ROBUST HAZE REMOVAL VIA JOINT DEEP TRANSMISSION AND SCENE PROPAGATION

Risheng Liu[*⋆†] *Shiqi Li*[⋆†] *Long Ma*[⋆†] *Xin Fan*[⋆†] *Haojie Li*[⋆†] *Zhongxuan Luo*[⋆†]

[⋆]DUT-RU International School of Information Science & Engineering, Dalian University of Technology
[†]Key Laboratory for Ubiquitous Network and Service Software of Liaoning Province

ABSTRACT

Haze is one of the most important factors which reduce the outdoor image quality. Existing approaches often aim to design their models based on principles of hazes. However, even with exactly modeled haze distribution, it is still a challenging task due to factors in real scenario, such as noises, halos and artifacts. To address limitations of existing approaches for real-world hazy removal problem, this paper proposes a novel framework to incorporate deep residual architectures into a propagation scheme to jointly estimate transmission and clean scene. We evaluate the proposed framework on both widely used benchmarks and real-world low-quality hazy images. Extensive experimental results demonstrate that our method performs favorably against approaches designed only based on haze cues and achieves the state-of-the-art results, compared with both conventional shallow models and deep dehzaing networks.

***Index Terms*—** Robust haze removal, transmission map estimation, deep residual learning, image propagation

1. INTRODUCTION

Haze is a common atmospheric phenomenon which causes visibility limited and reduce the quality of outdoor images. The reason of such phenomenon is that the light reflecting from the object is scattered by haze and only a part of light can arrive at lens finally. Since outdoor images are very important in computer vision, hazy images removal is highly desired. Haze images removal is a challenging ill-posed problem, because hazy images have unknown depth. A variety of methods for hazy images removal are proposed based on multiple information of images and depth. Most methods have been proposed via atmospheric scattering model which is first proposed by McCartney *et al.* [1] and further developed by Narasimhan *et al.* [2] and Nayar *et al.* [3]. In general, scene transmission estimation by the depth of images and surrounding scattered light [4, 5, 6] is used to build atmospheric scattering model.

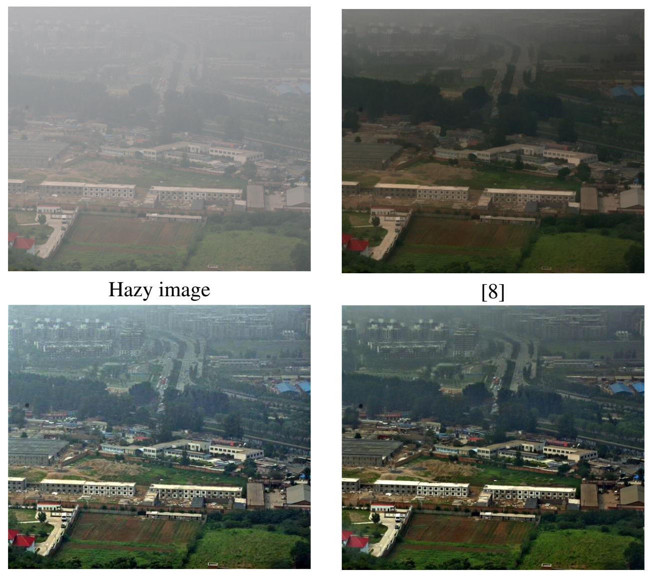


Fig. 1. Visual comparisons of our method with state-of-the-art robust dehazing approaches (e.g., [7, 8]) on a real low-quality hazy image. The result of [7] is over smooth and the method in [8] may hard to remove the artifacts. In contrast, the result of our method is more clear and with less artifacts.

