan, S. Xiang, and C. Pan, "Efficient image dehazing with boundary constraint and contextual regularization," in *IEEE International Conference on Computer Vision (ICCV)*, 2013, pp. 617–624.
- [12] K. B. Gibson, D. T. Vo, and T. Q. Nguyen, "An investigation of dehazing effects on image and video coding," *IEEE Transactions on Image Processing*, vol. 21, no. 2, pp. 662–673, 2012.
- [13] J.-P. Tarel and N. Hautiere, "Fast visibility restoration from a single color or gray level image," in *IEEE International Conference on Computer Vision*, 2009, pp. 2201–2208.
- [14] J. Yu, C. Xiao, and D. Li, "Physics-based fast single image fog removal," in *IEEE International Conference on Signal Processing (ICSP)*, 2010, pp. 1048–1052.
- [15] K. He, J. Sun, and X. Tang, "Guided image filtering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 6, pp. 1397–1409, 2013.
- [16] A. Levin, D. Lischinski, and Y. Weiss, "A closed-form solution to natural image matting," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 2, pp. 228–242, 2008.
- [17] K. Tang, J. Yang, and J. Wang, "Investigating haze-relevant features in a learning framework for image dehazing," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014, pp. 2995–3002.
- [18] Q. Zhu, J. Mai, and L. Shao, "A fast single image haze removal algorithm using color attenuation prior," *IEEE Transactions on Image Processing*, vol. 24, no. 11, pp. 3522–3533, 2015.
- [19] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.
- [20] C. Ding and D. Tao, "Robust face recognition via multimodal deep face representation," *IEEE Transactions on Multimedia*, vol. 17, no. 11, pp. 2049–2058, 2015.
- [21] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014, pp. 580–587.
- [22] C. J. Schuler, H. C. Burger, S. Harmeling, and B. Scholkopf, "A machine learning approach for non-blind image deconvolution," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2013, pp. 1067–1074.
- [23] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, no. 99, pp. 1–1, 2015.
- [24] D. Eigen, D. Krishnan, and R. Fergus, "Restoring an image taken through a window covered with dirt or rain," in *IEEE International Conference on Computer Vision (ICCV)*, 2013, pp. 633–640.
- [25] Z. Wu, D. Lin, and X. Tang, "Adjustable bounded rectifiers: Towards deep binary representations," *arXiv preprint arXiv:1511.06201*, 2015.
- [26] E. J. McCartney, "Optics of the atmosphere: scattering by molecules and particles," *New York, John Wiley and Sons, Inc.*, vol. 1, 1976.
- [27] S. K. Nayar and S. G. Narasimhan, "Vision in bad weather," in *IEEE International Conference on Computer Vision*, vol. 2, 1999, pp. 820–827.
- [28] S. G. Narasimhan and S. K. Nayar, "Contrast restoration of weather degraded images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 6, pp. 713–724, 2003.
- [29] C. O. Ancuti, C. Ancuti, C. Hermans, and P. Bekaert, "A fast semi-inverse approach to detect and remove the haze from a single image," in *Computer Vision—ACCV 2010*, 2011, pp. 501–514.
- [30] I. Goodfellow, D. Warde-farley, M. Mirza, A. Courville, and Y. Bengio, "Maxout networks," in *Proceedings of the 30th International Conference on Machine Learning (ICML-13)*, 2013, pp. 1319–1327.
- [31] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 1–9.
- [32] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein *et al.*, "Imagenet large scale visual recognition challenge," *International Journal of Computer Vision*, pp. 1–42, 2014.
- [33] I. Lampl, D. Ferster, T. Poggio, and M. Riesenhuber, "Intracellular measurements of spatial integration and the max operation in complex cells of the cat primary visual cortex," *Journal of neurophysiology*, vol. 92, no. 5, pp. 2704–2713, 2004.
- [34] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
- [35] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.
- [36] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, 2010, pp. 807–814.
- [37] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, "Caffe: Convolutional architecture for fast feature embedding," in *Proceedings of the ACM International Conference on Multimedia*. ACM, 2014, pp. 675–678.
- [38] J.-P. Tarel and N. Hautiere, "Fast visibility restoration from a single color or gray level image," in *IEEE International Conference on Computer Vision*, 2009, pp. 2201–2208.
- [39] M. Sulami, I. Glatzer, R. Fattal, and M. Werman, "Automatic recovery of the atmospheric light in hazy images," in *IEEE International Conference on Computational Photography (ICCP)*, 2014, pp. 1–11.
- [40] J. Mai, Q. Zhu, D. Wu, Y. Xie, and L. Wang, "Back propagation neural network dehazing," *IEEE International Conference on Robotics and Biomimetics (ROBIO)*, pp. 1433–1438, 2014.
- [41] C. Dong, C. C. Loy, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution," in *Computer Vision—ECCV 2014*. Springer, 2014, pp. 184–199.
- [42] T.-H. Chan, K. Jia, S. Gao, J. Lu, Z. Zeng, and Y. Ma, "Pcanet: A simple deep learning baseline for image classification?" *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 5017–5032, 2015.
- [43] B. Scholkopf and K.-R. Mullert, "Fisher discriminant analysis with kernels," *Neural networks for signal processing IX*, vol. 1, p. 1, 1999.
- [44] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *International journal of computer vision*, vol. 47, no. 1-3, pp. 7–42, 2002.
- [45] —, "High-accuracy stereo depth maps using structured light," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1. IEEE, 2003, pp. 1–195.
- [46] D. Scharstein and C. Pal, "Learning conditional random fields for stereo," in *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2007, pp. 1–8.
- [47] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image

quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

- [48] J. L. Mannos and D. J. Sakrison, "The effects of a visual fidelity criterion of the encoding of images," *IEEE Transactions on Information Theory*, vol. 20, no. 4, pp. 525–536, 1974.