IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received May 24, 2021, accepted July 20, 2021, date of publication July 30, 2021, date of current version August 10, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3101319

A Model-Driven Deep Dehazing Approach by Learning Deep Priors

DONG YANG AND JIAN SUN
School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an 710049, China
Corresponding author: Jian Sun (jiansun@mail.xjtu.edu.cn)

This work was supported in part by the NSFC under Grant U20B2075, Grant 11690011, Grant 11971373, Grant U1811461, Grant 12026605, and Grant 12090021; and in part by the National Key Research and Development Program under Grant 2018AAA0102201.

ABSTRACT Photos taken in hazy weather are usually covered with white masks and lose important details. Haze removal is a fundamental task and a prerequisite to many other vision tasks. Single image dehazing is an ill-posed inverse problem that has attracted much attention in recent years. Generally, current single dehazing methods can be categorized into the traditional prior-based methods and the data-driven deep learning methods that respectively investigate haze-related image priors and deep architectures. In this paper, we propose a novel model-driven deep learning approach that combines the advantages of both kinds of methods. First, we build an energy model for single image dehazing with physical constraints in both color image space and haze-related feature space (implemented as dark channel space in this work), regularized by haze-related image priors. Then, we design an iterative optimization algorithm for solving the proposed dehazing energy model based on the half-quadratic splitting algorithm, and the priors are transformed to their corresponding proximal operators. Finally, inspired by the optimization algorithm, we design a deep dehazing neural network, dubbed as *proximal dehaze-net*, by learning the proximal operators for haze-related image priors using CNNs. Our network incorporates physical model constraints of hazes and haze-related prior learning into a novel deep architecture. Extensive experiments show that our method achieves promising performance for single image dehazing.

INDEX TERMS Image dehazing, deep learning, neural networks, model-driven, prior-learning.

I. INTRODUCTION

Haze is an atmospheric phenomenon where dust, smoke, or dry particles obscure the clarity of a scene. Hazes usually degrade the quality of photos by reducing the total contrast in color and occluding objects in photos. It is necessary to reduce the haze effect on these photos to make them visually pleasing and appealing. Moreover, many vision tasks in practical life depend on clean images, such as face detection and automatic license plate recognition from monitoring images, scene analysis from satellite images, etc. However, real captured images even on sunny days often suffer from low visual quality with haze effect. It is therefore essential for a vision system to firstly remove hazes from the captured images and then conduct detection or recognition.

In hazy images, only a portion of the reflected light reaches the observer because of the absorption in the atmosphere. Based on this observation, the captured image I of a hazy scene can be modeled as a linear combination of the direct attenuation and the airlight [1]–[4]:

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

where x is the spatial coordinate, 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 $d(x)$. Depending on the number of given hazy images, the image dehazing task can be divided into single image dehazing and multi-image dehazing. In this paper, we focus on single image dehazing that requires to recover the unknown haze-free image J, atmospheric light A, and transmission T from a single input hazy image I. Single image dehazing is more challenging compared with multi-image dehazing, and it is essential to investigate effective haze-related priors to regularize this inverse problem. Previous works can be roughly categorized into

The associate editor coordinating the review of this manuscript and approving it for publication was Hong-Mei Zhang.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

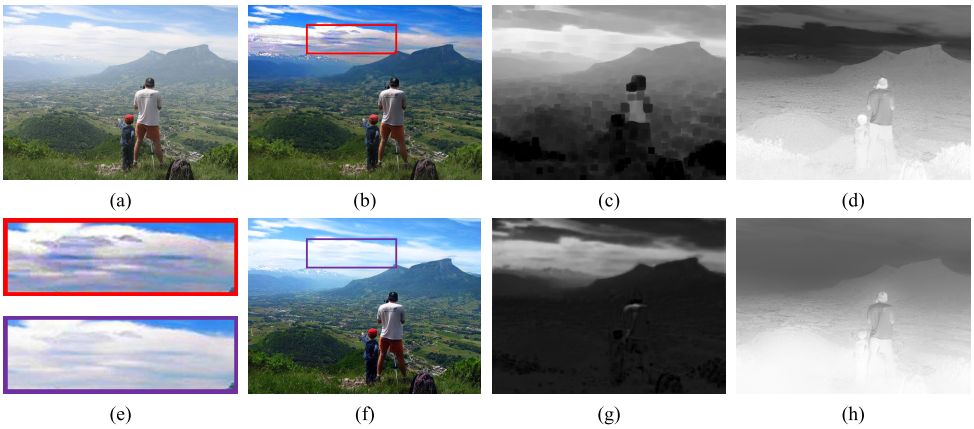


FIGURE 1. Single image dehazing results. (a) Input hazy image. (b) Recovered haze-free image using DCP. (c) Dark channel of the input image. (d) Transmission map by DCP. (f)∼(h) Recovered haze-free image, dark channel and transmission map by our network. (e) Comparisons on image blocks. Our method can better deal with sky regions.

the traditional methods that investigate various haze-related priors and the learning-based methods that build learning systems for dehazing.

The traditional image dehazing methods [1]–[12] have investigated various haze-related image priors. Tan [2] assumes 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 [3] uses independent component analysis for estimating the transmission in hazy scenes assuming that the transmission and surface shading are locally uncorrelated. He *et al.* [4] propose the dark channel prior to estimate the transmission map based on the observation that the local minimums of color channels of haze-free images are close to zero. Liu *et al.* [12] propose the rank-one prior based on the observation that in most of the regions except the light source area, the imaging scene is covered by spatially homogenous light. Polarization-based methods [13], [14] are also effective for haze removal, which are based on the fact that the airlight scattered by atmospheric particles is partially polarized. These methods are effective in image dehazing due to the full investigation of image prior knowledge and the understanding of physical mechanism of hazes. However, these priors are mainly based on human observations that would not always hold for diverse real-world images. For example, dark channel prior [4] is effective for most outdoor images but usually fails for those containing large areas of white scenery such as white walls or clouds in the sky, as shown in Figure 1 (b). Despite the effectiveness, polarization-based dehazing requires professional optical equipment and multiple image fusion, which is not suitable for single image dehazing.

In recent years, deep learning methods [15]–[29] have overwhelmed single image dehazing area. The early learning-based methods usually learn transmission maps or other haze-related variables. For instance, Ren *et al.* [16] and Cai *et al.* [17] design CNNs to learn transmission maps and then recover clean images, and Li *et al.* [18] propose to learn a K-module for the estimation of transmission T and atmospheric light A. More recent works [19], [23], [25], [26], [29]–[31] propose to learn a direct mapping between hazy images and clean images by developing new CNN architectures or investigating novel effective loss functions. Besides, to achieve more realistic dehazing results, generative adversarial networks [20], [21], [25], [28], [32] are also widely used in image dehazing task. 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 image priors to constrain the mapping space compared with the traditional methods.

Our main motivation is to combine haze imaging mechanism and deep learning in a novel model-driven deep learning framework [33], which takes the advantages of both the prior-based methods and the deep learning-based methods. Compared with the traditional prior-based methods, our approach is free of parameter tuning after training and runs at a higher speed. Most importantly, combined with the data-driven method, the dehazing performance is significantly improved. Compared with the deep learning-based methods, our approach integrates physical mechanism constraint into the deep learning framework and can produce stable dehazing results (see Figure 10). Moreover, our method learns to explicitly predict atmospheric light and transmission maps. They may be helpful for model explanation and downstream tasks such as semantic segmentation on hazy images.

We build our model-driven learning approach by the following steps. First, based on the haze imaging model, we formulate the inverse problem of single image dehazing as an energy model with physical constraints in image space and haze-related feature space (dark channel space in this paper), regularized by haze-related image priors. Second, we design an iterative optimization algorithm for minimizing the dehazing energy function using the half-quadratic

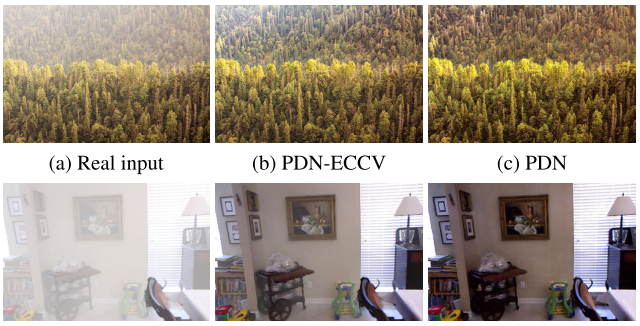


FIGURE 2. A comparison of the extended PDN model and the original PDN model. We can see that the extended PDN model achieves dramatic improvement in PSNR (17.77 to 28.78) and removes haze more effectively.

splitting algorithm, with proximal operators for modeling the regularization terms. Third, we propose a deep neural network based on the iterative algorithm, dubbed as *proximal dehaze-net*, to implicitly learn these image priors by learning their corresponding proximal operators using convolutional neural networks.

In summary, our work makes three main contributions. First, we propose a novel energy model for single image dehazing, which investigates the haze imaging model constraints in both image space and haze-related feature space. Second, based on the iterative algorithm for minimizing the energy model, we design a multi-stage deep neural network, by discriminatively learning haze-related image priors, including dark channel prior, transmission prior and clean image prior, saving the effort of manually designing them. We also learn to predict the atmospheric light instead of estimating it with traditional methods. Third, extensive experiments show the effectiveness of learning haze-related image priors, and the proposed proximal dehaze-net achieves promising results on both synthetic and real-world hazy images.

This paper is an extension of our previous work [22]. In this paper, we extend the contents of our work in the following three aspects. First, we reformulated our dehazing model by introducing an additional clean image prior learning module. We also learned to predict the atmospheric light instead of estimating it with the traditional methods. Second, we improved the performance of our model by replacing lightweight sub-networks with more powerful CNN backbones. Moreover, we treated hyper-parameters that need manual adjusting as learnable variables. Third, for fair comparisons, we evaluated our method on multiple public datasets, including SOTS and NTIRE 2018 datasets, and our method achieves promising results on both synthetic images and real-world images. As a comparison between the original PDN model [22] (denoted as PDN-ECCV) and the current extended PDN model, we show in Figure 2 two examples of synthetic and real hazy image dehazing. We can see that the extended PDN model improves the dehazing ability both quantitatively and qualitatively.

