

Physics-based Feature Dehazing Networks

Jiangxin Dong¹ and Jinshan Pan²

¹ Max Planck Institute for Informatics

² Nanjing University of Science and Technology

Abstract. We propose a physics-based feature dehazing network for image dehazing. In contrast to most existing end-to-end trainable network-based dehazing methods, we explicitly consider the physics model of the haze process in the network design and remove haze in a deep feature space. We propose an effective feature dehazing unit (FDU), which is applied to the deep feature space to explore useful features for image dehazing based on the physics model. The FDU is embedded into an encoder and decoder architecture with residual learning, so that the proposed network can be trained in an end-to-end fashion and effectively help haze removal. The encoder and decoder modules are adopted for feature extraction and clear image reconstruction, respectively. The residual learning is applied to increase the accuracy and ease the training of deep neural networks. We analyze the effectiveness of the proposed network and demonstrate that it can effectively dehaze images with favorable performance against state-of-the-art methods.

Keywords: Image dehazing, physics model, feature dehazing unit, deep convolutional neural networks.

1 Introduction

Haze is a common atmospheric phenomenon. Images captured in hazy environments usually lose the color fidelity and contrast. Restoring clear images from hazy ones has been an active research effort in the computational photography and vision community within the last decade. As low-quality hazy images usually interfere with the subsequent image editing, analysis, and so on, it is of great interest to remove haze and restore high-quality images.

Mathematically, a hazy image I is usually modeled by [9, 13]

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

where J , T , and A denote the latent clear image, medium transmission, and global atmospheric light, respectively, and \mathbf{x} denotes the pixel position. This problem is highly ill-posed as only the hazy image I is available.

To make this problem well-posed, conventional methods usually develop kinds of priors based on the statistical properties of clear images, transmission, or atmospheric light to constrain the solution space. The commonly used priors include dark channel prior [13], color line prior [10], color attenuation prior [42],

sparse gradient prior [7], non-local prior [5], and so on. The methods based on these priors generate promising results. However, those aforementioned priors are manually designed based on specific observations, which do not always model the inherent properties of clear images, transmission, or atmospheric light. In addition, those aforementioned priors usually lead to highly non-convex problems that are difficult to solve.

Motivated by the success of deep learning in high-level vision tasks, the deep neural networks have been developed to overcome the limitation in image dehazing. Several approaches [6, 28] develop deep neural networks to estimate the transmission and then follow the conventional method [13] to restore clear images. However, these approaches do not correct the errors of the atmospheric light. To avoid complex estimations of the transmission and atmospheric light, end-to-end trainable deep neural networks have been developed [18, 20, 22, 23, 27]. Based on the end-to-end trainable framework, the physics model of the haze process has been utilized to constrain the networks [26, 38, 39]. Although these methods achieve decent performance, they mainly consider the physics model in the raw image space and few of them explicitly utilize the physics model in the feature space, which does not fully explore the useful feature information for image dehazing.

To overcome these problems, we propose a physics-based feature dehazing network for image dehazing. The critical component is the proposed feature dehazing unit (FDU), which is derived based on the physics model of the haze process (i.e., (1)) in the feature space, so that more feature information that is useful for the clear image reconstruction can be effectively obtained with the constraint of the physics model. The FDU is then embedded into an encoder and decoder network architecture with residual learning, where the encoder module is first adopted to extract features from the hazy image and the decoder is further adopted to process the output of the physics-based feature dehazing block for the clear image reconstruction. The residual learning is used to increase the accuracy of image dehazing and ease the training of deep neural networks. Both quantitative and qualitative evaluation results demonstrate that the proposed approach performs favorably against state-of-the-art methods. The main contributions of this work are summarized as follows:

- We propose an effective feature dehazing unit, which is derived based on the physics model of the haze process in a feature space. We show that it can explicitly utilize the physics model in the feature space and learn useful information that is required in the derived physics model to better remove the haze.
- We develop a physics-based feature dehazing network (PFDN) that embeds the feature dehazing unit into an encoder and decoder network architecture with residual learning. The encoder module is first adopted for useful feature extraction and the decoder is adopted for the final clear image reconstruction.
- We analyze the effect of the proposed PFDN on image dehazing and demonstrate that the proposed approach achieves state-of-the-art performance on both the dehazing benchmarks and real-world images.

2 Related Work

Recent years have witnessed significant advances in single image dehazing [4]. Existing methods can be roughly categorized into three kinds: adaptive contrast enhancement [11, 33], model-based [5, 7, 9, 10, 13, 42], and data driven-based methods [8, 18, 20, 22, 23, 26, 27, 34, 37–39, 42].

The adaptive contrast enhancement method (e.g., [33]) removes haze by maximizing the local contrast of restored images. These approaches are able to remove haze to some extent, but usually suffer from visual artifacts.

The model-based methods rely on the physics model of the haze process and usually make assumptions on the clear images, transmission, or global atmospheric light. For example, Fattal [9] assumes that the transmission and the surface shading are locally uncorrelated. In [13], He et al. propose a novel dark channel prior to model the properties of the hazy-free images and use this prior to estimate the transmission. Fattal [10] observes that the pixels of the image patches typically exhibit a one-dimensional distribution and proposes a color-line prior to estimate the transmission. Berman et al. [5] find that pixels in a given cluster are often non-local and they use this constraint to solve image dehazing. The priors based on the transmission have been developed by Chen et al. [7]. Although promising results have been achieved, these priors used in image dehazing are designed based on some strong assumptions and do not always hold for some applications.

