



RCA-NET: Image Recovery Network with Channel Attention Group for Image Dehazing

Juan Du^(✉), Jiajia Zhang, Zhe Zhang, Wei Tan, Shangzhen Song, and Huixin Zhou

School of Physics and Optoelectronic Engineering, Xidian University, Xi'an 710071, China
juandu_xidian1107@outlook.com

Abstract. In this paper, we propose an image dehazing network with a channel attention model. Most existing methods try to resolve the dehazing problem through an atmospheric transmission model, but always fail to get promising results since the real-world physical imaging system is of high complexity. Therefore, we propose recovering a fog-free image from its foggy image using an end-to-end pipeline which can produce more realistic results. We apply a channel-wise attention model into our network and also employ the perceptual loss for supervision. Experimental results indicate that our method performs better than several state-of-the-art algorithms.

Keywords: Image dehazing · Image recovery network · Channel attention group · Transmission matrix · Atmospheric scattering model

1 Introduction

The existence of haze degrades the quality of the image captured by the surveillance system. Therefore, haze removal has become one of the hotspots in image processing; it can provide promising solutions for various tasks such as remote sensing, surveillance systems, and aerospace.

Originally, image dehazing was performed via enhancing image contrast to reduce the effects of haze on images. [1, 2] introduced a method to resolve the haze problem by maximizing local contrast of input images. A fast image defogging method was refined to estimate the amount of fog in an image using a locally adaptive wiener filter [3]. Subsequently, more algorithms were proposed based on the atmospheric scattering model theory. In [4], the albedo of a scene is estimated to aid in haze removal. [5–8] apply the dark channel prior to the atmospheric scattering model to calculate the transmission map. Further, algorithms have been proposed to improve the defogging by: enforcing the boundary constraint and contextual regularization [9], minimizing the nonconvex potential in the random field [10], enriching contextual information [11], or automating the defogging algorithm [12]. Recently, the prior information has been applied in dehazing to recover haze-free images. [13] proposed a method to dehaze in the same, or different, scale images using internal patches. [14] built a haze removal

model of scene depth, then trained this model with a color attenuation prior. The above algorithms reduce the image blur to some extent, but the restored image still has color distortion.

Recently, a deep convolutional neural network (CNN) method has become a hotspot for image dehazing [15, 16]. [17] introduced a strategy which generated first a coarse transmission matrix and then tried to obtain an accurate transmission matrix through refinement. The Dehazenet uses hazy images to train the network for the medium transmission matrix and then uses the atmospheric scattering model to recover haze-free images [18]. Most methods are based on the transmission matrix estimation which, unfortunately, cannot generate the defogging image directly. Furthermore, existing CNN-based networks usually treat each channel function equally, and hence varying information features and important details are missing [19–22].

To resolve these problems, we consider using channel-wise information as a weight in the normal convolution layer for excellent feature extraction. In this paper, we propose an image recovery network with channel attention (RCA-Net) to extract, adaptively, channel-wise features. Firstly, the transmission model $M(x)$ is drawn from an end-to-end network by minimizing the reconstruction errors. Secondly, the $M(x)$ is gradually optimized through an image recovery network with channel attention. Finally, more realistic color and structural details can be drawn from the recovery network.

The remainder of this paper is organized as follows: the proposed algorithm is described in Sect. 2. The experimental results and analysis are reported in Sect. 3. Finally, the conclusion is given in Sect. 4.

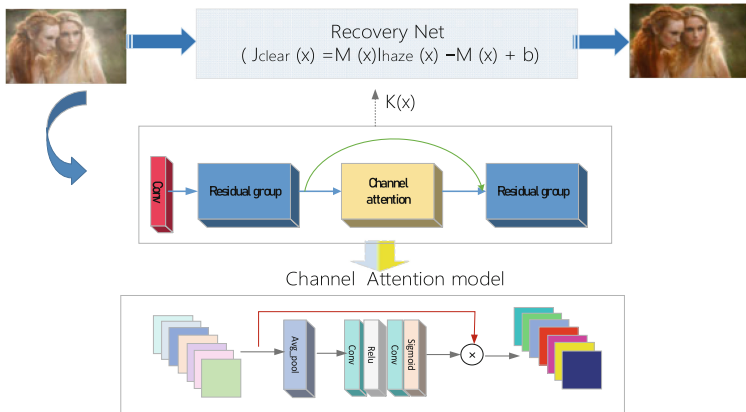


Fig. 1. Structure of recovery network with channel attention model.

