

# Single Image Defogging Method based on Deep Learning

Baoping Yuan <sup>a</sup>, Yong Yang <sup>b</sup> and Baofu Zhang <sup>c</sup>

School of Computer Science and Technology, Chongqing University of Posts and  
Telecommunications, Chongqing 400065, China

<sup>a</sup>hedaybp@163.com, <sup>b</sup>yangyong@cqupt.edu.cn, <sup>c</sup>zhangbfxo@163.com

**Keywords:** Image Defogging; Deep Learning; Convolutional Neural Network; Transmission Map; Atmospheric Scattering Model.

**Abstract.** Single image defogging is a challenging ill-posed problem. Current image defogging methods usually get defogging solutions based on various priors or assumption, which is hardly satisfied in practice. In this paper, a single image defogging method based on deep learning is proposed, in which the priors and assumption do not hold. Firstly, the prediction of transmission map is progressively refined by using three scales convolutional neural networks. Secondly, the fog-free image can be recovered by the atmospheric scattering model after transmission map is got. Experiment results show that the proposed deep learning based defogging algorithm achieves superior performance over state-of-art algorithms on both natural foggy images and synthetic images.

## 1. Introduction

Foggy weather has been gotten the continued concern in recent years. A cause of vehicles accidents is the existence of fog. In recent years, there is a great progress for research on the single image defogging, but most image defogging methods use various priors or assumption to get defogging solutions, which is hardly satisfied in practice. Tan *et al.* [1] assume that the local contrast ratio of fog-free images is higher than that of foggy images. Fattal *et al.* [2] assume reflectance in the local region is constant and used independent component analysis to remove fog. He *et al.* [3] find the dark channel priority theory with statistically analysis a large number of fog-free images, and then use the channel priority theory to defog. However, above mentioned algorithms have some drawbacks. For example, Tan *et al.*'s method easily causes oversaturation, Fattal *et al.*'s method isn't suitable for dense fog images, and He *et al.*'s can't solve the scene with sky areas.

Recently, some researchers have begun to use machine learning frameworks. Tang *et al.* [4] use random forest to combine the four features of foggy image to obtain the transmission map. Zhu *et al.* [5] construct a linear model for the depth map, and use a supervised learning algorithm to train parameters of the model. Despite great progresses have made in above mentioned algorithms, they still depend on various priors or constraints, and still have limitations.

In this paper, an algorithm to get the transmission map by data driving is proposed. The main contributions of this paper are as follows. First, an end-to-end image defogging algorithm is proposed, which does not depend on the fog-relevant properties or statistic prior of foggy images. Second, multi-scale convolutional neural networks model which can predict the global information and local details of the transmission map at the same time is designed.

## 2. Image Defogging Based on MS-CNN

In this paper, an image defogging method based on MS-CNN is proposed as follows. Firstly, the atmospheric scattering model is introduced, which suggests that the estimation of transmission map is the main step to recover a fog-free image. Secondly, a Multi-scale convolutional neural network model which learns the mapping relationships between foggy images and corresponding transmission maps is designed. Then, the loss function for training with SGD is proposed. Finally, the operation to get the

fog-free image from the estimated transmission map under the instructions of atmospheric scattering model [6] will be presented in this section. The flow chart of image defogging is shown as Fig. 1.

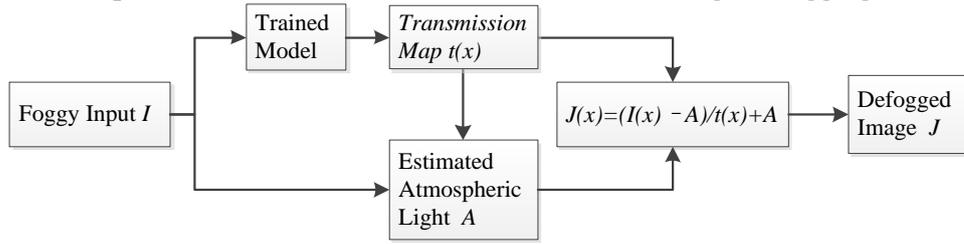


Fig. 1 Defogging flow chart

**2.1 Atmospheric Scattering Model.**

Atmospheric scattering model is first proposed by McCartney [6], and then developed by Nayer and Narasimhan. McCartney's atmospheric scattering model can be defined as

$$I(x) = J(x)t(x) + A(1 - t(x)) \tag{1}$$

Where  $I(x)$  is the observed foggy image,  $J(x)$  is the fog-free scene image,  $t(x)$  is the transmission map,  $A$  is the global atmospheric light,  $x$  is the position of the pixel in the image.  $t(x)$  represents the ability that the scene reaches to the sensor,  $t(x)$  can be written as

