Gated Contiguous Memory U-Net for Single Image Dehazing

Lei Xiang[1,2(✉)], Hang Dong[1], Fei Wang[1], Yu Guo[2], and Kaisheng Ma[3]

[1] National Engineering Laboratory for Vision Information Processing and Application, Xi'an Jiaotong University, Xi'an 710049, Shaanxi, China
xianglei96@stu.xjtu.edu.cn, dhunter1230@gmail.com, wfx@mail.xjtu.edu.cn
[2] School of Software Engineering, Xi'an Jiaotong University, Xi'an 710049, Shaanxi, China
yu.guo@xjtu.edu.cn
[3] Tsinghua University, Beijing 100084, China
kaisheng@mail.tsinghua.edu.cn

Abstract. Single image dehazing is a challenging problem that aims to recover a high-quality haze-free image from a hazy image. In this paper, we propose an U-Net like deep network with contiguous memory residual blocks and gated fusion sub-network module to deal with the single image dehazing problem. The contiguous memory residual block is used to increase the flow of information by feature reusing and a gated fusion sub-network module is used to better combine the features of different levels. We evaluate our proposed method using two public image dehazing benchmarks. The experiments demonstrate that our network can achieve a state-of-the-art performance when compared with other popular methods.

Keywords: Single image dehazing · U-Net like deep network · Contiguous memory resblock · Gated fusion sub-network

1 Introduction

Hazy images are generated by scattering and absorption of the turbid medium (e.g., particles, water-droplets) in the atmosphere. As shown in Fig. 1, hazy images lose contrast and color fidelity, which will bring difficulties to many automated computer vision applications today. Image dehaze aims to recover clean images from hazy input, which has received significant attention as images need to be first enhanced before other high-level vision tasks.

Image quality degradation caused by haze can be roughly mathematically formulated [2,19,21,29] as:

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

where I is the observed hazy image, x is the pixel coordinates, J is the true scene radiance or clear image, $t(x) = e^{-\rho d(x)}$ is the transmission map [6], A is

T. Gedeon et al. (Eds.): ICONIP 2019, LNCS 11954, pp. 117–127, 2019.
https://doi.org/10.1007/978-3-030-36711-4_11

Fig. 1. Examples of realistic hazy images.

the global atmospheric light which indicates the intensity of ambient light. As shown in the formula, only the observed image $I(x)$ is known, recovering the scene radiance $J(x)$ is an ill-posed inverse problem.

The traditional single image dehazing methods [3,7,8,15,18,22,23] have investigated various image prior to estimate transmission maps and atmospheric lights. However, the image prior assumptions may not be valid in all cases. For example, He et al. [8] assumes that value of the dark channel in clear images is close to zero and then use this assumption to estimate the transmission map. However, this assumption may not work well when the scene objects are similar to atmospheric light.

There are also some deep learning-based methods which have been proposed to estimate the transmissions instead of using image priors. However, inaccuracies in the estimation of transmission map lead to low quality dehazed result. To avoid the aforementioned problems, we adopt an end-to-end trainable neural network like [4,14,17,26] which directly restores hazy-free images without estimating the transmission and atmospheric light.

PFFNet [17] is the method that performs best on our datasets in existing methods, but it produces undesired artifacts on the planar region when applied on some images. In order to improve the performance of PFFNet [17], our proposed network adds extra contiguous memory resblocks and gated fusion subnetwork based on PFFNet [17]. The encoder convolves input image into feature maps. The decoder then recovers image details from the encoded feature. We add contiguous memory residual blocks to each scale of the encoder and decoder in order to extract the features on each scale. Inspired by fact that fusing different levels of features are often beneficial for both low-level and high-level tasks [4,16,31], we use a gated fusion sub-network [4,26,33] to determine the importance of different levels and then fuse them based on their corresponding importance weights. To evaluate our proposed network, we test it on two public image dehazing benchmarks. Compared with several popular methods, the experimental results have shown that our proposed network outperforms all the previous methods.

