Learning Deep Priors for Image Dehazing

Yang Liu [1,2] Jinshan Pan[3,*] Jimmy Ren[2] Zhixun Su[1,4]
[1]Dalian University of Technology [2]SenseTime Research [3]Nanjing University of Science and Technology
[4]Guilin University of Electronic Technology
https://lewisyangliu.github.io/projects/LDP

Abstract

Image dehazing is a well-known ill-posed problem, which usually requires some image priors to make the problem well-posed. We propose an effective iteration algorithm with deep CNNs to learn haze-relevant priors for image dehazing. We formulate the image dehazing problem as the minimization of a variational model with favorable data fidelity terms and prior terms to regularize the model. We solve the variational model based on the classical gradient descent method with built-in deep CNNs so that iteration-wise image priors for the atmospheric light, transmission map and clear image can be well estimated. Our method combines the properties of both the physical formation of image dehazing as well as deep learning approaches. We show that it is able to generate clear images as well as accurate atmospheric light and transmission maps. Extensive experimental results demonstrate that the proposed algorithm performs favorably against state-of-the-art methods in both benchmark datasets and real-world images.

1. Introduction

Single image dehazing aims to estimate a haze-free image from a hazy image. It is a classical image processing problem, which has been an active research topic in the vision and graphics communities within the last decade. As numerous real-world tasks (e.g., traffic detection and environmental monitoring) require high-quality images, and the hazy environment usually leads to deprecated images, it is of great interest to develop an effective algorithm to recover haze-free images.

Haze is a complex atmospheric phenomenon. Images with haze may lose color fidelity and visual contrast as a result of light scattering through the haze particles. Mathematically, the hazing process can be simplified by

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

where I is the hazy image, J is the haze-free image, and t is the medium transmission map which describes the relative portion of the light that reaches the camera sensor from scene surfaces without being scattered. While A is the atmospheric light and x denotes pixel coordinate. If the atmospheric light A is homogeneous, the transmission map t can be expressed as $t(x) = e^{-\beta d(x)}$, where β denotes scattering coefficient of the atmospheric light and $d(x)$ is the scene depth. As only I is available, we need to recover A, t, J simultaneously. This problem is highly ill-posed because many different pairs of A, t and J give rise to the same I, e.g., a delta matrix for A, an all-ones matrix for t and a hazy image for J.

(a) Input image (b) Ren et al. [15]
(c) Yang and Sun [24] (d) Ours

Figure 1. A real-world example. The proposed algorithm combines the advantages of the variational model and deep CNNs and can generate much clearer results.

To make image dehazing problem well-posed, most existing algorithms make assumptions on atmospheric light, transmission maps, or clear images [7, 4, 28, 1]. For example, He et al. [7] propose a dark channel prior based on statistical properties of clear images to estimate transmission map. Fattal [4] develops color lines for image dehazing. The non-local image prior has been used in [1]. Although image dehazing methods using such image priors have shown good performance, those priors mentioned above are mainly based on the observations of specific image properties, which do not always reflect inherent properties of natural images.

Recently, image dehazing methods based on deep learning have been proposed [14, 2, 10, 15, 25, 24]. These al-

*Corresponding author: sdluran@gmail.com

gorithms usually perform better than conventional hand-crafted priors based methods by a large margin. One of the main reasons is that deep CNNs can extract informative features from huge amounts of images and often shown strong capability of generalization. However, most of these approaches use deep feed-forward CNNs to directly estimate clear images from hazy images, or predict atmospheric light and transmission maps and then calculate the clear images according to (1). The learning process makes it like a black box and less domain knowledge is involved. In [24], Yang and Sun combine deep learning techniques and the half-quadratic optimization method to solve image dehazing. This method uses deep CNNs to learn the dark channel prior and adopts a conventional method [7] to compute atmospheric light. However, the atmospheric light estimation method by [7] is less effective as pointed by [22], which usually leads to the images with significant color distortion.

Different from these methods, we propose an effective iteration algorithm with deep CNNs to learn iteration-wise priors for atmospheric light, transmission maps and clear images simultaneously. We formulate the image dehazing problem as a as the minimization of a variational model with favorable data fidelity terms and prior terms. The data fidelity terms consistently regularize the forward and backward physical formation of image dehazing problem, while the prior terms learn haze-relevant priors for the atmospheric light, transmission map and clear image. We solve the variational model with a novel gradient descent method based iteration scheme with an improved Gauss-Seidel iteration strategy which converges within several iterations. At each iteration, the iteration-wise priors for the atmospheric light, transmission map and clear image are learned via effective deep CNNs. We show that the proposed algorithm is able to generate clear images as well as accurate atmospheric light and transmission maps.

The contributions of this work are as follows:

- We formulate the image dehazing problem as the minimization of a variational model and introduce favarable data fidelity terms and prior terms to regularize the model.
- We propose an effective gradient descent method based iteration algorithm with improved Gauss-Seidel iteration to solve the variational model, where effective deep CNNs are embedded to learn haze-relevant priors for atmospheric light, transmission maps and clear images simultaneously.
- We analyze the effect of the iteration algorithm and learned priors and show that the proposed algorithm performs favorably against state-of-the-art methods on both benchmark datasets and real-world images.

2. Related Work

In recent years, we have witnessed significant advances in single image dehazing. In this section, we discuss the methods most related to this work within proper contexts.

Image prior-based Methods: Single image dehazing is a classical ill-posed problem, various methods have been proposed to tackle this problem. As the hazing model involves the estimation of atmospheric light, transmission map and haze-free image, statistical priors based methods usually make some assumptions on them. In [3], Fattal develops a refined image formation model for surface shading and scene transmission. The DCP method [7] has shown remarkable effectiveness for haze removal, which proposes the dark channel prior to estimate the transmission map. In [4], color-line prior is proposed to characterize the haze-free image based on the 1D distribution of pixels within small image patches in RGB color space. Similarly, Berman et al. [1] introduce a haze-line prior formed by non-local pixels to describe the hazy image. Zhu et al. [28] propose a fast single image haze removal method based on hand-crafted features. These priors are powerful and show strong effectiveness in helping haze removal. However, they are designed under the observation of specific image properties, which may not reflect inherent properties of natural images, e.g., DCP [7] performs less effective for the scenes where the color is similar to that of the atmospheric light.

