

Proximal Dehaze-Net: A Prior Learning-Based Deep Network for Single Image Dehazing

Dong Yang and Jian Sun

Xi'an Jiaotong University, Xi'an, 710049, China

yangdong2010@stu.xjtu.edu.cn, jiansun@xjtu.edu.cn

Abstract. Photos taken in hazy weather are usually covered with white masks and often lose important details. In this paper, we propose a novel deep learning approach for single image dehazing by learning dark channel and transmission priors. First, we build an energy model for dehazing using dark channel and transmission priors and design an iterative optimization algorithm using proximal operators for these two priors. Second, we unfold the iterative algorithm to be a deep network, dubbed as *proximal dehaze-net*, by learning the proximal operators using convolutional neural networks. Our network combines the advantages of traditional prior-based dehazing methods and deep learning methods by incorporating haze-related prior learning into deep network. Experiments show that our method achieves state-of-the-art performance for single image dehazing.

Keywords: single image dehazing, prior learning, deep neural network

1 Introduction

Haze is an atmospheric phenomenon that dust, smoke, or dry particles obscure the clarity of a scene. With hazes, only a portion of reflected lights reach the observer as a result of absorption in the atmosphere. Based on this observation, the captured image I of a hazy scene can be modeled as a linear combination of direct attenuation and airlight contributions [7, 11, 27, 28]:

$$I(x) = J(x)T(x) + A(1 - T(x)), \quad (1)$$

where I is the image degraded by hazes, J is the scene radiance or haze-free image, A is the global atmospheric light and $T(x) = \exp(-\eta d(x))$ is the media transmission along the cone of vision which depends on scattering coefficient η and scene depth $d(x)$. Single image dehazing is an ill-posed inverse problem that requires to recover the unknown haze-free image J , atmospheric light A and transmission T from a single input image I . Therefore, it is essential to investigate effective haze-related priors to regularize this inverse problem.

The traditional single image dehazing methods [2, 5, 7, 11, 18, 27, 28, 30, 31, 38] have investigated various image priors. Tan et al. [28] assume that the contrast of hazy images is lower than haze-free images and propose to maximize the contrast of hazy images under the MRF framework. Fattal [7] uses independent component analysis for estimating the transmission in hazy scenes assuming that the transmission and surface shading are

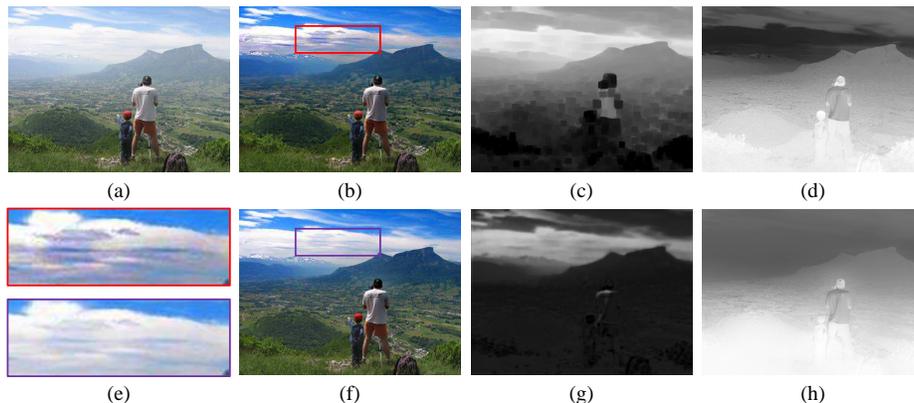


Fig. 1. Single image dehazing results. (a) Input hazy image. (b) Recovered haze-free image using DCP. (c) Dark channel of input image. (d) Transmission map by DCP. (f)~(h) Recovered haze-free image, dark channel and transmission by our network. (e) Comparisons on image blocks.

locally uncorrelated. He et al. [11] propose dark channel prior to estimate transmission based on the observation that the local minimum of color channels of haze-free images is close to zero. These prior-based methods are effective in single image dehazing due to the investigations of prior knowledge and understandings of the physical mechanism for hazes. However, these priors are mainly based on human observations and would not always hold for diverse real world images. For example, dark channel prior [11] is effective for most outdoor images but usually fails for those containing white scenery such as white walls or clouds in sky region, as shown in Fig. 1 (b).

Recently, learning-based methods [4, 14, 15, 20, 21, 29, 34] have been introduced to image dehazing. Ren et al. [20] propose a multi-scale convolutional neural network (MSCNN) for learning the transmission map of the hazy image. It consists of a coarse-scale network predicting a holistic transmission map and a fine-scale network for refining the map. Inspired by haze-relevant features, Cai et al. [4] propose a trainable end-to-end system called DehazeNet for transmission estimation, with specially designed feature extraction layers. Instead of estimating transmission map and atmospheric light separately, Li et al. [14] propose AOD-Net by embedding the estimation of T and A into learning a new K -module. It directly generates the clean image through a light-weight CNN. The learning-based methods have shown promising results for single image dehazing. However, these methods usually take CNNs to learn a mapping from input hazy images to the transmissions or haze-free images, without considering haze-related priors to constrain the mapping space compared with the traditional methods.

In this paper, we propose a novel deep learning-based method that integrates haze imaging model constraints and image prior learning into a single network architecture. *First*, based on the haze imaging model, we formulate the inverse problem of single image dehazing as an energy model with haze imaging constraints in color and dark channel spaces, regularized by dark channel and transmission priors. *Second*, we design an iterative optimization algorithm using half-quadratic splitting, jointly estimating the

transmission map and latent haze-free image. *Third*, we unfold this iterative algorithm to be a deep architecture, dubbed as *proximal dehaze-net*, to implicitly learn the transmission and dark channel priors by learning their corresponding proximal operators using convolutional neural networks. We justify the effectiveness of learned dark channel and transmission priors by ablation study and the learned proximal dehaze-net leads to state-of-the-art results on real and synthetic datasets.

