

Dual-Path in Dual-Path Network for Single Image Dehazing

Aiping Yang¹, Haixin Wang¹, Zhong Ji^{1*}, Yanwei Pang¹, Ling Shao^{1,2}

¹ School of Electrical and Information Engineering, Tianjin University, China

² Inception Institute of Artificial Intelligence, Abu Dhabi, UAE

{yangaiping, wanghaixin199503, jizhong, pyw}@tju.edu.cn, ling.shao@ieee.org

Abstract

Recently, deep learning-based single image dehazing method has been a popular approach to tackle dehazing. However, the existing dehazing approaches are performed directly on the original hazy image, which easily results in image blurring and noise amplifying. To address this issue, the paper proposes a DPDP-Net (Dual-Path in Dual-Path network) framework by employing a hierarchical dual path network. Specifically, the first-level dual-path network consists of a Dehazing Network and a Denoising Network, where the Dehazing Network is responsible for haze removal in the structural layer, and the Denoising Network deals with noise in the textural layer, respectively. And the second-level dual-path network lies in the Dehazing Network, which has an AL-Net (Atmospheric Light Network) and a TM-Net (Transmission Map Network), respectively. Concretely, the AL-Net aims to train the non-uniform atmospheric light, while the TM-Net aims to train the transmission map that reflects the visibility of the image. The final dehazing image is obtained by nonlinearly fusing the output of the Denoising Network and the Dehazing Network. Extensive experiments demonstrate that our proposed DPDP-Net achieves competitive performance against the state-of-the-art methods on both synthetic and real-world images.

1 Introduction

Single image dehazing has become a promising research area, which can be widely used to object detection [Xiaogang *et al.*, 2014] and object recognition [Zheng *et al.*, 2012]. However, it is still a challenging task due to its uncertainty of scene transmission map and atmospheric light.

Recently, the renaissance of deep learning greatly promotes the development of single image dehazing technique. Early studies employ Convolutional Neural Networks (CNNs) to estimate the transmission map and then follow the conventional method to estimate the atmospheric light to recover clear images [Cai *et al.*, 2016; Ren *et al.*, 2016; Xi *et al.*,

2018]. A major drawback for this line of work is the neglect of atmospheric light. If the transmission map is not well estimated, it will lead to a wrong estimation of atmospheric light. To this end, it is of great interest to jointly estimate the transmission map and atmospheric light [Zhang and Patel, 2018; Yang *et al.*, 2017], or recovering the dehazing image from the hazy image with end-to-end methods [Li *et al.*, 2017; Ren *et al.*, 2018; Liu *et al.*, 2018; Li *et al.*, 2019].

However, the existing approaches suffer from the following drawbacks. First, few of them consider the noise in the hazing image. They perform the dehazing operation directly on the original hazy image, which easily leads to the blurring and noise amplifying in the dehazing image [Cai *et al.*, 2016; Ren *et al.*, 2016; Li *et al.*, 2017; Ren *et al.*, 2018; Zhang and Patel, 2018; Yang *et al.*, 2017; Liu *et al.*, 2018; Li *et al.*, 2019], as shown in Figure 1. Second, they usually estimate both the transmission map and global atmospheric light with one integrated loss function, which is prone to slow training and unpredictable errors [Zhang and Patel, 2018; Yang *et al.*, 2017]. The reason lies in the fact that the atmospheric light is independent of the transmission map in the atmospheric scattering model (as shown in Eq.(1)), and there is no necessary connection between the two variables. Third, these approaches generally estimate the atmospheric light as a constant. In fact, this assumption is unsuitable for hazy images, particularly for those with large sky regions, since it may cause unacceptable brightness imbalance and color distortion in dehazing images, as shown in Figure 2.