The data driven-based approaches mainly use some well-known learning methods to learn the most hazy-relevant features for image dehazing. In [34], Tang et al. develop a random forest algorithm to learn priors for the transmission estimation. Zhu et al. [42] formulate the depth estimation using a linear model and learn the color attenuation prior for haze removal. These approaches are able to learn more reliable priors but still heavily depend on hand-crafted priors. Recently, the deep convolutional neural network (CNN) as an effective learning method has been developed to solve image dehazing. In [28, 6], they use deep CNNs to estimate the transmission and then use the conventional method [13] to estimate clear images. However, these approaches do not correct the errors of the atmospheric light. To avoid complex estimations of transmission and atmospheric light, several methods [18, 20, 22, 27] develop efficient deep CNNs to solve image dehazing in an end-to-end fashion. Furthermore, to make the deep end-to-end trainable CNNs more compact, the physics model of the haze process has been used [26, 38, 39]. These deep CNNs-based methods outperform the adaptive contrast enhancement approaches and model-based approaches by large margins. However, few of them discriminatively explore the features of deep CNNs in image dehazing and most of them usually consider the physics model in the raw image space instead of fully exploring the feature information in the feature space. Thus, these methods may generate results with some color distortions and artifacts.

Discriminatively exploring the features of deep CNNs has been demonstrated to be effective in some applications, such as image classification [35], image restoration [41], image super-resolution [40], etc. These methods use the gen-

eral attention mechanism to discriminatively learn useful features. However, we note that directly applying the attention mechanism in image dehazing does not effectively remove haze. Different from the attention mechanism used in these methods, we develop a physics-based feature dehazing network based on the physics model of the haze process to discriminatively learn useful information in the feature space for image dehazing.

3 Physics-based Feature Dehazing Network

Most end-to-end trainable deep neural network-based image dehazing methods take the hazy image I as the input and directly output the dehazed result. To better constrain the solution space, some image dehazing methods first estimate the key components (e.g., T and A in (1)) and then compute the final clear image using the physics model (e.g., (1)). These approaches usually utilize the physics model in the raw image space, but rarely consider it in the feature space. To overcome this problem, we develop an effective feature dehazing unit (FDU), which is able to effectively utilize the physics model of the haze process in a feature space for better image dehazing. Then the proposed FDU is embedded into an encoder and decoder architecture with residual learning. Figure 1 shows the network architecture of the proposed physics-based feature dehazing network (PFDN). In the following, we present the details of each component in the PFDN.

3.1 Feature dehazing unit

To make full use of the feature information for image dehazing, we develop a feature dehazing unit (FDU). The proposed FDU is motivated by the physics model of the haze process (1). We note that the clear image J can be obtained by

$$J(\mathbf{x}) = I(\mathbf{x})\frac{1}{T(\mathbf{x})} + A(1 - \frac{1}{T(\mathbf{x})}) \quad (2)$$

according to (1). For simplicity, let k denote a feature extractor, e.g., the filter kernel in a deep CNN. By applying k to (2), we can get

$$k \otimes J = k \otimes (I \odot \frac{1}{T}) + k \otimes A(1 - \frac{1}{T}), \quad (3)$$

where \otimes denotes the convolution operator and \odot denotes the element-wise product operation. By taking a few algebraic operations using matrix-vector forms, we can get

$$\mathbf{KJ} = \mathbf{KT}_d\mathbf{I} + \mathbf{KA}_t, \quad (4)$$

where \mathbf{K} , \mathbf{J} , \mathbf{I} , and \mathbf{A}_t denote the matrix-vector forms of k , J , I , and $A(1 - \frac{1}{T})$, respectively; \mathbf{T}_d denotes the diagonal matrix, where the i -th diagonal element corresponds to the i -th element of the vector form of $\frac{1}{T}$. As we can decompose the matrix \mathbf{KT}_d into the product of two matrices $\mathbf{F}_1\mathbf{F}_2$, (4) can be rewritten as

$$\mathbf{KJ} = \mathbf{F}_1(\mathbf{F}_2\mathbf{I}) + \mathbf{KA}_t, \quad (5)$$

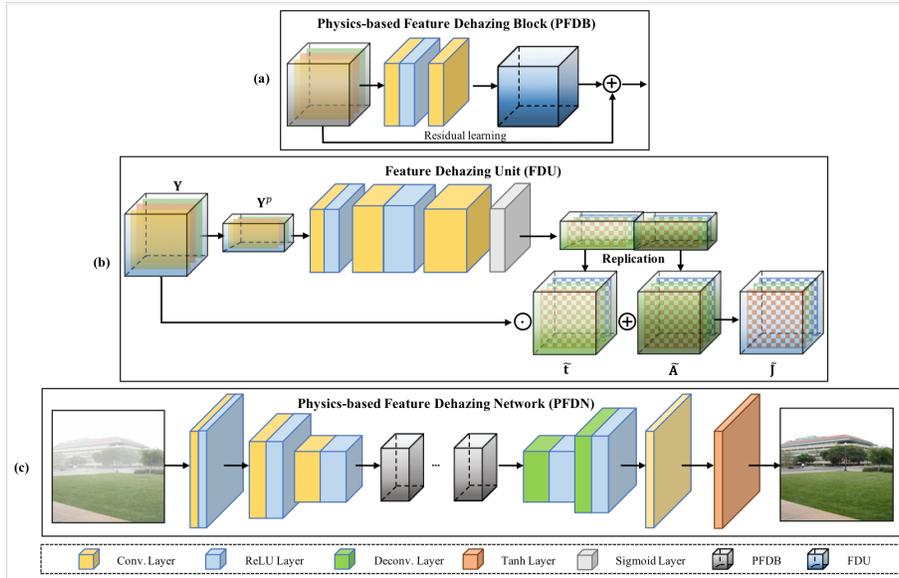


Fig. 1. Proposed network architecture for image dehazing. The whole image dehazing network PFDN in (c) is based on an encoder and decoder architecture with the proposed PFDBs in (a). The proposed PFDB consists of an FDU with a residual learning architecture, where the proposed FDU can make full use of the physics model in the feature space for better image dehazing. Please see text for more details.

where \mathbf{F}_2 can be regarded as the feature extraction operation. The equation (5) presents the relation between the clear image and the hazy image in a feature space, and is based on (4), which assumes that k is a linear operator. Note that a deep CNN with piece-wise linear activation functions (e.g., ReLU) is inherently locally linear [17, 24]. Due to the strong representation ability of deep neural networks, we propose a feature dehazing unit and adopt deep CNNs to approximate the features corresponding to \mathbf{F}_1 and $\mathbf{K}\mathbf{A}_t$. Therefore, the discriminatively useful features can be better estimated for the clear image reconstruction.