To the best of our knowledge, this is the first work that uses CNNs to learn both dark channel and transmission priors for single image dehazing. Compared with traditional prior-based methods, our approach can discriminatively learn haze-related priors from training data instead of using hand-crafted priors. Compared with deep learning-based dehazing methods, our approach incorporates both haze imaging constraint and haze-related priors learning into the network architecture, which may provide a promising learning-based approach for solving the challenging inverse problem of dehazing.

2 Related Work

2.1 Dark channel

The most related work to ours is the dark channel prior [11] (DCP) method. The dark channel of an image is defined as:

$$I^{dk}(x) = \min_{c \in \{r, g, b\}} \left(\min_{y \in \Omega(x)} (I^c(y)) \right), \quad (2)$$

where I^c is a color channel of I and $\Omega(x)$ is a local patch centered at x . The dark channel prior assumes that, in most non-sky patches, at least one color channel of a haze-free outdoor image has very low intensity at some pixels. According to dark channel prior, the transmission can be estimated as:

$$T = 1 - \omega \min_{c \in \{r, g, b\}} \left(\min_{y \in \Omega(x)} \left(\frac{I^c(y)}{A^c} \right) \right). \quad (3)$$

DCP is effective for dehazing but may fail when the scene color is close to atmospheric light. Instead of constraining dark channel to be close to zero as in DCP, we learn dark channel prior by learning its corresponding proximal mapping from training data using a convolutional neural network, potentially being able to well approximate dark channels of haze-free images as shown in Fig. 1.

2.2 Learning CNNs for image inverse problems

Recently, there have been several works to solve image inverse problems in deep learning frameworks [17, 22, 33, 35, 36]. Zhang et al. [36] train a set of effective denoisers and plug them in the scheme of half-quadratic splitting algorithm as a modular. Meinhardt et al. [17] solve the inverse problem in image processing using primal-dual hybrid gradient method and replace the proximal operator by a denoising neural network. In [22, 33, 35], the linear inverse problems are solved by learning proximal operators in the scheme of iterative optimization algorithms. These methods can well solve the

linear inverse problems such as denoising, super-resolution, inpainting, non-blind deconvolution, compressive sensing MRI, etc.

Compared with these works, we focus on single image dehazing which is a challenging inverse problem with more unknown variables in imaging model. Instead of using common linear inverse models in these works, we specify single image dehazing as a non-linear inverse problem with dark channel and transmission priors (will be discussed in section 3). We then propose an iterative solver and build a deep architecture incorporating the prior learning for single image dehazing. This proposed energy model and dehazing network are novel and the learned network achieves state-of-the-art results in single image dehazing.

3 Dehazing as An Inverse Problem

In this section, we first build an energy function with dark channel and transmission priors and then propose an iterative solver for energy minimization based on half-quadratic splitting (HQS) algorithm. This proposed energy model and its iterative optimization algorithm are the basis for building our proximal dehaze-net as discussed in section. 4.

3.1 Dehazing energy model

Considering the haze imaging model in Eqn. (1), given a hazy image $I \in \mathbb{R}^{M \times N \times 3}$, we assume a known global atmospheric light $A \in \mathbb{R}^3$ and divide both sides of Eqn. (1) by A in each color channel:

$$\frac{I^c(x)}{A^c} = \frac{J^c(x)}{A^c} T(x) + (1 - T(x)), c \in \{r, g, b\}. \quad (4)$$

For simplicity, we denote I^c/A^c by P^c and J^c/A^c by Q^c . Then P , Q represent the scaled hazy image and latent haze-free image respectively. Thus Eqn. (4) can be rewritten in a concise form as:

$$P^c = Q^c \circ T + (1 - T), \quad (5)$$

where \circ is the element-wise product for matrices. We further assume that, within a local patch, the transmission map T is constant as in [11], then it holds that

$$P^{dk} = Q^{dk} \circ T + 1 - T, \quad (6)$$

where P^{dk} , Q^{dk} are dark channels of P , Q .

By enforcing Eqn. (5) in color space and Eqn. (6) in dark channel space as loss terms, we design a dehazing energy function:

$$E(Q, T) = \frac{\alpha}{2} \sum_{c \in \{r, g, b\}} \|Q^c \circ T + 1 - T - P^c\|_F^2 + \frac{\beta}{2} \|Q^{dk} \circ T + 1 - T - P^{dk}\|_F^2 + f(T) + g(Q^{dk}), \quad (7)$$

where α and β are coefficients for data terms. $f(T)$ and $g(Q^{dk})$ are regularization terms modeling the priors on transmission map T and dark channel Q^{dk} . The optimized

haze-free image Q^* and transmission map T^* can be obtained by solving the following optimization problem:

$$\{Q^*, T^*\} = \arg \min_{Q, T} E(Q, T). \quad (8)$$

Regularization terms of f and g . These two terms respectively model the transmission prior and dark channel prior. Multiple image priors can be taken for them, e.g., g for dark channel can be taken as L_0 or L_1 regularizer enforcing the dark channel close to zero. The transmission map is closely related to depth which is piecewise linear, and its regularizer f can be modeled by MRF [8, 19] or TGV [5], etc. However, instead of using these hand-designed regularizations, we set them as uncertainties and aim to learn them from data using a deep learning approach.

Relationship to DCP method. We find that DCP method is a special case of our model. When $\alpha = 0$, $f(T) = 0$ and $g(Q^{dk})$ strictly enforces that $Q^{dk} = 0$, then Eqn. (7) has the solution of transmission $T^* = 1 - P^{dk}$, which is same as DCP in Eqn. (3) when $\omega = 1$. The post-processing procedure of soft matting in DCP acts as an implicit transmission regularization.

Given this energy model, we next introduce how to design its optimization algorithm and deduce a deep architecture in sections 3.2 and 4 respectively.