Since accurate transmission plays very important role for recovering the clear image, a local contrast maximization method [9] is proposed based on the assumption that the local contrast of haze-free images is much higher than the hazy images in Markov random field framework. However, the results of this method are often over-saturated. The work in [10] can remove most hazes from corrupted observations. But it is hard to handle the dense-haze situations as it always assumes that the shading and transmission are locally and statistically uncorrelated. Dark channel prior (DCP) is a classical method [4] based on the assumption that pixels at the local regions (expect for the sky) in most outdoor scenes are very low at least one color channel. However, DCP cannot solve such condition that images do not have shadow and the pixels values of the object are similar to the airlight. Berman

This work is partially supported by the National Natural Science Foundation of China (Nos. 61672125, 61733002, 61572096, 61772108 and 61632019), and the Fundamental Research Funds for the Central Universities. Dr. Liu is also a visiting researcher with Shenzhen Key Laboratory of Media Security, Shenzhen University, Shenzhen 518060.

*Corresponding author. E-mail: rsliu@dlut.edu.cn

et al. [11] defines the concept of haze-line which is the tight clusters of a few hundred distinct colors in RGB space, this method can hardly detect the haze line when the airlight is significantly brighter than the scene. To address limitations in DCP, the work in [7] builds an optimization problem with contextual regularization to estimate transmission and it helps to further reduce the ambiguity between color and depth.

With the developments in deep learning, convolutional neural networks (CNNs) based methods [12, 13, 14, 15] have been proposed in recent years. Different from traditional shallow approaches, which are designed based on different haze cues, CNN based methods often discord the principles of atmospheric scattering model. The methods [16, 17] also directly establish the correlations between transmission maps and clear images. Moreover, these methods are also sensitive to noises and low-quality images (see Fig. 1).

In summary, most existing shallow methods may hard to handle complex real scenario, while the performance of deep networks may heavily rely on the scale and quality of training data. To address limitations in these approaches, we develop a novel framework to incorporate deep residual architectures into a propagation scheme to simultaneously estimating transmission maps and recovering the clean images. We will demonstrate that the proposed joint propagation can provide us an efficient and robust haze removal method, especially for real-world scenario.

2. JOINT DEEP TRANSMISSION AND SCENE PROPAGATION

In this section, we first build the basic energy model about haze removal with the implicit prior term based on atmospheric scattering model. Then we solve the energy model by half quadratic optimization to obtain the iterative scheme. Next, we exploit joint deep transmission and scene propagation to solve the subproblems in the iterative scheme.

2.1. The Fundamental Haze Removal Model

Based on atmospheric scattering model, the general physical haze removal model can be described as follows:

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

where $\mathbf{I}$ is the observed hazy image, $\mathbf{J}$ is the latent clear image, $\mathbf{A}$ is the global atmospheric light, and t is the scene transmission describing the portion of the light that is not scattered and reaches the camera. The goal of haze removal is to solve three variables including $\mathbf{J}$, $\mathbf{A}$ and t.

Inspired by the prior regularization ideas in existing image processing formulations, we reformulate Eq. (1) as using maximum a posteriori framework with abstract prior terms and then consider the following energy minimization model:

$$\min_{\mathbf{J},t} \mathcal{F}(\mathbf{J}(\mathbf{x}), t(\mathbf{x})) + \Phi(\mathbf{J}(\mathbf{x})) + \Psi(t(\mathbf{x})), \tag{2}$$

where $\mathcal{F}(\mathbf{J}, t) = \frac{1}{2}\|\mathbf{I} - t \odot \mathbf{J} - (1-t) \odot \mathbf{A}\|_2^2$ is the fidelity term, $\odot$ represents the matrix dot product. We also introduce $\Phi(\cdot)$ and $\Psi(\cdot)$ to represent the regularization term of the haze removal image $\mathbf{J}$ and the regularization term about the transmission t, respectively. Till now, the energy model on the task of single image haze removal has been built. In the following, we will introduce a novel framework to incorporate deep network architectures to optimize Eq. (2).