To that end, the proposed FDU consists of two parts. The first part is used to learn the features that mostly approximate the key components \mathbf{F}_1 and $\mathbf{K}\mathbf{A}_t$ associated with the haze formation. The second part is then used to estimate the features $\mathbf{K}\mathbf{J}$ for the clear image reconstruction based on (5).

As the transmission map T is related to the scene depth [13] with the piece-wise constant property and the atmospheric light is usually assumed to be homogeneous, we use the global average pooling (GAP) to remove redundant information in the feature space and remain useful values in the approximated features that correspond to \mathbf{F}_1 and $\mathbf{K}\mathbf{A}_t$. Specifically, let $\mathbf{Y} = \{y_i\}_{i=1}^N$ denote the input of FDU, which has N features with the size of $h \times w$ pixels. We first apply the GAP to \mathbf{Y} and obtain $\mathbf{Y}^p = \{y_i^p\}_{i=1}^N$, each element of which is defined

as

$$y_i^p = \frac{1}{h \times w} \sum_{q=1}^{h \times w} y_i(q), \quad (6)$$

where q denotes the pixel position and \times denotes the product operation. Then, motivated by the success of the encoder and decoder network architecture in feature exploration, we apply the similar downsampling and upsampling operations [40] to the features in \mathbf{Y}^p by

$$\tilde{\mathbf{Y}} = \mathcal{R}(\mathcal{C}^N(\mathcal{R}(\mathcal{C}^{\frac{N}{16}}(\mathbf{Y}^p)))), \quad (7)$$

where \mathcal{C}^N denotes the convolution operation with the filter kernel size of 1×1 pixel and N filters; \mathcal{R} denotes the ReLU activation function. With $\tilde{\mathbf{Y}}$, we further apply the feature upsampling operation \mathcal{C}^{2N} with the Sigmoid function to get the intermediate feature. We respectively use the replication of the first N features ($\tilde{\mathbf{t}}$) and the remaining N features ($\tilde{\mathbf{A}}$) to approximate the features corresponding to \mathbf{F}_1 and \mathbf{KA}_t . Note that by extracting and remaining the most useful information with GAP, the features $\tilde{\mathbf{t}}$ and $\tilde{\mathbf{A}}$ are channel-specific. Thus, we generate the output of the FDU by

$$\tilde{\mathbf{J}} = \mathbf{Y} \odot \tilde{\mathbf{t}} + \tilde{\mathbf{A}}. \quad (8)$$

Based on (8), we can discriminatively learn reliable features $\tilde{\mathbf{J}}$ from \mathbf{Y} for the clear image reconstruction. In Section 5, we demonstrate that using $\tilde{\mathbf{J}}$ instead of \mathbf{Y} is able to help image dehazing. Figure 1(b) shows the network architecture of the FDU.

3.2 Residual learning

As residual learning [14] has been demonstrated to be effective in lots of vision tasks, we use it in the PFDB to increase the accuracy of image dehazing and ease the training of deep neural networks. Specifically, each PFDB has two convolutional layers with the filter kernel size of 3×3 pixels, where the first one is followed by the ReLU as the activation function and the second one is followed by the FDU. The detailed network architectures of the residual learning and the proposed PFDB are shown in Figure 1(a).

3.3 PFDN for image dehazing

Since the proposed PFDB is performed in the feature space, we embed it into the encoder and decoder network architecture to solve image dehazing. The encoder module is adopted to extract useful features from the hazy image, which contains three scale convolutional blocks. Each convolutional block has one convolutional layer followed by a ReLU layer. The stride value is 1 for the first convolutional layer and 2 for the remaining two convolutional layers. The decoder module is adopted to further process the output of the PFDB and reconstruct the final

clear image. It consists of two transposed convolutional blocks and one convolutional block. Each transposed convolutional block has a transposed convolutional layer with stride 2 and a ReLU layer. The stride value of the final convolutional layer is 1. For the parameters in the convolutional and the transposed convolutional layers, we use the same settings as [15]. The network architecture for the proposed PFDN is shown in Figure 1(c). The detailed parameters of the proposed network are included in the supplemental material.

3.4 Implementation details

We use the ADAM optimizer [16] with the momentum parameters $\beta_1 = 0.5$ and $\beta_2 = 0.999$. The initial learning rate is set as 2×10^{-4} and we follow the decay strategy as [15]. In the training stage, we use the L_1 -norm as the loss function to constrain the network output and the ground truth. We implement our network based on the PyTorch using a machine with an NVIDIA GTX 2080Ti GPU. Our un-optimized code takes about 0.09s to dehaze an image of 512×512 pixels on average.

4 Experimental Results

We evaluate the proposed approach using the publicly available benchmark datasets [19, 30, 32] and compare it with the state-of-the-art single image dehazing methods.

4.1 Datasets

NYU & Make3D datasets. For fair comparisons, we first follow the standard protocols adopted by existing methods (e.g., [20, 28, 29, 39]) to generate 2,413 hazy/clear image pairs using the NYU depth v2 dataset [32] and the Make3D dataset [30]. This synthetic dataset contains hazy/clear image pairs of both indoor and outdoor scenes. We randomly choose 2,172 images for training and the remaining 241 images for test, where the training images and test images do not overlap. The image patch size used in the training process is set to be 512×512 pixels.

