LEARNING DEEP TRANSMISSION NETWORK FOR SINGLE IMAGE DEHAZING

Zhigang Ling [1] *, Guoliang Fan* [2]*, Yaonan Wang* [1] *and Xiao Lu* [1]

College of Electrical and Information Engineering, Hunan University, China [1]
School of Electrical and Computer Engineering, Oklahoma State University, USA [2]

ABSTRACT

State-of-the-art single image dehazing algorithms have some challenges to deal with images captured under complex weather conditions because their assumptions usually do not hold in those situations. In this paper, we develop a deep transmission network for robust single image dehazing. This deep transmission network simultaneously copes with three color channels and local patch information to automatically explore and exploit haze-relevant features in a learning framework. We further explore different network structures and parameter settings to achieve tradeoffs between performance and speed, which shows that color channels information is the most useful haze-relevant feature rather than local information. Experiment results demonstrate that the proposed algorithm outperforms state-of-the-art methods on both synthetic and real-world datasets.

Index Terms— Image dehazing, convolutional neural networks, deep transmission network

1. INTRODUCTION

In hazy or foggy weather, dust, water droplets or other particles in the atmosphere greatly scatter and absorb the radiance from objects in the scene and blend with the airlight, thus only a small percentage of the reflected light reaches observers, which often yields low contrast, obscures the clarity of the sky, fades the objects colors, and severely degrades the visibility of the captured scene. Therefore, image dehazing or removing atmospheric effects from outdoor hazy images has captured researchers' interest for many applications [1][2][3], such as, visual surveillance [4], driving assistance [5], etc.

Single image dehazing algorithms are recently attracting more and more attention from many researchers [6] because the haze is highly related to the scene depth, but additional depth information [7] or multiple images [8] for the depth recovery are not available in many situations. Single image dehazing methods usually involve prior assumptions or constraints. For example, Fattal [9] assumes that the transmission and surface shading are locally uncorrelated, and estimates haze by independent component analysis. Tan [10] removes haze by maximizing the local contrast of the restored image since that haze-free images have higher contrast than hazy ones. The results of these methods are visually compelling. He et al. [6] proposed an interesting dark channel prior (DCP) for natural outdoor images, which assumes that at least one color channel of each pixel in a haze-free image should have a small intensity value. Based on this prior, the transmission can be easily estimated and then refined for haze removal. Due to its simplicity and high efficiency, DCP has been improved and used for different applications [11][12].

Despite the fact that the remarkable progress on single image haze removal has been made, the results generated by state-of-the-art methods may not be satisfactory, because some prior assumptions made in previous methods may not hold in many situations. For example, the assumption proposed by Fattal [9] could be too strong for a variety of images. The DCP may not be appropriate when an input image contains highly saturated structures like white walls, sky or heavy haze, etc. All of these factors may lead to over-estimate the thickness of haze, cause degradation or color distortions into resulted images, especially in sky regions. Tang et al. [13] investigated different features by extracting multi-scale dark channel, local max contrast, saturation and so on, in a learning framework to identify the best feature combination for image dehazing. However, these haze-related features are hand-crafted and still very hard to excavate the most informative haze-related features and information. It would be desirable to automatically learn useful haze-related features.

In this paper, we develop a novel deep transmission network for image dehazing via a convolutional neural network (CNN). To the best of our knowledge, it is the first attempt of using deep learning for image dehazing problems. The deep transmission network can automatically discovers high order relationships between haze and image features for accurate transmission estimation. From the output of deep transmission network, we confirm that the information from three color channels information is the most useful feature for haze estimation. Experimental results show that our approach achieves better results than state-of-the-art methods.

2. BACKGROUND

In computer vision and computer graphics, an atmospheric scattering model proposed by Koschmieder [14] has been

This work was supported by National Natural Science Foundation of China (No.61471166 and 61271382).