2.2. Optimization with Deep Network Architecture

In general, based on half quadratic optimization [18] and the alternating direction type in [19], we set the auxiliary variables as $\tilde{\mathbf{J}} = \mathbf{J}$ and $\tilde{t} = t$. As thus, we can split these two regularization terms. We can solve Eq. (2) via half-quadratic optimization to obtain the original iterative scheme as follows:

$$\begin{aligned}\{\mathbf{J}^{k+1}, t^{k+1}\} = \arg\min_{\mathbf{J},t} \frac{1}{2}\|(\mathbf{J} - \mathbf{A}) \odot t + (\mathbf{A} - \mathbf{I})\|_2^2 \\ + \frac{\alpha_k}{2}\|t - \tilde{t}^k\|_2^2 + \frac{\beta_k}{2}\|\mathbf{J} - \tilde{\mathbf{J}}^k\|_2^2,\end{aligned} \tag{3}$$

$$\tilde{\mathbf{J}}^{k+1} = \arg\min_{\tilde{\mathbf{J}}} \frac{\beta_k}{2}\|\mathbf{J}^{k+1} - \tilde{\mathbf{J}}\|_2^2 + \Phi(\tilde{\mathbf{J}}), \tag{4}$$

$$\tilde{t}^{k+1} = \arg\min_{\tilde{t}} \frac{\alpha_k}{2}\|t^{k+1} - \tilde{t}\|_2^2 + \Psi(\tilde{t}). \tag{5}$$

where α, β are penalty parameters and should keep the tendency of increasing in the iteration process theoretically.

Because of the Eq. (3) is differentiable, we can get a closed-form solution. The specific formulations as

$$t^{k+1} = \frac{[(\mathbf{I} - \mathbf{A}) \odot (\mathbf{J}^k - \mathbf{A}) + \alpha_k \tilde{\mathbf{t}}^k]}{[(\mathbf{J}^k - \mathbf{A}) \odot (\mathbf{J}^k - \mathbf{A}) + \alpha_k]}. \tag{6}$$

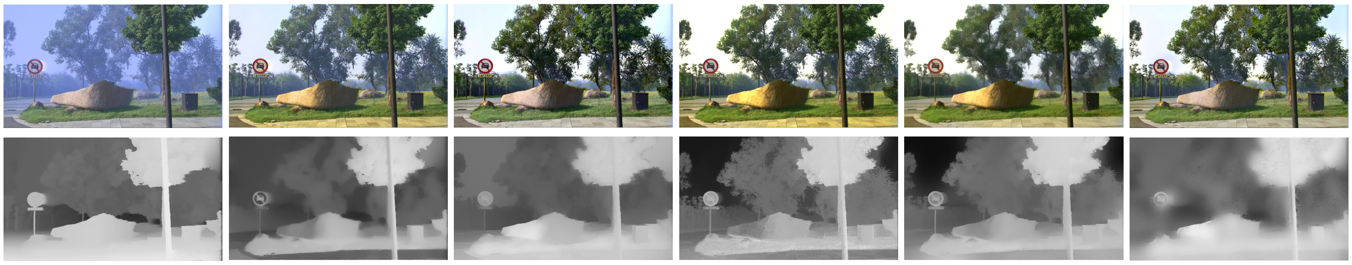
$$\mathbf{J}^{k+1} = \frac{\beta_k \tilde{J}^k + \mathbf{A} \odot t^{k+1} \odot t^{k+1} + (\mathbf{I} - \mathbf{A}) \odot t^{k+1}}{t^{k+1} \odot t^{k+1} + \beta_k}. \tag{7}$$

Due to the fact that $\Phi(\tilde{\mathbf{J}})$ and $\Psi(\tilde{t})$ are coupled, we cannot directly obtain our solution. Therefore, we employ a plug-and-play mechanism to solve these two subproblems by jointly performing deep transmission and scene propagation. We set $\mathcal{N}_{\mathbf{J}}(\mathbf{J}^{k+1}; \Theta_{\mathbf{J}})$ denoting the network for latent image estimation and $\Theta_{\mathbf{J}}$ denoting parameters appeared in $\mathcal{N}_{\mathbf{J}}$. Similarly, we set the network for transmission estimation as $\mathcal{N}_t(t^{k+1}; \Theta_t)$, Θ_t denotes parameters appeared in $\mathcal{N}_t$. Then Eq. (4) and Eq. (5) can be rewritten as

$$\tilde{\mathbf{J}}^{k+1} = \mathbf{J}^{k+1} - \mathcal{N}_{\mathbf{J}}(\mathbf{J}^{k+1}; \Theta_{\mathbf{J}}). \tag{8}$$

$$\tilde{t}^{k+1} = t^{k+1} - \mathcal{N}_t(t^{k+1}; \Theta_t). \tag{9}$$

In fact, it is evident that the designed networks resemble to give a descent direction by observing Eq. (8) and Eq. (9). And it can further explain the efficiency of the networks in the theoretical sense. Now, we can summarize the final iterative scheme to original images as shown in Alg. 1 from above all.