II. RELATED WORK

A. HAZE-RELATED IMAGE PRIORS

Most traditional dehazing methods assume an image prior on haze-free images or latent transmission maps based on human experiences. Researchers have proposed various effective priors for single image dehazing. The most related work to ours is the dark channel prior (DCP) [4]. The dark channel of a color image is defined as the minimum of local image patches:

$$I^d(x) = \min_{c\in\{r,g,b\}} \Big(\min_{y\in\Omega(x)} \big(I^c(y)\big)\Big), \tag{2}$$

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

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

where ω is a constant for keeping aerial perspective.

Since image priors usually rely on human observations and experiences, the traditional dehazing methods are not always applicable to diverse scenes. For instance, DCP is effective for dehazing but may fail when the scene color is close to the atmospheric light, e.g., sky regions in the wild environment and light color walls in cityscapes. 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 the dark channels of haze-free images as shown in Figure 1.

B. DEHAZING BY ENERGY MINIMIZATION

It is a common practice to build energy functions for various image restoration and reconstruction problems. For the problem of single image dehazing, there are also several methods proposed based on image priors and energy minimization [2], [7], [10], [34]–[38]. The main idea of these methods is to firstly find effective image priors for describing transmission maps or haze-free images, and then build energy functions with these image priors as regularization terms. The energy function is then minimized for optimal transmission maps or haze-free images in an iterative manner.

However, it highly relies on expert experiences to find or design effective dehazing priors. It is often time-consuming for most energy-minimization methods to dehaze a single image due to a quantity of optimization steps. There also exist hyper-parameters in the energy models that usually have to be carefully adjusted in order to get the best visual effect for each individual image. As a comparison, our proposed model can adaptively learn haze-related image priors. Through discriminative learning, we can reduce the number of iterations of the optimization algorithm to only 2 or 3, which greatly reduces the time cost. In the meanwhile, hyper-parameters in

the energy model are also learned during training, saving the effort of manually tuning.

C. DEEP UNFOLDING NETWORKS

Recently, there have been several works to solve image inverse problems under the iterative deep learning framework [22], [24], [33], [39]–[45]. Zhang *et al.* [40] train a set of effective denoisers and plug them in the scheme of the half-quadratic splitting algorithm as modules. Meinhardt *et al.* [41] solve the inverse problem in image processing using the primal-dual hybrid gradient method, and replace the proximal operator with a denoising neural network. In [39], [42], [43], the linear inverse problems are solved by learning proximal operators in the scheme of iterative optimization algorithms. These methods can well solve linear inverse problems such as denoising, super-resolution, 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 the imaging model. Instead of using common linear inverse models in these works, we specify single image dehazing as a non-linear inverse problem with regularization terms for haze-related features. We propose to discriminatively learn effective image priors by learning proximal mappings for the regularization terms using CNNs. The most related works to ours are [24], [45]. In [24], the authors learned deep priors for single image dehazing, but did not investigate image priors on haze-related features. In [45], the authors learned image priors for deraining problem based on the model-driven approach. To the best of our knowledge, our work [22] is the first to learn haze-related priors for image dehazing task.

III. DEHAZING AS AN INVERSE PROBLEM

In this section, we first build an energy function with physical model constraints in both image space and feature space, and then design an iterative algorithm for energy minimization based on the half-quadratic splitting (HQS) algorithm.

A. 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 subtract A from both sides of Eqn. (1) in each color channel:

$$I^c(x) - A^c = (J^c(x) - A^c)T(x),\ c \in \{r, g, b\}, \tag{4}$$

where c is the color channel. For simplicity, let $P^c = I^c - A^c$ and $Q^c = J^c - A^c$. Then P and Q represent the normalized hazy image and clean image respectively. Thus, Eqn. (4) can be rewritten in a concise form as:

$$P^c = Q^c \circ T,\ c \in \{r, g, b\}, \tag{5}$$

where $\circ$ is the Hadamard product for matrices. This is a physical constraint in image color space.

Now we consider physical constraint in haze-related feature space. Let Φ_h be a transformation from image space to any haze-related feature space, and we simply apply Φ_h on both sides of Eqn. (5):

$$\Phi_h(P) = \Phi_h(Q \circ T), \tag{6}$$

in which we let T element-wisely multiply Q channel by channel. There are various choices for transformation Φ_h, such as dark channel, local max contrast, hue disparity and local max saturation, as mentioned by Tang *et al.* [15]. However, dark channel has been proved to be the most effective [15], we therefore take Φ_h as the dark channel of an image. Thus, we physically constrain our model in the dark channel feature space. According to He [4], we can further assume that T is locally constant, then we have

$$P^d = Q^d \circ T, \tag{7}$$

where P^d, Q^d are the dark channels of P, Q.

By enforcing Eqns. (5) and (7) as data fidelity terms, we design a dehazing energy function:

$$E(Q, T) = \frac{\alpha}{2}\sum_c \|P^c - Q^c \circ T\|_F^2 + \frac{\beta}{2}\|P^d - Q^d \circ T\|_F^2 + f(Q^d) + g(T) + h(Q), \tag{8}$$

where α and β are coefficients for data terms, $\|\cdot\|_F$ is the Frobenius norm, and $f(Q^d)$, $g(T)$ and $h(Q)$ are regularization terms modeling the priors on dark channel Q^d, transmission map T and clean image Q. The optimal haze-free image Q^* and transmission map T^* can be obtained by solving the following optimization problem:

$$\{Q^*, T^*\} = \arg\min_{Q,T} E(Q, T). \tag{9}$$

Regularization Terms: We have three regularization terms f, g and h that respectively model dark channel prior, transmission prior and clean image prior. Multiple image priors can be taken for them. For example, f for the dark channel can be taken as ℓ_0 or ℓ_1 regularizer, enforcing the dark channel to be sparse and close to zero. The transmission map is closely related to the latent scene depth, which is piecewise-smooth and edge-aligned with the depth, thus its regularizer g can be modeled by MRF [46], [47], or TGV [10], [48]. For clean image prior h, we can use common image priors like TV [49].

B. MODEL OPTIMIZATION

It is non-trivial to directly solve optimization problem Eqn. (9), so we turn to the half-quadratic splitting (HQS) algorithm to break it into easier sub-problems. The HQS algorithm has been widely used to solve image inverse problems [50]–[54]. By introducing an auxiliary variable U to substitute Q^d, i.e., the dark channel of the latent haze-free image, we derive the augmented energy function:

$$E(Q, T, U) = \frac{\alpha}{2}\sum_c \|Q^c \circ T - P^c\|_F^2 + \frac{\beta}{2}\|U \circ T - P^d\|_F^2 + \frac{\gamma}{2}\|U - Q^d\|_F^2 + f(U) + g(T) + h(Q), \tag{10}$$

in which γ is a penalty weight, and when $\gamma \to \infty$, the solution of minimizing Eqn. (10) converges to that of minimizing Eqn. (8). We minimize Eqn. (10) by alternately updating U, T and Q while fixing the other two variables. We initialize $Q_0 = P$ and all elements of T_0 are ones, then for the n-th iteration of the HQS algorithm, we successively solve the following sub-problems.

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_n for the n-th iteration is updated as:

$$U_n = \arg\min_U \frac{\beta}{2}\|U \circ T_{n-1} - P^d\|_F^2 + \frac{\gamma}{2}\|U - Q_{n-1}^d\|_F^2 + f(U), \quad (11)$$

from which we can derive

$$U_n = \text{prox}_{\frac{1}{u_n}f}(\hat{U}_n), \quad (12)$$

where $\hat{U}_n$ is an intermediate variable defined as:

$$\hat{U}_n = \frac{\beta T_{n-1} \circ P^d + \gamma Q_{n-1}^d}{\beta T_{n-1} \circ T_{n-1} + \gamma}, \quad (13)$$

and $u_n = \beta T_{n-1} \circ T_{n-1} + \gamma$. The proximal operator prox [55] we used is defined as:

$$\text{prox}_{\lambda f}(V) = \arg\min_X \frac{1}{2}\|X - V\|_F^2 + \lambda f(X), \quad (14)$$

assuming that $f(X)$ is separable for different elements in a matrix X such that $f(X) = \sum_i f(x_i)$. This assumption is reasonable for many common regularizations such as ℓ_1 or ℓ_2. In practice, we relax this constraint and extend it to general regularizations. In addition, we also extend λ to be a matrix with the same size as X.

Update T: We next update the transmission map T_n. Given Q_{n-1} and U_n, T_n is computed as:

$$T_n = \arg\min_T \frac{\alpha}{2}\sum_c \|Q_{n-1}^c \circ T - P^c\|_F^2 + \frac{\beta}{2}\|U_n \circ T - P^d\|_F^2 + g(T). \quad (15)$$

Then we can derive

$$T_n = \text{prox}_{\frac{1}{t_n}g}\left(\hat{T}_n\right), \quad (16)$$

where $\hat{T}_n$ is an intermediate variable defined as:

$$\hat{T}_n = \frac{\alpha \sum_c Q_{n-1}^c \circ P^c + \beta U_n \circ P^d}{\alpha \sum_c Q_{n-1}^c \circ Q_{n-1}^c + \beta U_n \circ U_n}, \quad (17)$$

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

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

$$Q_n = \arg\min_Q \frac{\alpha}{2}\sum_c \|Q^c \circ T_n - P^c\|_F^2 + \frac{\gamma}{2}\|Q^d - U_n\|_F^2 + h(Q). \quad (18)$$

Since computing the dark channel of an image is to extract the smallest value from the local color patch around each pixel, the second term of Eqn (18) only constrains on limited pixels in the original image Q, which may cause unstable results. Therefore we ignore the second data term, and the clean image Q_n is updated as:

$$Q_n = \text{prox}_{\frac{1}{q_n}h}(\hat{Q}_n), \quad (19)$$

where $\hat{Q}_n = \frac{P}{\max(T_n, \epsilon)}$, $q_n = T_n$, and ϵ is a constant to prevent extreme low transmission values.

After N iterations, the final haze-free image J can be derived by adding Q_N with A in each color channel: $J^c = Q_N^c + A^c$. In summary, the iterative procedure for optimizing the proposed energy model with the HQS algorithm is shown as Algorithm 1.