3.2 Model optimization

We now solve the optimization problem of Eqn. (8) using half-quadratic splitting (HQS) algorithm. The HQS algorithm has been widely used in solving image inverse problems [9, 10, 13, 32, 35]. By introducing an auxiliary variable U to substitute the dark channel Q^{dk} of latent haze-free image, we derive the augmented energy function:

$$\begin{aligned} E(Q, T, U) = & \frac{\alpha}{2} \sum_{c \in \{r, g, b\}} \|Q^c \circ T + 1 - T - P^c\|_F^2 \\ & + \frac{\beta}{2} \|U \circ T + 1 - T - P^{dk}\|_F^2 + \frac{\gamma}{2} \|U - Q^{dk}\|_F^2 + f(T) + g(U), \end{aligned} \quad (9)$$

in which γ is a penalty weight, and when $\gamma \rightarrow \infty$, the solution of minimizing Eqn. (9) converges to that of minimizing Eqn. (7). We initialize $Q_0 = P$ and all elements of T_0 are ones, then for an iteration n of HQS algorithm, minimizing Eqn. (9) can be achieved by solving three sub-problems for alternately updating U , T and Q .

Update U . Given the estimated haze-free image Q_{n-1} and transmission map T_{n-1} at iteration $n - 1$, the auxiliary variable U can be updated as:

$$U_n = \arg \min_U \frac{\beta}{2} \|U \circ T_{n-1} + 1 - T_{n-1} - P^{dk}\|_F^2 + \frac{\gamma}{2} \|U - Q_{n-1}^{dk}\|_F^2 + g(U), \quad (10)$$

from which we can derive

$$U_n = \text{prox}_{\frac{1}{b_n}g}(\hat{U}_n), \quad (11)$$

where

$$\hat{U}_n = \frac{1}{b_n} [\beta T_{n-1} \circ (P^{dk} + T_{n-1} - 1) + \gamma Q_{n-1}^{dk}], \quad (12)$$

and $b_n = \beta T_{n-1} \circ T_{n-1} + \gamma$. The proximal operator is defined as:

$$\text{prox}_{\lambda g}(V) = \arg \min_X \frac{1}{2} \|X - V\|_F^2 + \lambda g(X), \quad (13)$$

assuming that $g(X)$ is separable for different elements in matrix X .

Update T . We next update the transmission map T . Given Q_{n-1} and U_n ,

$$\begin{aligned} T_n = \arg \min_T & \frac{\alpha}{2} \sum_c \|Q_{n-1}^c \circ T + 1 - T - P^c\|_F^2 \\ & + \frac{\beta}{2} \|U_n \circ T + 1 - T - P^{dk}\|_F^2 + f(T). \end{aligned} \quad (14)$$

Then we derive

$$T_n = \text{prox}_{\frac{1}{c_n} f}(\hat{T}_n), \quad (15)$$

where

$$\hat{T}_n = \frac{1}{c_n} [\sum_c \alpha (Q_{n-1}^c - 1) \circ (P^c - 1) + \beta (U_n - 1) \circ (P^{dk} - 1)], \quad (16)$$

and $c_n = \sum_c \alpha (Q_{n-1}^c - 1) \circ (Q_{n-1}^c - 1) + \beta (U_n - 1) \circ (U_n - 1)$.

Update Q . Given T_n and U_n , the haze-free image Q is updated as:

$$Q_n = \arg \min_Q \frac{\alpha}{2} \sum_c \|Q^c \circ T_n + 1 - T_n - P^c\|_F^2 + \frac{\gamma}{2} \|Q^{dk} - U_n\|_F^2. \quad (17)$$

Computing dark channel is to extract the smallest value from local color patch around each pixel. This operation can be implemented by a matrix D with value of one indicating the position of extracted minimal value, i.e., $\overrightarrow{Q}^{dk} = D \overrightarrow{Q}$, where \overrightarrow{Q} is the vectorized Q . We further denote $\mathcal{T}_n \in \mathbb{R}^{M \times N \times 3}$ as a matrix with each color channel as T_n . Then Eqn. (17) can be rewritten as:

$$\overrightarrow{Q}_n = \arg \min_{\overrightarrow{Q}} \frac{\alpha}{2} \|\overrightarrow{Q} \circ \overrightarrow{\mathcal{T}}_n + 1 - \overrightarrow{\mathcal{T}}_n - \overrightarrow{P}\|_2^2 + \frac{\gamma}{2} \|D \overrightarrow{Q} - \overrightarrow{U}_n\|_2^2, \quad (18)$$

which has the solution:

$$\overrightarrow{Q}_n = \frac{\alpha (\overrightarrow{P} + \overrightarrow{\mathcal{T}}_n - 1) \circ \overrightarrow{\mathcal{T}}_n + \gamma D^\top \overrightarrow{U}_n}{\alpha \overrightarrow{\mathcal{T}}_n \circ \overrightarrow{\mathcal{T}}_n + \gamma \text{diag}(D^\top D)}. \quad (19)$$

The updated haze-free image Q_n can be derived by reshaping \overrightarrow{Q}_n back to a matrix with the same size of input image. Note that the divisions in the Eqns. (12), (16), (19) are all element-wise operations. The detailed conduction of above equations can be found in supplementary material.