Deep Learning-based Methods: Deep learning based methods have been made remarkable progress in image dehazing problem. In [2, 14], end-to-end CNNs are introduced to estimate the transmission map. These two methods perform well but do not consider the atmospheric light or haze-free image. Li et al. [10] propose an all-in-one dehazing network which need to compute an intermediate variable integrating both atmospheric light and transmission map. Ren et al. [15] pre-processes the hazy image to generate multiple inputs, by which color distortions may be introduced.

Recently, several state-of-the-art method [25, 13] introduce dual CNNs structure to directly predict the atmospheric light and transmission map from the hazy image, respectively. Then the haze-free image can be calculated via inversing equation (1). We only take this approach as an initialization step for our iterative optimization algorithm. Different from the blind learning process, the atmospheric light, transmission map as well as the haze-free image will be solved iteratively with favorable prior knowledge in the following iteration units.

The most related work to ours is PDN [24]. As feed-forward deep CNNs may be less effective due to the ill-posed nature [6], PDN [24] combines deep CNNs and half-quadratic optimization method to iteratively learn dark channel prior. However, this method is limited to learn dark channel prior and take a conventional method [7] to calculate atmospheric light. Our method is different from PDN [24] in model formulation, optimization method and prior learning process. We takes a variational model and gradient descent method based optimization algorithm, on which the atmospheric light, transmission map and haze-free im-

age are all learned simultaneously with no limitation to any hand-crafted prior.

3. Variational model and optimization

To better motivate the proposed method, we start from the classical prior-based variational model. Based on the hazing model (1), image dehazing problem can be achieved by solving the following minimization problem:

$$\begin{cases} E(A,t) & = \mathcal{D}(A,t,J) + \lambda\phi(A) + \gamma\psi(t), \\ s.t. \quad J & = \frac{I-A}{t} + A, \end{cases} \tag{2}$$

where $\phi(\cdot)$ and $\psi(\cdot)$ are the priors w.r.t. A and t, respectively; λ and γ are positive weight parameters; $\mathcal{D}(A,t,J)$ denotes the data fidelity term, which is usually defined as

$$\mathcal{D}(A,t,J) = \mathcal{F}(Jt + A(1-t) - I), \tag{3}$$

where $\mathcal{F}$ is taken as an L_2-norm in this paper.

If $\mathcal{D}(A,t,J)$ and $\phi(\cdot)$, $\psi(\cdot)$ are differentiable, problem (2) can be solved via the gradient descent method. At each iteration, we need to solve

$$\begin{cases} A^{n+1} = A^n - \alpha_A(\nabla_A\mathcal{F}^{n+1} + \lambda\nabla_A\phi(A^n)), \\ t^{n+1} = t^n - \alpha_t(\nabla_t\mathcal{F}^{n+1} + \gamma\nabla_t\psi(t^n)), \\ J^{n+1} = \frac{I-A^{n+1}}{t^{n+1}} + A^{n+1}, \end{cases} \tag{4}$$

where α_A and α_t denotes the step size for A and t, respectively; ∇ denotes the gradient operator; $n = 1, 2, 3 \cdots K-1$ denotes the iteration number and K is the maximum iteration number.

Another question is, how to obtain the initial value of A^1 and t^1 for this iteration process? Some state-of-the-art methods [25, 13] take a one-step end-to-end manner to directly obtain A and t from I, then calculating J via (1). As this dual CNNs structure can generate reasonable results, we take this approach as an initialization method for the proposed iteration method (4) which is formulated as:

$$\begin{cases} A^1 = N^1(I), \\ t^1 = N^1(I), \\ J^1 = \frac{I-A^1}{t^1} + A^1, \end{cases} \tag{5}$$

where A^1, t^1 and J^1 are the initial value for (4), while N^1 is a CNN for this initialization process.

As above, image dehazing problem can be achieved by iteratively solve the variational model (2) with the gradient descent based optimization method (4) and the initialization method (5). We further explore a prior term for haze-free image J and a favorable data fidelity term to regularize the variational model in Section 4, where an effective deep CNNs based iterative learning algorithm is proposed for single image dehazing problem.

4. Proposed Algorithm

Prior term on J. The objective function (2) considers prior terms on A and t, and a closed-form solution of J can be directly obtained. However, even when A and t can be approximated well the haze-free image J still lacks of proper regulations due to the highly ill-posed nature of image dehazing problem, e.g., Fattal [4] proposes the color-line prior to model haze-free images. We further introduce a prior term on J to regularize the predicted haze-free image J and reformulate the objective function (2) as:

$$\begin{cases} E(A,t) & = \mathcal{D}(A,t,J) + \lambda\phi(A) + \gamma\psi(t) + \eta\omega(J), \\ s.t. \quad J & = \frac{I-A}{t} + A, \end{cases} \tag{6}$$

where $\omega(\cdot)$ is the prior on J and η is a positive weight parameter. We note that it is not trivial to define the prior term on J in (6). In addition, J will not have a close-form solution as in (4). We will give a iteration method to solve J after we introducing a novel data term for $\mathcal{D}(A,t,J)$.