Algorithm 1 Energy Minimization for Single Image Dehazing With Half-Quadratic Splitting

Input: Hazy image I, iteration number N
Output: Transmission T, haze-free image J

```
Estimate atmospheric light A from I.
Let P^c = I^c − A^c, c ∈ {r, g, b}.
Let Q_0^c(x) := P^c(x), T_0(x) := 1, γ := 1.
for n = 1 : N do
    Update U_n based on Eqn. (12).
    Update T_n based on Eqn. (16).
    Update Q_n based on Eqn. (19).
    Update γ := δγ, δ > 1.
end for
return A, T = T_N, J^c = Q_N^c + A^c.
```

For an instance of our energy model, we concretize these regularization terms in Eqn (10). Specifically, for dark channel U, we use ℓ_1 regularization for sparse and low pixel values. For transmission T, we use ATGV [10], [48] regularization for smooth and edge-aligned maps. For clean image Q, we do not add any constraints, i.e.,

$$f(U) = \|U\|_1,$$
$$g(T) = \text{ATGV}(T),$$
$$h(Q) = 0.$$

ATGV is the anisotropic total generalized variation:

$$\min_{T,V} \alpha_1\|D^{\frac{1}{2}}(\nabla T - V)\|_1 + \alpha_0\|\nabla V\|_1 + \mathbf{w}\|T - \hat{T}\|_1,$$

where $\hat{T}$ is the initial transmission map, $D^{\frac{1}{2}}$ is the anisotropic diffusion tensor decided by guide image P. The dehazing result is shown as *Ours-EM* (energy minimization) in Figure 3. We also demonstrate several other energy-based methods as comparisons, and we can see that our energy model is effective in removing hazes.

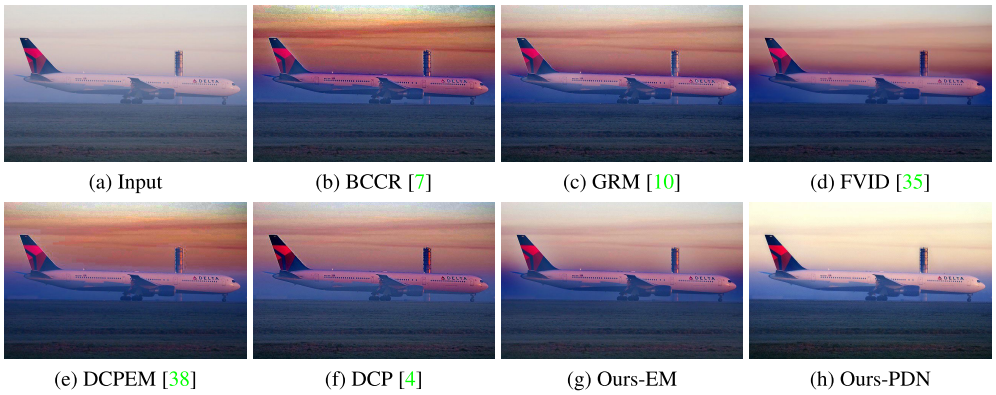


FIGURE 3. Dehazing results on a real-world hazy image of our energy-based method as well as four energy minimization based methods. We also show the results of the original dark channel prior and our proximal dehaze-net as comparisons. Ours-EM can achieve satisfactory result while our learning-based model further improves the visual effect and processing speed.

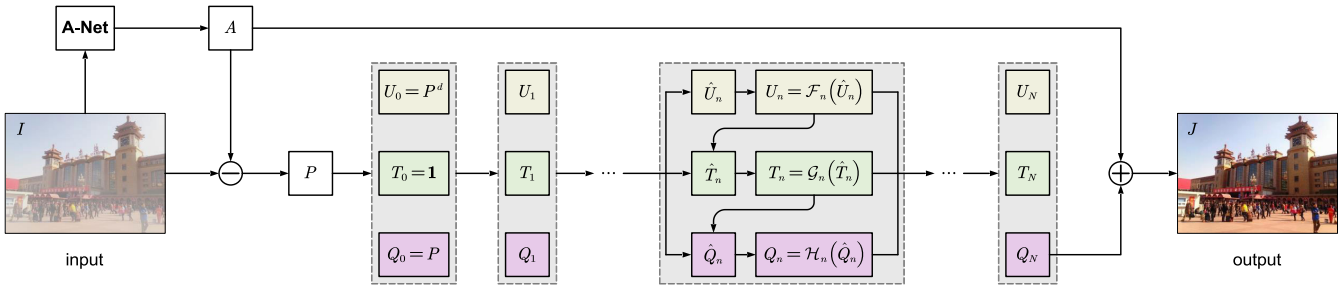


FIGURE 4. The architecture of proximal dehaze-net. We first estimate the atmospheric light A from input hazy image I by A-Net, then subtract A from I in each color channel to get the scaled hazy image P ($P^c = I^c - A^c$), which will be sent into a multi-stage learning framework to predict the haze-free image. For each stage of the learning framework, we first calculate the auxiliary variables $\hat{U}_n$, $\hat{T}_n$ and $\hat{Q}_n$ following Algorithm 1 and then learn corresponding proximal mappings with CNNs. $\mathcal{F}_n$, $\mathcal{G}_n$, and $\mathcal{H}_n$ in the gray dashed box are sub-modules to be learned at the n-th stage. After N stages, we get the final predictions U_N, T_N and Q_N. We add A back to Q_N to reconstruct the clean image J. The whole framework can be trained end-to-end.

IV. PROXIMAL DEHAZE-NET

Although our energy-based method can remove hazes effectively, we can further improve it in visual performance and processing speed by learning haze-related image priors. Instead of designing image priors by hand according to human experiences, we model haze-related priors with convolutional neural networks via learning proximal mappings appeared in Section III-B. Note that the above introduced optimization process requires the atmospheric light A to be known in advance. In the conference version of our work [22], we refer to [4] to pre-estimate the atmospheric light A. In this paper, to estimated A more accurately, we introduce an extra convolutional neural network to predict the atmospheric light. Combining the learning of A and the proposed iterative optimization algorithm, we build a deep learning framework for single image dehazing, denoted by *proximal dehaze-net*, as illustrated in Figure 4.

The core part of this framework is an architecture with N stages implementing N iterations of Algorithm 1 for solving Eqn. (10). As shown in Figure 4, the atmospheric light A is first estimated by A-Net, then we subtract A from I in each color channel to obtain P, which is an normalized hazy image within the range of $[-1, 1]$. P is sent into an N-stage learning framework. For the n-th stage, it takes the outputs of the previous stage U_{n-1}, T_{n-1}, Q_{n-1} as inputs, and learns to predict U_n, T_n and Q_n for the current stage.

As mentioned above, instead of designing image priors by hand, we model them by using CNNs to learn their corresponding proximal operators $\text{prox}_{\frac{1}{u_n}f}$, $\text{prox}_{\frac{1}{t_n}g}$ and $\text{prox}_{\frac{1}{q_n}h}$ for updating U_n, T_n and Q_n in each stage, i.e.,

$$\begin{aligned} U_n &= \text{prox}_{\frac{1}{u_n}f}(\hat{U}_n) \triangleq \mathcal{F}_n(\hat{U}_n), \\ T_n &= \text{prox}_{\frac{1}{t_n}g}(\hat{T}_n) \triangleq \mathcal{G}_n(\hat{T}_n), \\ Q_n &= \text{prox}_{\frac{1}{q_n}h}(\hat{Q}_n) \triangleq \mathcal{H}_n(\hat{Q}_n), \end{aligned} \tag{20}$$

where $\mathcal{F}_n$, $\mathcal{G}_n$ and $\mathcal{H}_n$ are sub-modules to be learned for representing the corresponding proximal operators at the n-th stage. In this way, we design an end-to-end training architecture, dubbed as proximal dehaze-net (PDN). As shown in Figure 4, each stage of the proximal dehaze-net implements one iteration of model optimization discussed in Section III-B, and the proximal operators are substituted by convolutional neural networks as in Eqn. (20). Note that in Eqn. (14), we assume that the regularization functions are separable, while in this section, we relax this constraint to make full use of the representation ability of convolutional neural networks. We now introduce the structures of these sub-modules for each stage.

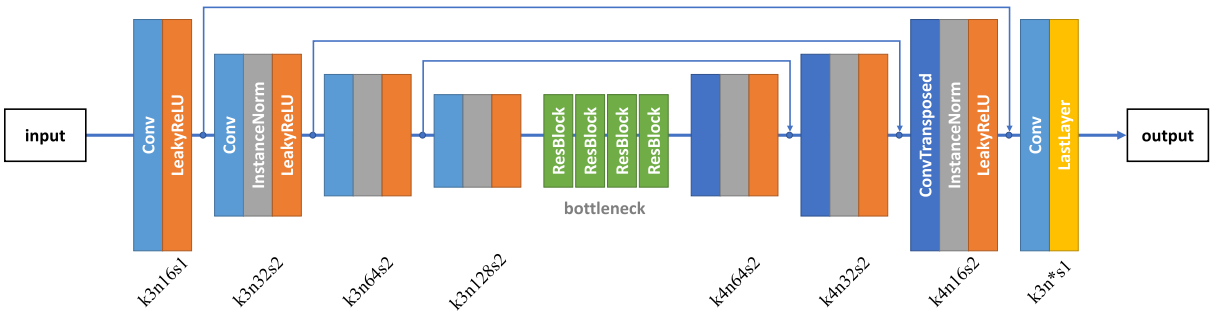


FIGURE 5. Sub-network structure. We adopt residual encoder-decoder (RED) as the base architecture of sub-networks in our model. We use 2-stride convolution for down-sampling and 2-stride transposed convolution for up-sampling. For A-Net and T-Net, the last layer is Sigmoid. For U-Net and Q-Net, the last layer is Tanh. The bottleneck of RED is stacked residual blocks.

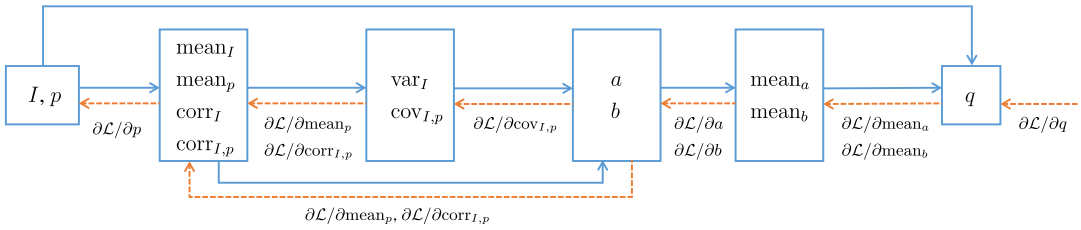


FIGURE 6. Computation graph of GIF block. Based on [56], we construct its differentiable computation graph. GIF block takes image $\boldsymbol{p}$ and guide image $\boldsymbol{I}$ as inputs, outputs the filtered image $\boldsymbol{q}$. In our paper, we set $\boldsymbol{p}$:= T-Net($\hat{\boldsymbol{T}}$), $\boldsymbol{I}$:= $\boldsymbol{P}$ and $\boldsymbol{q}$:= $\boldsymbol{T}$.