To summarize, the contributions of this paper are three-fold:

- We propose a new trainable U-Net like end-to-end network for image dehazing, which can perform well on both indoor and outdoor images.

- The proposed architecture uses the contiguous memory resblocks to better extract the features on each scale, and a gated fusion sub-network is applied to fuse the features of different levels.
- Experiments show that our method is better than all previous single image dehazing methods. In addition, our proposed network can directly process ultra-high definition realistic haze images with up to 4K resolution at a reasonable speed and memory usage. To validate the importance and necessity of each module, we also provide comprehensive ablation studies.

2 Related Works

As image dehazing is not an easy problem, early methods [11,20,21,27,28,30] usually require multiple images to solve this problem. These methods make the same assumption that there are multiple images from the same scene. However, we can only get one image for a specific scene in most cases. So image priors based methods have been proposed to solve the problem of single image dehazing at early stages. Tan et al. [29] maximized the contrast to enhance the visibility of hazy images. He et al. [8] proposed a dark channel prior (DCP) based method to estimate transmission map, which is based on the observation that the local minimum of dark channels of haze-free images is close to zero. But sometimes these priors that mainly relies on human observations does not work well.

With the successful application of convolutional neural networks in the field of computer vision, recent algorithms [1,24,25,31,32] directly learn transmission map using a deep convolutional neural network. Ren et al. [25] propose a Multi-Scale CNN (MSCNN) to estimate the transmission map. DCPDN [31] and [32] joint transmission map estimation and dehazing using deep networks. However, inaccurate transmission map estimation often harms dehazing results [31]. Instead of estimating the transmission map, the AOD-Net [12] introduces a new intermediate variable to recover the final haze-free image, but it does not bring good results. Similar to the algorithms in [4,14,17,26], we use an end-to-end network to recover the haze-free image directly without estimating the transmission map.

Although the existing single image dehazing algorithm can remove the haze on the image to a certain extent, they all cannot get the expected high-quality hazy-free image. Compared with several existing single image dehazing methods, the PFFNet [17] shows the best performance on RESIDE [13] datasets. The PFFNet [17] can recover most of the outdoor hazy images as expected, but when applied to indoor hazy images, it produces undesired artifacts. In order to solve the problem that the PFFNet [17] does not perform well on indoor hazy images, we have made some improvements based on it to make our method work well both indoors and outdoors.

3 Proposed Method

3.1 Overview

In this part, we will introduce the main modules of our method in detail. The overall network architecture is shown in Fig. 2. The goal of our network is to restore the hazy-free image corresponding to the hazy image. Our network consists of four main parts: (1) an encoder with contiguous memory resblocks; (2) a base feature extraction module to extract structural and contour; (3) an extra gated fusion sub-network; (4) a decoder with contiguous memory resblocks.