Data fidelity term $\mathcal{G}$. As only I is known and A, t, J are unknown in the hazing model (1), the data fidelity term $\mathcal{F}$ in (3) helps constrain the predicted A^n, t^n, J^n to satisfy (1) so that $J^n t^n + A^n(1-t^n)$ should as similar as possible to the ground truth I in the iteration process. Motivated by a recently proposed work CycleGAN [27] which introduces a cycle-consistency loss to regularize the forward and backward mapping, we propose another data fidelity term to better regularize A and t. In a backward view of (1), if the ground truth J is known, then $\frac{I-A^n}{t^n} + A^n$ should as similar as possible to the ground truth J. Although the ground truth J is unknown in image dehazing problem, the predicted J^n is a good approximation of J, and therefore it can be used to better predicting A^{n+1} and t^{n+1} in the next iteration. In this way, we propose the data fidelity term $\mathcal{G}$ on A and t:

$$\mathcal{G}(A,t) = \mathcal{G}(\frac{I-A}{t} + A - J), \tag{7}$$

where the superscripts n and $n+1$ are omitted for simplicity. $\mathcal{G}$ is taken as an L_2-norm similar to $\mathcal{F}$. Formally, the data term $\mathcal{D}(A,t,J)$ defined in (3) becomes:

$$\begin{aligned} &\mathcal{D}(A,t,J) = \\ &\quad \mathcal{F}(Jt + A(1-t) - I) + \mathcal{G}(\frac{I-A}{t} + A - J), \end{aligned} \tag{8}$$

Base on the reformulated objective function (6) and data term (8), we iteratively solve J with gradient descent method and reformulate (4) as:

$$\begin{cases} A^{n+1} = A^n - \alpha_A(\nabla_A\mathcal{F}^{n+1} + \nabla_A\mathcal{G}^{n+1} + \lambda\nabla_A\phi(A^n)), \\ t^{n+1} = t^n - \alpha_t(\nabla_t\mathcal{F}^{n+1} + \nabla_t\mathcal{G}^{n+1} + \gamma\nabla_t\psi(t^n)), \\ J^{n+\frac{1}{2}} = \frac{I-A^{n+1}}{t^{n+1}} + A^{n+1}, \\ J^{n+1} = J^{n+\frac{1}{2}} - \alpha_J\eta\nabla_J\omega(J^{n+\frac{1}{2}}), \end{cases} \tag{9}$$

Similar to the classical Gauss-Seidel iteration strategy [5], instead of directly optimizing J^{n+1} from J^n, we make use of the latest A^{n+1} and t^{n+1} on the $n+1$-th iteration to get an intermediate close-form solution $J^{n+\frac{1}{2}}$ with the physical model of the dehazing process to better constrain J^{n+1}. The detailed conduction of above iteration method can be found in the supplementary material on the project homepage.

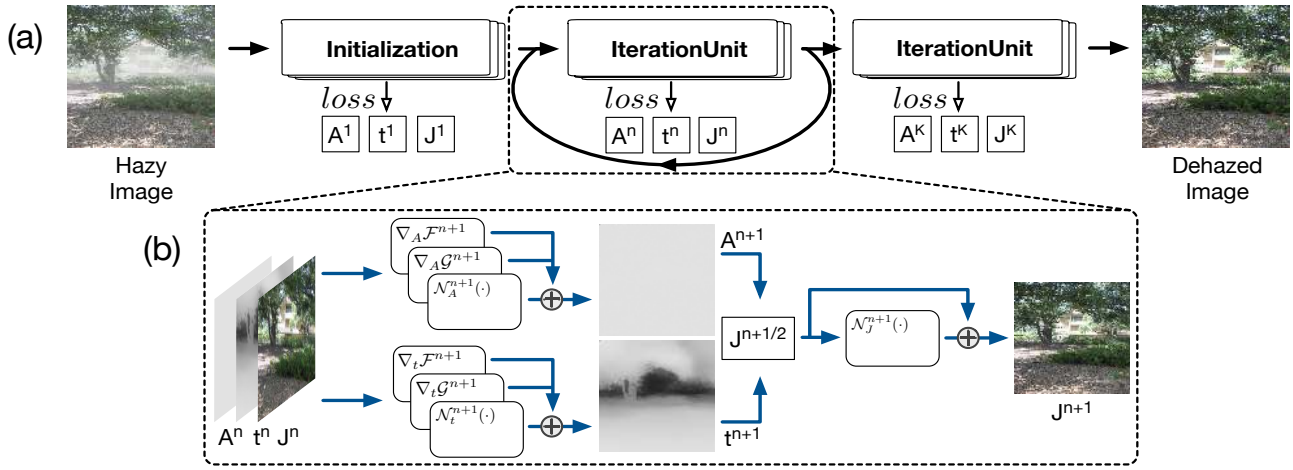


Figure 2. An overview of the proposed method. Our method takes an iterative optimization scheme with an effective neural network-based learning manner. (a) Starting from an initialization value with (5), the proposed algorithm iteratively optimize intermediate solutions A^n, t^n and J^n under the supervision of stage-wise loss functions (13) and (14) until it converges to a good state. (b) The basic structure of the iteration algorithm (9). The learning process to the solutions A^{n+1}, t^{n+1} and J^{n+1} of next step is regularized by the physics-based data fidelities ((3) and (7)) and the learned iteration-wise priors for A^n, t^n and J^n ((10), (11) and (12)).

Table 1. Network parameters. We use SC, LR, I, R and TC denotes strided convolutional layer, LeakyReLU, instance normalization [23], ReLU and transposed convolutional layer, respectively.