To address these issues, this paper proposes a DPDP-Net (Dual-Path in Dual-Path network) framework by employing a hierarchical dual-path network, which tackles the dehazing and denoising steps with two individual networks. Specifically, to restore more details during the operation of dehazing as well as remove the noise, the first dual-path network decomposes the hazy image into the structural layer and the texture layer, respectively. One path is designed to perform dehazing on the structure layer, and the other is constructed to perform denoising on the texture layer. The second-level dual-path network is designed for the purpose of avoiding interactive interference between the transmission map and the atmospheric light. It lies in the Dehazing Network and consist of two independent CNNs with their corresponding loss functions to estimate the transmission map and the spatial variant atmospheric light, respectively. The overview of the proposed

*Corresponding author: Zhong Ji, email: jizhong@tju.edu.cn

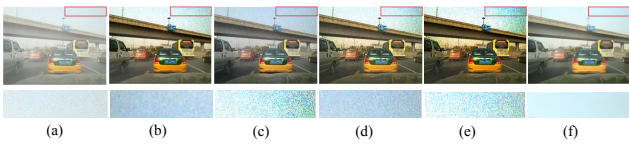


Figure 1: Dehazing results on the noisy image. (a) Hazy image. (b) DCP. (c) BCCR. (d) DehazeNet. (e) MSCNN. (f) DPDP-Net (Ours). The bottom row shows the detail of the marked regions, from which we could observe that the competing methods result in noise amplification in the sky regions, while our proposed DPDP-Net has natural sky region, and well preserves the details.

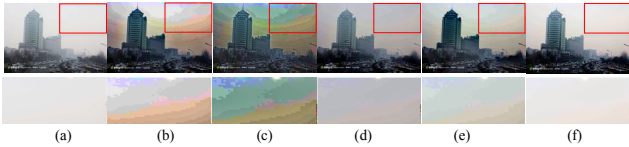


Figure 2: Comparison of dehazing effects on image with large sky region. (a) Hazy image. (b) DCP. (c) BCCR. (d) DehazeNet. (e) MSCNN. (f) DPDP-Net (ours). The bottom row shows the detail of the marked regions, from which we could observe that the competing methods result in brightness imbalance and color distortion.

DPDP-Net is illustrated in Figure 3.

1.1 The Contributions Are Summarized as Follows

(1) We present a novel Dual-Path in Dual-Path network (DPDP-Net) to address the image denoising issue for single image dehazing. As far as we know, it is the first deep architecture to tackle this issue. Specifically, Dehazing Network and Denoising Network constitute the first-level dual-path network, which is responsible for haze removal in the structural layer and noise removal in the textural layer, respectively. This design guarantees the purpose of haze removal without amplifying the noise and damaging the details.

(2) We propose to estimate the atmospheric light and transmission map with independent loss function. The two operations are respectively realized in the Dehazing Network with Atmospheric Light Network (AL-Net) and Transmission Map Network (TM-Net), which constitute the second-level dual-path network. Concretely, the AL-Net aims to estimate the non-uniform atmospheric light, while the TM-Net aims to learn the transmission map that reflects the visibility of the image. The two networks share the same visual features, but have their independent loss functions, which avoids their interactive interference and guarantees better estimations.

(3) The AL-Net is built to acquire the non-uniform atmospheric light value to avoid the color shift issue that is easy to occur in the sky part of the hazy image.

2 Related Work

Image dehazing technique has witnessed many achievements in last decades. It can be roughly classified into four categories: additional information-based methods [Narasimhan and Nayar, 2002; Nayar, 1999], contrast enhancement-based methods [Narasimhan and Nayar, 2003], prior-based methods [He *et al.*, 2009; Qingsong *et al.*, 2015; Tan, 2008; Berman *et al.*, 2016; Meng *et al.*, 2014], and learning-based