Special case. If we discard transmission regularization term by setting $f(T) = 0$ and set the dark channel regularization term as L_1 -norm, i.e., $g(Q^{dk}) = \sum_{x \in \Omega} |Q^{dk}(x)|$ enforcing the dark channel sparse and close to zero, then the corresponding proximal operators in Eqns. (11) and (15) are defined as:

$$\text{prox}_{\frac{1}{b_n} g}(\hat{U}_n) = \text{softThresh}(\hat{U}_n, \frac{1}{b_n}), \quad (20)$$

$$\text{prox}_{\frac{1}{c_n} f}(\hat{T}_n) = \hat{T}_n, \quad (21)$$

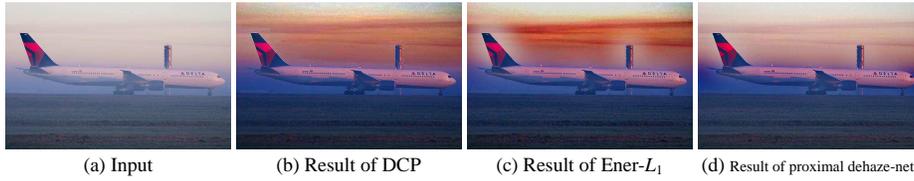


Fig. 2. An example of real image dehazing.

respectively, where $\text{softThresh}(\cdot, 1/b_n)$ is a soft thresholding function [6] with threshold $1/b_n$. We take this as the baseline of our method, denoted as *Ener-L₁*. An example of *Ener-L₁* is shown in Fig. 2 (c). *Ener-L₁* is indeed effective for haze removal which is comparable to DCP. Through prior learning as will be discussed in next section, our proximal dehaze-net can better recover haze-free image as shown in Fig. 2 (d).

4 Proximal Dehaze-Net

Based on the above iterative optimization algorithm, we build a deep neural network for single image dehazing as illustrated in Fig. 3 (a). The network is a structure with N stages implementing N iterations in the iterative optimization algorithm for solving Eqn. (8). Each stage takes outputs of the previous stage U_{n-1} , T_{n-1} and Q_{n-1} (representing dark channel, transmission map and haze-free image respectively) as inputs and computes updated U_n , T_n and Q_n .

Instead of setting by hand the regularization terms of $g(\cdot)$ and $f(\cdot)$ (modeling dark channel and transmission priors) in the energy function of Eqn. (7), we use deep CNNs to learn their corresponding proximal operators $\text{prox}_{\frac{1}{b_n}g}$ and $\text{prox}_{\frac{1}{c_n}f}$ for updating U_n and T_n in each stage n :

$$\begin{aligned} U_n &= \text{prox}_{\frac{1}{b_n}g}(\hat{U}_n) \triangleq \mathcal{G}(\hat{U}_n), \\ T_n &= \text{prox}_{\frac{1}{c_n}f}(\hat{T}_n) \triangleq \mathcal{F}(\hat{T}_n), \end{aligned} \quad (22)$$

where \mathcal{G} and \mathcal{F} are deep CNNs to be learned for representing the corresponding proximal operators. In this way, we design an end-to-end training architecture, dubbed as *proximal dehaze-net*. We will next introduce the network structure.

4.1 Network design

As shown in Fig. 3 (a), each stage of the proximal dehaze-net implements one iteration of model optimization discussed in section 3.2, and the proximal operators are substituted by convolutional neural networks as in Eqn. (22).

We now introduce the network structure for each stage. Please refer to Fig. 3 (b) for better understanding. For the n -th stage, \hat{U}_n is first computed by Eqn. (12), then sent

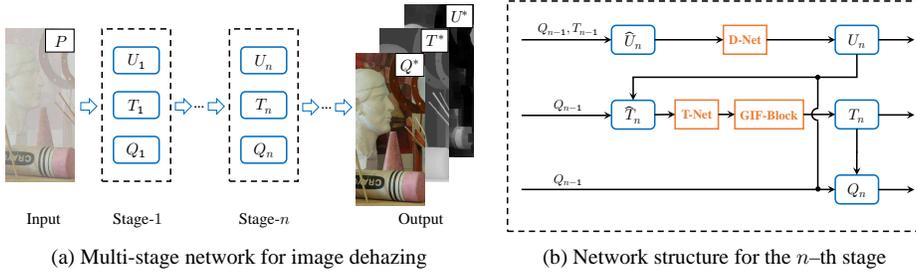


Fig. 3. Architecture of proximal dehaze-net. (a) The network consists of multiple stages. The outputs U^* , T^* , Q^* are estimated dark channel, transmission map and dehazed image. (b) The detailed architecture for the n -th stage, in which the variables U_n , T_n and Q_n are successively updated by the algorithm described in section. 3.2. D-Net is to learn proximal mapping \mathcal{G} for dark channel prior. T-Net and GIF-Block are to learn the proximal mapping \mathcal{F} for transmission prior.

into a convolutional neural network, i.e., *D-Net*, to perform proximal mapping $\text{prox}_{\frac{1}{b_n}g}$. The updated dark channel is:

$$U_n = \mathcal{G}(\hat{U}_n) \triangleq \text{D-Net}(\hat{U}_n, P), \quad (23)$$

in which we concatenate \hat{U}_n with hazy image P as input.

Similarly, \hat{T}_n is first computed using Eqn. (16), then concatenated with P and sent into another convolutional neural networks, *T-Net* and a *GIF-Block*, to perform proximal mapping $\text{prox}_{\frac{1}{c_n}f}$. The updated transmission map is:

$$T_n = \mathcal{F}(\hat{T}_n) \triangleq \text{GIF-Block}(\text{T-Net}(\hat{T}_n, P)), \quad (24)$$

T-Net and GIF-Block are respectively responsible for transmission estimation and guided image filtering (GIF) for better edge alignment with image edges.

Finally, with Q_{n-1} , U_n and T_n , we get the updated Q_n using Eqn. (19). After N stages, the final estimated haze-free image J is with channels of $J^c = Q_N^c A^c$ for $c \in \{r, g, b\}$. We estimate the atmospheric light A according to [11]. We next introduce the structures of D-Net, T-Net and GIF-Block.