SSIM / L_1 Error	0.946 / 0.054	0.970 / 0.031	0.968 / 0.031	0.897 / 0.039	**0.972 / 0.027**
Haze	[7]	[11]	[4]	[8]	Ours

Fig. 2. Haze removal results on the *Road1* image. **Top row:** Dehazing results. **Bottom row:** The estimated transmission maps. The scores of SSIM and L_1 Error are also reported accordingly.

Algorithm 1 Joint Deep Transmission and Scene Propagation

Input: $\mathbf{I}$, $\mathbf{J}^0$, t^0, α_0, β_0, γ, η, $k_{max} > 1$

for $k = 1, \cdots, k_{max}$ **do**
  Update t by Eq. (6).
  Update $\mathbf{J}$ by Eq. (7).
  $\tilde{\mathbf{J}}^{k+1} = \mathbf{J}^{k+1} - \mathcal{N}_{\mathbf{J}}(\mathbf{J}^{k+1}; \Theta_{\mathbf{J}})$.
  $\tilde{t}^{k+1} = t^{k+1} - \mathcal{N}_t(t^{k+1}; \Theta_t)$.
  $\alpha_{k+1} = \gamma\alpha_k$.
  $\beta_{k+1} = \eta\beta_k$.
end for

Output: $\mathbf{J}^*, t^*$

3. EXPERIMENTAL RESULTS

In order to make the verification of the proposed method better, we first conduct the experiments on the widely used Fattal's dataset [20] and adopt the method of [6] to estimate the global atmospheric light. We evaluate our method via experiments on distinct datasets. The first section reports the network architecture and training data. The second experiment reports the result of different methods in synthetic hazy images without noises generated on Fattal's dataset [20]. Then we conduct the third experiment on the performance of different methods in synthetic hazy images with 10% Gaussian noises. Finally, we test the visual quality of different methods in high and low quality images.

3.1. Network Architecture and Training data

In the designed method, the architecture of the network really matters. In this paper, we adopt the same basic architecture for different tasks that consists of 7 dilated convolution layers, 7 corresponding ReLu nonlinearity layers and 5 batch normalization layers following the work in [21].

In regard to the training data of the latent images estimation network, we collect 800 images combining 400 images from Berkeley segmentation dataset [22] and 400 images from ImageNet database [23]. In addition, we add 12% Gaussian noises as the inputs of network. As for training transmission estimation network, we randomly collect 400 clear images with the size of 624×464 and corresponding depth maps from NYU Depth dataset [24]. Specifically, we first generate hazy images by real transmissions based on atmospheric scattering model. Then we generate the input transmissions by [7] and add 10% Gaussian noises. In order to train the suitable networks, the learning rate decreases from $1e-1$ to $1e-4$ for the 50 epochs based on these training data during the training process.

3.2. Synthetic Hazy Images

We first compared the performance of our method with other state-of-the-art dehazing methods [4, 8, 7, 11] on the widely used Fattal's dataset [20]. Averaged quantitative scores, such as PSNR, SSIM and L_1 error, are reported in Table 1. We observed that our method achieved better results than other dehazing methods. Moreover, we illustrate visual comparisons in Fig. 2. It also can be seen that our method can successfully wipe out the haze in the distance and thus our result has the better visual expression in hue.

Table 1. Averaged quantitative results on Fattal's dataset without noises.

	[7]	[8]	[11]	[4]	Ours
PSNR	26.13	26.47	26.09	27.11	**27.32**
SSIM	0.951	0.926	0.953	0.957	**0.959**
L_1 Error	0.040	0.037	0.041	0.034	**0.033**

Table 2. Averaged quantitative results on Fattal's dataset with 10% Gaussian noises.