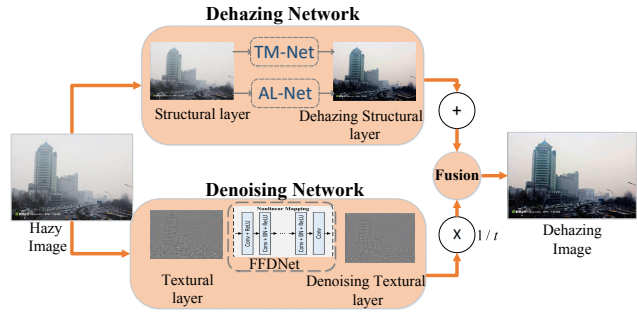


Figure 3: The proposed DPDP-Net framework. The Dehazing-Net and the Denoising-Net constitute the first-level dual-path, TM-Net and the AL-Net make up the second-level dual-path.

methods [Cai *et al.*, 2016; Ren *et al.*, 2016; Li *et al.*, 2017; Ren *et al.*, 2018; Zhang and Patel, 2018; Yang *et al.*, 2017; Liu *et al.*, 2018; Li *et al.*, 2019]. The first two categories have their obvious shortcomings. The additional information-based method is hard to apply in cases where additional information or multiple images are not available. And the contrast enhancement-based methods do not consider the degradation cause of the haze image, leading to the results often having severe color distortion.

Prior-based methods [He *et al.*, 2009; Qingsong *et al.*, 2015; Lai *et al.*, 2015; Berman *et al.*, 2016; Meng *et al.*, 2014] are proposed to address the ill-posed issue caused by the unknown transmission map and global atmospheric light. The dark channel prior (DCP) presented by [He *et al.*, 2009] is one of the most famous ones. [Meng *et al.*, 2014] enhanced it by optimizing the inherent boundary constraint with weighted L1-norm contextual regularization to estimate the transmission map. [Qingsong *et al.*, 2015] proposed a color attenuation prior (CAP) to recover depth information by creating a linear model and learning the parameters of the model with a supervised learning method. In contrast, [Berman *et al.*, 2016] introduced a non-local haze-line prior, which is based on an approximation of the entire dehazing images including a few hundred distinct colors.

Recently, learning-based methods have attracted great attention, which employ CNNs or GANs (Generative Adversarial Network) for single image dehazing. Specifically, CNN-based methods mainly focus on estimating transmission map or atmospheric light [Cai *et al.*, 2016; Ren *et al.*, 2016; Li *et al.*, 2017; Ren *et al.*, 2018] to restore haze-free images via atmospheric scattering model. For example, [Ren *et al.*, 2016] proposed a Multi-Scale CNN (MSCNN) model, which consists of coarse-scale and fine-scale networks to estimate the transmission map. [Cai *et al.*, 2016] proposed a DehazeNet approach, which is an end-to-end CNN network for estimating the transmission map with a novel BReLU unit. In addition, All-In-One Dehazing Network (AOD-Net) [Li *et al.*, 2017] reformulated a new atmospheric scattering model from the classic one by leveraging a linear transformation to integrate both the transmission map function and the atmospheric light into a unified map. More recently, [Zhang and Patel, 2018] proposed a multi-task method with three components, including a densely connected encoder-decoder net-

work to estimate the transmission map, a U-net network to predict the atmospheric light, and a GAN to discriminate the generated images and the real ones, respectively.

It can be known that the existing prior-based and learning-based methods estimate the transmission map and atmospheric light directly from the hazy image. However, few existing methods take the image noise into consideration, which may lead to image blurring and noise amplifying in dehazing images.

3 The Dual-Path in Dual-Path Network (DPDP-Net)

To have a better understanding of our work, we first briefly review the atmospheric scattering model. Then a detailed introduction of our proposed DPDP-Net, including the first-level and the second level dual-path networks is presented. Finally, we illustrate how to use the estimated transmission map and non-uniform atmospheric light to restore the dehazing image.