used in image dehazing work. According to this model, the radiance which reaches the observer or camera is composed of two main additive components: direct attenuation and veiling light. Narasimhan et al. [15] further developed this model and presented that the brightness at any pixel recorded by a monochrome camera is given by

$$I(x) = I_\infty \rho(x) e^{-\beta(\lambda)d(x)} + I_\infty (1 - e^{-\beta(\lambda)d(x)}), \quad (1)$$

where I_∞ denotes sky or airlight intensity. $\rho(x)$ is the normalized radiance of a scene point and it only depends on the scene point reflectance spectrum and spectral response of the camera, rather than weather condition. The scene transmission $t(x) = e^{-\beta(\lambda)d(x)}$ relates to the scene depth from the observer and wavelength λ. The direct attenuation $D(x) = I_\infty \rho(x) e^{-\beta(\lambda)d(x)}$ represents how the scene radiance is attenuated due to medium properties. The veiling light $V = I_\infty (1 - e^{-\beta(\lambda)d(x)})$ is the main cause of the color shifting. Generally, scattering coefficient $\beta(\lambda)$ depends on the wavelength and their relationship may be described as

$$\beta(\lambda) \propto \frac{1}{\lambda^\gamma}, \quad (2)$$

where $0 \leq \gamma \leq 4$ depends on the size distribution of scattering particles. For pure air, the particle (molecules) sizes are very small compared to the wavelength of light and hence, there is a strong wavelength dependence of scattering and $\gamma = 4$; short (blue) wavelengths dominate and we see the clear blue sky. For fog, the particle (water droplets) sizes are large compared to the wavelength of light, thus the scattering coefficient does not depend on wavelength. All wavelengths are scattered equally while $\gamma = 0$ and we see white fog. A wide gamut of atmospheric conditions arise from aerosols whose particle sizes range between minute air molecules ($10^{-4}\mu m$) and large fog droplets ($1 - 10\mu m$). Such aerosols (e.g., mild haze) show significant wavelength selectivity.

The goal of the haze removal is to recover $\rho(x)$ for each pixel in hazy images. Practically, while no additional information about depth and airlight is given, haze removal is an ill-posed problem. Recently, many strong assumptions or constraints are developed to estimate the transmission $t(x)$ for haze removal.

3. DEEP NETWORKS FOR IMAGE DEHAZING

3.1. Learning the deep transmission model

Instead of extracting hand-crafted features, we attempt to directly learn a deep transmission model from a large data set via a convolutional neural network [16][17] to automatically capture haze-relevant features for image dehazing, which may overcome the complex design and inefficiency of hand-crafted features. Given a hazy image I, we first compute airlight I_∞ , and normalize this hazy image as $I_n = 1 - I/I_\infty$. Thus, Eq.(1) may be rewritten as follows:

$$I_n(x) = 1 - I(x)/I_\infty = (1 - \rho(x))t(x) = P(x)t(x). \quad (3)$$

Obviously, the normalized hazy image $I_n(x)$ only depends on transmission value $t(x)$ while the scene and the spectral response of the camera are given. Our goal is to recover $t(x)$ from the normalized hazy image I_n. In other word, we wish to learn a projection space $P(x)$ and further obtain the representation of the normalized image I_n on space $P(x)$, which consists of three operations, as shown in Fig. 1.

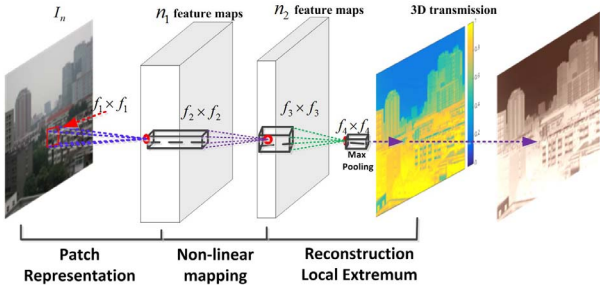


Fig. 1. The deep transmission network structure map.

3.1.1. *Patch representation*

Patch extraction and representation has been widely used to extract features in pattern recognition and image restoration. Different from extracting multiple haze-relevant features in Ref. [13] to represent local patch, we densely extract local patches and then represent them by a set of filters, which not only sufficiently extracts the structure information between pixels in the local patch but also excavates the spectrum relevance between RGB channels of the objects reflectance by optimizing the network. Formally, our first layer is expressed as an operation F_1:

$$F_1(I_n) = \min(\max(0, W_1 * I_n + B_1), 1), \quad (4)$$

where W_1 and B_1 represent the filters and biases, respectively. Here W_1 is of a size $c \times f_1 \times f_1 \times n_1$, where c is the number of channels in the input image, f_1 is the spatial size of a filter, and n_1 is the number of filters. Intuitively, W_1 applies convolutions on the image, and each convolution has a kernel size $c \times f_1 \times f_1$. The first layer output is composed of n_1 feature maps. B_1 is a n_1-dimensional vector, whose each element is associated with a filter. We apply an improved Rectified Linear Unit $\min(\max(0, x)), 1)$ on the filter responses because the transmission value is in the range of [0, 1].