RESIDE dataset. The RESIDE dataset [19] is a large-scale benchmark dataset, which contains hazy/clear image pairs in both indoor and outdoor scenarios. The subsets ITS and OTS are used as the training dataset. The subset SOTS is used for test, which contains 500 indoor hazy images and 500 outdoor hazy ones.

4.2 Comparisons with the state of the art

To evaluate the performance of the proposed approach, we first compare it against the state-of-the-art methods based on statistical priors [13, 41] and deep CNNs [6, 20, 22, 26–29, 39] on the synthetic NYU & Make3D test dataset. For

Table 1. Quantitative evaluations with the state-of-the-art methods on the synthetic NYU & Make3D test dataset. \uparrow and \downarrow denote that the better results should achieve higher and lower values of this metric.

| Methods | DCP [13] | Nonlocal [41] | MSCNN [28] | DehazeNet [6] | GFN [29] | DCPDN [39] | eGAN [20] | DualCNN [26] | EPDN [27] | GDN [22] | Ours |
|------------------------|----------|---------------|------------|---------------|----------|------------|-----------|--------------|-----------|----------|---------------|
| PSNR \uparrow | 18.04 | 16.69 | 17.73 | 20.24 | 20.79 | 16.71 | 26.12 | 21.30 | 22.97 | 26.34 | 27.26 |
| SSIM \uparrow | 0.7852 | 0.7397 | 0.7725 | 0.8176 | 0.8064 | 0.7676 | 0.8831 | 0.8170 | 0.8265 | 0.8816 | 0.9047 |
| CIEDE2000 \downarrow | 10.45 | 12.80 | 10.56 | 7.87 | 7.65 | 12.34 | 4.49 | 6.85 | 6.01 | 4.35 | 3.87 |

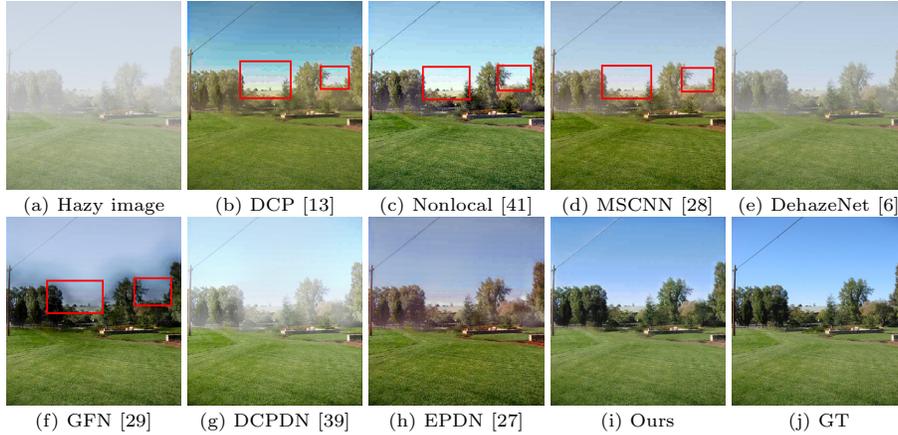


Fig. 2. Image dehazing results from the synthetic NYU & Make3D test dataset. The parts enclosed in red boxes in (b)-(d) and (f) contain color distortions and artifacts. The result in (h) contains color distortions, while the results in (e) and (g) still have haze residual. The proposed method generates a much clearer image that is visually close to the ground truth image (best viewed on high-resolution display with zoom-in).

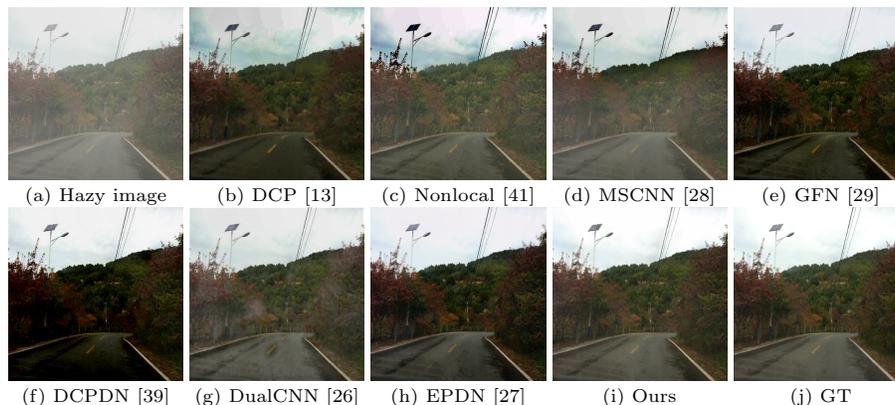
fair comparisons with deep learning-based methods, we fine-tune them using the proposed training dataset to achieve the best performance. We use PSNR, SSIM [36], and CIEDE2000 [31] to evaluate the quality of each recovered image.

Table 1 summarizes the quantitative results on the synthetic NYU & Make3D test dataset. The proposed approach performs favorably against state-of-the-art methods, where the average PSNR by our method is at least 0.92dB higher than those by other image dehazing methods.