Layers	SCLR_1	SCILR_1	SCILR_2	SCILR_3	SCR_1	TCIR_1	TCIR_2	TCIR_3	TCIR_4	TC_1
Filter size	4	4	4	4	4	4	4	4	4	4
Filter numbers	8	16	32	64	64	64	32	16	8	3
Stride	2	2	2	2	2	2	2	2	2	2

Learning prior terms via CNNs. Some related works [20] approximate regularization functions using a series of parameterized mapping functions. As modern CNNs have good capability in modeling informative image features and approximating mapping functions, we develop deep CNNs to learn the image prior terms of A, t and J in a data-driven manner and put the networks in a proper place of the iterative algorithm. Formally, the differentiable prior terms $\lambda\nabla\phi(A^n)$, $\gamma\nabla\phi(t^n)$ and $\eta\nabla\omega(J^n)$ in (9) are learned by:

$$\mathcal{N}_A^{n+1}(A^n) = \lambda\nabla\phi(A^n), \tag{10}$$

$$\mathcal{N}_t^{n+1}(t^n) = \gamma\nabla\phi(t^n), \tag{11}$$

$$\mathcal{N}_J^{n+1}(J^n) = \eta\nabla\omega(J^n), \tag{12}$$

where $\mathcal{N}_A^{n+1}$, $\mathcal{N}_t^{n+1}$ and $\mathcal{N}_J^{n+1}$ are CNNs taking A^n, t^n, and J^n as the input, respectively. Positive weight parameters λ, γ and η are naturally learned together.

Based on above considerations, we can use existing network architectures to define the networks. In this paper, we use the U-Net [16] with an encoder-decoder architecture for $\mathcal{N}_A^n$, $\mathcal{N}_t^n$, and $\mathcal{N}_J^n$. We use Leaky ReLU with negative slope 0.2 for the encoder part and ReLU for the decoder part except for the last layer. Four skip connections are added by concatenating corresponding feature maps to avoid gradient vanishing and to accelerate training. In addition, the instance normalization [23] is also adopted. Table 1 shows the details of the network parameters. The deep networks $\mathcal{N}_A^n$, $\mathcal{N}_t^n$, and $\mathcal{N}_J^n$ share the same network architecture but do not share weights parameters during training and test. As the matrix size of $\mathcal{N}_A^n(A^n)$, $\mathcal{N}_t^n(t^n)$, and $\mathcal{N}_J^n(J^n)$ respectively equals to $\lambda\nabla\phi(A^n)$, $\gamma\nabla\phi(t^n)$ and $\eta\nabla\omega(J^n)$, these networks can be plugged into (9) naturally and trained with modern deep learning techniques jointly.

4.1. Cascaded training

To effectively solve the networks $\mathcal{N}_A^n$, $\mathcal{N}_t^n$ and $\mathcal{N}_J^n$ at each iteration, we train them in a cascaded manner.

Let Θ^n denotes the network parameters of $\mathcal{N}_A^n$, $\mathcal{N}_t^n$ and $\mathcal{N}_J^n$ at iteration n, $\{A_m^n, \overline{A}_m^n; t_m^n, \overline{t}_m^n; J_m^n, \overline{J}_m^n\}$ denotes the training set with M training samples, where $\overline{A}$, $\overline{t}$, $\overline{J}$ denote the ground truth atmospheric light, transmission map, and clear image, respectively. We learn the iteration-dependent model parameters Θ^n from $\{A_m^n, \overline{A}_m^n; t_m^n, \overline{t}_m^n; J_m^n, \overline{J}_m^n\}$ by minimizing the cost function:

$$\mathcal{J}(\Theta^n) = \sum_{m=1}^{M} \ell(\tilde{A}_m^n, \overline{A}_m^n; \tilde{t}_m^n, \overline{t}_m^n; \tilde{J}_m^n, \overline{J}_m^n), \tag{13}$$

where $\tilde{A}_m^n$, $\tilde{t}_m^n$, and $\tilde{J}_m^n$ are the outputs by solving the iteration algorithm (9) at the n-th iteration, corresponding to A^{n+1}, t^{n+1} and J^{n+1} of (9), respectively. Note that the networks $\mathcal{N}_A^n$, $\mathcal{N}_t^n$, and $\mathcal{N}_J^n$ are plugged into (9) via (10), (11) and (12) so that the learning process of iteration algorithm (9) is of a jointly end-to-end training manner. As using L_1 norm as the loss function is able to generate good results [12], we define $\ell(\tilde{A}_m^n, \overline{A}_m^n; \tilde{t}_m^n, \overline{t}_m^n; \tilde{J}_m^n, \overline{J}_m^n)$ as

Algorithm 1 Proposed algorithm.

Require: hazy image I, step size α_A, α_t and α_J, maximum iteration number K.

$A^1 = N_A^1(I)$,
$t^1 = N_t^1(I)$,
$J^1 = \frac{I - A^1}{t^1} + A^1$,
Update $\mathcal{N}_A^1$ and $\mathcal{N}_t^1$ with (13) and (14),
for n = 1, 2, 3 $\cdots$ K-1 **do**
$A^{n+1} = A^n - \alpha_A(\nabla_A \mathcal{F}^{n+1} + \nabla_A \mathcal{G}^{n+1} + \mathcal{N}_A^{n+1}(A^n))$,
$t^{n+1} = t^n - \alpha_t(\nabla_t \mathcal{F}^{n+1} + \nabla_t \mathcal{G}^{n+1} + \mathcal{N}_t^{n+1}(t^n))$,
$J^{n+\frac{1}{2}} = \frac{I - A^{n+1}}{t^{n+1}} + A^{n+1}$,
$J^{n+1} = J^{n+\frac{1}{2}} - \mathcal{N}_J^{n+1}(J^{n+\frac{1}{2}})$,
Update $\mathcal{N}_A^{n+1}$, $\mathcal{N}_t^{n+1}$ and $\mathcal{N}_J^{n+1}$ with (13) and (14).
end for

(a) Input image (b) Iteration #1 (c) Iteration #4 (d) Ground truth

Figure 3. An example result of the iteration algorithm. There still exist haze residues in the initialization step, while the predicted haze-free image of the 4-th iteration is close to the ground truth.

$$\ell(\tilde{A}_m^n, \overline{A}_m^n; \tilde{t}_m^n, \overline{t}_m^n; \tilde{J}_m^n, \overline{J}_m^n) = \|\tilde{A}_m^n - \overline{A}_m^n\|_1 + \|\tilde{t}_m^n - \overline{t}_m^n\|_1 + \|\tilde{J}_m^n - \overline{J}_m^n\|_1, \quad (14)$$

We minimize (13) and discriminatively learn the model parameters Θ^n stage by stage from $n = 1, ..., K$.

