An End-to-End Image Dehazing Method Based on Deep Learning

Yi Zhang\textsuperscript{1}, Hongbing Huang\textsuperscript{1}, Junyi Liu\textsuperscript{1}, Chao Fan\textsuperscript{1}, Yanyan Wang\textsuperscript{1}, Qing Cai\textsuperscript{1}, Yingying Ruan\textsuperscript{2,} and Xiaojin Gong\textsuperscript{2}

\textsuperscript{1}State Grid Zhejiang Electric Power Co., Ltd, Hangzhou, Zhejiang, China
\textsuperscript{2}College of Information Science & Electronic Engineering, Zhejiang University, Hangzhou, Zhejiang, China

*E-mail: 3140102269@zju.edu.cn

Abstract. Image dehazing is a classic problem in computer vision. Most traditional methods use human-engineered features, such as dark channel prior, for dehazing. Recently, deep learning based approaches have been developed to solve this problem, but most of them are not end-to-end. In this paper, we propose an end-to-end learning method. This network consists of three parts to, respectively, estimate the transmission map, predict the global atmospheric light, and perform dehazing based on the estimated parameters. In order to train this network, we use Virtual KITTI dataset and NYU depth dataset to synthesize a training set composed of haze images and their corresponding transmission maps and global atmospheric light. Experiments demonstrate that our approach can obtain good performance on both synthetic and real haze images; moreover, the dehazed images have natural color and light contrast.

1. Introduction

With the development of industry and rapid economic production, a large amount of particles is emitted, which aggravates air pollution problems. In the outdoor haze scene, the captured image is affected by the haze particles, and there are problems such as poor contrast and low visibility. These problems will affect the classification, segmentation, detection and tracking of images. Therefore, it is very meaningful to design an effective image dehazing method which can recover clear dehazed images from haze images.

The image dehazing methods proposed earlier are very complicated, and some methods need multiple images or additional information to dehaze. Based on the polarization phenomenon, Schechner et al. \cite{1} propose a dehazing method using multiple images of different degrees of polarization of the same scene. Narasimhan et al. \cite{2} propose an atmospheric scattering model, which using several images under different weather to dehaze. However, this model is impossible to dehaze the objects similar to haze effectively. Then, Narasimhan et al. \cite{3} improve the model and propose a monochromatic atmospheric scattering model. However, in most cases, it is very difficult to get two images that are the same except for the weather conditions, so the application of the above methods is very limited. Narasimhan and Nayar \cite{4} obtain rough depth of the scene by manually specifying the maximum and minimum depths, and then obtained a dehazed map. Hautiere et al. \cite{5} use the optical sensor system to get depth as additional information for dehazing. Kopf et al. \cite{6} use geographic digital terrain to dehaze with the known depths of the scene. These three methods require additional
depth information to dehaze. However, in most cases, the depth information of scene is unknown, so these methods have significant limitations.

Since dehazing methods which need multiple images or additional depth information have too many restrictions. A variety of methods for single image dehazing are proposed, which is obviously faster and more practical. Kim et al. [7] use the histogram equalization to enhance the contrast of the haze image. However, this method cannot restore the real scene. Based on the physical model, some single image dehazing methods by estimating the transmission map are proposed. In [8], Independent Component Analysis (ICA) is used to estimate the transmission map and then perform dehazing. But this method cannot deal with dense-haze images well. He et al. [9] propose Dark Channel Prior (DCP) for image dehazing, and solve the halo phenomenon with soft matting. In the following research, guided filtering is proposed in [10], which improved the efficiency. Zhu et al. [11] observe Color Attenuation Prior (CAP) for estimating the transmission map and use it to dehaze.

In recent years, some dehazing methods based on deep learning are proposed. For example, Cai et al. [12] propose the DehazeNet and Ren et al. [13] propose a multi-scale convolutional network. After Generative Adversarial Networks (GAN) proposed, Li et al. [14] propose a method using GAN. But the above methods only obtain transmission map through network. In view of this, this paper proposes an image dehazing network that not only estimates transmission map, but also estimates atmospheric light. The network is based on the physical model and Convolutional Neural Network (CNN), and it is divided into three parts: the transmission map estimation part, the atmospheric light estimation part, and the dehazed image restoration part. Because it is difficult to get the haze image and corresponding dehazed image, we select the Virtual KITTI data set and the NYU depth data set to form a training set composed of haze images and corresponding transmission map and atmospheric light.