3.1 The Atmospheric Scattering Model with Noise

The atmospheric scattering model Eq.(1) for the hazy image is expressed as:

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

where $I(x)$ is the hazy image, and $J(x)$ is the dehazing image. The atmospheric light vector $A(x)$ describes the intensity of the ambient light for a particular scene set to be location related in this paper. And $t(x)$ is a transmission matrix defined as:

$$t(x) = e^{-\beta d(x)}, 0 < t(x) < 1, \quad (2)$$

where β is the scattering coefficient of the atmosphere, and $d(x)$ is the distance between the object and the camera. Denote $n(x)$ be the noise, then the hazy image and the dehazing image are respectively expressed as:

$$I(x) = J(x)e^{-\beta d(x)} + A(1 - e^{-\beta d(x)}) + n(x). \quad (3)$$

$$J(x) = I(x)e^{\beta d(x)} + A(1 - e^{-\beta d(x)}) - n(x)e^{-\beta d(x)}. \quad (4)$$

3.2 The First-level Dual-path Network

The first-level dual-path network includes a Dehazing Network and a Denoising Network, where the Dehazing Network is responsible for haze removal in the structural layer, and the Denoising Network deals with noise in the textural layer, respectively. In this subsection, we first introduce how to decompose an image into a structural layer and a textural layer, and then provide the details of the Denoising Network.

Image Structure-texture Decomposition

A common drawback in the existing single image dehazing methods is that they ignore the presence of noise, which is actually inevitable in natural images. According to [Li *et al.*, 2014], haze or fog is mostly related to the structural layer and noise or artifact is related to the textural layer in an image. To preserve more details as well as suppress the noise and artifacts, we propose to address the image noise in single image

dehazing by considering dehazing and denoising simultaneously. First, we decompose the hazy image into a structural layer and a textural layer as follows:

$$I = I_T + I_S, \quad (5)$$

where I_S is the structural layer corresponding to the objects with large gradient magnitudes in the image I , and I_T is the textural layer reflecting the fine details. Based on the Total Variation (TV) regularization technique [Li *et al.*, 2014], the structural layer is obtained by minimizing the following objective function:

$$\min_{I_S} \sum_i (I_S - I_i)^2 + \lambda |\nabla I_S|, \quad (6)$$

where i represents the position of a pixel, λ is the regularization parameter and ∇ is the gradient operator. The textural layer is obtained by calculating the difference between the input image and its structural layer:

$$I_T = I - I_S. \quad (7)$$

The Denoising Network

The structure-texture decomposition explores the fact that the structural layer is related to large gradient magnitudes, while the textural layer captures the smaller gradient magnitudes, exhibiting the fine image details, noise and some artifacts. To preserve more details and suppress the noise and artifacts, we apply FFDNet [Zhang *et al.*, 2017] in the textural layer. FFDNet is a fast and flexible denoising net, it can handle many types of noises, such as camera noise, video noise, JPEG compression noise and some spatially variant noise.

3.3 The Second-Level Dual-Path Network (The Dehazing Network)

The second-level dual-path network lies in the Dehazing Network. It consists of two branches with cascaded convolutional layers, as shown in Figure 4. Specifically, the TM-Net branch learns relationship between the hazy image and the transmission map, while the AL-Net branch learns the relationship between the hazy image and the atmospheric light. Compared with the previous methods, our Dehazing Network is directly applied to the structural layer of the hazy image, which is devoid of noise and artifacts. In addition, we estimate the transmission map and non-uniform atmospheric light respectively in a cascaded CNN model through independent loss function, which guarantees an accurate dehazing performance and improves the flexibility of the network.