D-Net and T-Net. D-Net and T-Net have similar structures. They both include three cascaded convolutional blocks. Each block consists of a convolutional layer, a ReLU layer, a pooling layer and an upsampling layer. The convolutional layers in these blocks have nine 7×7 filters, nine 5×5 filters and nine 3×3 filters respectively. The last block is followed by another convolutional layer with one 1×1 filter. Then the output map is finally sent to a ReLU layer for D-Net or a sigmoid layer for T-Net. The ReLU layer aims to keep the output dark channel U non-negative, while the sigmoid layer is to keep the output transmission map T within $[0, 1]$.

GIF-Block. GIF-Block performs standard guided image filtering [12]. This block enforces the transmission map to be well aligned with image along edges. As shown in Fig. 3 (b), GIF-Block takes the input image P as guidance and performs guided image

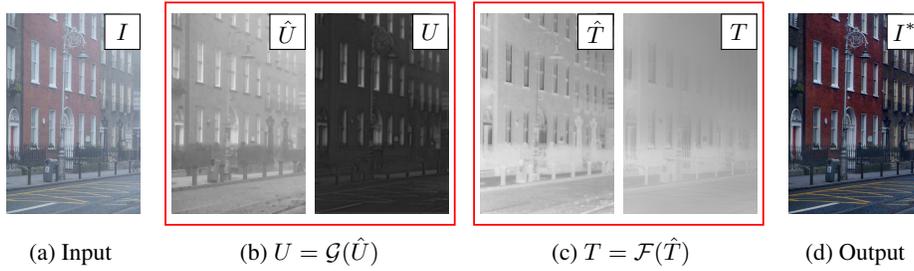


Fig. 4. Effectiveness of using CNNs to learn proximal mappings for dark channel and transmission priors. The left and right images in (b) (c) respectively represent the inputs and outputs of proximal mappings \mathcal{G} and \mathcal{F} for dark channel and transmission. (d) Our final dehazing result.

filtering on the output of T-Net:

$$\text{GIF-Block} \left(\text{T-Net}(\hat{T}_n, P) \right) \triangleq \text{GIF}_P \left(\text{T-Net}(\hat{T}_n, P) \right), \quad (25)$$

using the guided filtering operator GIF_P with guidance of P . The GIF-block is a differentiable block implementing guided filtering represented by a computation graph. It is included as a part of our end-to-end trainable system more than just a post-processing step. More details on the algorithm and computation graph of GIF-Block can be found in the supplementary material.

To illustrate what were learned for these proximal mappings, in Fig. 4, we show an example of our learned proximal mappings for dark channel and transmission using the learned proximal dehaze-net (we will introduce network training in section 4.2). Fig. 4 (b), (c) respectively show the dark channels and transmission maps before and after the proximal mappings of \mathcal{G} and \mathcal{F} . We can observe that the learned proximal mapping \mathcal{G} produces reasonable dark channel with low values but retaining the high values for white windows. The learned proximal mapping \mathcal{F} produces a smooth transmission map consistent with the underlying scenery depth.

4.2 Network training

The training loss for each training image is defined as the sum of pixel-wise L_1 distances between the outputs of proximal dehaze-net $\{Q^*, T^*, U^*\}$ and the ground truths $\{Q^{\text{gt}}, T^{\text{gt}}, U^{\text{gt}}\}$:

$$\ell = \sum_{O \in \{Q, T, U\}} \sum_x \|O^*(x) - O^{\text{gt}}(x)\|_1. \quad (26)$$

When training the proximal dehaze-net, we compute the gradients of loss w.r.t. the inputs of Eqns. (11), (15), (19) for back-propagation. Due to space limit, please refer to supplementary material for these gradients. In our implementation, the parameters of nets (including D-Net and T-Net) in different stages are not shared. Parameters of α and β in energy function are set to 1 and 5 respectively. We implement and train our network using MatConvNet¹ framework. We choose Adam solver with a learning rate

¹ <http://www.vlfeat.org/matconvnet/>

of 0.001. We use a batch size of 10 and it usually takes 13.6 hours to train a single stage network for 50 epochs on a Titan X GPU with 12 GB memory.

5 Experiments

In this section, we evaluate our method on both synthetic and real datasets and compare with the other state-of-the-art dehazing methods proposed in recent years.

5.1 Datasets

Training dataset. Our training dataset consists of three RGB-D datasets including NYU-Depth [26], Middlebury-Stereo [23–25] and MPI-Sintel [3]. We sampled 9000 pairs of RGB-D image patches in size of 240×240 from the training dataset. According to haze imaging model Eqn. (1), given A and η , the hazy image and transmission are generated by the clean image and its corresponding depth map. For each pair of RGB-D image patches, we randomly select $A \in [0.7, 1]$ and $\eta \in [0.5, 1.5]$ to generate the ground truth transmission map and hazy image, then compute the dark channel from the clean image. Therefore we construct a set of 9000 training samples including hazy images, ground truth haze-free images, transmission maps and dark channels.

Test datasets. To quantitatively evaluate the dehazing methods, we established two benchmark datasets – TestA and TestB for evaluating dehazing performance. For TestA, we use the RGB-D images from NYU, Middlebury and Sintel datasets (without overlapping with training images) to generate 548 full sized hazy images. We also establish a more realistic dataset TestB including 128 images by applying different hazes on HazeRD [37] dataset. In TestA, images are either taken indoors or computer generated, thus are different from the real-world outdoor hazy images. HazeRD provides real outdoor images with high-quality depth information, therefore enables us to better simulate real outdoor images taken under hazy condition.