Figure 2 shows the visual comparison results on the synthetic NYU & Make3D test dataset by the evaluated methods. The results by the statistical priors-based methods [13, 41] contain artifacts in the regions of the sky. The methods by [6, 28] develop CNNs to estimate the transmission and then use the conventional method [13] to estimate the clear images. However, these approaches do not correct the errors that are caused by the imperfect atmospheric light estimation. In [29], Ren et al. develop a gated neural network for image dehazing. However, as this method uses hand-crafted features to constrain the network, the quality of dehazed results is limited by these hand-crafted features. We note that both DCPDN [39] and EPDN [27] methods develop end-to-end trainable networks for image dehazing. However, the DCPDN method does not remove the haze from

Table 2. Quantitative evaluations with state-of-the-art methods on the indoor scenes of SOTS dataset [19].

| | DCP [13] | Nonlocal [41] | MSCNN [28] | DehazeNet [6] | AOD-Net [18] | GFN [29] | DCPDN [39] | DualCNN [26] | EPDN [27] | GDN [22] | Ours |
|-----------------|----------|---------------|------------|---------------|--------------|----------|------------|--------------|-----------|----------|--------------|
| PSNR \uparrow | 16.61 | 17.30 | 19.84 | 19.82 | 20.51 | 24.91 | 15.85 | 22.25 | 25.06 | 32.16 | 32.68 |
| SSIM \uparrow | 0.8546 | 0.7768 | 0.8327 | 0.8209 | 0.8162 | 0.9186 | 0.8175 | 0.8751 | 0.9232 | 0.9836 | 0.9760 |

**Fig. 3.** Image dehazing results from the SOTS dataset. The colors of the dehazed images in (b), (e), (f), and (h) look darker than those of the ground truth image. The dehazed images in (c), (d), and (g) contain haze residual. The proposed method generates a much clearer image that is visually close to the ground truth image (best viewed on high-resolution display with zoom-in).

the input image, while the EPDN method generates the result with obvious color distortions. In contrast, the proposed approach explicitly considers the physics model of the haze process to discriminatively learn the useful information in the feature space, which accordingly generates high-quality images.

Then, we evaluate the proposed approach against state-of-the-art methods on the SOTS dataset [19]. For fair comparisons, we retrain the proposed method on the training dataset by [19] according to their protocols. Table 2 shows the evaluation results on the indoor scenes of the SOTS dataset, where the proposed approach performs favorably against state-of-the-art methods in terms of PSNR and SSIM.

Figure 3 shows the dehazed results of the outdoor scene on the SOTS dataset by the evaluated methods. The results by [13, 41] have severe color distortions in the sky. The dehazed images generated by [26, 28] still have significant haze residuals. The results obtained by [29, 39] look too dark. In contrast, the proposed method recovers a clearer image than state-of-the-art methods. More experimental results on the datasets [1–3, 19] are included in the supplemental material.

Real examples. We further evaluate the proposed approach using real images in Figure 4. The state-of-the-art methods [20, 27, 29] tend to over-estimate the colors of the restored images as shown in Figure 4(b), (c), and (h). The method [26]



Fig. 4. Image dehazing results on real examples. The colors of the dehazed images in (b), (c), and (h) look darker. The part enclosed in the red box of (d) contains significant artifacts, e.g., the coat. The result in (g) has severe color distortions as shown in the red box. The image in (i) still contains some haze residuals. The proposed approach generates much clearer images (best viewed on high-resolution display with zoom-in).

generates the result with significant artifacts as shown in Figure 4(d). The result by the method [22] still contains the haze residual. In contrast, our approach generates much clearer and brighter images than those by the state-of-the-art methods as shown in Figure 4 (e) and (j). More visual comparisons on real-world images are included in the supplemental material.

5 Analysis and Discussions

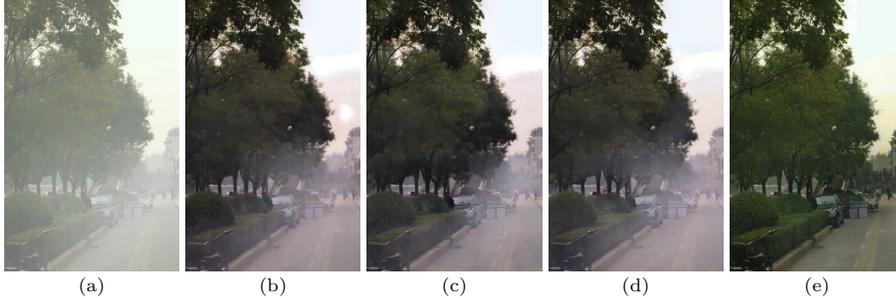
We have shown that using PFDB is able to remove haze and outperforms state-of-the-art methods. To better understand the proposed approach, we perform further analysis and compare with related methods.

Effectiveness of FDU. As the proposed PFDB consists of an FDU and residual learning architecture, one may wonder whether the performance gains merely come from the use of residual learning architecture [14]. To answer this question, we remove the FDU from our network architecture and train this baseline method using the same settings for fair comparisons. We note that the proposed method without using FDU reduces to the one that directly uses the features \mathbf{Y} for the clear image reconstruction in (8) (i.e., “w/o FDU & w/ 9RBs” in Table 3). The comparisons in Table 3 demonstrate that it is more effective to explicitly consider the physics model in the feature space and use the FDU to discriminatively learn useful features from \mathbf{Y} , which generates higher-quality images than directly using the features \mathbf{Y} .

In addition, as each FDU contains three convolutional layers, one may also wonder whether using more ResBlocks [14] instead of FDU can generate better results. To answer this question, we remove the FDU from our network architecture and adopt more ResBlocks to train the baseline method using the same

Table 3. Effectiveness of FDU on image dehazing on the synthetic NYU & Make3D test dataset. RBs denotes ResBlocks.

| Methods | w/o FDU & w/ 9RBs | w/o FDU & w/ 12RBs | w/o FDU & w/ 15RBs | Ours |
|------------|-------------------|--------------------|--------------------|---------------|
| Avg. PSNRs | 26.33 | 26.49 | 26.75 | 27.26 |
| Avg. SSIMs | 0.8851 | 0.8897 | 0.8956 | 0.9047 |

**Fig. 5.** Effectiveness of FDU on image dehazing. (a) Hazy image. (b)-(d) denote the results by the baseline methods w/o FDU & w/ 9RBs, w/o FDU & w/ 12RBs, and w/o FDU & w/ 15RBs, respectively. (e) Our result. The methods without using FDU generate the results with obvious haze residual as shown in (b)-(d). In contrast, the proposed approach with the FDU generates a much clearer image in (e).

settings for fair comparisons. Specifically, the baseline models contain 9, 12, and 15 ResBlocks, respectively, which are denoted as “w/o FDU & w/9 RBs”, “w/o FDU & w/12 RBs”, and “w/o FDU & w/15 RBs” in Table 3. Table 3 shows that using more ResBlocks does not significantly improve the performance. In contrast, the proposed method with FDU performs better than purely stacking ResBlocks due to the discriminatively learned features by the FDU.