3.1.2. *Non-linear mapping*

The first layer extracts a n_1-dimensional feature for each patch. However, the output features may contain some redundant information. Moreover according to Eq.1, the $\rho(x)$ only

depends on scene point reflectance rather than local spatial information around scene point. Therefore, we adopt a trivial spatial support $f_2 \times f_2$ to map each of these n_1 dimensional vectors into an n_2-dimensional features, respectively. The operation of the second layer is:

$$F_2(I_n) = min(max(0, W_2 * F_1(I_n) + B_2), 1), \quad (5)$$

where W_2 is of a size $n_1 \times f_2 \times f_2 \times n_2$ and B_2 is a n_2-dimensional vector.

3.1.3. Reconstruction and Local Extremum

According to the classical architecture of CNN, the pooling operation may be used to achieve spatial invariance. Moreover, the medium transmission may be assured to be locally constant, which may commonly overcome the noise of transmission estimation. Thus, we use a local extremum operation in the third layer of CNN network as follows:

$$F_4^c(I_n(x)) = \max_{y \in \Omega(x)} F_3^c(I_n(y)),$$
$$F_3(I_n) = min(max(0, W_3 * F_2(I_n) + B_3), 1), \quad (6)$$

Here W_3 is of a size $n_2 \times f_3 \times f_3 \times c$, and B_3 is a c-dimensional vector, the $\Omega(x)$ is a $f_4 \times f_4$ local window centered at x and the c is the channel number.

3.2. Preparing training data

Due to the haze depends on the scene depth which is hardly obtained from single image, we use some real haze-free images to build our training dataset for the transmission estimation. Here, given a haze-free image, first, we randomly sample many small patches with the size of $r \times r$, and assume image pixels in each patch tend to have similar depth values. Then, we take the following preparation procedure to build a hazy image patches: (1) randomly produce a visibility distance V in the ranges [0.1, 50]km and construct extinction coefficient $\beta(\lambda)$ via the following function:

$$\beta(\lambda) = \frac{3.91}{V}(\frac{\lambda}{0.55})^{\gamma} \quad (7)$$

$$\gamma = \begin{cases} 0.585V^(1/3) & V < 6km \\ 1.3 & 6km \leq V \leq 50km \\ 1.6 & V > 50km, \end{cases} \quad (8)$$

where $\beta(\lambda)$ is in 1/km, the wavelength λ is in μm. (2) Randomly produce a local depth value d, then we can build the transmission $t^c(x)$ where the wavelengths λ^c of the red, green and blue light are set as 700μm, 500μm, 435μm, respectively. (3) We set the atmospheric light I_∞=[255,255,255] to reduce the uncertainty of variables. Thus, we synthesize a hazy image while a uniform scattering effect is assumed. By this method, we obtain 500000 pairs image patches as the training dataset, in which haze-free natural images are collected from the Internet. Lastly, the deep transmission network may be trained through minimizing the loss function between the output of the deep transmission network and the corresponding ground truth transmission, in which Mean Squared Error is adopted as the loss function.

3.3. Airlight estimation and image dehazing

We first set I_∞ as [255,255,255] to normalize a hazy image I as I_n, and put I_n into the learned deep transmission network to obtain the transmission map. Then, we pick up the bottom 0.01% darkest pixels in the transmission map, and compute the average intensity of these picked pixels as the estimate of I_∞. Once the I_∞ is given, this hazy image is re-normalized and put into the deep transmission network for transmission estimation. In order to reduce computation burden and color distortion, a minimum operation is executed to estimate the transmission map, i.e., $t(x) = \min_{c=r,g,b} t^c(x)$, and then a guided filter [18] is used to refine the transmission.

Furthermore, since the scene radiance is usually not as bright as the atmospheric light, the dehazed image looks dim. Therefore, we introduce an adaptive exposure scaling method proposed by Tang et al. [13] to improve the exposure of dehazed images for better visual quality.

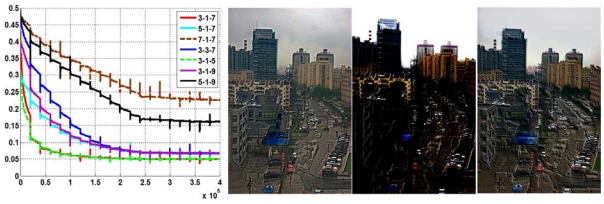
4. EXPERIMENTAL RESULTS

In order to comprehensively demonstrate the effectiveness of this proposed algorithm, this paper first analysis the models parameters and dehazing performance of deep transmission network. Then, this study compares this proposed algorithm with several state-of-the-art methods on the synthetic hazy images and a group of real hazy images. The parameters of deep network f_1, f_2, f_3, f_4, n_1, n_2 and c are set as 3, 1, 5, 7, 32, 64 and 3, respectively. The r is set to 13.