Based on above analysis, we summarize the proposed algorithm for image dehazing in Algorithm 1. An overview of the proposed method is presented in Figure 2, where we denote (5) at the 1-st iteration as an initialization method, while the iteration algorithm (9) at the n-th (n = 2, 3 $\cdots$ K) iteration as an iteration unit. Starting from the initialization value, the proposed deep image dehazing algorithm converges within several iterations (four iterations is enough), Figure 3 shows an example result.

5. Experimental Results

We compare our method against state-of-the-art image dehazing methods on the proposed synthetic dataset as well as publicly available benchmark datasets. Due to the comprehensive experiments performed, we only show a small portion of the results in the main paper. Please visit the project homepage for more and larger results.

5.1. Implementation details

In the learning process, we use the ADAM optimizer [9] with parameters $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 10^{-4}$. The minibatch size is set to be 20. The learning rate is initialized as 10^{-4} for the first 10 epochs which is halved every 10 epochs.The parameters are initialized randomly and the network converges well after 50 epochs. We empirically set the maximum iteration number $K = 4$ as a trade-off between accuracy and speed. The step size α_A, α_t, and α_J are experimentally set to be 0.01 for the second iteration and are reduced to 0.001 for the following iterations.

Table 2. Quantitative evaluations for the state-of-the-art dehazing methods on the benchmark datasets in terms of PSNR and SSIM.

PSNR/SSIM	Synthetic	SOTS	HazeRD
DCP [7]	18.04/0.8497	18.22/0.8484	14.64/0.7761
CAP [28]	17.99/0.8031	18.63/0.8101	14.15/0.7434
NLD [1]	17.13/0.7906	17.71/0.8154	14.58/0.8100
MSCNN [14]	19.51/0.8437	18.34/0.8380	15.62/0.8179
DehazeNet [2]	21.37/0.8703	22.18/0.8839	15.30/0.7859
AOD-Net [10]	19.21/0.8455	19.71/0.8663	15.64/0.8014
GFN [15]	23.05/0.8882	22.18/0.8757	13.73/0.6686
DCPDN [25]	20.09/0.8454	18.04/0.8515	15.86/0.7736
PDN [24]	19.09/0.8480	17.33/0.8164	14.48/0.7499
Ours	**24.71/0.8917**	**22.46/0.8844**	**17.51/0.8461**

For the training data, we use the clear images and their corresponding depth images to synthesize the hazy images according to (1), where the clear images and depth images are obtained from the NYU Depth dataset [21] and the Make3D dataset [17, 18, 19], which contain indoor and outdoor images, respectively. To generate better hazy images, we randomly select atmospheric light $A \in [0.7, 1.0]$ and $\rho \in \{0.5, 1.0, 1.5\}$ for each image according to [14]. In addition, the guided image filter [8] is used to fill the holes in the depth images when synthesizing hazy images. We randomly synthesize 7,000 hazy images of size 512×512 pixels for training. To evaluate the proposed algorithm, we generate another 300 images for testing, where the test data and training data do not overlap.

For a RGB image of size $h \times w \times c$, the size of the hazy image I, haze-free image J, transmission map t, and atmospheric light A is $h \times w \times c$, $h \times w \times c$, $h \times w \times 1$, $h \times w \times 1$, respectively. The mathematical operations and CNNs used in the proposed method do not change the dimensions.

5.2. Comparisons with the state-of-the-arts

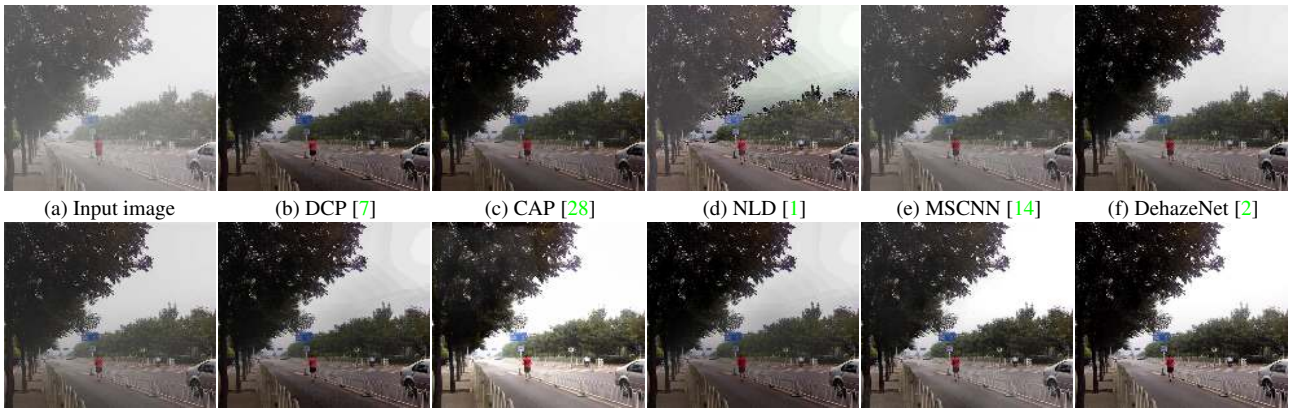
To evaluate the performance of the proposed algorithm, we compared it against state-of-the-art algorithms including DCP [7], CAP [28], NLD [1], MSCNN [14], DehazeNet [2], AOD-Net [10], GFN [15], DCPDN [25] and PDN [24]. In addition to the synthetic test dataset, we employ the benchmark datasets SOTS dataset [11]and HazeRD dataset [26] as well as real-world images to evaluate the performance. SOTS dataset [11] is a large-scale dataset containing lots of challenging examples. HazeRD dataset [26] is an outdoor dataset which simulates different weather conditions. To evaluate the quality of the restored images, we use PSNR and SSIM as the metrics.