The visual comparison results in Figure 5(b)-(e) further demonstrate the benefit of using the FDU in generating clearer images. We note that the proposed method without using FDU does not effectively remove haze. The generated results still contain significant haze residual (Figure 5(b)-(d)). In contrast, the proposed method with FDU generates a much clearer image in Figure 5(e).

As the proposed FDU performs image dehazing in the feature space, one may also wonder how it generates useful features for haze removal. To better demonstrate the effect of the proposed FDU intuitively, we show some intermediate features from the proposed approach. As the feature space is more complex than the RGB space, we map the principal components of the intermediate features to the principal components of the RGB space according to the visualization method [21]. Figure 6 shows that the features learned after FDU (i.e., $\tilde{\mathbf{J}}$ in (8)) have a better color contrast than those learned before FDU, suggesting that the proposed FDU is able to remove haze and thus facilitates the clear image restoration. All the above results demonstrate that the proposed FDU is able to help image dehazing.

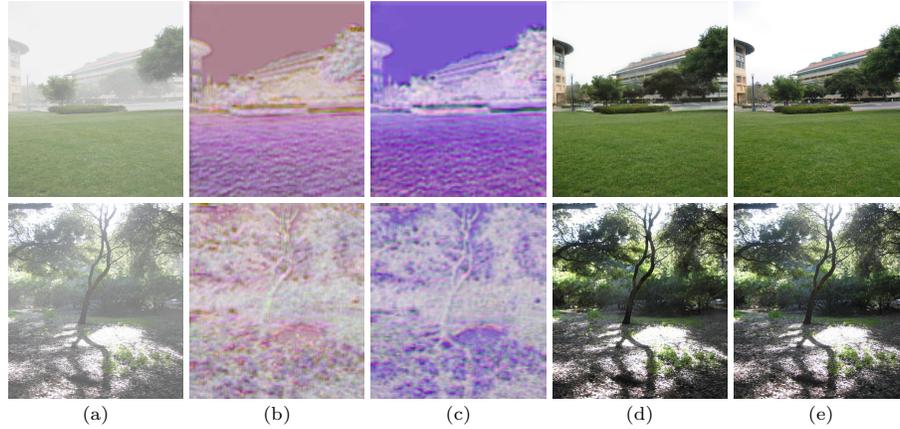


Fig. 6. Visualizations of the features from the proposed approach. (a) Hazy images. (b) and (c) denote the visualization of the features learned before and after FDU. (d) Our dehazed results. (e) Ground truth images. The principal components of the features are mapped to the principal components of the RGB space for visualization. The features learned after FDU (i.e., $\tilde{\mathbf{J}}$) have a better color contrast than those before FDU, indicating that the proposed FDU is able to help haze removal (best viewed on high-resolution display with zoom-in).

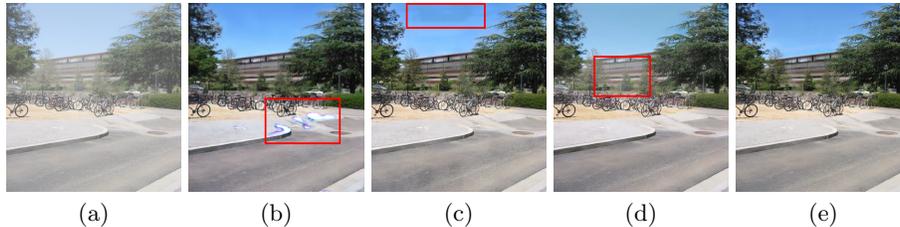
Effectiveness of residual learning in PFDB. As stated in Section 3.2, we use residual learning in PFDB to increase the accuracy of image dehazing and ease the training of the deep CNNs. To illustrate the effect of residual learning, we compare with the proposed methods without using residual learning in PFDB. We retrain this baseline method and use the same settings as the proposed approach for fair comparisons. Both Table 4 and Figure 7 show that using residual learning in PFDB is able to generate high-quality images.

Effectiveness of GAP in PFDB. As stated in Section 3.1, the proposed FDU uses the GAP operation to maintain the most important information for the features that are related to \mathbf{F}_1 and $\mathbf{K}\mathbf{A}_t$. To demonstrate the effectiveness of the GAP, we compare with the proposed network without using the GAP operation and evaluate this baseline method on the synthetic NYU & Make3D test dataset. Table 4 indicates that using the GAP operation is able to maintain useful information and thus facilitates image dehazing. Although the method without using GAP can also estimate these features, it does not significantly improve the performance. In addition, the comparisons in Figure 7(c) and (e) demonstrate that using GAP is able to remove artifacts.