	[7]	[8]	[11]	[4]	Ours
PSNR	23.24	25.24	23.60	24.51	**25.81**
SSIM	0.699	0.755	0.677	0.703	**0.820**
L_1 Error	0.055	0.043	0.053	0.047	**0.040**

Fig. 3. Dehazing results on the *mansion* image (with 10% Gaussian noises). **Top row:** Dehazing results. **Bottom row:** The estimated transmission maps. The scores of SSIM and L_1 Error are also reported accordingly.

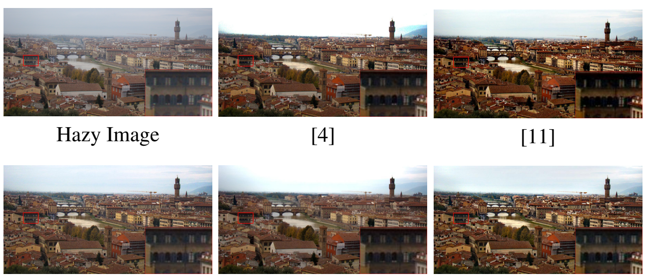


Fig. 4. Visual comparison on real hazy image.

Fig. 5. Dehazing results on low-quality images. **Top row:** Hazy image with artifacts in real scenario. **Bottom row:** Hazy image in bad weather condition.

3.3. Hazy Images with Noises

We generate low-quality hazy images by adding 10% Gaussian noises on Fattal's dataset [20] and then compare our methods with state-of-the-art approaches on this new dataset.

Quantitative evaluation results are shown in Table 2 and the result by our method has the outstanding performance in contrast. Fig. 3 shows the dehazing results of different methods. It is apparent that most of the noises have been eliminated in the result of our method. In addition, our result has the highest SSIM and the lowest L_1 error, we can declare that our method can recover the nature of images utmostly.

3.4. Real Low-quality Hazy Images

In general, the experiments conducted on real images are more meaningful to evaluate the performance of the proposed algorithm. In this part, we consider three types of real images including high-quality hazy images, hazy images with many artifacts and hazy images in bad weather condition.

Fig. 4 shows the comparative results about high-quality haze images. As shown in the zoomed region, it is obvious that our method generates the more clean image with fine texture. The dehazing results about lower quality images as shown in Fig. 5. The top row of the Fig. 5 shows the results about real hazy image with many artifacts, and the result by our method keeps the plain details of the background by contrast. The results in regard to hazy image hampered by severe weather are shown in the bottom row of the Fig. 5. It is prominent that our method recovers more luminous results than all the other compared methods.

4. CONCLUSION

In this paper, we proposed a novel method to recover clean images from inferior hazy images. We first built haze removal energy model via atmospheric scattering model and implicit prior term. Then we obtained the iterative scheme by solving the energy model using half quadratic optimization. Next, we exploited joint deep transmission and scene propagation to solve the subproblems in the iterative scheme. Such a design can reduce the number of iterations of the whole iterative. Results showed that our method had the outperformance than other state-of-the-art methods on different typological datasets including synthetic dataset with and without noise. And we further conducted the experiments in different real scenarios and our method had the best visual quality.

5. REFERENCES

[1] Earl J McCartney, "Optics of the atmosphere: scattering by molecules and particles," in *New York, John Wiley and Sons, Inc., 1976. 421 p.*, 1976.

[2] Srinivasa G. Narasimhan and Shree K. Nayar, "Contrast restoration of weather degraded images," *IEEE TPAMI*, vol. 25, no. 6, pp. 713–724, 2003.

[3] Shree K. Nayar and Srinivasa G. Narasimhan, "Vision in bad weather," in *ICCV*, 1999, pp. 820–827.

[4] Kaiming He, Jian Sun, and Xiaoou Tang, "Single image haze removal using dark channel prior," *IEEE TPAMI*, vol. 33, no. 12, pp. 2341–2353, 2011.