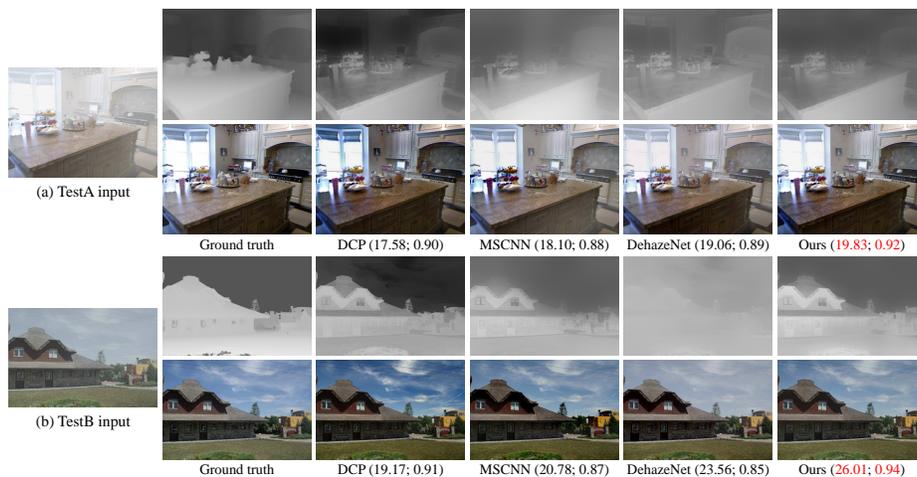
5.2 Results on synthetic datasets

We first compare our network with recent methods for single image dehazing on both synthetic datasets – TestA and TestB. The compared methods include dark channel prior (DCP) [11], fast visibility restoration (FVR) [30], boundary constraint and contextual regularization (BCCR) [18], gradient residual minimization (GRM) [5], color attenuation prior (CAP) [38], non-local dehazing (NLD) [2], multi-scale CNN (MSCNN) [20], DehazeNet [4] and AOD-Net [14]. Among these methods, MSCNN, DehazeNet and AOD-Net are learning-based methods. For quantitative evaluation, we show the average peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) index between the recovered images and ground truths.

As shown in Table 1, our proximal dehaze-net achieves best results in both PSNR and SSIM values on TestA and significantly improves the DCP [11] method by nearly 1 dB in PSNR. Since all learning-based methods do not include images in TestB as training data, it is fair to compare them on TestB. As shown in Table 1, our method

Table 1. Dehazing results on synthetic datasets TestA and TestB. We show the average PSNR and SSIM of the recovered images by the compared methods.

Methods	DCP [11]	FVR [30]	BCCR [18]	CAP [38]	NLD [2]	GRM [5]	MSCNN [20]	DehazeNet [4]	AOD-Net [14]	Ours	
TestA	PSNR	18.32	15.18	16.00	16.63	17.87	18.45	17.37	19.17	18.00	19.31
	SSIM	0.8244	0.7599	0.7556	0.7050	0.7978	0.8043	0.7564	0.7888	0.7928	0.8388
TestB	PSNR	17.66	16.17	16.31	18.56	18.82	17.47	19.10	19.53	18.13	20.14
	SSIM	0.8430	0.8472	0.8337	0.8256	0.8355	0.7921	0.8540	0.8498	0.8266	0.8932

**Fig. 5.** Dehazing results on two examples from TestA and TestB. We show the recovered images and corresponding transmission maps by different methods. PSNR and SSIM values of dehazed images are shown in brackets.

achieves highest PSNR and SSIM on TestB and exceeds the second best learning-based method DehazeNet [4] by 0.61dB in PSNR.

In Fig. 5, we show two examples of dehazing results and corresponding transmission maps from TestA and TestB. Compared with other methods, our proximal dehaze-net can better estimate transmission maps and produce more visually pleasant haze-free images with highest PSNR and SSIM values. Though DehazeNet [4] achieves relatively high PSNR as shown in Table 1, its results still contain hazes as shown in Fig. 5. On the other hand, DCP [11] can effectively remove the hazes, but sometimes may over-enhance the color contrast in these results. As a comparison, our method can well control the amount of removed hazes and produces visually natural results.

5.3 Results on real datasets

In Fig. 6, we also evaluate and compare our network with recent state-of-the-art methods [4, 11, 14, 20, 38] on real-world hazy images. The real-world example images are collected from Internet and previous works. For traditional methods, such as DCP [11] and CAP [38], the hazes are significantly removed and the results are with high color



Fig. 6. Dehazing results on real images. Please zoom in for better illustration.

contrast. However, CAP sometimes blurs image textures and causes over-saturation in color, as observed in the 2nd and 4th images of Fig. 6. DCP can not properly deal with sky regions and is likely to introduce artifacts as shown in the 4th, 5th and 7th images of Fig. 6. It is interesting that the learning-based methods [4, 14, 20] trained on synthetic dataset generalize well to produce visually pleasant results for real image dehazing. However, as shown in the 4th and 6th images of Fig. 6, MSCNN [20] sometimes cause color distortion, which makes the recovered images seem unnatural. DehazeNet [4], although achieves high PSNR values on synthetic datasets, does not remove hazes as effectively as other methods, such as the 1st, 3rd and 7th images of Fig. 6. AOD-Net [14] usually slightly reduces image brightness and sometimes causes faded scene of foreground as shown in the 3rd and 6th images of Fig. 6. Our proximal dehaze-net, integrating haze imaging model with deep learning, can effectively remove hazes in different amounts while still keeping the results visually natural and pleasant.