4.1. Analyze on models parameters and performance

Based on the basic network settings, we will progressively modify some of these parameters to study the relations between dehazing performance and model's parameters, specially for filter size.

Fig.2(a) shows the changes of the convergence errors and dehazed results with different filter number, where 5-1-7 shows the filter size of the first, second and third layers is 5, 1, 7, respectively. The larger filer size of the first layer may increase convergence error and time, and even darken the dehazed result, which shows local patch information is not efficient for transmission estimation. Moreover, the large filter in the second layer may produce ring effect around the edges, because the normalized reflection coefficients of different pixels should be independent each other, thus if many pixels in local patch are used to estimate the normalized reflection coefficient, the interference between these pixels may



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

Fig. 2. The convergence error and dehazed results using different filter size. (a) The convergence errors and dehazed results for image *City* using filter size: (b)5-1-7, (c)7-1-7, (d)5-3-7 for the first third filters.

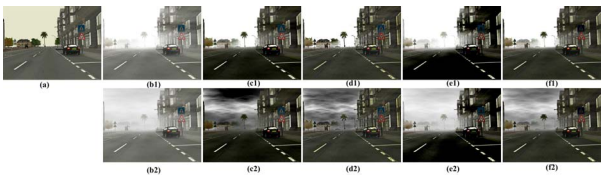
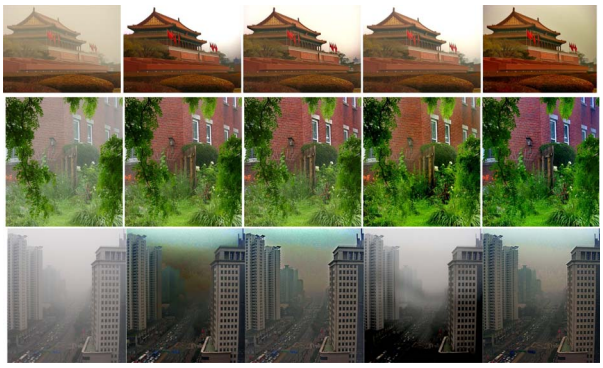


Fig. 3. Dehazed results for a synthetic image. (a) Haze-free image. (b1) Image with uniform fog. (c1)-(f1) Dehazed results by He's, Meng's, Tang's and our method for the image (b1). (b2) Image with cloudy heterogeneous fog. (c2)-(f2) Dehazed results by He's, Meng's, Tang's and our method for the image (b2).

be introduced so that the ring effect is produced, especially around the edges, as shown in Fig.2(d). Meanwhile, It is clear that the convergence speed could decrease with increasing the filter number and the layer number of deep network. However, if a fast dehazing speed is desired, a small filter number or layer number is preferred, which could still achieve the similar performance to ones with larger filter number and layer number.

4.2. Experiment on synthetic hazy images

Fig.3 shows a experiment result on Foggy Road Image Database (FRIDA). Obviously, Meng's [11] method has the best dehazing performance and produces less dark regions in the sky than He's [6] method because it adopts boundary constraint and contextual regularization to estimate the transmission, so that Meng's method produces high transmission values in the far regions for suppressing over-dark appearance. Tang's method [13] recovers over-dark road regions. Meanwhile, Tang's method slightly removes haze in the far regions. In contrast, our method not only has the similar capability of dehazing to Meng's method in the far regions but also sufficiently suppress over-dark looking in the road and sky regions of images with cloudy heterogeneous fog.



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

Fig. 4. Dehazing results for image *Tam, House, Street*. (a) Input image and dehazing results by (b) He's, (c) Meng's, (d) Tang's and (e) our method.

4.3. Experiment on real hazy images

We also perform image dehazing tests on a large dataset of real hazy images and compare our proposed method with some existing methods including He's [6], Meng's [11], Tang's [11]. Some experimental results are shown in Fig.4. He's and Meng's method can effectively remove haze. However, they also introduce degradation including halo artifacts, noise and lose some original spectral information in the sky patches where DCP doesn't hold. Tang's method not only removes little haze in the faraway regions but also produces over-dark looking in nearby regions of image *street*. Compared to other methods, our method can remove haze in the nearby regions like He's and Meng's method, and recover the original spectral information of sky patches with less degradation.