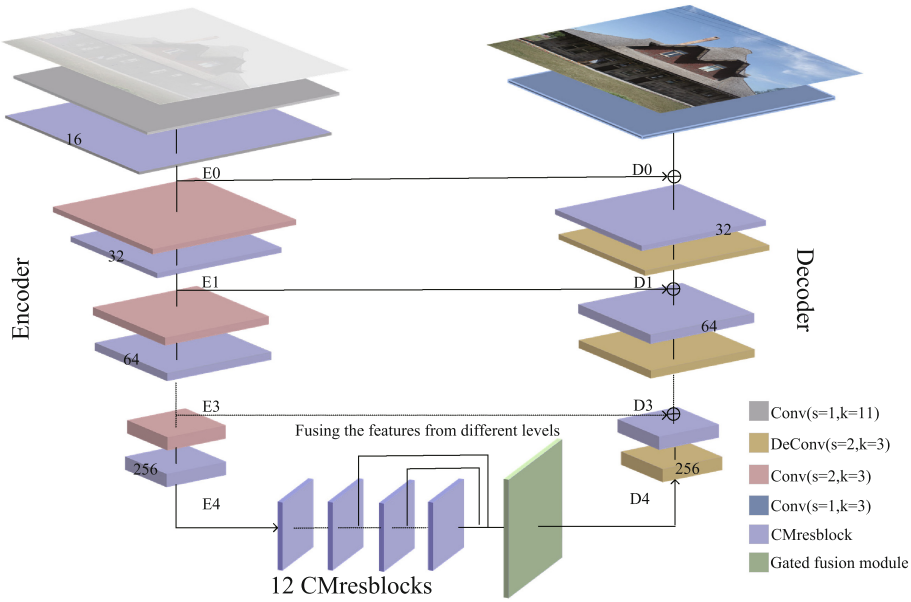

Fig. 2. The architecture of our proposed network for image dehazing.

3.2 Network Structure

The encoder-decoder network architecture which has been used for many tasks [4,5,9,17,26] is also applied in our network to recover the haze-free image. Unlike the encoder-decoder network used in PFFNet [17], we add a CMres behind the convolutional and deconvolution layers of each scale for better features extracting. If the size of the input hazy image is $w \times h \times 3$, then the size of the output of the encoder module is $\frac{1}{16}w \times \frac{1}{16}h \times 256$, where w and h are width and height respectively.

After the encoder module, we use 12 contiguous memory resblocks as the feature extraction module to learn features. Then we use the gated fusion sub-network to fuse the features extracted by the feature extraction module. Furthermore, we employ skip connections between corresponding layers of different level from encoder and decoder.

As shown in Fig. 2, the proposed network contains four strided convolutional layers and four strided deconvolutional layers. The contiguous memory resblock includes two residual blocks with the filter size set as 3×3 and the bottleneck convolutional layer with filter size set as 1×1. The filter size is set as 11×11 pixels in the first convolutional layer in the encoder module and 3×3 in all other convolutional and deconvolutional layers.

3.3 Contiguous Memory Resblock

Our contiguous memory resblock (CMres) consists of two common resblocks with kernel size 3×3 and a convolutional layer with kernel size 1×1. As shown in Fig. 3, contiguous memory mechanism [35] is realized by the operation similar to denseblock which increases the flow of information by feature reusing. In order to reduce memory usage and runtime, the concatenation is only used between each ordinary resblock rather than each convolution layer. Let F_{d-1} and F_d be the input and output of the d-th CMres respectively, the output of d-th CMres can be formulated as

$$F_d = H_d[F_{d-1}, F_{d,1}, F_{d,2}] \tag{2}$$

where H denotes the bottleneck layers. $F_{d,1}$ and $F_{d,2}$ are the feature-maps produced by the ordinary residual blocks 1 and 2 in the d-th scale.

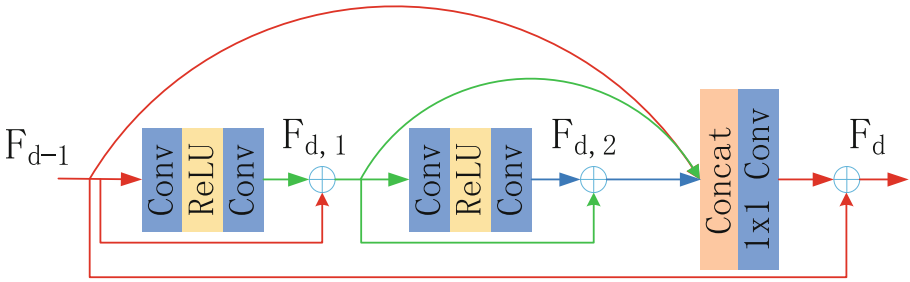

Fig. 3. Contiguous memory resblock architecture.