Relations with attention-based methods. Recently, the attention mechanism has been used to solve image super-resolution [40]. This method develops the channel attention strategy to learn useful features for image super-resolution. In contrast, our approach learns the features using FDU which is based on the

Table 4. Effectiveness of residual learning and GAP on image dehazing on the synthetic NYU & Make3D test dataset.

| | w/o residual learning | w/o GAP | Attention-based method [40] | Ours |
|------------|-----------------------|---------|-----------------------------|---------------|
| Avg. PSNRs | 23.55 | 26.56 | 26.64 | 27.26 |
| Avg. SSIMs | 0.7935 | 0.8973 | 0.8846 | 0.9047 |

**Fig. 7.** Effectiveness of the proposed PFDB for discriminatively learning features on image dehazing. (a) Hazy image. (b)-(d) denote the results generated by the proposed method without using residual learning, the proposed method without using GAP, and the feature learning method based on the attention mechanism [40], respectively. (e) Our result. The parts enclosed in red boxes in (b)-(c) contain significant artifacts, while the part enclosed in the red box in (d) still contains some haze residual. In contrast, the proposed approach generates a much clearer image in (e) (best viewed on high-resolution display with zoom-in).

physics model of the haze process in a feature space. To further demonstrate the effectiveness of the proposed method, we retain the attention-based method [40] using the proposed training dataset in the same settings for fair comparisons. We evaluate the proposed method against this method on the synthetic NYU & Make3D test dataset. Table 4 shows that directly using the attention mechanism does not always facilitate haze removal. In contrast, explicitly considering the physics model of the haze process in the feature space is able to generate high-quality images. The comparison results in Figure 7(d) and (e) further demonstrate that the proposed approach is more effective for haze removal than directly using the attention mechanism.

Relations with deep physics model-based methods. We note that several notable methods [12, 18, 25, 26, 38, 39] use the physics model of the haze process to constrain the deep neural network for image dehazing. The DualCNN method [26] develops a network based on two branches to estimate the transmission and atmospheric light. In [38], Yang et al. develop a disentangled dehazing network based on the physics model of the haze process to solve image dehazing using unpaired images. Zhang et al. [39] develop a new dense network for image dehazing based on the physics model (1). This method first uses the deep CNNs to estimate the transmission and atmospheric light and then reconstructs clear images based on (1). Similar to [39], Li et al. [18] estimate the clear image based on a re-formulated atmospheric scattering model. Guo et al. [12] use different

Table 5. Ablation study on the number of PFDBs using the synthetic NYU & Make3D test dataset.

| Number of PFDBs | 7 PFDBs | 9 PFDBs | 13 PFDBs | 15 PFDBs | 17 PFDBs |
|-----------------|---------|---------|----------|----------|----------|
| Avg. PSNRs | 27.17 | 27.26 | 27.32 | 27.58 | 27.33 |
| Avg. SSIMs | 0.9050 | 0.9047 | 0.9068 | 0.9082 | 0.9080 |

Table 6. Comparisons of model sizes against the state-of-the-art methods.

| Methods | DCPDN [39] | cGAN [20] | GDN [22] | EPDN [27] | Ours |
|------------|------------|-----------|----------|-----------|------|
| Model size | 134M | 140M | 1M | 17M | 12M |

CNN models to separately estimate the transmission map, atmospheric light, and the latent clear image. Then the final results are generated based on the physics model. As these methods mainly consider the physics model in the raw image space and do not fully explore the physics information in the feature space, the final estimated dehazed images contain haze residual and artifacts as shown in Figures 2-4. In contrast, the proposed approach develops the FDU to explicitly consider the physics model in a feature space, which is able to effectively learn the useful features for image dehazing. Thus, the haze can be well removed and textures of the images are well recovered (see both quantitative and qualitative evaluations in Tables 1-2 and Figures 2-4).

Analysis on the number of PFDBs. The proposed network contains several PFDBs. We further evaluate the effect of the number of PFDBs by setting the number of PFDBs from 7 to 17. Table 5 shows that using more PFDBs does not significantly improve the performance. We empirically use 9 PFDBs as a trade-off between accuracy and efficiency.

Model size. We evaluate the model size of the proposed approach against state-of-the-art methods. Table 6 shows that the proposed approach has competitive model parameters against state-of-the-art methods.

6 Conclusions

We have presented an effective PFDB for image dehazing. The critical component PFDB consists of an FDU with a residual learning architecture. The FDU is developed to fully explore the useful features for image dehazing based on the physics model of the haze process. The residual learning architecture is applied to the FDU to increase the accuracy and ease the training of deep neural networks. The proposed PFDB is embedded as a backbone into an encoder and decoder network architecture in an end-to-end fashion for image dehazing. We have analyzed the effect of the proposed PFDB on image dehazing. Both quantitative and qualitative results show that the proposed approach performs favorably against state-of-the-art methods.