Transmission Map Network (TM-Net)

The task of predicting transmission map is to learn a pixel-wise non-linear mapping from a given input image to the corresponding transmission map by minimizing the loss between them. The TM-Net includes three layers and each layer consists of filters with different sizes and numbers. Concretely, the first layer and the second layer have 96 filters of size 9×9 and 64 filters of size 5×5 , respectively. Both take structural layer of hazy images as input to extract the features. And the last layer sets up a convolutional layer to convert the input data into a feature map with a single channel image with

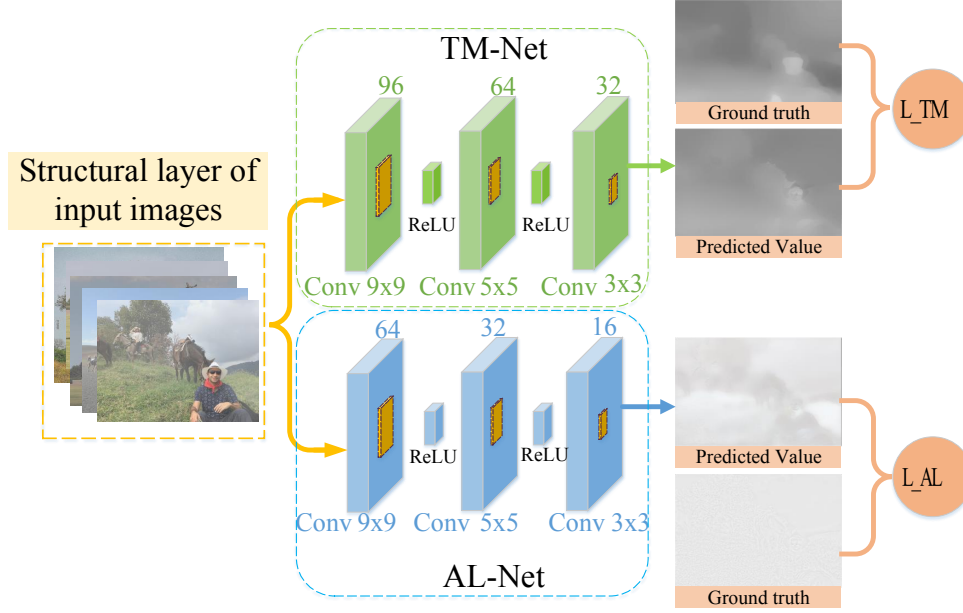


Figure 4: The illustration of the proposed Dehazing Network.

32 filters of size 3×3 . Meanwhile, the non-linear activation function ReLU is used after the first two convolutional layer. Zero padding is applied to the convolutional layer to ensure the same input and output sizes. Besides, we use the filtered transmission map by the guide filtering [He *et al.*, 2010] as the ground truth to compute the loss.

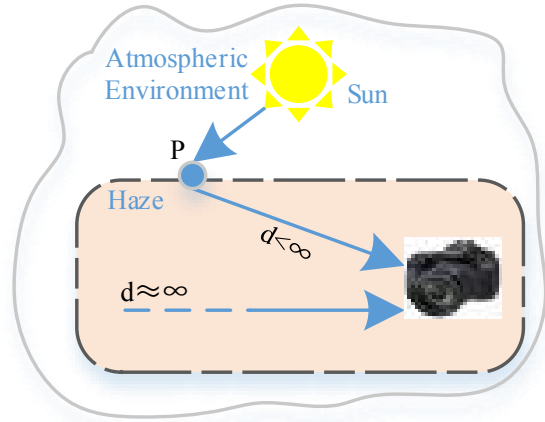
Non-uniform Atmospheric Light Network (AL-Net)

Most of the exiting dehazing methods focus on the estimation of transmission map, paying less attention to the prediction of atmospheric light. For example, [He *et al.*, 2009; Meng *et al.*, 2014; Ren *et al.*, 2016] estimates the atmospheric light by predicted transmission map and consider atmospheric light to be a global constant. However, as shown in Figure 5, suspended particles in the atmosphere such as haze aerosols are always distributed near the surface of the earth. Under the assumption that there is a constant atmospheric light L_∞ , the airlight component at p in Figure 5 is calculated as

$$L_a = L_\infty(1 - e^{-\beta d(p)}) + L_{sun}(p) \quad (8)$$

where L_{sun} is the brightness of direct sunlight. The closer to the sun, the stronger brightness it has. Because $d(p) < \infty$, then $L_a \neq L_\infty$. Only if L_{sun} can be ignored, we have $L_a < L_\infty$. If not, there will be $L_a > L_\infty$. It implies that when an image has a large sky region, the restored image will have color distortion due to the improper assumption of atmospheric light.