3.4 Gated Fusion Sub-network

Motivated by this idea [16,31] that fusing the features from different levels usually brings the improvement of experimental results, we adopt a gated fusion sub-network ∂. We feed the feature maps extracted from different levels (F_4, F_8, F_{12})

into the gated fusion sub-network. The output of the gated fusion sub-network are the weights (W_4, W_8, W_{12}) respectively corresponding to each feature level. Once the weights calculated, they multiplied by the three feature maps to get the final feature map.

$$(W_4, W_8, W_{12}) = \partial(F_4, F_8, F_{12}) \tag{3}$$

$$F_o = W_4 * F_4 + W_8 * F_8 + W_{12} * F_{12} \tag{4}$$

Then we feed the combined feature map F_o into the decoder to get the target hazy-free image.

4 Experiments

4.1 Implementation Details

In this section, we will describe the parameter settings of our proposed network in image dehazing. We use Parametric Rectified Linear Unit (PReLU) as the activation function. When training our network we use the mean square error (MSE) as the loss function to constrain the network output and ground truth. We use the ADAM solver [10] with $\beta_1 = 0.9$ and $\beta_2 = 0.999$ to optimize the network. We use an initial learning rate of 0.0001 with a decay rate of 0.1 every 50 epochs and the epoch is set to be 80. And we use a batch size of 16 and increase the training data by random rotation and horizontal flipping. All the training and evaluation processes are conducted on an NVIDIA GeForce GTX 1080Ti graphics card. The source code used in the paper are publicly available at the website: https://github.com/xianglei96/CRF-GFN.

4.2 Dataset

In order to train our network, we need to feed pairs of hazy and hazy-free images to the network. But the question is that it is difficult to get a large number of such pairs of images, so we use synthetic images based on formula 1 to train the network.

Recently, an image dehazing benchmark RESIDE [13] has been proposed, which contains a lot of synthetic training and testing hazy image pairs from depth and stereo datasets. In order to get the training dataset, we randomly select 5005 outdoor hazy image pairs in 35 different haze concentrations and 5000 indoor hazy image pairs in 10 different haze concentrations from RESIDE training sets. We then crop image pairs selected into 256×256 patches with a stride of 128 to obtain hazy-free patches I and hazy patches J.

We evaluate our methods on the RESIDE [13] and HazeRD [34] datasets. The test set of the RESIDE dataset [13] named as SOTS, consists of 500 indoor image pairs and 500 outdoor image pairs. HazeDR [34] consists of 75 hazy images produced by simulation in 15 different scenes and all these images are close to 4K

resolution level. Due to resource constraints, most current dehazed algorithms cannot be tested on HazeDR [34]. For guarantee absolute fairness, we resize all the images in HazeDR [34] to the same size of 512×512.

4.3 Performance Evaluation

We compare our method with several single image dehazing methods, including methods based on hand-crafted features (DCP [8]) and deep convolutional neural networks (AOD [12], DcGAN [14], GFN [26], GCANet [4], and PFFNet [17]). We use the MSE, PSNR and SSIM metrics to evaluate the quality of each restored image. As most existing methods based on deep models are not trained on the datesets we used, we re-train the models of GFN [26], GCANet [4] and PFFNet [17] on our training dataset for fair comparisons.

Table 1. Quantitative evaluations on the benchmark dehazing datasets. Red texts and blue texts indicate the best and second-best performance respectively.

Methods		DCP [8]	AODNet [12]	DcGAN [14]	GFN* [26]	GCANet* [4]	PFFNet* [17]	Ours
SOTS	MSE	0.0178	0.0160	0.0038	0.0059	0.0048	0.0019	0.0013
	PSNR	17.01	18.84	25.46	23.51	26.39	27.76	29.82
	SSIM	0.8213	0.8330	0.9167	0.8935	0.9410	0.9045	0.9592
HazeRD	MSE	0.0377	0.0242	0.0332	0.0271	0.0355	0.0352	0.0183
	PSNR	14.83	16.54	16.20	16.94	15.63	16.11	19.23
	SSIM	0.7805	0.8085	0.8263	0.8173	0.8190	0.8050	0.8417
	CIEDE2000	14.83	13.23	12.02	11.59	14.45	12.89	9.625

As shown in Table 1, the methods based on hand-crafted features [8] or learning based method accompanied by intermediate estimates [12] do not perform well. The learning based methods [4,14,17,26] which direct estimation recover the haze-free image have better performance. And our method outperforms almost all the existing methods on the two datasets.