References

1. Ancuti, C.O., Ancuti, C., Sbert, M., Timofte, R.: Dense-Haze: A benchmark for image dehazing with dense-haze and haze-free images. In: IEEE ICIP. pp. 1014–1018 (2019)
2. Ancuti, C.O., Ancuti, C., Timofte, R., De Vleeschouwer, C.: O-HAZE: a dehazing benchmark with real hazy and haze-free outdoor images. In: IEEE CVPR Workshops. pp. 754–762 (2018)
3. Ancuti, C.O., Ancuti, C., Timofte, R., Vleeschouwer, C.D.: I-HAZE: a dehazing benchmark with real hazy and haze-free indoor images. In: International Conference on Advanced Concepts for Intelligent Vision Systems. pp. 620–631 (2018)
4. Ancuti, C., Ancuti, C.O., Timofte, R.: NTIRE 2018 Challenge on image dehazing: Methods and results. In: IEEE CVPR Workshops. pp. 891–901 (2018)
5. Berman, D., Treibitz, T., Avidan, S.: Non-local image dehazing. In: IEEE CVPR. pp. 1674–1682 (2016)
6. Cai, B., Xu, X., Jia, K., Qing, C., Tao, D.: Dehazenet: An end-to-end system for single image haze removal. IEEE TIP **25**(11), 5187–5198 (2016)
7. Chen, C., Do, M.N., Wang, J.: Robust image and video dehazing with visual artifact suppression via gradient residual minimization. In: ECCV. pp. 576–591 (2016)
8. Chen, D., He, M., Fan, Q., Liao, J., Zhang, L., Hou, D., Yuan, L., Hua, G.: Gated context aggregation network for image dehazing and deraining. In: IEEE WACV. pp. 1375–1383 (2019)
9. Fattal, R.: Single image dehazing. ACM TOG **27**(3), 72:1–72:9 (2008)
10. Fattal, R.: Dehazing using color-lines. ACM TOG **34**(1), 13:1–13:14 (2014)
11. Galdran, A., Vazquez-Corral, J., Pardo, D., Bertalmío, M.: Enhanced variational image dehazing. SIAM J. Imaging Sciences **8**(3), 1519–1546 (2015)
12. Guo, T., Li, X., Cherukuri, V., Monga, V.: Dense scene information estimation network for dehazing. In: IEEE CVPR Workshops (2019)
13. He, K., Sun, J., Tang, X.: Single image haze removal using dark channel prior. IEEE TPAMI **33**(12), 2341–2353 (2011)
14. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: IEEE CVPR. pp. 770–778 (2016)
15. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: IEEE CVPR. pp. 5967–5976 (2017)
16. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. In: ICLR (2015)
17. Lee, G.H., Alvarez-Melis, D., Jaakkola, T.S.: Towards robust, locally linear deep networks. In: ICLR (2019)
18. Li, B., Peng, X., Wang, Z., Xu, J., Feng, D.: AOD-Net: All-in-one dehazing network. In: IEEE ICCV. pp. 4780–4788 (2017)
19. Li, B., Ren, W., Fu, D., Tao, D., Feng, D., Zeng, W., Wang, Z.: Benchmarking single-image dehazing and beyond. IEEE TIP **28**(1), 492–505 (2019)
20. Li, R., Pan, J., Li, Z., Tang, J.: Single image dehazing via conditional generative adversarial network. In: IEEE CVPR. pp. 8202–8211 (2018)
21. Liu, C., Yuen, J., Torralba, A.: SIFT flow: Dense correspondence across scenes and its applications. IEEE TPAMI **33**(5), 978–994 (2011)
22. Liu, X., Ma, Y., Shi, Z., Chen, J.: GridDehazeNet: Attention-based multi-scale network for image dehazing. In: IEEE ICCV (2019)
23. Mei, K., Jiang, A., Li, J., Wang, M.: Progressive feature fusion network for realistic image dehazing. In: ACCV (2018)

24. Montufar, G.F., Pascanu, R., Cho, K., Bengio, Y.: On the number of linear regions of deep neural networks. In: NIPS. pp. 2924–2932 (2014)
25. Pan, J., Dong, J., Liu, Y., Zhang, J., Ren, J., Tang, J., Tai, Y.W., Yang, M.H.: Physics-based generative adversarial models for image restoration and beyond. IEEE TPAMI (2020)
26. Pan, J., Liu, S., Sun, D., Zhang, J., Liu, Y., Ren, J.S.J., Li, Z., Tang, J., Lu, H., Tai, Y.W., Yang, M.H.: Learning dual convolutional neural networks for low-level vision. In: IEEE CVPR. pp. 3070–3079 (2018)
27. Qu, Y., Chen, Y., Huang, J., Xie, Y.: Enhanced pix2pix dehazing network. In: IEEE CVPR (2019)
28. Ren, W., Liu, S., Zhang, H., Pan, J., Cao, X., Yang, M.H.: Single image dehazing via multi-scale convolutional neural networks. In: ECCV. pp. 154–169 (2016)
29. Ren, W., Ma, L., Zhang, J., Pan, J., Cao, X., Liu, W., Yang, M.H.: Gated fusion network for single image dehazing. In: IEEE CVPR. pp. 3253–3261 (2018)
30. Saxena, A., Sun, M., Ng, A.Y.: Make3D: Learning 3D scene structure from a single still image. IEEE TPAMI **31**(5), 824–840 (2009)
31. Sharma, G., Wu, W., Dalal, E.N.: The CIEDE2000 color-difference formula: Implementation notes, supplementary test data, and mathematical observations. Color Research & Application **30**(1), 21–30 (2005)
32. Silberman, N., Hoiem, D., Kohli, P., Fergus, R.: Indoor segmentation and support inference from RGBD images. In: ECCV. pp. 746–760 (2012)
33. Tan, R.T.: Visibility in bad weather from a single image. In: IEEE CVPR (2008)
34. Tang, K., Yang, J., Wang, J.: Investigating haze-relevant features in a learning framework for image dehazing. In: IEEE CVPR. pp. 2995–3002 (2014)
35. Wang, F., Jiang, M., Qian, C., Yang, S., Li, C., Zhang, H., Wang, X., Tang, X.: Residual attention network for image classification. In: IEEE CVPR. pp. 6450–6458 (2017)
36. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE TIP **13**(4), 600–612 (2004)
37. Yang, D., Sun, J.: Proximal dehaze-net: A prior learning-based deep network for single image dehazing. In: ECCV. pp. 729–746 (2018)
38. Yang, X., Xu, Z., Luo, J.: Towards perceptual image dehazing by physics-based disentanglement and adversarial training. In: AAAI. pp. 7485–7492 (2018)
39. Zhang, H., Patel, V.M.: Densely connected pyramid dehazing network. In: IEEE CVPR. pp. 3194–3203 (2018)
40. Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B., Fu, Y.: Image super-resolution using very deep residual channel attention networks. In: ECCV. pp. 294–310 (2018)
41. Zhang, Y., Li, K., Li, K., Zhong, B., Fu, Y.: Residual non-local attention networks for image restoration. In: ICLR (2019)
42. Zhu, Q., Mai, J., Shao, L.: A fast single image haze removal algorithm using color attenuation prior. IEEE TIP **24**(11), 3522–3533 (2015)