At the n-th stage, $\hat{U}_n$ is first computed by Eqn. (13), then sent into a convolutional neural network, U-Net$_n$, to learn sub-module $\mathcal{F}_n$. The updated dark channel is:

$$U_n = \mathcal{F}_n(\hat{U}_n) \triangleq \text{U-Net}_n([\hat{U}_n, P]), \tag{21}$$

in which we concatenate $\hat{U}_n$ with hazy image P as the input of U-Net$_n$ to prevent information loss.

Similarly, $\hat{T}_n$ is first computed by Eqn. (17), concatenated with P and then sent into a convolutional neural network, T-Net$_n$, to learn sub-module $\mathcal{G}_n$. Note that T-Net$_n$ is followed by a GIF block (performing guided image filtering) to ensure edge alignment between the transmission map and the original image. The updated transmission map is computed as:

$$\begin{aligned} \hat{T}_{n'} &= \text{T-Net}_n\big([\hat{T}_n, P]\big), \\ T_n &= \text{GIF}\big(\hat{T}_{n'}, P\big), \end{aligned} \tag{22}$$

where $\hat{T}_{n'}$ is the output of T-Net$_n$, and GIF$(\hat{T}_{n'}, P)$ is to perform guided image filtering [56] on image $\hat{T}_{n'}$ with P as the guidance image. Thus, the sub-module $\mathcal{G}_n$ can be represented as the composition of T-Net$_n$ and the GIF block:

$$T_n = \mathcal{G}_n(\hat{T}_n) \triangleq \text{GIF}\big(\text{T-Net}_n([\hat{T}_n, P]), P\big). \tag{23}$$

Finally, for Q_n, the intermediate variable $\hat{Q}_n$ is first computed using Eqn. (19), and then sent into a convolutional neural network Q-Net$_n$ concatenated with P to perform proximal mapping $\text{prox}_{\frac{1}{q_n}h}$. The updated clean image is:

$$Q_n = \mathcal{H}_n(\hat{Q}_n) \triangleq \text{Q-Net}_n([\hat{Q}_n, P]), \tag{24}$$

where Q_n is the estimation of the haze-free image with A subtracted at the n-th stage. U_n, T_n and Q_n then serve as the inputs of the $(n+1)$-th stage of our proximal dehaze-net.

After N stages, we obtain the final outputs U_N, T_N and Q_N, and we use Q_N and the predicted atmospheric light A to reconstruct the haze-free image as shown in Figure 4.

Sub-Networks: Our proximal dehaze-net includes four kinds of sub-networks, i.e., A-Net, U-Net, T-Net and Q-Net, and they share similar structures. As shown in Figure 5, we adopt the commonly used residual encoder-decoder (RED) as the base architecture for these sub-networks. Accordingly, these sub-networks consist of several stacked down-sampling convolution blocks for the encoder part and up-sampling convolution blocks for the decoder part. The bottleneck is made up of stacked residual blocks [57], [58]. Skip connections between the encoder and the decoder are used to prevent from losing spatial information. For the last layer of an RED, we use Sigmoid for A-Net and T-Net, since the outputs are within the range of [0, 1], and we use Tanh for U-Net and Q-Net since their outputs are within the range of $[-1, 1]$.

GIF Block: GIF block stands for guided filtering [56] computation block within our proximal dehaze-net. GIF block enforces the transmission map learned by T-Net to be well aligned with the image in edges. It takes the hazy image P as guidance and performs guided image filtering on the output of T-Net. As stated in [56], the process of GIF consists of a series of average filtering operations and simple element-wise operations. The computation graph of GIF block is shown in Figure 6, and more details of the GIF algorithm can be

found in [56]. Many previous works [16], [17], [56] use GIF as a post-processing of the estimated transmission map. On the contrary, we include the process of guided image filtering within our end-to-end trainable system. For GIF block implementation, we adopt the work of Wu *et al.* [59] and we will discuss the necessity of GIF block later.

Loss Functions: To train our proximal dehaze-net, we introduce commonly used loss functions, including ℓ_1 reconstruction loss, total variation (TV) loss and structure similarity (SSIM) loss on network outputs and the corresponding ground truths. Specifically, the atmospheric light loss is:

$$L_A = \|A^* - A^{gt}\|_1 + \varepsilon_1 \, TV(A^*) + \varepsilon_2 \, L_s(A^*, A^{gt}),$$

where A^* is the atmospheric light map estimated by A-Net and $L_s(\cdot,\cdot) = 1 - \text{SSIM}(\cdot,\cdot)$ defines the structure similarity loss. As for U, T and Q, the loss function is defined as:

$$L_z = \sum_{n=1}^{N} \|z_n - z^{gt}\|_1 + \varepsilon_1 \, TV(z_n) + \varepsilon_2 \, L_s(z_n, z^{gt}),$$

where $z \in \{U, T, Q\}$, $Q_n^{gt} = J^{gt} - A^{gt}$, N is the number of PDN stages, ε_1 and ε_2 are coefficients for TV loss and SSIM loss. Thus the total loss function is:

$$L = \lambda_A L_A + \lambda_U L_U + \lambda_T L_T + \lambda_Q L_Q,$$

where λ_A, λ_U, λ_T, λ_Q are weights for these loss terms. Empirically, for all experiments, the loss coefficients ε_1 for total variation loss is set to 0.1 for L_A, 0.01 for L_U and L_T, and 0.001 for L_Q. Coefficients ε_2 for SSIM loss are set to 0.01 for all loss terms. Finally we set the coefficients $\lambda_A = \lambda_U = \lambda_T = \lambda_Q = 1$.

V. EXPERIMENTS

To verify the effectiveness of the proposed proximal dehaze-net, we evaluate our method on different datasets and compare it with other single image dehazing methods.

A. DATASETS

We evaluate our proximal dehaze-net on multiple benchmark datasets for single image dehazing, including RESIDE dataset [60] and NTIRE 2018 single image dehazing challenge [61]. Both RESIDE and NTIRE 2018 datasets consist of indoor and outdoor subsets.

1) RESIDE DATASET

RESIDE dataset consists of indoor and outdoor datasets. The indoor dataset contains 13990 generated hazy/clean training images and 500 test images. The outdoor dataset contains about 8400 haze-free images with depth maps that can be used to synthesize training pairs and 500 hazy images as the test set. Note that for the outdoor dataset, we first remove redundant images from the training set that are overlapped with the test set. For each dataset, we randomly crop 64000 patches of 256 × 256 as the training set and apply horizontal/vertical flipping and random rotation as data augmentation. Considering the natural gap between indoor and outdoor images, we individually train two models for indoor and outdoor situations.

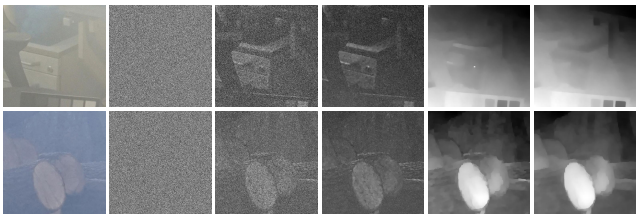


FIGURE 7. Intermediate results of transmission maps produced by our proposed data-preprocessing.

2) NTIRE 2018 DATASET

NTIRE 2018 single image dehazing challenge also consists of two subsets, I-Hazy (indoor hazy images) and O-Hazy (outdoor hazy images). I-Hazy contains 25 training images and 5 test images while O-Hazy contains 35 training images and 5 test images. Similar to RESIDE, we train two models separately for indoor and outdoor images.

3) DATA PRE-PROCESSING

To effectively train our PDN model, we need ground truths $\{A^{gt}, U^{gt}, T^{gt}, Q^{gt}\}$. For RESIDE dataset, A^{gt}, T^{gt} and J^{gt} are directly provided, $Q^{gt} = J^{gt} - A^{gt}$, and we compute the ground truth U^{gt} as the dark channel of $J^{gt} - A^*$. However, for I-Hazy and O-Hazy, only J^{gt} is available. To handle this problem, we first treat A and T as unknowns and minimize the following energy function to obtain A and T:

$$E(A, T) = \frac{1}{2} \sum_c \|J^c \circ T + A^c(1 - T) - I^c\|_F^2 + \lambda \text{TV}(T), \tag{25}$$

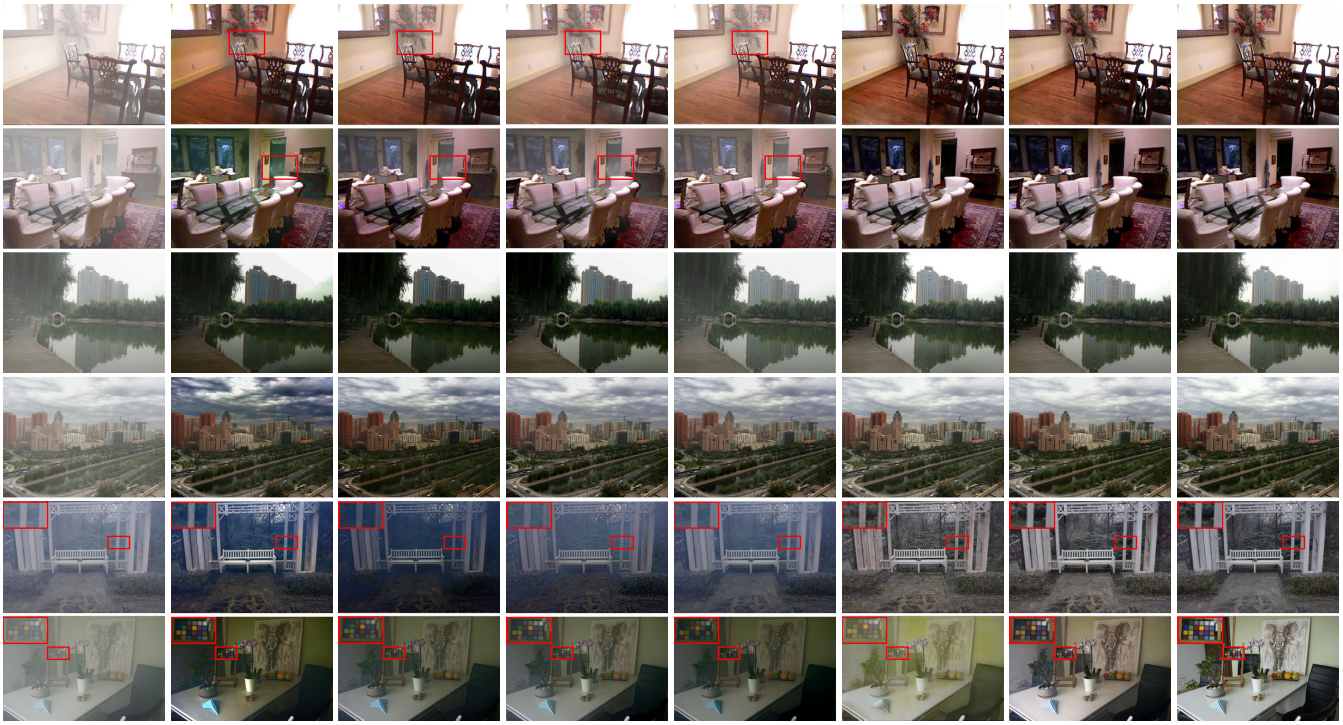
where $A \in \mathbb{R}^3$ and $T \in \mathbb{R}^{p \times p}$, which means that we compute A and T within a $p \times p$ local patch ($p = 512$ in our experiments). To show the effectiveness of the proposed pre-processing method, in Figure 7, we illustrate two examples of the intermediate results of the recovered transmission maps. Although the accuracy of recovered transmission maps is restricted since real-world hazy images do not abide strictly by Eqn. (1), we can claim that the results are reasonable to a certain extent and sufficient to train our dehazing model. We use simple gradient descent algorithm to solve the above problem, and after we have A and T, we prepare training data with the same setting as RESIDE.