Fig. 4. Visual comparisons with state-of-the-art methods on SOTS.

Figure 4 shows two examples from the SOTS dataset. We can see that the AOD-Net [12] cannot remove the haze clearly as expected and the DCP [8] may

lead to color distortion. The learning based methods GFN [26] and PFFNet [17] produce undesired artifacts on the planar region. However, the proposed method generates better results on both indoors and outdoors images.

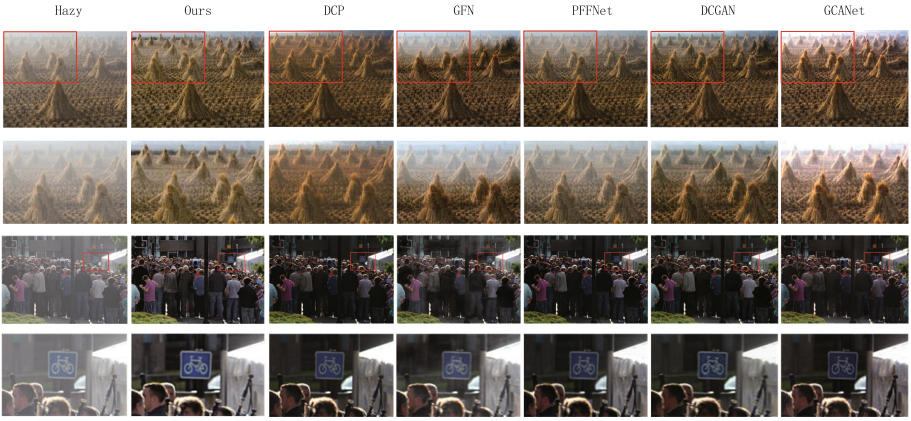

Fig. 5. Visual comparison on the real-world image. The second and fourth lines of images are the enlargement of the red box in the first and third lines of images, respectively. (Color figure online)

To further evaluate the performance of our algorithm, we also test on real images. Figure 5 shows a real hazy image and dehazed results from the state-of-the-art methods. We can see that other methods cannot remove haze clearly or have color distortion. However, the proposed method generates a much better dehazed image.

4.4 Ablation Study

We analyze how the proposed method performs and demonstrates its effectiveness for image dehazing with ablation studies. All the baseline methods mentioned above are trained using the same settings as the proposed algorithm for fair comparisons.

As shown in Fig. 6(a), replacing CMresblocks with denseblocks or resblocks will result in the final result getting worse. So we finally use a structure between denseblocks and resblocks which we call as CMresblocks.

For choosing the appropriate amount of CMresblocks for the base feature extraction module, we did some additional experiments. We can see from Fig. 6(b) that 6CMres and 9CMres are not good enough, and 12CMres can achieve almost the same effect as 15CMres under more resource-saving conditions.

In order to demonstrate the improvements obtained for each module introduced in the proposed network, we conducted an ablation study involving the

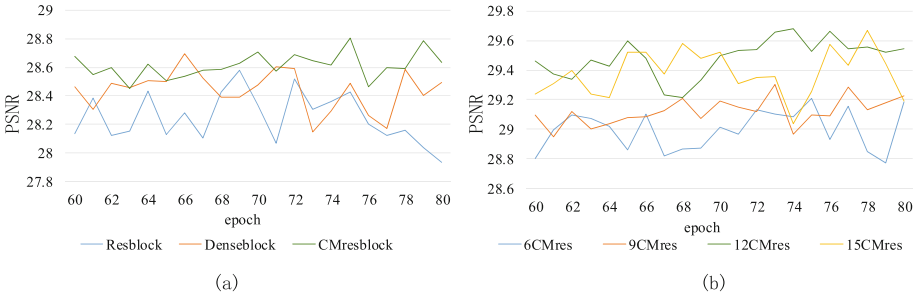

Fig. 6. (a): The testing performance comparisons on SOTS of the network using different blocks. (b): The testing performance comparisons on SOTS in different CMresblock size of the base feature extraction module.