2. Physical model of haze image

In 1976, McCartney [15] proposed a physical model of atmospheric scattering. It can describe the formation of a haze image taken with a camera on a haze day, as shown in figure 1. Narasimhan and Nayar [4, 16] develop this model and the model can be formally written as

\[
I(x) = I_\infty \rho(x)e^{-\beta d(x)} + I_\infty (1 - e^{-\beta d(x)})
\]

(1)

Where \(I(x)\) is the hazy image, \(I_\infty\) is the global atmospheric light, \(\rho(x)\) represents the radiance normalized by the point \(x\) relative to the sky, \(\beta\) represents the scattering coefficient of the atmospheric light, \(d(x)\) is the distance between the point \(x\) in the scene and the camera.

![Figure 1. Physical model of atmospheric scattering in haze](image)

In [9], the equation (1) is further simplified as following

\[
I(x) = I(x)t(x) + A(1 - t(x))
\]

(2)

\[
t(x) = e^{-\beta d(x)}
\]

(3)

Where \(t(x)\) is the transmission map, and \(A\) is the atmospheric light.

3. The proposed image dehazing network

The goal is to obtain dehazed images from haze images. It is an operation for images, and CNN has achieved good results in many fields of image processing in recent years. Therefore, we use CNN to construct the dehazing network. DehazeNet [12] and multi-scale convolutional network [13] only estimate the portion of the transmission map required for dehazing. Therefore, in the design of the image dehazing network, not only the estimation of the transmission map portion but also the
atmospheric light estimation portion is added, thereby obtaining a neural network for end-to-end dehazing.

Based on the physical model of haze image, it is necessary to obtain the transmission map and the atmospheric light. Therefore, the network is composed of three parts: the transmission map estimation part, the atmospheric light estimation part, and dehazed image recovery part. The entire network is shown in figure 2. The input of this network is the RGB channels of the haze images.

3.1. Transmission map estimation
The purpose of this part is to predict the transmission map corresponding to the haze image. It is composed of three operation layers: convolutional layer, pooling layer and upsampling layer.

The convolutional layer contains the convolution kernel and the bias. The response function is:

\[ f_{i+1} = ReLU\left(\sum f_j * k_{j,i} + b_{i+1}\right) \]  (4)

Where \( f_{i+1} \) is the i-th feature map of the current layer, \( f_j \) is the j-th feature map of the previous layer, \( k_{j,i} \) is the kernel, \( b_{i+1} \) is the bias to get the i-th feature map of current layer. The convolution operation can extract useful features. In order to add nonlinearity, the activation function used by the first three convolutional layers is the ReLU function. For the last convolutional layer, since the transmission map that needs to be output is one-dimensional, but the feature map dimension of the previous layer is 10, we use a convolution kernel with a scale of 1x1. It not only can retain the planar structure of the previous layer, but also can play a role in dimensionality reduction. Considering that the value of the transmission map should be between 0 and 1, the activation function of this layer uses the Sigmoid function to ensure the range of the output value.

The pooling layer uses the maximum pooling, and the scale is 2x2. The maximum pooling can well maintain the translation and rotation invariance of the feature map, but at the same time it will reduce feature map’s size to half, but transmission map we need is the same size as the input image, so we add an upsampling layer after the maximum pooling layer to keep the size unchanged. The upsampling layer uses nearest neighbor interpolation method. If the maximum pooling layers and upsampling layers are not included in the network, the size is the same, but this will reduce the nonlinearity of the entire network.

3.2. Atmospheric light estimation
The purpose of this part of the network is to obtain the global atmospheric light corresponding to the haze image.

Since the global atmospheric light to be obtained in this part is related to the whole image, the convolution kernel used in the convolutional layer has a large scale, and its response function is consistent with the transmission map estimation part. The characteristics depended on are relatively simple, so the number of convolution layers in this part is set to 2 layers. The pooling layer selects the maximum pooling. On the basis of maintaining the invariance, it plays the role of dimensionality reduction, and the size is 2x2. The fully connection layer is set to 3 layers, respectively reducing the dimension to 256, 10, and finally to 1, and estimating the final atmospheric light value.
3.3. Dehazed image recovery
According to the equation (2) in physical model of haze image, the following formula can be obtained:

\[ J(x) = \frac{t(x) - A}{\max(t(x), t_0)} + A \]  

(5)

Where \( t_0 \) is the minimum threshold and it is set to 0.1, to avoid \( J(x) \) is too large due to the influence of noise. After obtaining transmission map and atmospheric light through the first two parts of the network, a dehazed image can be recovered by using equation (5).

4. Experiment

4.1. Synthesis of training set
Since it is very difficult to obtain haze and dehazed images of the same scene, this training set is obtained by using the physical model of haze image to generate haze images from clear images. We use equation (2) and (3) to synthesize the haze images. It needs to get the depth map corresponding to the clear image. Here, we select 4500 images in the NYU depth dataset containing the depth information and 3300 images in the KITTI dataset. The former is an indoor scene, and the latter is an outdoor scene.