Table 2 shows the quantitative evaluations on different benchmark datasets, where the results of the state-of-the-art methods are obtained using the corresponding publicly available codes for fair comparisons. The proposed method performs favorably against the state-of-the-art methods in terms of PSNR and SSIM.

(a) Input image (b) DCP [7] (c) CAP [28] (d) NLD [1] (e) MSCNN [14] (f) DehazeNet [2]

(g) AOD-Net [10] (h) GFN [15] (i) DCPDN [25] (j) PDN [24] (k) Ours (l) Ground truth

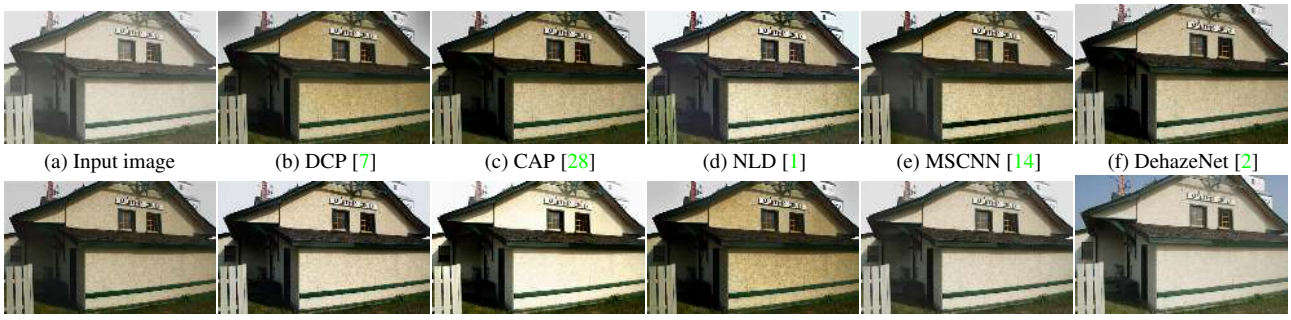
Figure 4. Example results on the synthetic dataset. Our method can remove haze and generate images visually closer to the ground truth.



(a) Input image (b) DCP [7] (c) CAP [28] (d) NLD [1] (e) MSCNN [14] (f) DehazeNet [2]

(g) AOD-Net [10] (h) GFN [15] (i) DCPDN [25] (j) PDN [24] (k) Ours (l) Ground truth

Figure 5. Example results on the SOTS dataset. Our method is able to remove haze and generates a much clearer image.



(a) Input image (b) DCP [7] (c) CAP [28] (d) NLD [1] (e) MSCNN [14] (f) DehazeNet [2]

(g) AOD-Net [10] (h) GFN [15] (i) DCPDN [25] (j) PDN [24] (k) Ours (l) Ground truth

Figure 6. Example results on the HazeRD dataset. Our method generates a clearer image visually closer to the ground truth image.

Figure 4 shows some dehazed results from the synthetic test dataset by the evaluated methods. As the methods [7, 28, 1, 14] usually assume that the atmospheric light is constant and they usually choose the brightest pixels from the hazy images according to the estimated transmission maps, the results by these methods contain significant color distortion. The deep learning based-methods, e.g., DehazeNet [2], AOD-Net [10], GFN [15], DCPDN [25], use end-to-end trainable networks to directly estimate the clear images. Haze residues still exist as shown in Figure 4 (f)-(i). PDN [24] use a deep CNN to learn dark channel prior [7] and develop an efficient algorithm based on the half-quadratic optimization. However, the atmospheric light estimation of this method is based on [7]. Thus, the colors of the generated images look unnatural. In contrast, our method develops deep CNNs to learns the priors for atmospheric light, transmission maps and clear images, which facilitates the hazy removal (Figure 4 (k)). Figures 5 and 6 show dehazed results from the SOTS dataset [11] and the HazeRD dataset [26]. As shown in Figure 5 (b)-(j), the state-of-the-art methods suffer from various degree of ripple-like artifacts or color distortions, while our method generates more natural results. The comparisons in Figure 6 further demonstrate that our method can recover clear im-

(a) Input image (b) DCP [7] (c) NLD [1] (d) MSCNN [14] (e) DehazeNet [2]

(f) AOD-Net [10] (g) GFN [15] (h) DCPDN [25] (i) PDN [24] (j) Ours

Figure 7. Example results on real-world image datasets. Our method removes haze and generates clearer images with vivid colors.

(a) Input image (b) DCP [7] (c) NLD [1] (d) MSCNN [14] (e) DehazeNet [2]

(f) AOD-Net [10] (g) GFN [15] (h) DCPDN [25] (i) PDN [24] (j) Ours

Figure 8. Example results on real-world image datasets. Our method removes haze and generates clearer images with vivid colors.

Table 3. Running time comparisons (in seconds) on images of size 512 x 512 pixels.

Methods	AOD-Net [10]	GFN [15]	PDN [24]	Ours
Platform	Pycaffe	Matcaffe	MatConvNet	PyTorch
Time	0.04	0.08	1.99	0.19

ages which are visually closer to the ground truth image.

Real examples. We further evaluate our method against state-of-the-art algorithms on real images from publicly available datasets. Figures 7 and 8 shows the results. Our algorithm generates clear images with vivid colors, which there are still exist haze residues or color distortion in the images of other methods.

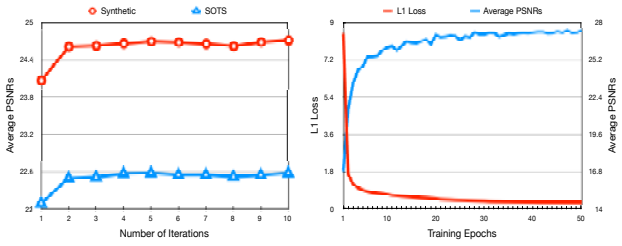
Running time. We evaluate the running time in the synthetic test dataset with randomly selected 10 images of size 512 $\times$ 512 pixels on an Intel Xeon E5-2620 CPU and an NVIDIA TITAN Xp GPU. Table 3 shows that our method takes slightly more running time compared to the feed-forward models [10, 15] due to the iteration strategy but is about 10 times faster than PDN [24].