Table 2. Analysis on key components in the proposed network. All the models are trained on the RESIDE dataset with the same hyper-parameters.

Modifications	Baseline	Model-1	Model-2	Model-3
resblock		✓		
CMresblock			✓	✓
gate fusion				✓
PSNR	27.76	28.58	29.68	29.82

following four experiments as shown in Table 2: (1) using ordinary resblocks without contiguous memory mechanism and omitting the gated fusion sub-network (Model-1), (2) using contiguous memory resblocks but omitting the gated fusion sub-network (Model-2), (3) the proposed network with contiguous memory resblocks and the gated fusion sub-network (Model-3).

As shown in Table 2, adding resblocks to Baseline (PFFNet [17]) has a nearly one point increase in PSNR. Based on Model-1, when the original resblocks replaced by CMresblocks, there will be some improvement in the result. Comparisons of Model-2 and Model-3 show that the gated fusion sub-network fusing the feature from different levels can also bring about an improvement in results.

5 Conclusion

In this paper, we propose an U-Net like deep network for single image dehazing which accompanied by contiguous memory residual blocks and gated fusion sub-network. The ablation study validates that the contiguous memory resblocks and the gated fusion sub-network module for feature level fusion can boost the performance of the proposed network. In addition, our dehazing network with U-Net like encoder-decoder architecture, has efficient memory usage and can directly recover the images close to 4K resolution level. Because it is difficult to get real-world pairs of hazy and hazy-free images, we will consider unsupervised learning in the future.

Acknowledgments. This work was supported in part by the National Major Science and Technology Projects of China grant under number 2018ZX01008103, National Natural Science Foundation of China (61603291), Natural Science Basic Research Plan in Shaanxi Province of China (Program No. 2018JM6057) and the Fundamental Research Funds for the Central Universities.

References

1. 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** (2016). https://doi.org/10.1109/TIP.2016.2598681
2. Cantor, A.: Optics of the atmosphere-scattering by molecules and particles. IEEE J. Quantum Electron. **14**(9), 698–699 (1978)
3. 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
4. Chen, D., et al.: Gated context aggregation network for image dehazing and deraining. arXiv preprint arXiv:1811.08747 (2018)
5. Fan, Q., Yang, J., Hua, G., Chen, B., Wipf, D.: A generic deep architecture for single image reflection removal and image smoothing. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 3238–3247 (2017)
6. Fattal, R.: Single image dehazing. ACM Trans. Graph. **27**(3), 1–9 (2008)
7. Hautiàre, N., Tarel, J.P., Aubert, D.: Towards fog-free in-vehicle vision systems through contrast restoration. In: IEEE Conference on Computer Vision & Pattern Recognition (2007)
8. He, K., Jian, S., Tang, X.: Single image haze removal using dark channel prior. In: IEEE Conference on Computer Vision & Pattern Recognition (2009)
9. Johnson, J., Alahi, A., Fei-Fei, L.: Perceptual losses for real-time style transfer and super-resolution. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9906, pp. 694–711. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46475-6_43
10. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
11. Kopf, J., et al.: Deep photo:model-based photograph enhancement and viewing. ACM Trans. Graph. **27**(5), 1–10 (2008)
12. Li, B., Peng, X., Wang, Z., Xu, J., Dan, F.: AOD-Net: All-in-One dehazing network. In: IEEE International Conference on Computer Vision (2017)
13. Li, B., et al.: Benchmarking single image dehazing and beyond. IEEE Trans. Image Process. 1 (2018). https://doi.org/10.1109/TIP.2018.2867951
14. Li, R., Pan, J., Li, Z., Tang, J.: Single image dehazing via conditional generative adversarial network, pp. 8202–8211 (2018). https://doi.org/10.1109/CVPR.2018.00856
15. Li, Z., Ping, T., Tan, R.T., Zou, D., Zhou, S.Z., Cheong, L.F.: Simultaneous video defogging and stereo reconstruction. In: Computer Vision & Pattern Recognition (2015)
16. Lin, T.Y., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S.: Feature pyramid networks for object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2117–2125 (2017)