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

$d(x)$  represents the distance from the scene object to the sensor, and  $\beta$  is the scattering coefficient. If the scene is infinite,  $d(x) = \infty$ , then  $A$  will be equal to  $I(x)$ . When  $d(x)$  is long enough,  $t(x)$  will be close to zero, and atmospheric light is equal to the observed value of the pixel. We have

$$A = \max I(y) \quad s.t. \ y \in \{x | t(x) \leq t_0\} \tag{3}$$

Thus, the defogged image  $J(x)$  can be obtained after the transmission map  $t(x)$  and atmospheric light  $A$  are estimated.

**2.2 Multi-scale Convolution Neural Networks Model.**

Although the traditional CNN is robust to translation, rotation and scaling of images, Sermanet proves that the size of the input image has a great influence on results in traditional CNN model.

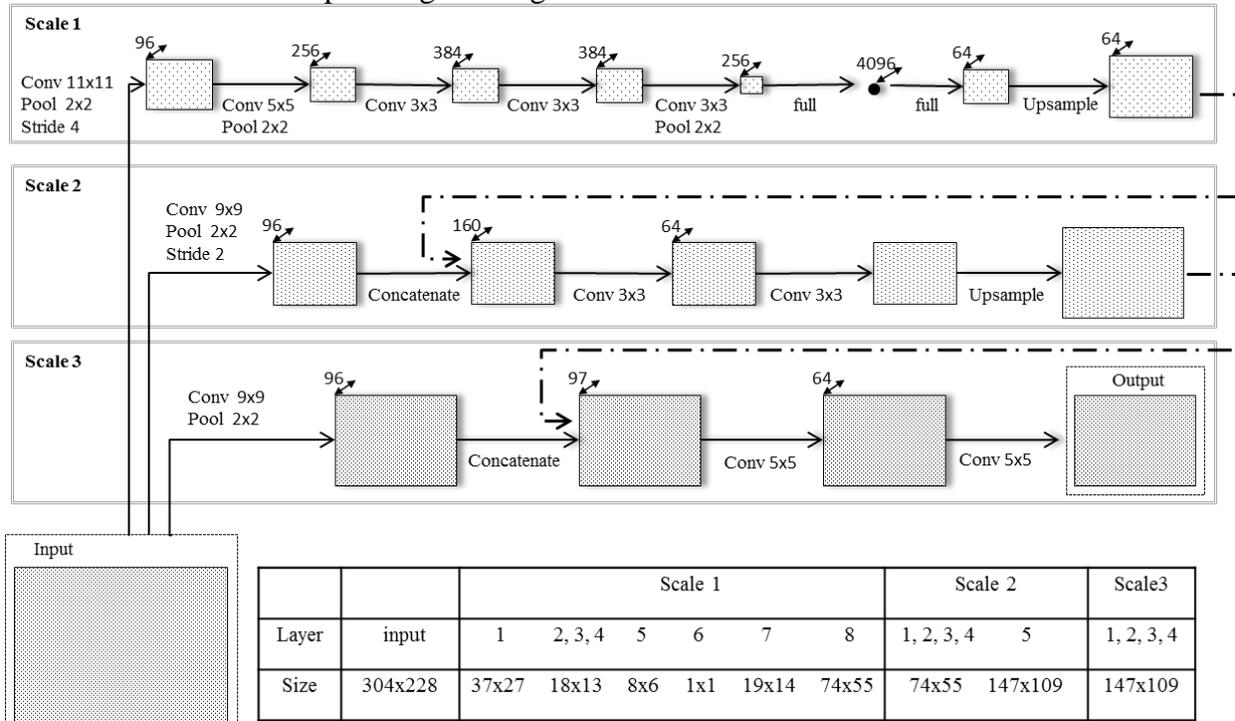


Fig. 2 Multi-scale convolutional neural network model

Multi-scale convolutional neural networks method is adopted to enhance the robustness of the model. In this paper, the multi-scale convolutional neural networks are used to obtain the transmission map  $t(x)$ , structural model is shown in Fig. 2. The scale1 network is used to obtain the holistic transmission map

which is then refined by the scale2 network. The scale3 network makes the model be applied to higher resolution foggy images.

The scale1 network is a deformation of AlexNet network model [7] which contains five convolution layers, two max-pooling layers, and two fully connect layers and one up-sample layer, as show in Fig. 2. As the top two layers of the network are fully connected, each pixel of the output feature maps contains global information of the entire image. However, the full connection leads to missing image edge information which will result in a fuzzy holistic transmission map.