6. Analysis and Discussions

We have shown that the proposed algorithm is able to generate better results compared to the state-of-the-art algorithms. In this section, we further analyze the effectiveness of the components via ablation studies.

6.1. Effectiveness of the iteration scheme

Our method is composed of the initialization method (5) and several iteration units (9). The proposed method without using the iteration scheme will reduce to a totally end-to-end feed-forward CNN method. To demonstrate the effects of the iteration scheme, we train the network with 10 iterations as presented in Figure 9 (a). The iteration scheme consistently improves the results on both the synthetic test dataset and SOTS dataset, especially from the initialization step to the second iteration. The iteration algorithm converges after 4 iterations and more iterations do not significantly improve the results, so we set the maximum iteration number $K = 4$ considering the trade-off between accuracy and speed. Another natural question is whether the deep CNNs based iteration method converges with deep learning techniques during training. The quantitative evaluation in Figure 9 (b) further shows that the proposed method converges steadily after 50 epochs on the validation dataset.



(a) Effects of iteration number (b) Effects of training epochs

Figure 9. The proposed iteration method converges well w.r.t. the iteration numbers and training epochs w.r.t. L1 loss and average PSNR. The results of (a) and (b) are evaluated on the synthetic test dataset, SOTS dataset and the validation dataset, respectively.

(a) Input image (b) DCP [7] (c) MSCNN [14] (d) DehazeNet [2] (e) DCPDN [25] (f) PDN [24] (g) Ours (h) Ground truth

Figure 10. Example resulting images of estimating atmospheric light A (top row) and transmission map t (bottom row). Our method can generate more accurate atmospheric light and transmission maps closer to the ground truth.

Table 4. Quantitative evaluations against the state-of-the-art methods in estimating A and t on the synthetic dataset in terms of PSNR/SSIM. The proposed algorithm achieves higher PSNR and SSIM value than other methods.

Methods	DCP [7] PSNR/SSIM	MSCNN [14] PSNR/SSIM	DehazeNet [2] PSNR/SSIM	DCPDN [25] PSNR/SSIM	PDN [24] PSNR/SSIM	Ours PSNR/SSIM
A	23.25/0.9907	23.86/0.9913	22.39/0.9899	**26.59/0.9944**	23.23/0.9910	26.34/0.9495
t	16.85/0.7951	16.37/0.8635	18.77/0.8972	14.58/0.8266	17.84/0.8608	**22.88/0.9229**
J	18.04/0.8497	19.51/0.8437	21.37/0.8703	20.09/0.8454	19.09/0.8480	**24.71/0.8917**

Table 5. Ablation study on the data-fidelity terms and prior terms. Compared to the baselines, our method can achiever better results in terms of PSNR and SSIM.

PSNR/SSIM	Synthetic	SOTS	HazeRD
w/o $\mathcal{F}$	24.69/0.8889	22.40/0.8836	17.36/0.8460
w/o $\mathcal{G}$	24.67/**0.8929**	22.35/0.8802	17.42/**0.8469**
w/o $\mathcal{F}$&$\mathcal{G}$	24.60/0.8851	22.32/0.8793	17.31/0.8455
w/o $\mathcal{N}_A$	24.70/0.8869	22.38/0.8806	17.29/0.8432
w/o $\mathcal{N}_t$	24.49/0.8815	22.24/0.8762	17.35/0.8341
w/o $\mathcal{N}_A$&$\mathcal{N}_t$	24.46/0.8841	22.21/0.8762	17.27/0.8397
w/o $\mathcal{N}_J$	24.25/0.8843	22.03/0.8780	17.45/0.8453
w/o $\mathcal{N}_A$&$\mathcal{N}_t$&$\mathcal{N}_J$	24.24/0.8794	21.94/0.8691	17.24/0.8329
Ours	**24.71**/0.8917	**22.46/0.8844**	**17.51**/0.8461

6.2. Effectiveness of the data-fidelity terms

We employ two data-fidelity terms $\mathcal{F}$ and $\mathcal{G}$ to regularize the variational model. $\mathcal{F}$ models the similarity between the predicted and ground truth hazy image while $\mathcal{G}$ helps estimate A and t during the iteration process. We conduct an ablation study to demonstrate the effectiveness of these two terms, where $\mathcal{F}$ or $\mathcal{G}$ is removed from the iteration unit (9) accordingly. Quantitative results in Table 5 show that $\mathcal{F}$ and $\mathcal{G}$ help achieve better performance.

6.3. Effectiveness of the learned image priors

Our proposed dehazing network simultaneously learns haze-relevant image priors for A, t and J in a data-driven manner. These stage-wise priors are embedded in the gradient descent iteration to help the restoration of A, t and J. Figure 10 shows some resulting images of the estimated A and t against state-of-the-art methods, where our method can achieve accurate restoration results on both A and t. Table 4 presents the quantitative evaluations against the compared methods in estimating A, t and J on the synthetic test dataset, the proposed method can achieve comparable results on A and higher PSNR and SSIM value on t and J. As demonstrate in Section 5.2, clear haze-free images are obtained with these learned priors. We further make an ablation study by removing the prior terms (10), (11) and (12) from (9), but retain the N^1 in (5). As shown in Table 5, the proposed algorithm with networks $\mathcal{N}_A$, $\mathcal{N}_t$ and $\mathcal{N}_J$ for learning image priors could generate better results than the other baselines.