2 Proposed Algorithm

In this paper, an end-to-end network with channel attention is proposed to achieve a high-level restoration of foggy images. Specifically, the atmospheric scattering model in RCA-Net has been simplified during the image restoration process. Furthermore, the channel attention model focuses on more significant information to improve haze removal.

2.1 Physical Model

An atmospheric scattering model is a foundation of defogging algorithms. The atmospheric scattering model can be expressed as

$$I_{haze}(x) = J_{clear}(x)t(x) + A(1 - t(x)) \quad (1)$$

$$J_{clear}(x) = \frac{1}{t(x)}I_{haze}(x) - \frac{A}{t(x)} + A \quad (2)$$

$$J_{clear}(x) = M(x)I_{haze}(x) - M(x) + b \quad (3)$$

where, $I_{haze}(x)$ denotes observed foggy image, $J_{clear}(x)$ is the fog-free image needed to be recovered.

According to the atmospheric scattering model, the final expression of $M(x)$ is

$$M(x) = \frac{\frac{1}{t(x)}(I_{haze}(x) - A) + B}{I_{haze}(x) - 1} \quad (4)$$

wherein the $M(x)$ parameter is a parameter that combines A and $t(x)$, B is the value of $A-b$, and the final value depends only on the model input $I_{haze}(x)$. A and $t(x)$ represent the global atmospheric light and the transmission matrix, respectively.

2.2 Dehaze Network Architecture

The previous models are obtained from a hazy image to the transmission matrix, but RCA-Net is the ultimate model acquired from the hazy image to the hazy-free image, which trains the end-to-end network to achieve defogging. The previous model of the defogging method from a hazy image to a transmission map is different from an end-to-end network of RCA-Net from hazy images to no hazy images. The proposed RCA-Net contains two parts as shown in Fig. 1: one part is an $M(x)$ estimation module that estimates $M(x)$ using five convolutional layers with channel attention model integrated. Another part is a recovery net that has 21 layers including multiplication layers and addition layers; A clean image can be restored through Eq. (3). The channel attention function is obtained through “Average pool \rightarrow Conv \rightarrow ReLU \rightarrow Conv \rightarrow Sigmoid” to get a channel-wise feature.

The matrix $M(x)$ is trained on NYU2 Depth Database [14]. The test image is then used as input to obtain a clear image. The architecture includes five steps in model parameters estimation, that is, the output of RCA-Net can be reconstructed in sequence through convolution feature extraction, a residual group, a channel attention model, and a residual group to the recovery image. The foggy image $I_{haze}(x)$ and the clear image $J_{clear}(x)$ is the input and output of RCA-Net model respectively and the network parameters are optimized by minimizing the loss. The image processing goes through three operations, which is shown in the following equations:

$$F_{con}(x) \rightarrow H_{conv}(I_{haze}(x)) \quad (5)$$

$$F_{rd}(x) \rightarrow H_{Residual}(F_{con}(x)) \quad (6)$$

$$F_{ca}(x) \rightarrow H_{channel}(F_{rd}(x)) \quad (7)$$

where H_{Conv} represents the convolution operation, and $F_{con}(x)$ is then used for deep feature extraction; $H_{Residual}$ denotes a deep convolution group with a long skip connection, $F_{rd}(x)$ is the furthest feature from the residual group; $H_{channel}$, the operation of channel attention function, which includes three steps: global average pooling, the sigmoid function, and ReLU function. $F_{ca}(x)$ is the feature map obtained from the channel attention model, which is multiplied by the channel attention feature, and then added to $F_{rd}(x)$.

The next part is the estimation of $M(x)$, which can be given by

$$M(x) \rightarrow H_{Residual}(F_{ca}(x)) \quad (8)$$

$$J_{clear}(x) = M(x) * I_{haze}(x) - M(x) + b = H_{Recovery}(I_{haze}(x)) \quad (9)$$

The loss function minimizes the error between the target image and the input image used to achieve network optimization. The common loss function in deep networks are the L1 loss and the L2 loss, or a combination of L1 loss and L2 loss. This is closely related to Mean Squared Error (MSE), but the visual effect is not good enough. To resolve this problem, a perceptual loss is introduced in RCA-Net for the purpose of making the feature extraction level of SR consistent with that of the original image. In this way, we can achieve perfect visual effects and more details close to the target image.