B. IMPLEMENTATION DETAILS

We implement and train our PDN model with PyTorch [62] framework. Note that hyper-parameters $[\alpha, \beta, \gamma]$ are also trained simultaneously and they are initialized with 1. Since all hyper-parameters should be positive, we learn the exponent of them instead of themselves. We choose the Adam

TABLE 1. Dehazing results on SOTS (indoor and outdoor) and NTIRE 2018 (I-hazy and O-hazy) datasets. We show the average peak signal-to-noise ratio (PSNR) and structure similarity (SSIM) of the recovered images by the compared methods. Best results are shown in red and the second best are shown in blue.

Method	DCP [4]	CAP [8]	NLD [9]	GRM [10]	MSCNN [16]	DehazeNet [17]	AODNet [18]	DcGAN [21]	GDN [23]	FDGAN [27]	DADN [28]	PDN
SOTS-In	16.62	19.05	17.29	16.63	17.57	21.41	19.56	25.58	32.16	28.21	27.76	33.55
	0.818	0.836	0.749	0.705	0.810	0.847	0.831	0.938	0.984	0.949	0.930	0.982
SOTS-Out	17.54	15.18	18.09	17.47	18.06	24.75	23.07	24.62	30.86	23.32	26.45	30.97
	0.848	0.760	0.772	0.820	0.809	0.930	0.915	0.889	0.982	0.924	0.896	0.973
I-Haze	11.79	14.01	12.28	15.03	13.82	14.56	15.41	16.06	18.53	17.61	17.66	21.23
	0.449	0.524	0.474	0.593	0.591	0.549	0.722	0.789	0.777	0.716	0.773	0.882
O-Haze	14.79	15.40	15.32	17.93	16.49	16.65	17.86	19.53	21.64	18.07	22.51	24.20
	0.367	0.349	0.421	0.419	0.434	0.413	0.646	0.850	0.733	0.730	0.747	0.868



(a) Input (b) DCP [4] (c) CAP [8] (d) DehazeNet [17] (e) AODNet [18] (f) GDN [23] (g) Ours (h) Clear

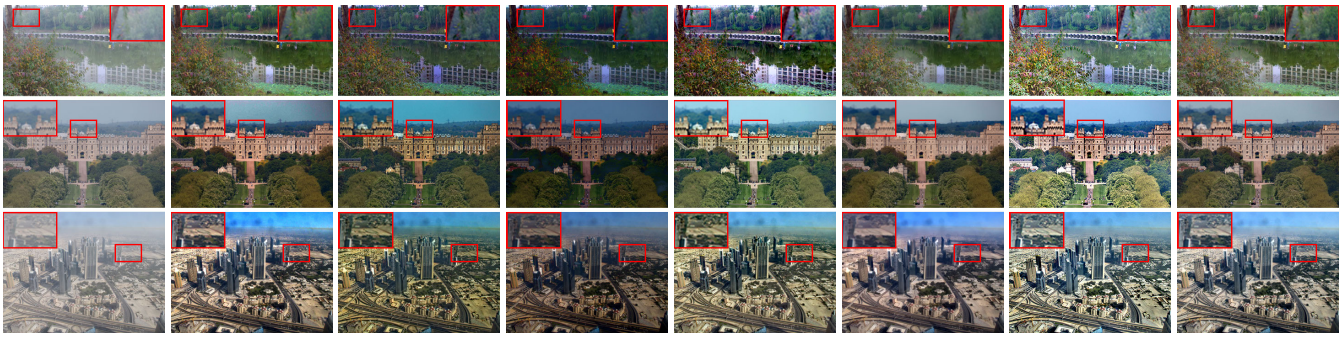
FIGURE 8. Dehazing results on synthetic images from multiple benchmark datasets.

optimizer with default parameters, and the initial learning rate is set to 10^{-4}, which will be decreased by multiplying a factor of 0.75 every 10 epochs. We use a batch size of 16 and it takes about 3 days to train a single-stage network for 100 epochs on a Titan X GPU. For multi-stage networks, we first initialize the weights with the pre-trained one-stage network and then continue the training.

C. RESULTS ON SYNTHETIC DATASETS

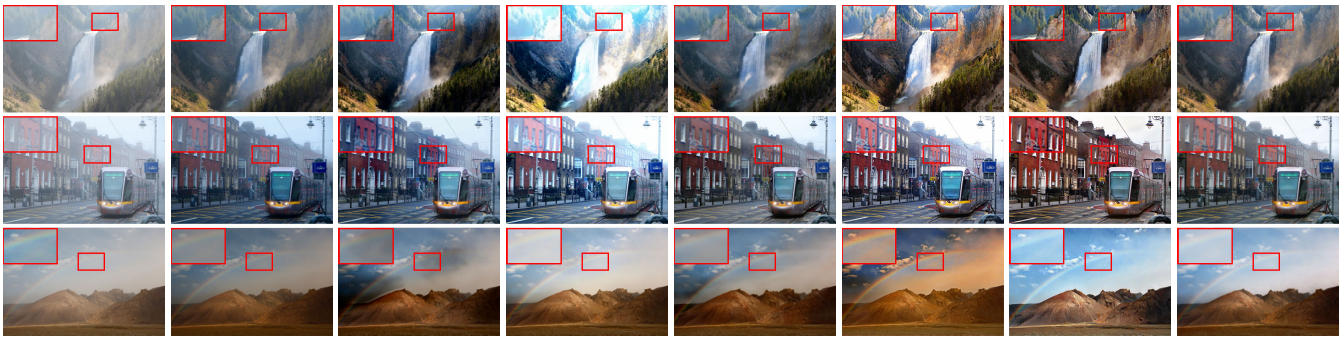
We first evaluate our proximal dehaze-net on synthetic datasets and compare it with other single image dehazing methods. We select some representative single image dehazing methods, including DCP [4], CAP [8], NLD [9], GRM [10], IDE [11], MSCNN [16], DehazeNet [17], AODNet [18], DcGAN [21], GDN [23], DADN [28] and FDGAN [27]. Among these methods, DCP, CAP, NLD, GRM and IDE are traditional image processing methods based on image priors. MSCNN, DehazeNet, AODNet, DcGAN, GDN, DADN and FDGAN are deep learning methods that predict either transmission maps or clean images. For fair comparisons, we *retrained* these learning-based methods on the corresponding datasets if the training codes are provided by authors. We evaluate these methods on SOTS and NTIRE 2018 datasets. Both datasets consist of indoor and outdoor subsets. We report the peak-signal-to-noise ratio (PSNR) and structural similarity (SSIM) as performance metrics.

As shown in Table 1, the learning-based methods are usually quantitatively superior to the traditional prior-based methods. Our proposed PDN model surpasses most methods and achieves competitive PSNR and SSIM values on multiple benchmark datasets. Specifically, on SOTS indoor and SOTS outdoor datasets, we achieve the highest PSNRs and



(a) Input (b) DCP [4] (c) BCCR [7] (d) CAP [8] (e) NLD [9] (f) GRM [10] (g) IDE [11] (h) Ours

FIGURE 9. Dehazing results of the prior-based methods on real-world images.



(a) Input (b) AODNet [18] (c) GFN [19] (d) DCPDN [20] (e) DcGAN [21] (f) FDGAN [27] (g) DADN [28] (h) Ours

FIGURE 10. Dehazing results of the learning-based methods on real-world images.

comparable SSIMs with state-of-the-art method GDN [23]. However, on realistic dehazing datasets, NTIRE 2018 (I-Hazy and O-Hazy in Table 1), we achieve the best PSNRs and SSIMs among compared methods on both indoor and outdoor cases. The dehazing results on NTIRE datasets also verify the effectiveness and rationality of the proposed data pre-processing algorithm to generate atmospheric lights and transmission maps from given hazy/clean image pairs.

As a visualization, we show some dehazing examples from these benchmark datasets in Figure 8. The 1st and 2nd examples are from SOTS indoor dataset, the 3rd and 4th examples are from SOTS outdoor dataset, and the last two examples are from I-Hazy and O-Hazy datasets respectively. From Figure 8, we can see that learning-based methods usually behave better than the traditional prior-based methods in visual effect. The traditional methods (DCP, CAP) sometimes over dehaze the images and cause darkened images and color distortion. On the other hand, the learning-based methods can produce dehazed images that are closer to ground truths. However, the learning-based methods sometimes fail to effectively remove all hazes on an image, such as DehazeNet and AODNet. Recent methods, such as GDN, are powerful in removing hazes from synthetic images but behave not that well on realistic images, e.g., the last two examples in Figure 8. As a comparison, our method can remove hazes effectively while keeping dehazed images visually pleasing.