The scale2 network consists of convolution layers and upsample layers, as shown in Fig. 2. The scale2 network generates a medium transmission map by combining the output of scale1 and narrow field original input image. The scale1 and scale2 networks are trained together using stochastic gradient descent with loss function.

The structure of scale3 network is similar to scale2. Scale3 network makes the model handle higher resolution images. The input image is pooled by a smaller stride and then concatenate with the output of scale2 network as input to the second layer. The concatenated feature maps for second layer result in a refined transmission map model which has rich details and can adapt to high resolution situations. The size of output transmission map is the half of network input.

### 2.3 Loss function.

By minimizing the loss function between the predicted transmission map  $T$  and the ground truth image  $T^*$ , the mapping relationship between foggy image and corresponding transmission map is learned. The loss function can be written as

$$L(T, T^*) = \frac{1}{n} \sum_i d_i^2 + \frac{1}{n} \sum_i [(\nabla_x d_i)^2 + (\nabla_y d_i)^2] \quad (4)$$

Where  $d_i = T - T^*$ ,  $i$  is the pixel point and  $n$  is the number of all the pixels in the training set.  $\nabla_x d_i$  and  $\nabla_y d_i$  denote gradient differences of  $d_i$  in horizontal direction and vertical direction, respectively. The second term on the left side of the equation is the difference between the predicted transmission map and ground truth on the first order gradient, which ensures that the predicted transmission map is not only close to the ground truth but also has similar local structures with ground truth.

### 2.4 Recovering Fog-free Image by Atmospheric Scattering Model.

From equation 1, if we want to remove the image fog, the transmission map and atmospheric light should be known first. From equation 3, when  $t(x)$  tends to zero,  $A = I(x)$ . The distance of scene ranges from zero to infinity, and  $t(x) = 0$  when  $d(x)$  approaches infinity. In this paper, we select pixels value of the foggy image whose corresponding pixels in the transmission map are top 0.1% darkest as the atmospheric light  $A$ .

After  $A$  and  $t(x)$  are obtained, the fog-free image  $J(x)$  can be obtained by equation 1. Since when  $t(x)$  tends to 0,  $J(x)$  will be too large and cause the whole image to be whitened, so here we set a minimum value for the transmittance. The final defogging equation is

$$J(x) = \frac{I(x) - A}{\max\{0.1, t(x)\}} + A \quad (5)$$

## 3. Experiments

### 3.1 Experiment Database.

A large labeled dataset is required to train a convolutional neural network model. In this paper, the NYU depth image dataset [8] is used to generate foggy image and the corresponding transmission map, because there are no ready-made natural foggy images and corresponding transmission maps group. In this paper, 120,000 clear fog-free images and corresponding depth images are randomly selected from 464 different indoor scenes. The foggy images and the corresponding transmission maps dataset are constructed according to equation1, 2. When the transmission maps are generated using  $t(x) = e^{-(\beta d)}$ , the scattering coefficient  $\beta$  is set to 0.75-1.5. Because if  $\beta$  is too small, the fog of the synthesized foggy image is thin and noisy; if too large, the transmission map will become zero [9]. The parameter  $k$  of the atmospheric light  $A = (k, k, k)$  can be set to a random value between 0.7 and 1.0. Each group of depth

images randomly select four  $\beta$  values from 0.75-1.5, and 480,000 transmission maps and foggy images are obtained. Then the Middlebury stereo datasets [10] are processed in the same way, and 50 foggy images are generated as the validation set. All images are rescaled to 320x240.

### 3.2 Experiments on Natural Foggy Images.

In order to verify the validity of the proposed method, the defogging results of He [3], Meng [11], Tarel [12], Zhu [5] on natural foggy images are compared. As all of above mentioned de-fog algorithms have strong de-fogging ability on ordinary foggy image, which is difficult to distinguish the advantages and disadvantages of above mentioned algorithms. Therefore, this paper chooses challenging foggy images with large gray or white area as test images. Fig. 3 shows the results of the proposed algorithm and other algorithms on natural foggy images. Fig. 3(a) are the input fog images, Fig. 3(b-f) show He's, Meng's, Tarel's, Zhu's and our results, respectively.



Fig. 3 Comparison on natural foggy images. (a) Input natural foggy image. (b) He et al.'s results. (c) Meng et al.'s results. (d) Tarel et al.'s results. (e) Zhu et al.'s results. (f) Our results.