17. Mei, K., Jiang, A., Li, J., Wang, M.: Progressive feature fusion network for realistic image dehazing. CoRR abs/1810.02283 (2018). http://arxiv.org/abs/1810.02283
18. Meng, G., Ying, W., Duan, J., Xiang, S., Pan, C.: Efficient image dehazing with boundary constraint and contextual regularization. In: IEEE International Conference on Computer Vision (2014)
19. Narasimhan, S.G., Nayar, S.K.: Vision and the atmosphere. Int. J. Comput. Vis. **48**(3), 233–254 (2002)
20. Narasimhan, S.G., Nayar, S.K.: Contrast restoration of weather degraded images. IEEE Trans. Pattern Anal. Mach. Intell. **25**(6), 713–724 (2003)
21. Narasimhan, S.G., Nayar, S.K.: Chromatic framework for vision in bad weather. In: IEEE Conference on Computer Vision & Pattern Recognition (2002)
22. Pei, S.C., Lee, T.Y.: Nighttime haze removal using color transfer pre-processing and dark channel prior. In: IEEE International Conference on Image Processing (2012)
23. Qingsong, Z., Jiaming, M., Ling, S.: A fast single image haze removal algorithm using color attenuation prior. IEEE Trans. Image Process. **24**(11), 3522–3533 (2015)
24. Ren, W., Cao, X.: Deep video dehazing. In: Zeng, B., Huang, Q., El Saddik, A., Li, H., Jiang, S., Fan, X. (eds.) PCM 2017. LNCS, vol. 10735, pp. 14–24. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-77380-3_2
25. 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
26. Ren, W., et al.: Gated fusion network for single image dehazing. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3253–3261 (2018)
27. Schechner, Y.Y., Narasimhan, S.G., Nayar, S.K.: Instant dehazing of images using polarization. In: Proceedings IEEE Conference Computer Vision & Pattern Recognition, vol. 1, p. 325 (2001)
28. Shwartz, S., Namer, E., Schechner, Y.Y.: Blind haze separation. In: IEEE Computer Society Conference on Computer Vision & Pattern Recognition (2006)
29. Tan, R.T.: Visibility in bad weather from a single image. In: IEEE Conference on Computer Vision & Pattern Recognition (2008)
30. Treibitz, T., Schechner, Y.Y.: Polarization: Beneficial for visibility enhancement? In: IEEE Conference on Computer Vision & Pattern Recognition (2009)
31. Zhang, H., Patel, V.M.: Densely connected pyramid dehazing network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3194–3203 (2018)
32. Zhang, H., Sindagi, V., Patel, V.M.: Joint transmission map estimation and dehazing using deep networks. arXiv preprint arXiv:1708.00581 (2017)
33. Zhang, X., Dong, H., Hu, Z., Lai, W., Wang, F., Yang, M.: Gated fusion network for joint image deblurring and super-resolution. CoRR abs/1807.10806 (2018). http://arxiv.org/abs/1807.10806
34. Zhang, Y., Li, D., Sharma, G.: HazeRD: an outdoor scene dataset and benchmark for single image dehazing. In: IEEE International Conference on Image Processing (2018)
35. Zhang, Y., Tian, Y., Kong, Y., Zhong, B., Fu, Y.: Residual dense network for image super-resolution. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2472–2481 (2018)