D. RESULTS ON REAL DATASETS

In Figures 9 and 10, we respectively show the dehazing results of our method on real-world hazy images by comparing with the traditional methods and the learning-based methods. On one hand, as we can see from Figure 9, the traditional methods are usually effective in removing hazes due to the investigation of useful haze-related image priors. However, they tend to overly enhance the hazy images, which causes over-saturation or color distortion, such as BCCR [7], CAP [8], NLD [9] and IDE [11]. DCP [4] are likely to produce undesirable artifacts, especially in the sky regions. GRM [10] can well suppress artifacts but will lose detailed textures due to its smooth regularization term. On the other hand, the learning-based methods are usually able to produce more visually pleasing results, as shown in Figure 10. However, MSCNN [16] and DehazeNet [17] are not always as effective as the traditional methods in haze removal. AODNet [18] often produces darken images. DCPDN [20] predicts inexact transmission maps in areas with high intensity. GFN [19] and DcGAN [21] directly predict haze-free images and sometimes cause unexpected dark artifacts due to the lack of haze imaging model constraint. FDGAN [27] and DADN [28] can produce more visual-pleasing results, but they both rely on photo-realistic synthetic training data. By learning haze-related image priors, our method combines the advantages of the traditional methods and the learning-based meth-

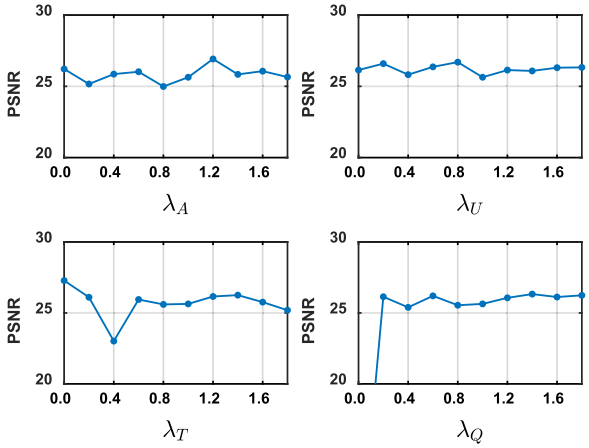


FIGURE 11. Performance of our PDN model on validation dataset with different parameter settings.

ods, being able to effectively remove hazes and keep the dehazed images natural in the meantime.

VI. DISCUSSION AND ANALYSIS

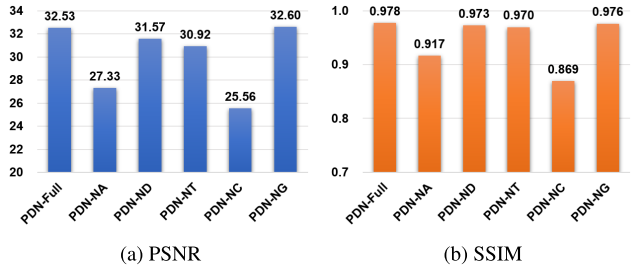
We now discuss the parameter selection, effectiveness of learning haze-related image priors as well as the necessity of GIF block. We then analyze the effect of stage numbers on the performance of our model. We also compare the running speed with the recent dehazing methods.

A. EFFECTS OF PARAMETERS FOR LOSS TERMS

As we mentioned in Section IV, we utilize a combined loss function for training our network, introducing multiple hyper-parameters for different loss terms, i.e., ε_1 for TV loss, ε_2 for SSIM loss and λ_z, $z \in \{A, U, T, Q\}$. First, we observe in experiments that TV loss and SSIM loss affect slightly the performance of our model. By removing TV and SSIM losses, the PSNR on SOTS indoor dataset decreases from 32.53 to 32.52 for 1-stage PDN, and from 33.55 to 33.41 for 2-stage PDN. Thus, we set the values of ε_1 and ε_2 empirically. Second, to evaluate the effect of λ_z, we adjust one of them from 0 to 1.8 with the step of 0.2 while fixing the other three as 1, and train for 20 epochs to investigate the performance on validation dataset (no overlapping with test dataset). As we can see from Figure 11, when $\lambda \geq 1$, our PDN model reaches a relative steady state. Thus, for simplicity, we set $\lambda_Q = \lambda_T = \lambda_U = \lambda_Q = 1$ in our experiments.

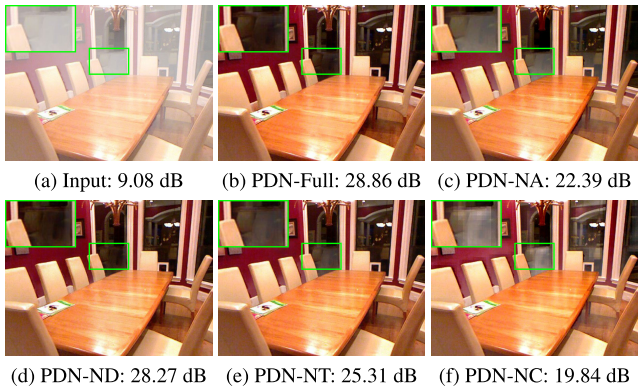
B. LEARNING IMAGE PRIORS

Our network learns multiple haze-related image priors. In this section, we discuss the effect of learning each image prior. To do so, we respectively remove dark channel prior f, transmission prior g and clean image prior h from the dehazing energy function in Eqn. (8). We then deduce new proximal dehaze-nets without learning these priors, dubbed as *Net-ND* (without learning dark channel prior), *Net-NT* (without learning transmission prior) and *PDN-NC* (without learning clean image prior). We also discuss the effect of



(a) PSNR (b) SSIM

FIGURE 12. Effectiveness of learning haze-related priors. The full PDN model with learning all haze related priors achieves the best performance.



(a) Input: 9.08 dB (b) PDN-Full: 28.86 dB (c) PDN-NA: 22.39 dB

(d) PDN-ND: 28.27 dB (e) PDN-NT: 25.31 dB (f) PDN-NC: 19.84 dB

FIGURE 13. Dehazing results of the variants of our model measured in PSNR. Omitting these prior learning modules causes that the hazes are not removed effectively, and the full model achieves the highest PSNR.

learning atmospheric light A by dropping out A-Net and pre-compute A with [4], denoted as *PDN-NA*. We train PDN-NA, PDN-ND, PDN-NT and PDN-NC with one stage from scratch on RESIDE indoor training set, and evaluate them on SOTS indoor test dataset. The quantitative results are shown in Figure 12. We can see that the PDN model is promoted by learning these image priors with CNNs. Notably, learning clean image prior and atmospheric light contribute most to the improvement by about 7 dB and 5 dB in PSNR. Learning dark channel prior and transmission prior also brings considerable performance improvements to our PDN model by about 1 dB and 1.5 dB respectively. This proves the effectiveness of learning haze-related image priors for the single image dehazing task.

As a visualization, we show an example of the dehazing results of these variants of our model in Figure 13. We can see that our full model achieves the highest PSNR, and learning all these image priors will help to remove hazes more effectively. Without learning dark channel prior or transmission prior, our model can also effectively dehaze images, but some slight hazes are still observable. Without learning clean image prior or the atmospheric light, the remaining hazes in the image are still quite obvious.

To illustrate what are learned for the proximal mappings $\mathcal{F}$, $\mathcal{G}$ and $\mathcal{H}$, in Figure 14, we show an example of the learned proximal mappings for dark channel, transmission

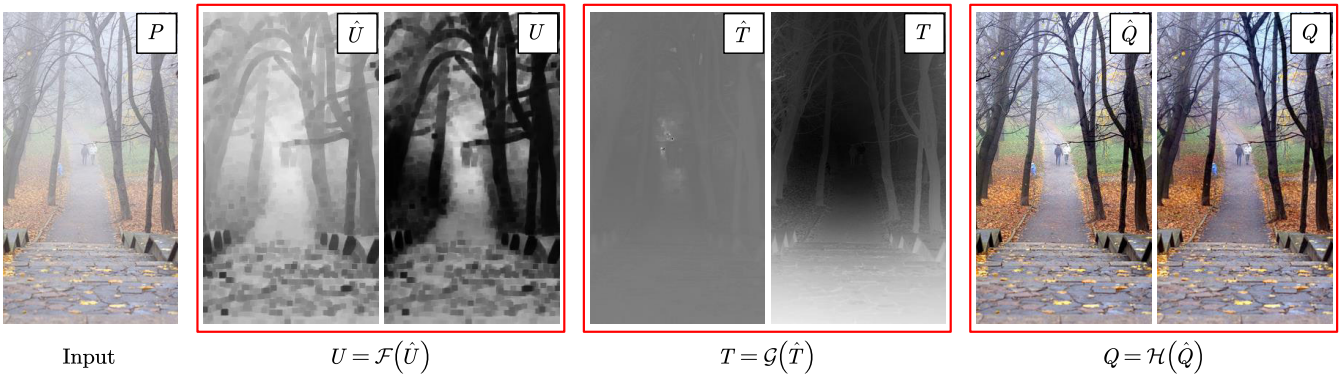


FIGURE 14. The learned hazed related image priors by our proximal dehaze-net. $\hat{U}$, $\hat{T}$, $\hat{Q}$ are dark channel, transmission and haze-free image before prior learning modules. U, T, Q are corresponding results after prior learning. We can see that the learned dark channel is darker, the learned transmission map is more accurate, and the learned haze-free image is visually better. Note that P and Q are obtained by subtracting A from I and J, and we add A back for better visualization.



FIGURE 15. Effectiveness of using GIF block. (a) Input hazy image. (b) Without using GIF block, there are obvious halo artifacts along image edges. (c) Using GIF block, our proximal dehaze-net produces haze-free image without artifacts. Please zoom in for better visualization.

map and clean image prior of our proximal dehaze-net. In Figure 14, the two figures in each of the three red boxes denotes the input and output of these learned proximal mappings. We can observe that the learned proximal mapping $\mathcal{F}$ produces reasonable dark channel U with lower values than input dark channel $\hat{U}$. The learned proximal mapping $\mathcal{G}$ produces a piecewise-smooth transmission map T that is consistent with the underlying scenery depth, which is more accurate compared with the input transmission estimation $\hat{T}$. Final proximal mapping $\mathcal{H}$ will further remove haze residuals in the estimated image $\hat{Q}$ and produce a more clear image Q.

C. NECESSITY OF GIF BLOCK

GIF block is a part of the learned proximal mapping $\mathcal{G}$. As stated in Section IV, GIF block serves as a guided image filtering. Its effect is to force edge alignment between the estimated transmission map and the original image. To verify the effectiveness of the embedded GIF block, we train a PDN without GIF block (denoted as *PDN-NG*) and evaluate PDN-NG on SOTS indoor dataset. The performance is shown in Figure 12. We can see that PDN-NG without GIF block achieves almost the same quantitative performance compared with the full PDN model. We also found that, for most images, the PDN model without GIF block behaves as well as our full model. The reason for this may be the powerful learning ability of Q-Net. However, as shown in Figure 15, for some examples, the dehazed image has visible halo effect around image edges, which degrades the image visual quality. As a comparison, the result of our full model with GIF block is more clear and natural.