Fig. 3(b) shows that in He's results; fog can be well removed without halo artifacts. However, there still have color distortions in white areas. Fig. 3(c) shows that Meng's algorithm can remove most of fog, and well restore the details of the object. However, it can be seen that the images recovered in Fig.3 (c) are oversaturated and distorted. In particular, the waterfall of the first image is colorful. Tarel's results have similar problems. In contrast, the results of Zhu are better than those of the first three. Compared with the other four algorithms, the proposed algorithm not only can recover the color of the sky and cloud, but also can enhance the details of the mountain.

### 3.3 Experiments on Synthetic Foggy Images.

Fig. 4 shows the results of five different algorithms on the synthetic foggy images. Fig. 4 (a) are the synthetic foggy images, Fig. 4(b-f) are the defogged images obtained by above mentioned five algorithms, Fig. 4(g) are the ground truth images for comparison. These foggy images and corresponding ground truth images are obtained from the Middlebury stereo datasets [10]. From Fig. 4(c), it can be seen that the results of Meng are over-enhanced, for example, the mask in the third image. Obviously, Teral's results are much darker than the ground truth images. The results of Zhu are very close to the ground truth, but there are still some deviations, such as the color of the cones in the third image is brighter than the ground truth image. In contrast, our method doesn't have the problem of over-saturation, and maintain the color of objects.

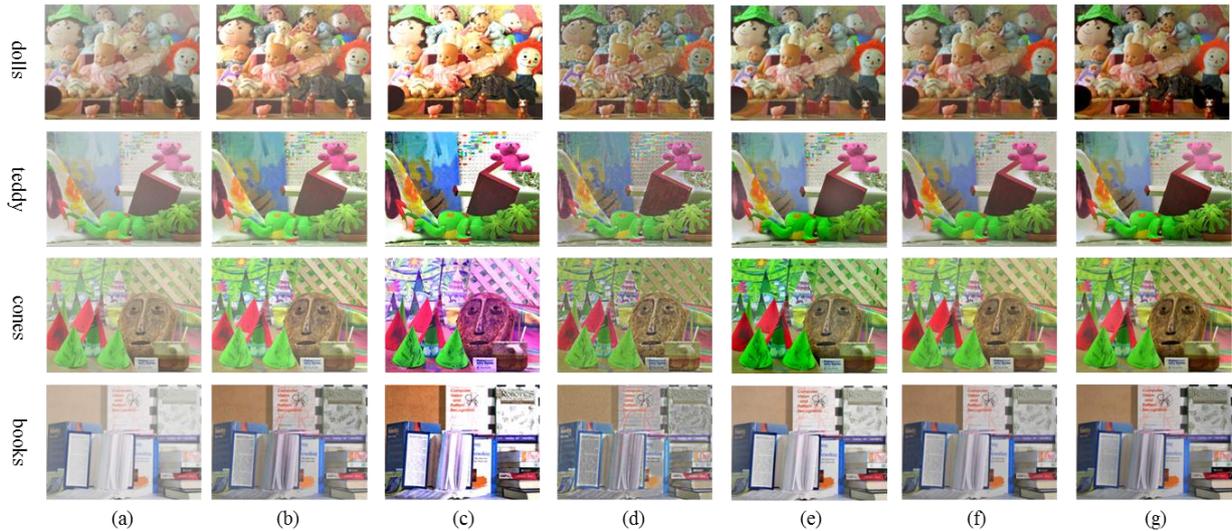


Fig. 4 Comparison on synthetic images. (a) Input natural fog image. (b) He et al.’s results. (c) Meng et al.’s results. (d) Tarel et al.’s results. (e) Zhu et al.’s results. (f) Our results. (g) Ground truth

In order to evaluate and quantify the result, the mean square error (MSE) and structural similarity (SSIM) are calculated on the results of Fig.4. The calculation formula of MSE is.

$$e = \sqrt{\frac{1}{3N} \sum_{c \in \{r,g,b\}} \|J^c - G^c\|^2} \tag{6}$$

$J$  is the defogged image obtained by the defogging algorithms,  $G$  is the ground truth image,  $J^c$  and  $G^c$  are the RGB channels of  $J$  and  $G$ , and  $N$  is the number of pixels of the image  $G$ .  $e$  is the mean square error between the defogged image  $J$  and the ground truth image  $G$ . The smaller the MSE is, the closer between the defogged image and the ground image. Fig.5 further shows the MSE results of different algorithms.

Structural similarity [13] is an image quality evaluation index which is used to evaluate the ability of structural information preservation. A high SSIM value indicates that the defogged image is highly similar to the ground truth image. Fig. 6 shows the SSIM value of different algorithms’ result. It can be seen that the defogged results of the proposed algorithm are superior to the others.