In the process of obtaining the transmission map through the depth map, we randomly generate the scattering coefficient \( \beta \), and the value of \( \beta \) ranges from 0.5 to 1.5. If the \( \beta \) is too large, the transmission map is too close to 0, and the \( \beta \) is too small, the concentration of the synthesized haze is too thin. In the process of synthesizing haze images, we randomly generate a global atmospheric light \( A \), where \( A \) has a value between 0.8 and 1. Resize all images, including synthetic haze images and generated transmission maps, to a uniform size of 120 x 160. A total of 7800 pairs of haze images and corresponding transmission maps are generated, with 4000 pairs of indoor scene and 3000 pairs of outdoor scenes as training sets, 500 pairs of indoor scene and 300 pairs of outdoor scene as test sets.

4.2. Training
The transmission map estimation part and the atmospheric light estimation part are trained separately. They are both trained through minimizing the corresponding loss function.

In transmission map estimation part, we use Mean Square Error (MSE) between the network-estimated transmission map and the real transmission map obtained during the synthesis, as in equation (6):

\[ L(t(x), t^*(x)) = \frac{1}{q} (\sum_{i=1}^{q} ||t_i(x) - t_i^*(x)||^2) \]  

(6)

Where \( t(x) \) is the estimated transmission map, \( t^*(x) \) is the real transmission map, and \( q \) is the number of images. The learning rate is set to 0.01, the batch size is set to 16, and when the iterations of trainings exceeds 100, the learning rate decreases 0.0001 every 10 iterations.

In atmospheric light estimation part, we use MSE between the atmospheric light estimated by the network and the atmospheric light used in the synthesis of the haze images, as in equation (7):

\[ L(A, A^*) = \frac{1}{q} (\sum_{i=1}^{q} ||A_i - A_i^*||^2) \]  

(7)

Where \( A \) is the estimated atmospheric light value, \( A^* \) is the atmospheric light value used in the synthesis of the haze images, and \( q \) is the number of images. The learning rate is set to 0.00005 and the batch size is set to 16.

4.3. Experimental results

4.3.1. Results of synthetic haze image. The dehazing results of different dehazing methods are analyzed for different scenes in the synthetic haze image, as shown in figure 3.

Observing the dehazing results of the first and second rows, we can find that the result of the DCP [9] is obviously dark at the ceiling. The result of the CAP [11] is slightly better in the ceiling portion than the DCP, but still darker than the method we proposed. For the wall part, as shown in the first and
the second row of figure 3(d), the results of CAP have colour difference. Our method obtains good performance on the wall and ceiling, no obvious darkness occurs. In the third and fourth rows, it can be clearly observed that the color of the road is darker in the results obtained by the DCP and CAP. The visual effect of the results of our method is better (see the road and the green plants in figure 3).

![Figure 3](image)

**Figure 3.** Results of synthetic haze images obtained by different dehazing methods

In the process of synthesis in outdoor scene, the distance of the sky is very far, so the corresponding pixel points’ transmission map value is close to 0. Therefore, when generating a haze image, the values of the three channels of these points are equal to the atmospheric light used in the synthesis. So the color information of the original sky is lost. The sky of the dehazing results of the outdoor scene cannot restore the original color (see the sky in figure 3).

4.3.2. *Results of real haze image.* The dehazing results of the actual haze image are shown in figure 4. Similar to the results in Section 5.3.1, the DCP has a large color difference in the sky. The results of CAP in non-sky area is not well. The dehazing results obtained by our method have no obvious colour difference in the sky area, and the details in the sky recover better (see the antenna of the first line in figure 4(b)).

![Figure 4](image)

**Figure 4.** Results of real haze images obtained by different dehazing methods
4.3.3. **Quantitative analysis.** We use SSIM for quantitative analysis, and the SSIM can measure the similarity of two images. The SSIM is defined as follows:

\[
SSIM(x, y) = \left( \frac{2\mu_x\mu_y+c_1}{\mu_x^2+\mu_y^2+c_1} \right) \left( \frac{2\sigma_{xy}+c_2}{\sigma_x^2+\sigma_y^2+c_2} \right)
\]

\[
c_1 = (k_1L)^2, c_2 = (k_2L)^2
\]

Where \(\mu_x\) and \(\mu_y\) are the mean value of the image, \(\sigma_x^2\) and \(\sigma_y^2\) are the variance of the image, \(\sigma_{xy}\) is the covariance of the two images, \(L\) is dynamic range of the image pixel value, \(k_1 = 0.01, k_2 = 0.03\). If \(x\) and \(y\) are identical, their SSIM value is 1.