5.4 Ablation study

To investigate the effect of learning dark channel and transmission priors for our network, we respectively discard dark channel regularization $g(U)$ and transmission regularization $f(T)$ in Eq. (9). We then denote our proximal dehaze-net without learning dark channel prior as *Net-ND* and without learning transmission prior as *Net-NT*. We train *Net-ND* and *Net-NT* with one stage from scratch and compare with *Ener- L_1* (see section. 3.2) and *Net-S1* (our proximal dehaze-net with one stage) on TestB. We show

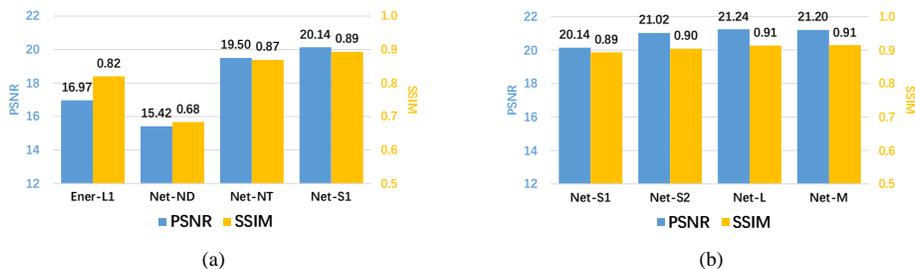


Fig. 7. Comparison of different network architectures on TestB dataset. (a) Results of our proximal dehaze-net and nets without prior learning. (b) Results of our nets in different complexities.

the PSNR and SSIM in Fig. 7 (a). Compared with Net-S1, Net-NT without learning transmission prior decreases the PSNR by 0.64 dB, and Net-ND without learning dark channel prior decreases the PSNR by 4.72 dB, even lower than Ener- L_1 . Therefore learning both priors, especially the dark channel prior, is essential for our approach.

To evaluate the effect of model complexity on performance, we trained proximal dehaze-nets with different filter sizes, filter numbers and number of stages. For *Net-L*, we use larger filter sizes by setting all convolutional kernels to be 7×7 . For *Net-M*, we use 64 rather than only 9 filters in each convolution layer. We also trained our proximal dehaze-net with 2 stages, denoted as *Net-S2*. We show the PSNR and SSIM on TestB in Fig. 7 (b), from which we can see that increasing network complexity promotes the PSNR by over 0.88 dB. However, we did not observe significant qualitative improvements visually using these more complex networks. Moreover, increasing network complexity increases the running time. To be specific, the running times on a single GPU for these networks on an image of 480×640 are 0.058s for Net-S1, 0.096s for Net-S2, 0.077s for Net-L and 0.143s for Net-M respectively. For the sake of efficiency, we adopt Net-S1 as our final model and all reported results are based on Net-S1.

5.5 Extension to more applications

Although our network is trained for image dehazing, we can also extend it to other tasks that are similar to dehazing. In Fig. 8 (a), we show an example of underwater image enhancement. Ignoring the forward scattering component, the simplified underwater optical model [1] has similar formulation with haze imaging model. Our network can effectively remove haze-like effect in this underwater image. Although halation has a different imaging model, it brings haze-like effect to image. Our proximal dehaze-net can be directly applied to anti-halation image enhancement without need of re-training, as shown in Fig. 8 (b). In Fig. 8 (c), we also show an example of our network applied to a haze-free image to test the robustness. Our network does not change the image much and the result still looks natural and clear.

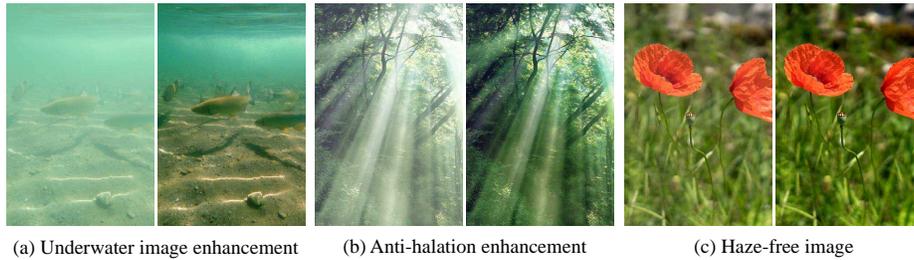


Fig. 8. Extension to other applications.



Fig. 9. Failure cases of our method.

5.6 Limitations

While our method behaves well on most natural images, it has limitations towards certain situations in which the photo is taken in heavy fog or at night-time. For the first case as shown in Fig. 9 (a), image information is seriously lost due to heavy fog and it is hard for us to recover satisfactory result. For the latter case, since night-time haze follows a different imaging model as described in [16], our method fails to effectively remove hazes in images taken at night-time as shown in Fig. 9 (b). However, if we change the data fidelity term of our dehazing energy function to fit night-time image haze model, there is a potential to improve the result.

6 Conclusion

In this paper, we proposed a novel proximal dehaze-net for single image dehazing. This network integrates haze imaging model, dark channel and transmission priors into a deep architecture. This is achieved by building an energy function using dark channel and transmission priors for single image dehazing, and learning these priors using their corresponding proximal operators in an optimization-inspired deep network. This energy function and proximal dehaze-net are novel for dehazing, and the learned network achieves promising results on both synthetic and real-world hazy images. In the future, we are interested in building realistic outdoor training dataset for dehazing or using outdoor clear images as supervision in a generative adversarial training framework.

Acknowledgement. This work is supported by National Natural Science Foundation of China under Grants 11622106, 61472313, 11690011, 61721002.