3 Preliminary Results

3.1 Experimental Setup

The Training Process: The weights are initialized using random numbers; the channel attention convolution group is more effective for a deep network; the decay parameter in model training is set to 0.0001;

The Database: We use the NYU2 Depth Database [14] which includes ground-truth images and depth meta-data. We set the atmospheric light $A \in [0.6, 1.0]$, and $\beta \in \{0.4, 0.6, 0.8, 1.0, 1.2, 1.4, 1.6\}$. For the NYU2 database, 27, 256 images are the training set and 3,170 images are the non-overlapping, called TestSet A.

3.2 Experiments and Results

To test the performance of the proposed algorithm, our method is compared with the traditional algorithms Boundary Constrained Context Regularization (BCCR) [9], Color Attenuation Prior (CAP) [14], Fast Visibility Restoration (FVR) [23], Gradient Residual Minimization (GRM) [24], Dark-Channel Prior (DCP) [4] and MSCNN [8].

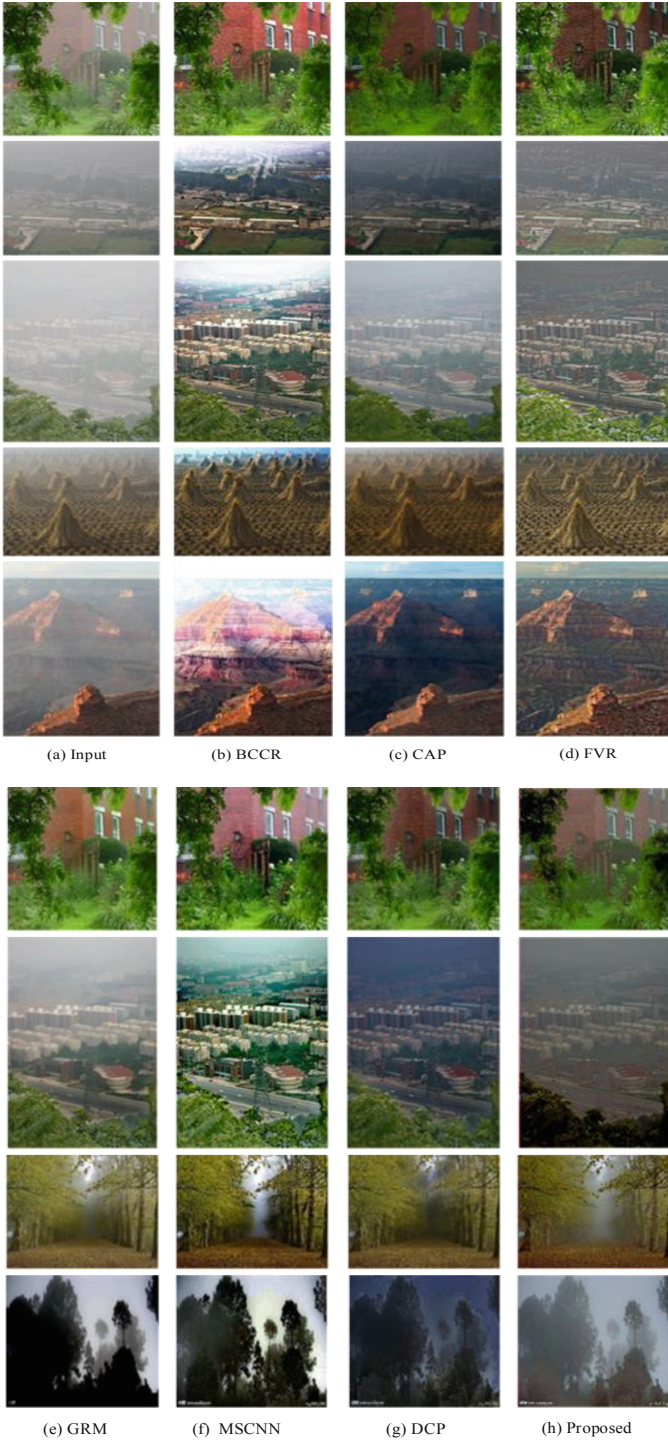


Fig. 2. Comparison of BCCR, CAP, FVR, GRM, MSCNN, DCP, and proposed RCA-Net

Figure 2 shows the defogging results of eight sets of images. The CNN-based algorithm is better than BCCR (0.9569), CAP (0.9757), FVR (0.9622), GRM (0.9249), and DCP (0.9449). The MSCNN and the proposed method, however, have higher SSIM, which are 0.9689 and 0.9792, respectively. Figure 3 shows the PSNR of the above defogging methods. The PSNR values of ATM, BCCR, FVR, NLD are between 15 dB and 20 dB. The PSNR of MSCNN, Dehazenet and the proposed method are higher than 20 dB. Compared with other methods, RCA-Net has perfect visual performance and a little higher PSNR.

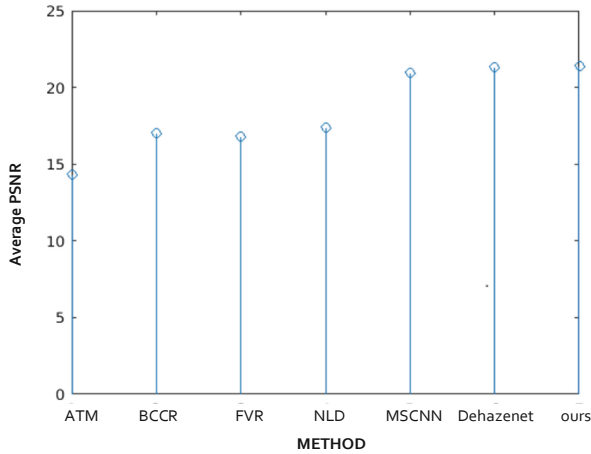


Fig. 3. Comparison of BCCR, ATM, FVR, NLD, MSCNN, DehazeNet and proposed RCA-Net on dehazing 800 synthetic images from the Middlebury stereo database.

Among to previous experiments, few quantitative results about the restoration quality were reported. Table 1 displays the average PSNR and SSIM results on TestSet A.