Figure 5 shows SSIM of different dehazing results. For conference room image, our method’s result is closer to real haze-free image than DCP and CAP. For library image, the most similar to real haze-free image is the result of CAP. Compared with the DCP, our method’s result is slightly higher.

**Figure 5.** SSIM of conference room image and library image dehazing results

4.3.4. **Compare with other networks.** DehazeNet [12] can get good dehazing results under normal scene. However, in figure 6, it can be observed that the result obtained by DehazeNet is dark, and the detail in the red box is seriously lost, but our method does not have this problem.

**Figure 6.** Comparison of our results with the results obtained by DehazeNet

Compare with multi-scale convolutional network (MSCNN) [13], the results are shown in figure 7. The color of the wall obtained by the MSCNN is not nature, but the results of our method are closer to the real haze-free images.

**Figure 7.** Comparison of our results with the results obtained by MSCNN
5. Conclusion
In this paper, an end-to-end image dehazing method based on deep learning is proposed. We design a network that can both estimate the transmission map and atmospheric light. We use the clear images with depth information to get the training set. This method is effective not only for the synthesized haze images, but also for the real haze images, and the visual effect is also good. Compared with the traditional image dehazing method, the proposed network can be applied to a wide range of scenes, and can obtain better results in sky region. Compared with some dehazing methods based on deep learning, our method can estimate the transmission map and atmospheric light simultaneously, which is more convenient. The method proposed in this paper still has some limitations. There exists improvement in the accuracy of the transmission map and atmospheric light estimation. Image dehazing based on deep learning still has development in future research.

References
[1] Y. Y. Schechner, S. G. Narasimhan, and S. K. Nayar. “Instant dehazing of images using polarization.” in Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR), vol. 1, Dec. 2001, pp. I-325–I-332.
[2] S. K. Nayar and S. G. Narasimhan. “Vision in bad weather.” in Proc. IEEE Conf. Comput. Vis., vol. 2, Sep. 1999, pp. 820–827.
[3] S. G. Narasimhan and S. K. Nayar, “Contrast restoration of weather degraded images,” IEEE Trans.Pattern Anal.Mach. Learn., vol. 25, no. 6, pp. 713–724, Jun. 2003.
[4] S. G. Narasimhan and S. K. Nayar, “Interactive (de)weathering of an image using physical models,” in Proc. IEEE Workshop Color Photometric Methods Comput. Vis., vol. 6. Oct. 2003, pp. 1–8.
[5] N. Hautiere, J.P. Tarel, D. Aubert, Towards fog-free in-vehicle vision systems through contrast restoration, in: Proceedings of the 2007 IE
[6] J. Kopf et al., “Deep photo: Model-based photograph enhancement and viewing,” ACM Trans. Graph., vol. 27, no. 5, p. 116, Dec. 2008.
[7] T. K. Kim, J. K. Paik, and B. S. Kang, “Contrast enhancement system using spatially adaptive histogram equalization with temporal filtering,” IEEE Trans. Consum. Electron., vol. 44, no. 1, pp. 82–87, Feb. 1998.
[8] R. Fattal. “Single image dehazing.” ACM Trans. Graph., vol. 27, no. 3, pp. 72:1–72:10, Aug. 2008.
[9] K. He, J. Sun, and X. Tang, “Single image haze removal using dark channel prior,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 33, no. 12, pp. 2341–2353, Dec. 2011.
[10] K. He, J. Sun, and X. Tang, “Guided image filtering,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 35, no. 6, pp. 1397–1409, Jun. 2013.
[11] 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.
[12] Bolun. Cai, Xiangmin. Xu, Kui. Jia, Chunmei. Qing, Dacheng. Tao, “DehazeNet: An End-to-End System for Single Image Haze Removal” IEEE Transactions on Image Processing, vol. 25, no. 11, pp. 5187-5198, 2016
[13] Wenqi. Ren, Si. Liu, Hua. Zhang, Jianshan. Pan, Xiaochun. Cao, Ming-Hsuan. Yang, “Single Image Dehazing via Multi-Scale Convolutional Neural Networks” ECCV 2016
[14] Li C, Zhao X, Zhang Z, et al. Generative adversarial dehaze mapping nets[J]. Pattern Recognition Letters, 2017.
[15] E. J. McCartney, Optics of the Atmosphere: Scattering by Molecules and Particles. New York, NY, USA: Wiley, 1976.
[16] S. G. Narasimhan and S. K. Nayar, “Removing weather effects from monochrome images,” in Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR), vol. 2. Dec. 2001, pp. II-186–II-193.