D. EFFECT OF MULTI-STAGE NETWORK

Our proximal dehaze-net model is a multi-stage learning architecture based on dehazing optimization algorithm, and more network stages are supposed to achieve higher performance on the benchmarks. However, since we use residual encoder-decoder (RED) backbone as sub-networks, which is powerful in universal image restoration tasks, we are able to achieve satisfactory results with only one stage both on quantitative performances and visual effects. As shown in Table 1 and Figure 12, we use a two-stage PDN to achieve the best PSNR on SOTS dataset. Compared with the one-stage PDN, the two-stage PDN improves the PSNR values on indoor and outdoor datasets by 1.02 dB and 0.42 dB respectively. The improvements of SSIM values are less than 0.01. On NTIRE I-Hazy and O-Hazy datasets, we achieve the best performances with a one-stage PDN. Adding more stages will not continue to improve qualitative results significantly. Furthermore, the visual effects are similar among PDN models with different stages. For simplicity, we use a two-stage PDN (PDN-S2) for SOTS and a one-stage PDN (PDN-S1) for NTIRE respectively in our paper.

E. RUNNING TIME

The complex part of our model in implementation is the dark channel computation and its backward process. In the conference version of our work [22], we realize this with low-level CUDA language for high-speed computation. In [32], dark channel and its backward are realized with the look-up table technique. In [63], dark channel is calculated by extracting local patches and finding the minimum. In this paper, we find that dark channel can be simply computed as the negative one-stride max-pooling operation, which densely extracts the local maximum of a negative image, i.e., $I^d = -\text{MP}(-\min(I), w, 1)$, where $\min(I)$ is the minimum of each

channel and MP is the max-pooling operator with kernel size w and stride 1. Therefore our implementation is purely PyTorch-based and our PDN model can process images at high speed on GPU devices. In Table 2, we show the average running time (in seconds) of single image dehazing methods on 500 images in 460×620 resolution, i.e., the average time of 500 evaluations. For a fair comparison, we compute the running times on both CPU and GPU devices if possible. For GPU time, we test on a Titan X GPU device with 12 GB memories. For CPU time, we test on an Intel Xeon E5-2650 CPU @ 2.20 GHz without parallel acceleration.

TABLE 2. Computation time for different methods to dehaze a color image of 460 × 620. PDN-S1 and PDN-S2 denote our PDN model with 1 and 2 stages. We report the running times on CPU and GPU devices for fair comparison.

Method	CPU Time (s)	GPU Time (s)
DCP [4]	1.17	-
CAP [8]	0.43	-
MSCNN [16]	0.68	0.24
AODNet [18]	0.17	0.04
DcGAN [21]	0.39	0.06
GDN [23]	1.25	0.18
PDN-S1	1.89	0.20
PDN-S2	3.46	0.39

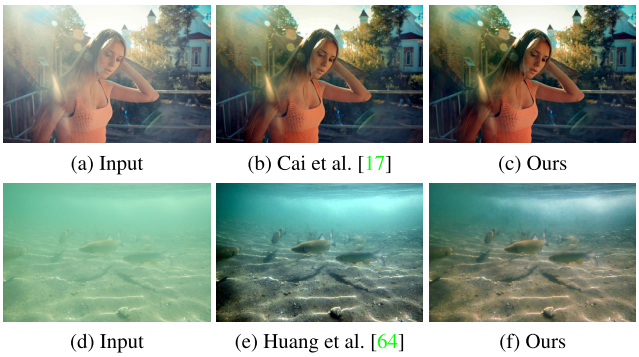
VII. OTHER APPLICATIONS AND LIMITATIONS

A. EXTENSIONS TO MORE APPLICATIONS

Though our network is trained for image dehazing, we can apply it to other similar tasks without the need to retrain on their corresponding datasets. Figure 16 (a)-(c) show an example of anti-halation enhancement using our method. Although halation has a different imaging model, it brings haze-like effects to image [17], [18]. Our proximal dehaze-net can be directly applied to anti-halation image enhancement. In Figure 16 (d)-(f), we show an example of underwater image enhancement. Ignoring the forward scattering component, the simplified underwater optical model has a similar formulation with haze imaging model [64], [65]. Compared with methods that are specifically designed for this task such as [64], our network can also effectively remove haze-like effects in this underwater image.

B. FAILED CASES

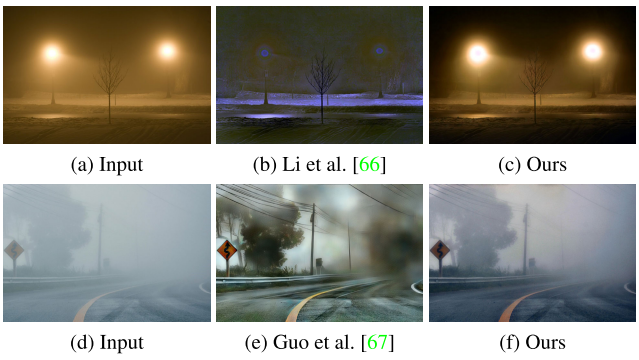
While our method behaves well on most natural images, it has limitations in certain situations where the photo is taken in the night or under heavy hazy weather. For night-time hazy images, there are usually multiple light and color sources, and the image quality is degraded by low light conditions. The commonly used haze imaging model is not sufficient to describe the night-time haze phenomenon, and our PDN model has no access to such kind of training data. As shown in Figure 17 (a)-(c), our method fails to remove all the hazes in the image. On the other hand, the method [66] designed for night-time dehazing is capable of removing hazes more



(a) Input (b) Cai et al. [17] (c) Ours

(d) Input (e) Huang et al. [64] (f) Ours

FIGURE 16. Extension to similar tasks. The first row shows an example of anti-halation enhancement. The second row shows an example of underwater image enhancement.



(a) Input (b) Li et al. [66] (c) Ours

(d) Input (e) Guo et al. [67] (f) Ours

FIGURE 17. Failure cases of our method. The first row shows an example of night-time image dehazing. The second row shows an example of heavy fog image dehazing.

adequately, but the visual quality is also lowered. As for images with very thick fog, too many details are overwhelmed by hazes, and it becomes quite difficult for current methods to achieve satisfactory results. As shown in Figure 17 (d)-(f), there still remains visible hazes in the dehazed image by our method. As a comparison, iPal-DH [67], trained on Dense Haze dataset [68], is able to dehaze images with dense haze more effectively, but it also fails to recover lost details covered by hazes.

VIII. CONCLUSION

In this paper, we propose a model-driven deep learning approach, proximal dehaze-net, for single image dehazing. We first build an energy function based on haze imaging model constraints in both image space and haze-related feature space, and then design an iterative algorithm for solving the energy model. We unfold the iterative algorithm into a multi-stage network by learning proximal operators using CNNs. The proposed proximal dehaze-net achieves promising results for single image dehazing.

Although the proposed PDN model is effective in most cases, our method shares some inherent drawbacks with other learning-based methods. First, the learning scheme is built upon the haze imaging model, which may limit the applications to more complex real-world situations, such

as non-uniform haze or night-time haze. This problem can be solved by considering imaging models that can better describe real situations and better techniques to simulate realistic datasets. Second, our method has difficulty in handling cross-domain image dehazing problems. As shown in Table 1, we have to train separate models to achieve the best performance on different benchmarks, and a model trained on one dataset may behave poorly on another dataset. This problem can be solved by the simple early stopping strategy to prevent over-fitting to a specific dataset but will decrease the performance on this dataset. There exists a trade-off between the performance on one dataset and the generalization ability of the trained model. In the future, we will consider domain adaptation to better handle this problem.

REFERENCES

[1] R. T. Tan, N. Pettersson, and L. Petersson, "Visibility enhancement for roads with foggy or hazy scenes," in *Proc. IEEE Intell. Vehicles Symp.*, Jun. 2007, pp. 19–24.

[2] R. T. Tan, "Visibility in bad weather from a single image," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2008, pp. 1–8.

[3] R. Fattal, "Single image dehazing," *ACM Trans. Graph.*, vol. 27, no. 3, p. 72, Aug. 2008.

[4] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 1956–1963.

[5] J.-P. Tarel and N. Hautiere, "Fast visibility restoration from a single color or gray level image," in *Proc. IEEE 12th Int. Conf. Comput. Vis.*, Sep. 2009, pp. 2201–2208.

[6] A. Tripathi and S. Mukhopadhyay, "Single image fog removal using anisotropic diffusion," *IET Image Processing*, vol. 6, no. 7, pp. 966–975, 2012.

[7] G. Meng, Y. Wang, J. Duan, S. Xiang, and C. Pan, "Efficient image dehazing with boundary constraint and contextual regularization," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 617–624.

[8] Q. Zhu, J. Mai, and L. Shao, "A fast single image haze removal algorithm using color attenuation prior," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3522–3533, Nov. 2015.

[9] D. Berman, T. Treibitz, and S. Avidan, "Non-local image dehazing," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1674–1682.

[10] C. Chen, M. N. Do, and J. Wang, "Robust image and video dehazing with visual artifact suppression via gradient residual minimization," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 576–591.

[11] M. Ju, C. Ding, W. Ren, Y. Yang, D. Zhang, and Y. J. Guo, "IDE: Image dehazing and exposure using an enhanced atmospheric scattering model," *IEEE Trans. Image Process.*, vol. 30, pp. 2180–2192, 2021.

[12] J. Liu, W. Liu, J. Sun, and T. Zeng, "Rank-one prior: Toward real-time scene recovery," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2021, pp. 14802–14810.

[13] Y. Y. Schechner, S. G. Narasimhan, and S. K. Nayar, "Polarization-based vision through haze," in *Proc. ACM SIGGRAPH ASIA Courses (SIGGRAPH Asia)*, 2008, p. 71.

[14] L. Shen, Y. Zhao, Q. Peng, J. C.-W. Chan, and S. G. Kong, "An iterative image dehazing method with polarization," *IEEE Trans. Multimedia*, vol. 21, no. 5, pp. 1093–1107, May 2019.

[15] K. Tang, J. Yang, and J. Wang, "Investigating haze-relevant features in a learning framework for image dehazing," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 2995–3000.

[16] W. Ren, S. Liu, H. Zhang, J. Pan, X. Cao, and M.-H. Yang, "Single image dehazing via multi-scale convolutional neural networks," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 154–169.