[5] Dana Berman, Tali Treibitz, and Shai Avidan, "Air-light estimation using haze-lines," in *ICCP*, 2017, pp. 1–9.

[6] Matan Sulami, Itamar Glatzer, Raanan Fattal, and Mike Werman, "Automatic recovery of the atmospheric light in hazy images," in *ICCP*, 2014, pp. 1–11.

[7] Gaofeng Meng, Ying Wang, Jiangyong Duan, Shiming Xiang, and Chunhong Pan, "Efficient image dehazing with boundary constraint and contextual regularization," in *ICCV*, 2013, pp. 617–624.

[8] Chen Chen, Minh N Do, and Jue Wang, "Robust image and video dehazing with visual artifact suppression via gradient residual minimization," in *ECCV*, 2016, pp. 576–591.

[9] Robby T. Tan, "Visibility in bad weather from a single image," in *CVPR*, 2008, pp. 1–8.

[10] Raanan Fattal, "Single image dehazing," *ACM ToG*, vol. 27, no. 3, pp. 72, 2008.

[11] Dana Berman, Tali Treibitz, and Shai Avidan, "Non-local image dehazing," in *CVPR*, 2016, pp. 1674–1682.

[12] Wenqi Ren, Si Liu, Hua Zhang, Jinshan Pan, Xiaochun Cao, and Ming Hsuan Yang, "Single image dehazing via multi-scale convolutional neural networks," in *ECCV*, 2016, pp. 154–169.

[13] Risheng Liu, Guangyu Zhong, Junjie Cao, Zhouchen Lin, Shiguang Shan, and Zhongxuan Luo, "Learning to diffuse: A new perspective to design pdes for visual analysis," *IEEE TPAMI*, vol. 38, no. 12, pp. 2457–2471, 2016.

[14] Xin Fan, Yi Wang, Xianxuan Tang, Renjie Gao, and Zhongxuan Luo, "Two-layer gaussian process regression with example selection for image dehazing," *IEEE TCSVT*, vol. PP, no. 99, pp. 1–1, 2016.

[15] Risheng Liu, Xin Fan, Minjun Hou, Zhiying Jiang, Zhongxuan Luo, and Lei Zhang, "Learning aggregated transmission propagation networks for haze removal and beyond," *arXiv preprint arXiv:1711.06787*, 2017.

[16] Bolun Cai, Xiangmin Xu, Kui Jia, Chunmei Qing, and Dacheng Tao, "Dehazenet: An end-to-end system for single image haze removal," *IEEE TIP*, vol. 25, no. 11, pp. 5187–5198, 2016.

[17] Hui Yang, Jinshan Pan, Qiong Yan, Wenxiu Sun, Jimmy Ren, and Yu Wing Tai, "Image dehazing using bilinear composition loss function," *arXiv preprint arXiv:1710.00279*, 2017.

[18] D Geman and Chengda Yang, "Nonlinear image recovery with half-quadratic regularization," *IEEE TIP*, vol. 4, no. 7, pp. 932–946, 1995.

[19] Risheng Liu, Xin Fan, Shichao Cheng, Xiangyu Wang, and Zhongxuan Luo, "Proximal alternating direction network: A globally converged deep unrolling framework," in *AAAI*, 2018.

[20] Raanan Fattal, "Dehazing using color-lines," *ACM ToG*, vol. 34, no. 1, pp. 13, 2014.

[21] Kai Zhang, Wangmeng Zuo, Shuhang Gu, and Lei Zhang, "Learning deep cnn denoiser prior for image restoration," in *CVPR*, 2017.

[22] Yunjin Chen and Thomas Pock, "Trainable nonlinear reaction diffusion: A flexible framework for fast and effective image restoration," *IEEE TPAMI*, vol. 39, no. 6, pp. 1256–1272, 2017.

[23] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *CVPR*, 2009, pp. 248–255.

[24] Nathan Silberman, Derek Hoiem, Pushmeet Kohli, and Rob Fergus, "Indoor segmentation and support inference from rgbd images," in *ECCV*, 2012, pp. 746–760.