5. CONCLUSION

This paper develops a novel deep transmission network for robust single image dehazing algorithm. Firstly, a deep transmission network is developed and further trained from a large set of hazy images and dehazed (or haze-free) ones. This learned network can simultaneously cope with three color channels to automatically explore and exploit haze-related features. Then, different network structures and parameter settings are explored to analyze the haze-related features and to achieve tradeoffs between performance and speed. It is shown that three color channels information rather than local contrast is the most useful haze-related feature. Experimental results on a variety of hazy images demonstrate the proposed method can robustly dehazed images in various realistic scenes.

6. REFERENCES

[1] S. Huang, B. Chen, and W. Wang, "Visibility restoration of single hazy images captured in real-world weather conditions," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 24, no. 10, pp. 1814–1824, 2014.

[2] N. Hautiere, J. P. Tarel, and D. Aubert, "Mitigation of visibility loss for advanced camera-based driver assistance," *IEEE Transactions on Intelligent Transportation Systems*, vol. 11, no. 2, pp. 474–484, 2010.

[3] Codruta Orniana Ancuti and Cosmin Ancuti, "Single image dehazing by multi-scale fusion," *IEEE Transactions on Image Processing*, vol. 22, no. 8, pp. 3271–3282, 2013.

[4] Tian Bin, Li Ye, Li Bo, and Wen Ding, "Rear-view vehicle detection and tracking by combining multiple parts for complex urban surveillance," *IEEE Transactions on Intelligent Transportation Systems,*, vol. 15, no. 2, pp. 597–606, 2014.

[5] M. Negru, S. Nedevschi, and R. I. Peter, "Exponential contrast restoration in fog conditions for driving assistance," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 4, pp. 2257–2268, 2015.

[6] Kaiming He, Jian Sun, and Xiaoou Tang, "Single image haze removal using dark channel prior," in *IEEE Conference on Computer Vision and Pattern Recognition(CVPR)*, 2009, vol. 1, pp. 1956–1963.

[7] Johannes Kopf, Boris Neubert, Billy Chen, Michael F. Cohen, Oliver Deussen, Konstanz, and Dani Lischinski, "Deep photo: Model-based photograph enhancement and viewing," *ACM Transactions on Graphics (TOG)*, vol. 27, no. 5, pp. 116, 2008.

[8] Srinivasa G. Narasimhan and Shree K. Nayar, "Contrast restoration of weather degraded images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 6, pp. 713 – 724, 2003.

[9] Raanan Fattal, "Single image dehazing," *ACM Transactions on Graphics*, vol. 27, no. 3, pp. 721–729, 2008.

[10] Robby T. Tan, "Visibility in bad weather from a single image," in *IEEE International Conference on Computer Vision (CVPR)*, New York, USA, 2008, pp. 2347–2354.

[11] Gaofeng Meng, Ying Wang, Jiangyong Duan, Shiming Xiang, and Chunhong Pan, "Efficient image dehazing with boundary constraint and contextual regularization," in *IEEE International Conference on Computer Vision (ICCV)*, Sydney, NSW, 2013, pp. 617–624.

[12] Y. Lai, Y. Chen, C. Chiou, and C. Hsu, "Single-image dehazing via optimal transmission map under scene priors," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 25, no. 1, pp. 1–14, 2015.

[13] Ketan Tang, Jianchao Yang, and Jue 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.

[14] H. Koschmieder, "Theorie der horizontaler sichtweite beitraege," *Physicae Freiberger Atmosphere*, , no. 12, pp. 33–55, 1925.

[15] Srinivasa G. Narasimhan and Shree K. Nayar, "Vision and the atmosphere," *International Journal on Computer Vision*, vol. 48, no. 3, pp. 233–254, 2002.

[16] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Backpropagation applied to handwritten zip code recognition," *Neural computation*, vol. 1, no. 4, pp. 541–551, 1989.

[17] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton, "Deep learning," *nature*, vol. 521, pp. 436–444, 2015.

[18] Kaiming He, Jian Sun, and Xiaoou Tang, "Guided image filtering," in *The 11th European Conference on Computer Vision (ECCV)*, Heraklion, Crete, Greece, 2010, vol. 6311, pp. 1–14, Springer.