Table 1. Average PSNR and SSIM results on TestSet A.

Metrics	BCCR	CAP	FVR	GRM	DCP	MSCNN	RCA-Net
PSNR	15.751	19.630	15.031	18.974	18.387	19.121	20.054
SSIM	0.771	0.837	0.743	0.823	0.833	0.829	0.841

To further compare the results, the PSNR of each method is shown in Fig. 3 on the same image. It can be seen that RCA-Net greatly improves the PSNR, which is about 1 dB higher than BCCR, ATM, FVR, and NLD. Therefore, the dehazing strategy in an appropriate model is confirmed to be more effective than other networks.

3.3 Time Consumption

The light-weight structure of RCA-Net leads to faster dehazing. We select 100 images from TestSet A to test all algorithms. All experiments are implemented in Matlab 2016a on a PC with GPU Titan V, 12 GB RAM in an Ubuntu 16.04 system. We crop each input image into patches with a size of $256 * 256$ for running time comparison. The per-image average running times of all models are shown in Table 2.

Table 2. Comparison of average running time (seconds).

Metrics	BCCR	CAP	FVR	GRM	DCP	MSCNN	RCA-Net
Time	0.523	0.446	3.02	2.18	9.87	0.920	0.471

4 Conclusion

In this paper, we propose an end-to-end dehazing network from hazy images to clean images using the channel attention model. Through giving a weight to normal convolution, the channel attention model contributes to further feature extraction. This approach simplifies the dehazing network and can effectively keep more realistic colors and structural details. Ultimately, this network can fully retain real details from the test image, which enables a state-of-art comparison of dehazing, both quantitatively and visually.

Acknowledgement. The authors are grateful to the anonymous reviews for their valuable comments and suggestion. This paper is supported by National Natural Science Foundation of China (61675160); 111 Project (B17035).