6.4. Relations with the most related methods

We note that a recent work PDN [24] develops deep CNNs to learn dark channel prior for transmission maps in a variational framework to solve image dehazing problem. As image dehazing is an highly ill-posed problem, favorable regularizations should be added on the atmosphere light, transmission map as well as the clear image. Compared with PDN [24], we approach the problem differently w.r.t. model formulation, the usage of deep CNNs and how we iteratively solve the optimization problem with the deep CNNs. Moreover, our algorithmic design facilitates incorporating domain knowledge of image dehazing in deep neural networks. Both the quantitative and qualitative results in Section 5.2 demonstrate that our algorithm performs favorably against PDN [24].

7. Conclusions

In this paper, we have proposed an efficient single image dehazing algorithm by iteratively optimizing a variational model with deep CNNs to learn proximal haze-relevant priors. The data fidelity terms and learned deep priors help achieve better estimation for the atmospheric light, transmission map and haze-free image. While the iteration scheme converges within several iterations with gradient descent method. Extensive qualitative and quantitative experiments on benchmark datasets show that the proposed algorithm performs favorably against state-of-the-art methods.

Acknowledgements: This work is supported in part by NSFC (Nos. 61572099, 61872421, and 61922043), NSF of Jiangsu Province (No. BK20180471).

References

[1] Dana Berman, Tali Treibitz, and Shai Avidan. Non-local Image Dehazing. In *CVPR*, pages 1628–1636, 2016.

[2] Bolun Cai, Xiangmin Xu, Kui Jia, Chunmei Qing, and Dacheng Tao. DehazeNet - An End-to-End System for Single Image Haze Removal. *IEEE TIP*, 25(11):5187–5198, 2016.

[3] Raanan Fattal. Single image dehazing. *ACM transactions on graphics (TOG)*, 27(3):72, 2008.

[4] Raanan Fattal. Dehazing Using Color-Lines. *ACM Trans. Graph.*, 34(1):1–14, 2014.

[5] Gene H Golub and Charles F Van Loan. *Matrix computations (3. ed.)*. Johns Hopkins University Press, 1996.

[6] Muhammad Haris, Greg Shakhnarovich, and Norimichi Ukita. Deep Back-Projection Networks For Super-Resolution. In *CVPR*, pages 1664–1673, 2018.

[7] Kaiming He, Jian Sun, and Xiaoou Tang. Single image haze removal using dark channel prior. In *CVPR*, pages 1956–1963, 2009.

[8] Kaiming He, Jian Sun, and Xiaoou Tang. Guided Image Filtering. *IEEE TPAMI*, 35(6):1397–1409, 2013.

[9] Diederik P Kingma and Jimmy Ba. Adam: A Method for Stochastic Optimization. *arXiv*, 2014.

[10] Boyi Li, Xiulian Peng, Zhangyang Wang, Jizheng Xu, and Dan Feng. AOD-Net - All-in-One Dehazing Network. In *ICCV*, pages 4780–4788, 2017.

[11] Boyi Li, Wenqi Ren, Dengpan Fu, Dacheng Tao, Dan Feng, Wenjun Zeng, and Zhangyang Wang. Benchmarking Single Image Dehazing and Beyond. *arXiv*, 2017.

[12] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced Deep Residual Networks for Single Image Super-Resolution. In *CVPR*, pages 136–144, 2017.

[13] Jinshan Pan, Sifei Liu, Deqing Sun, Jiawei Zhang, Yang Liu, Jimmy S J Ren, Zechao Li, Jinhui Tang, Huchuan Lu, Yu-Wing Tai, and Ming-Hsuan Yang. Learning Dual Convolutional Neural Networks for Low-Level Vision. In *CVPR*, pages 3070–3079, 2018.

[14] 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*, pages 154–169, 2016.

[15] Wenqi Ren, Lin Ma, Jiawei Zhang, Jinshan Pan, Xiaochun Cao, Wei Liu, and Ming-Hsuan Yang. Gated Fusion Network for Single Image Dehazing. In *CVPR*, pages 3253–3261, 2018.

[16] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. *arXiv*, 2015.

[17] Ashutosh Saxena, Sung H Chung, and Andrew Y Ng. Learning Depth from Single Monocular Images. In *NIPS*, pages 1161–1168, 2005.

[18] Ashutosh Saxena, Sung H Chung, and Andrew Y Ng. 3-D Depth Reconstruction from a Single Still Image. *IJCV*, 76(1):53–69, 2008.

[19] A Saxena, Min Sun, and A Y Ng. Make3D: Learning 3D Scene Structure from a Single Still Image. *IEEE TPAMI*, 31(5):824–840, 2009.

[20] Uwe Schmidt and Stefan Roth. Shrinkage Fields for Effective Image Restoration. In *CVPR*, pages 2774–2781, 2014.

[21] Nathan Silberman, Derek Hoiem, Pushmeet Kohli, and Rob Fergus. Indoor Segmentation and Support Inference from RGBD Images. In *ECCV*, pages 746–760, 2012.

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

[23] Dmitry Ulyanov and Andrea Vedaldi. Instance Normalization: The Missing Ingredient for Fast Stylization. *arXiv*, 2016.

[24] Dong Yang and Jian Sun. Proximal Dehaze-Net: A Prior Learning-Based Deep Network for Single Image Dehazing. In *ECCV*, pages 702–717, 2018.

[25] He Zhang and Vishal M Patel. Densely Connected Pyramid Dehazing Network. In *CVPR*, pages 3194–3203, 2018.

[26] Yanfu Zhang, Li Ding, and Gaurav Sharma 0001. HazeRD - An outdoor scene dataset and benchmark for single image dehazing. In *ICIP*, pages 3205–3209, 2017.

[27] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. *arXiv*, 2017.

[28] Qingsong Zhu, Jiaming Mai, and Ling Shao. A Fast Single Image Haze Removal Algorithm Using Color Attenuation Prior. *IEEE TIP*, 24(11):3522–3533, 2015.