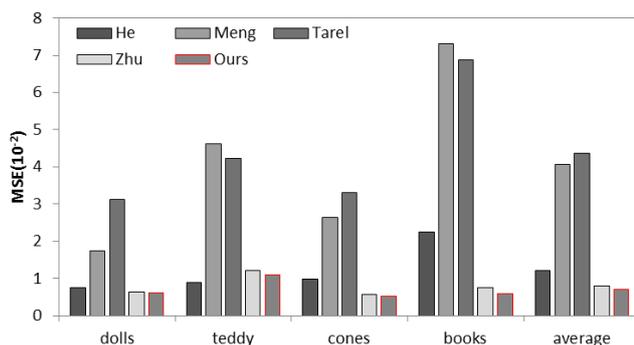


Fig. 5 Mean square error of different algorithms

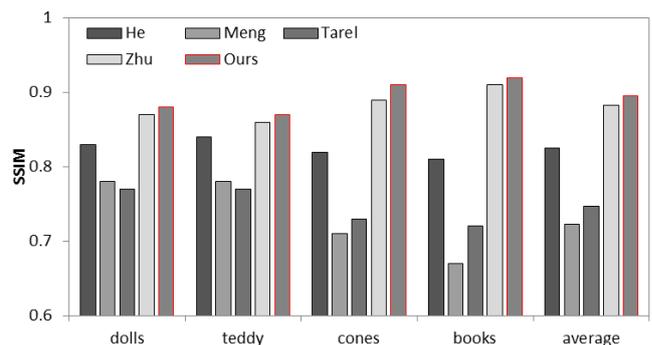


Fig. 6 Structural similarity of different algorithms

#### 4. Summary

In this paper, a single image defogging algorithm based on deep learning is proposed. Three different scales convolutional neural networks are used to train the transmission map model. The first scale convolutional neural network outputs rough transmission map which then enhanced details by the scale2 network. Finally, the scale3 network makes the model adaptable to high resolution images. Our algorithm does not depend on the fog-relevant properties or statistic prior of foggy images. It is proved that the defogging performance of the proposed algorithm is superior to the conventional defogging algorithms in both natural foggy images and synthetic images.

## References

- [1] Tan, R. T. (2008, June). Visibility in bad weather from a single image. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on* (pp. 1-8). IEEE.
- [2] Fattal, R. (2008). Single image dehazing. *ACM transactions on graphics (TOG)*, 27(3), 72.
- [3] He, K., Sun, J., & Tang, X. (2011). Single image haze removal using dark channel prior. *IEEE transactions on pattern analysis and machine intelligence*, 33(12), 2341-2353.
- [4] Tang, K., Yang, J., & Wang, J. (2014, June). Investigating haze-relevant features in a learning framework for image dehazing. In *2014 IEEE Conference on Computer Vision and Pattern Recognition* (pp. 2995-3002). IEEE.
- [5] Zhu, Q., Mai, J., & Shao, L. (2015). A fast single image haze removal algorithm using color attenuation prior. *IEEE Transactions on Image Processing*, 24(11), 3522-3533.
- [6] McCartney, E. J. (1976). *Optics of the atmosphere: scattering by molecules and particles*. New York, John Wiley and Sons, Inc., 1976. 421 p., 1.
- [7] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (pp. 1097-1105).
- [8] Silberman, N., Hoiem, D., Kohli, P., & Fergus, R. (2012, October). Indoor segmentation and support inference from RGBD images. In *European Conference on Computer Vision* (pp. 746-760). Springer Berlin Heidelberg.
- [9] Kim, T. K., Paik, J. K., & Kang, B. S. (1998). Contrast enhancement system using spatially adaptive histogram equalization with temporal filtering. *IEEE Transactions on Consumer Electronics*, 44(1), 82-87.
- [10] Hirschmuller, H., & Scharstein, D. (2007, June). Evaluation of cost functions for stereo matching. In *2007 IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1-8). IEEE.
- [11] Meng, G., Wang, Y., Duan, J., Xiang, S., & Pan, C. (2013). Efficient image dehazing with boundary constraint and contextual regularization. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 617-624).
- [12] Tarel, J. P., Hautiere, N., Caraffa, L., Cord, A., Halmaoui, H., & Gruyer, D. (2012). Vision enhancement in homogeneous and heterogeneous fog. *IEEE Intelligent Transportation Systems Magazine*, 4(2), 6-20.
- [13] Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4), 600-612.