References

1. Ancuti, C.O., Ancuti, C., De Vleeschouwer, C., Bekaert, P.: Color balance and fusion for underwater image enhancement. *IEEE Transactions on Image Processing* **27**(1), 379–393 (2018)
2. Berman, D., Treibitz, T., Avidan, S.: Non-local image dehazing. In: *IEEE Conference on Computer Vision and Pattern Recognition* (2016)
3. Butler, D.J., Wulff, J., Stanley, G.B., Black, M.J.: A naturalistic open source movie for optical flow evaluation. In: *European Conference on Computer Vision* (2012)
4. Cai, B., Xu, X., Jia, K., Qing, C., Tao, D.: Dehazenet: An end-to-end system for single image haze removal. *IEEE Transactions on Image Processing* **25**(11), 5187–5198 (2016)
5. Chen, C., Do, M.N., Wang, J.: Robust image and video dehazing with visual artifact suppression via gradient residual minimization. In: *European Conference on Computer Vision* (2016)
6. Donoho, D.L.: De-noising by soft-thresholding. *IEEE Transactions on Information Theory* **41**(3), 613–627 (1995)
7. Fattal, R.: Single image dehazing. *ACM Transactions on Graphics (TOG)* **27**(3), 72 (2008)
8. Fattal, R.: Dehazing using color-lines. *ACM transactions on graphics (TOG)* **34**(1), 13 (2014)
9. Geman, D., Reynolds, G.: Constrained restoration and the recovery of discontinuities. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **14**(3), 367–383 (1992)
10. Geman, D., Yang, C.: Nonlinear image recovery with half-quadratic regularization. *IEEE Transactions on Image Processing* **4**(7), 932–946 (1995)
11. He, K., Sun, J., Tang, X.: Single image haze removal using dark channel prior. In: *IEEE Conference on Computer Vision and Pattern Recognition* (2009)
12. He, K., Sun, J., Tang, X.: Guided image filtering. In: *European Conference on Computer Vision* (2010)
13. Krishnan, D., Fergus, R.: Fast image deconvolution using hyper-laplacian priors. In: *Advances in Neural Information Processing Systems* (2009)
14. Li, B., Peng, X., Wang, Z., Xu, J., Feng, D.: Aod-net: All-in-one dehazing network. In: *IEEE International Conference on Computer Vision* (2017)
15. Li, R., Pan, J., Li, Z., Tang, J.: Single image dehazing via conditional generative adversarial network. In: *IEEE Conference on Computer Vision and Pattern Recognition* (2018)
16. Li, Y., Tan, R.T., Brown, M.S.: Nighttime haze removal with glow and multiple light colors. In: *Proceedings of the IEEE International Conference on Computer Vision*. pp. 226–234 (2015)
17. Meinhardt, T., Moller, M., Hazirbas, C., Cremers, D.: Learning proximal operators: Using denoising networks for regularizing inverse imaging problems. In: *IEEE International Conference on Computer Vision* (2017)
18. Meng, G., Wang, Y., Duan, J., Xiang, S., Pan, C.: Efficient image dehazing with boundary constraint and contextual regularization. In: *IEEE International Conference on Computer Vision* (2013)
19. Nishino, K., Kratz, L., Lombardi, S.: Bayesian defogging. *International Journal of Computer Vision* **98**(3), 263–278 (2012)
20. Ren, W., Liu, S., Zhang, H., Pan, J., Cao, X., Yang, M.H.: Single image dehazing via multi-scale convolutional neural networks. In: *European Conference on Computer Vision* (2016)
21. Ren, W., Ma, L., Zhang, J., Pan, J., Cao, X., Liu, W., Yang, M.H.: Gated fusion network for single image dehazing. In: *IEEE Conference on Computer Vision and Pattern Recognition* (2018)

22. Rick Chang, J.H., Li, C.L., Poczos, B., Vijaya Kumar, B.V.K., Sankaranarayanan, A.C.: One network to solve them all – solving linear inverse problems using deep projection models. In: IEEE International Conference on Computer Vision (2017)
23. Scharstein, D., Szeliski, R.: High-accuracy stereo depth maps using structured light. In: IEEE Conference on Computer Vision and Pattern Recognition (2003)
24. Scharstein, D., Hirschmüller, H., Kitajima, Y., Krathwohl, G., Nešić, N., Wang, X., Westling, P.: High-resolution stereo datasets with subpixel-accurate ground truth. In: German Conference on Pattern Recognition (2014)
25. Scharstein, D., Pal, C.: Learning conditional random fields for stereo. In: IEEE Conference on Computer Vision and Pattern Recognition (2007)
26. Silberman, N., Hoiem, D., Kohli, P., Fergus, R.: Indoor segmentation and support inference from rgb-d images. In: European Conference on Computer Vision (2012)
27. Tan, R.T.: Visibility in bad weather from a single image. In: IEEE Conference on Computer Vision and Pattern Recognition (2008)
28. Tan, R.T., Pettersson, N., Petersson, L.: Visibility enhancement for roads with foggy or hazy scenes. In: IEEE Intelligent Vehicles Symposium (2007)
29. Tang, K., Yang, J., Wang, J.: Investigating haze-relevant features in a learning framework for image dehazing. In: IEEE Conference on Computer Vision and Pattern Recognition (2014)
30. Tarel, J.P., Hautiere, N.: Fast visibility restoration from a single color or gray level image. In: IEEE International Conference on Computer Vision (2009)
31. Tripathi, A., Mukhopadhyay, S.: Single image fog removal using anisotropic diffusion. *IET Image Processing* **6**(7), 966–975 (2012)
32. Wang, Y., Yang, J., Yin, W., Zhang, Y.: A new alternating minimization algorithm for total variation image reconstruction. *SIAM Journal on Imaging Sciences* **1**(3), 248–272 (2008)
33. Yang, Y., Sun, J., Li, H., Xu, Z.: Deep admm-net for compressive sensing mri. In: Advances in Neural Information Processing Systems (2016)
34. Zhang, H., Patel, V.M.: Densely connected pyramid dehazing network. In: IEEE Conference on Computer Vision and Pattern Recognition (2018)
35. Zhang, J., Pan, J., Lai, W.S., Lau, R.W.H., Yang, M.H.: Learning fully convolutional networks for iterative non-blind deconvolution. In: IEEE Conference on Computer Vision and Pattern Recognition (2017)
36. Zhang, K., Zuo, W., Gu, S., Zhang, L.: Learning deep cnn denoiser prior for image restoration. In: IEEE Conference on Computer Vision and Pattern Recognition (2017)
37. Zhang, Y., Ding, L., Sharma, G.: Hazerd: an outdoor scene dataset and benchmark for single image dehazing. In: IEEE International Conference on Image Processing (2017)
38. Zhu, Q., Mai, J., Shao, L.: A fast single image haze removal algorithm using color attenuation prior. *IEEE Transactions on Image Processing* **24**(11), 3522–3533 (2015)