[17] B. Cai, X. Xu, K. Jia, C. Qing, and D. Tao, "DehazeNet: An end-to-end system for single image haze removal," *IEEE Trans. Image Process.*, vol. 25, no. 11, pp. 5187–5198, Nov. 2016.

[18] B. Li, X. Peng, Z. Wang, J. Xu, and D. Feng, "AOD-Net: All-in-one dehazing network," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 4770–4778.

[19] W. Ren, L. Ma, J. Zhang, J. Pan, X. Cao, W. Liu, and M.-H. Yang, "Gated fusion network for single image dehazing," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 3253–3261.

[20] H. Zhang and V. M. Patel, "Densely connected pyramid dehazing network," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 3194–3203.

[21] R. Li, J. Pan, Z. Li, and J. Tang, "Single image dehazing via conditional generative adversarial network," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8202–8211.

[22] D. Yang and J. Sun, "Proximal dehaze-net: A prior learning-based deep network for single image dehazing," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 702–717.

[23] X. Liu, Y. Ma, Z. Shi, and J. Chen, "GridDehazeNet: Attention-based multi-scale network for image dehazing," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 7314–7323.

[24] Y. Liu, J. Pan, J. Ren, and Z. Su, "Learning deep priors for image dehazing," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 2492–2500.

[25] Y. Qu, Y. Chen, J. Huang, and Y. Xie, "Enhanced pix2pix dehazing network," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 8160–8168.

[26] Z. Deng, L. Zhu, X. Hu, C.-W. Fu, X. Xu, Q. Zhang, J. Qin, and P.-A. Heng, "Deep multi-model fusion for single-image dehazing," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 2453–2462.

[27] Y. Dong, Y. Liu, H. Zhang, S. Chen, and Y. Qiao, "FD-GAN: Generative adversarial networks with fusion-discriminator for single image dehazing," in *Proc. AAAI Conf. Artif. Intell.*, 2020, pp. 10729–10736.

[28] Y. Shao, L. Li, W. Ren, C. Gao, and N. Sang, "Domain adaptation for image dehazing," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 2808–2817.

[29] H. Dong, J. Pan, L. Xiang, Z. Hu, X. Zhang, F. Wang, and M.-H. Yang, "Multi-scale boosted dehazing network with dense feature fusion," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 2157–2167.

[30] H. Wu, Y. Qu, S. Lin, J. Zhou, R. Qiao, Z. Zhang, Y. Xie, and L. Ma, "Contrastive learning for compact single image dehazing," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2021, pp. 10551–10560.

[31] Z. Chen, Y. Wang, Y. Yang, and D. Liu, "PSD: Principled synthetic-to-real dehazing guided by physical priors," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 7180–7189.

[32] L. Li, Y. Dong, W. Ren, J. Pan, C. Gao, N. Sang, and M.-H. Yang, "Semi-supervised image dehazing," *IEEE Trans. Image Process.*, vol. 29, pp. 2766–2779, 2020.

[33] Z. Xu and J. Sun, "Model-driven deep-learning," *Nat. Sci. Rev.*, vol. 5, no. 1, pp. 22–24, Jan. 2017.

[34] Y. Wang and C. Fan, "Single image defogging by multiscale depth fusion," *IEEE Trans. Image Process.*, vol. 23, no. 11, pp. 4826–4837, Nov. 2014.

[35] A. Galdran, J. Vazquez-Corral, D. Pardo, and M. Bertalmío, "Fusion-based variational image dehazing," *IEEE Signal Process. Lett.*, vol. 24, no. 2, pp. 151–155, Feb. 2017.

[36] Q. Liu, X. Gao, L. He, and W. Lu, "Single image dehazing with depth-aware non-local total variation regularization," *IEEE Trans. Image Process.*, vol. 27, no. 10, pp. 5178–5191, Oct. 2018.

[37] Z. G. Li and J. H. Zheng, "Single image de-hazing using globally guided image filtering," *IEEE Trans. Image Process.*, vol. 27, no. 1, pp. 442–450, Jan. 2018.

[38] M. Zhu, B. He, and Q. Wu, "Single image dehazing based on dark channel prior and energy minimization," *IEEE Signal Process. Lett.*, vol. 25, no. 2, pp. 174–178, Feb. 2018.

[39] Y. Yang, J. Sun, H. Li, and Z. Xu, "Deep ADMM-Net for compressive sensing MRI," in *Proc. Adv. Neural Inf. Process. Syst.*, 2016, pp. 10–18.

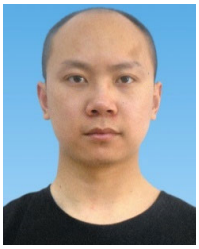
[40] K. Zhang, W. Zuo, S. Gu, and L. Zhang, "Learning deep CNN denoiser prior for image restoration," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 3929–3938.

[41] T. Meinhardt, M. Moeller, C. Hazirbas, and D. Cremers, "Learning proximal operators: Using denoising networks for regularizing inverse imaging problems," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 1781–1790.

[42] J. H. R. Chang, C.-L. Li, B. Poczos, B. V. K. V. Kumar, and A. C. Sankaranarayanan, "One network to solve them all—Solving linear inverse problems using deep projection models," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 5888–5897.

[43] J. Zhang, J. Pan, W.-S. Lai, R. W. H. Lau, and M.-H. Yang, "Learning fully convolutional networks for iterative non-blind deconvolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 3817–3825.

[44] L. Li, J. Pan, W.-S. Lai, C. Gao, N. Sang, and M.-H. Yang, "Learning a discriminative prior for blind image deblurring," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 6616–6625.
[45] H. Wang, Q. Xie, Q. Zhao, and D. Meng, "A model-driven deep neural network for single image rain removal," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 3103–3112.
[46] R. Fattal, "Dehazing using color-lines," *ACM Trans. Graph.*, vol. 34, no. 1, pp. 13:1–13:14, Dec. 2014.
[47] K. Nishino, L. Kratz, and S. Lombardi, "Bayesian defogging," *Int. J. Comput. Vis.*, vol. 98, no. 3, pp. 263–278, Jul. 2012.
[48] D. Ferstl, C. Reinbacher, R. Ranftl, M. Rüether, and H. Bischof, "Image guided depth upsampling using anisotropic total generalized variation," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 993–1000.
[49] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys. D, Nonlinear Phenomena*, vol. 60, nos. 1–4, pp. 259–268, 1992.
[50] D. Geman and G. Reynolds, "Constrained restoration and the recovery of discontinuities," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 3, pp. 367–383, Mar. 1992.
[51] D. Geman and C. Yang, "Nonlinear image recovery with half-quadratic regularization," *IEEE Trans. Signal Process.*, vol. 4, no. 7, pp. 932–946, Jul. 1995.
[52] D. Krishnan and R. Fergus, "Fast image deconvolution using hyper-Laplacian priors," in *Proc. Adv. Neural Inf. Process. Syst.*, 2009, pp. 1033–1041.
[53] Y. Wang, J. Yang, W. Yin, and Y. Zhang, "A new alternating minimization algorithm for total variation image reconstruction," *SIAM J. Imag. Sci.*, vol. 1, no. 3, pp. 248–272, Aug. 2008.
[54] L. Xu, C. Lu, Y. Xu, and J. Jia, "Image smoothing via ℓ_0 gradient minimization," in *Proc. SIGGRAPH Asia Conf.*, 2011, pp. 1–12.
[55] N. Parikh and S. Boyd, "Proximal algorithms," *Found. Trends Optim.*, vol. 1, no. 3, pp. 127–239, Jan. 2014.
[56] K. He, J. Sun, and X. Tang, "Guided image filtering," in *Proc. Eur. Conf. Comput. Vis.*, 2010, pp. 1–14.
[57] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
[58] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4700–4708.
[59] H. Wu, S. Zheng, J. Zhang, and K. Huang, "Fast end-to-end trainable guided filter," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 1838–1847.
[60] B. Li, W. Ren, D. Fu, D. Tao, D. Feng, W. Zeng, and Z. Wang, "Benchmarking single-image dehazing and beyond," *IEEE Trans. Image Process.*, vol. 28, no. 1, pp. 492–505, Jan. 2019.
[61] C. Ancuti, C. O. Ancuti, and R. Timofte, "NTIRE 2018 challenge on image dehazing: Methods and results," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2018, pp. 891–901.
[62] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, and A. Desmaison, "PyTorch: An imperative style, high-performance deep learning library," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 32, Dec. 2019, pp. 8026–8037.
[63] A. Golts, D. Freedman, and M. Elad, "Unsupervised single image dehazing using dark channel prior loss," *IEEE Trans. Image Process.*, vol. 29, pp. 2692–2701, 2019.
[64] D. Huang, Y. Wang, W. Song, J. Sequeira, and S. Mavromatis, "Shallow-water image enhancement using relative global histogram stretching based on adaptive parameter acquisition," in *Proc. Int. Conf. Multimedia Modeling*. Cham, Switzerland: Springer, 2018, pp. 453–465.
[65] C. O. Ancuti, C. Ancuti, C. De Vleeschouwer, and P. Bekaert, "Color balance and fusion for underwater image enhancement," *IEEE Trans. Image Process.*, vol. 27, no. 1, pp. 379–393, Jan. 2018.
[66] Y. Li, R. T. Tan, and M. S. Brown, "Nighttime haze removal with glow and multiple light colors," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 226–234.
[67] T. Guo, X. Li, V. Cherukuri, and V. Monga, "Dense scene information estimation network for dehazing," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2019, pp. 2122–2130.
[68] C. O. Ancuti, C. Ancuti, R. Timofte, L. V. Gool, L. Zhang, and M.-H. Yang, "NTIRE 2019 image dehazing challenge report," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2019, pp. 2241–2253.



DONG YANG received the bachelor's degree from the School of Mathematics and Statistics, Xi'an Jiaotong University, China, in 2014, where he is currently pursuing the Ph.D. degree. His research interests include image processing and deep learning.



JIAN SUN received the Ph.D. degree in applied mathematics from Xi'an Jiaotong University. He worked as a Visiting Student at Microsoft Research Asia, from November 2005 to March 2008, a Postdoctoral Researcher at the University of Central Florida, Orlando, FL, USA, from August 2009 to April 2010, and a Postdoctoral Researcher at Willow Team, Ecole Normale Superieure de Paris/INRIA, from September 2012 to August 2014. He is currently serves as a Professor with the School of Mathematics and Statistics, Xi'an Jiaotong University. His research interests include mathematics and machine learning-based approaches for image processing/recognition, and medical image analysis.