References

1. Tan, R.T.: Visibility in bad weather from a single image. In: 2008 Computer Vision and Pattern Recognition, pp. 1–8. IEEE, Anchorage (2008)
2. Guo, D., Liu, P., She, Y., Yang, H., Liu, D.: Ultrasonic imaging contrast enhancement using modified dehaze image model. *Electron. Lett.* **49**(19), 1209 (2013)
3. Gibson, K.B., Nguyen, T.Q.: Fast single image fog removal using the adaptive wiener filter. In: 2013 IEEE International Conference on Image Processing, pp. 714–718 (2013)
4. Fattal, R.: Single image dehazing. *ACM Trans. Graph. (TOG)* **27**(3), 72 (2008)
5. Du, Y., Li, X.: Recursive deep residual learning for single image dehazing. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 730–737 (2018)
6. Zhang, H., Sindagi, V., Patel, V.M.: Joint transmission map estimation and dehazing using deep networks. arXiv preprint [arXiv:1708.00581](https://arxiv.org/abs/1708.00581) (2017)
7. Tang, K., Yang, J., Wang, J.: Investigating haze-relevant features in a learning framework for image dehazing. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2995–3000 (2014)

8. Diao, Y., Zhang, H., Yadong, W., Chen, M.: Real-time image/video haze removal algorithm with color restoration. *J. Comput. Appl.* **34**(9), 2702–2707 (2014)
9. Meng, G., Wang, Y., Duan, J., Xiang, S., Pan, C.: Efficient image dehazing with boundary constraint and contextual regularization. In: *Proceedings of the IEEE International Conference on Computer Vision*, pp. 617–624 (2013)
10. Wang, Y.K., Fan, C.T.: Single image defogging by multiscale depth fusion. *IEEE Trans. Image Process.* **23**(11), 4826–4837 (2014)
11. Song, W., Deng, B., Zhang, H., Xiao, Q., Peng, S.: An adaptive real-time video defogging method based on context-sensitiveness. In: *2016 IEEE International Conference on Real-Time Computing and Robotics (RCAR)*, pp. 406–410 (2016)
12. Sulami, M., Glatzer, I., Fattal, R., Werman, M.: Automatic recovery of the atmospheric light in hazy images. In: *2014 IEEE International Conference on Computational Photography (ICCP)*, pp. 1–11 (2014)
13. Bahat, Y., Irani, M.: Blind dehazing using internal patch recurrence. In: *2016 IEEE International Conference on Computational Photography (ICCP)*, pp. 1–9 (2016)
14. Zhu, Q., Mai, J., Shao, L.: A fast single image haze removal algorithm using color attenuation prior. *IEEE Trans. Image Process.* **24**(11), 3522–3533 (2015)
15. Guan, J., Lai, R., Xiong, A.: Wavelet deep neural network for stripe noise removal. *IEEE Access.* **7**, 44544–44554 (2019)
16. Guan, J., Lai, R., Xiong, A.: Learning spatiotemporal features for single image stripe noise removal. *IEEE Access.* **7**, 144489–144499 (2019)
17. Ren, W., Liu, S., Zhang, H., Pan, J., Cao, X., Yang, M.-H.: Single image dehazing via multi-scale convolutional neural networks. In: *Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS*, vol. 9906, pp. 154–169. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46475-6_10
18. Cai, B., Xu, X., Jia, K., Qing, C., Tao, D.: Dehazenet: an end-to-end system for single image haze removal. *IEEE Trans. Image Process.* **25**(11), 5187–5198 (2016)
19. Yang, D., Sun, J.: Proximal dehaze-net: a prior learning-based deep network for single image dehazing. In: *Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds.) ECCV 2018. LNCS*, vol. 11211, pp. 729–746. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-01234-2_43
20. Tao, L., Zhu, C., Song, J., Lu, T., Jia, H., Xie, X.: Low-light image enhancement using CNN and bright channel prior. In: *2017 IEEE International Conference on Image Processing (ICIP)*, pp. 3215–3219 (2017)
21. Li, J., Li, G., Fan, H.: Image dehazing using residual-based deep CNN. *IEEE Access* **6**, 26831–26842 (2018)
22. Zhang, Y., et al.: An end-to-end image dehazing method based on deep learning. *J. Phys. Conf. Ser.* **1169**(1), 012046 (2019)
23. Tarel, J.P., Hautiere, N.: Fast visibility restoration from a single color or gray level image. In: *2009 IEEE 12th International Conference on Computer Vision*, pp. 2201–2208 (2009)
24. Chen, C., Do, M.N., Wang, J.: Robust image and video dehazing with visual artifact suppression via gradient residual minimization. In: *Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS*, vol. 9906, pp. 576–591. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46475-6_36