Therefore, we treat the atmospheric light as a non-uniform 3-channel matrix, which is able to reflect the spatial-varying intensity of the ambient light. The ground truth and the


 Figure 5: Non-uniform airlight due to haze inhomogeneity and influential sunlight. p is the point in the sky that is closer to the sun and d is the scene depth, i.e., the distance from the point p to the camera.

spatial-varying 3-channel atmospheric light image predicted by the DPDP-Net are shown in Figure 6. The constant atmospheric light, the true transmission map and the clear image are used to generate the hazy image in Figure 6(a). To back-propagate errors to the AL-Net, the ground truth of atmospheric light in Figure 6(b) is synthesized by the hazy structural layer, the clear structural layer, and the true transmission map. It can be observed that the intensity of the sky

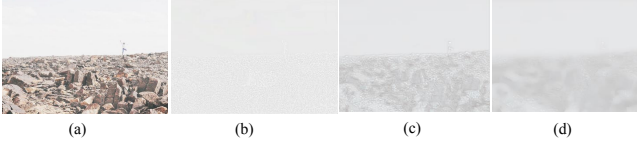


Figure 6: Atmospheric light predicted by DPDP-Net and filtered results. (a) Hazy image. (b) the ground truth of atmospheric light. (c) the predicted non-uniform atmospheric light of DPDP-Net. (d) the filtering result of (c).

region is stronger than the non-sky region. The Figure 6(c) is the predicted atmospheric light, it can be apparently seen that the sky region has a consistent and higher atmospheric light value. Then, to ensure its smoothness, we use guide filtering on the predicted atmospheric light in Figure 6(d). The specific results from different algorithm are compared as shown in Figure 2.

Specifically, the AL-Net includes 3 convolutional layers with different filter sizes followed by ReLU nonlinearity function. Different from the TM-Net, the AL-Net enjoys small amount of filters with lighter weights. Specifically, the first layer consists of 64 filters with the size of 9×9 , the second layer consists of 32 filters with the size of 5×5 , and the third layer has 16 filters with the size of 3×3 , which interacts with the loss layer by combing the 32 channels into feature maps. Zero padding is also performed to ensure the same size of input and output.

Loss Functions

Most of the existing learning-based methods tend to jointly optimize the estimation of atmospheric light and transmission map [Zhang and Patel, 2018; Yang *et al.*, 2017]. They expect to learn the atmospheric light and transmission map simultaneously and design a joint loss function by combining the atmospheric light estimation error and the transmission map estimation error. However, this leads to a result that the dehazing image is a combination of the atmospheric light and the transmission map estimated by their networks. When the error is back propagated, the joint error composed of atmospheric light and the transmission map will respectively back propagate to each of them to cause mutual influence. What's more, the ground truth of atmospheric light in [Yang *et al.*, 2017] is obtained by using hazy image, clear image and the estimated transmission map. When the error back propagates, the inaccuracy of the transmission map will affect the correctness of atmospheric light. That is to say, the errors in one subnetwork will affect the other, leading to the inaccurate estimation of atmospheric light and transmission map. To this end, we design two independent loss functions to respectively train the two subnetworks to eliminate the interactive error influences, as shown in Figure 4.

For the TM-Net, we minimize its MSE loss function L_{TM} between the predicted transmission $f(t)$ and the corresponding ground truth of the transmission map t_{gt} , which is expressed as:

$$L_{TM} = \frac{1}{NHW} \sum_{i=1}^N \|f(t)_i - t_{gt_i}\|^2, \quad (9)$$

where $H \times W$ is the dimension of the predicted transmission, H and W are the height and the width of the image, respectively. And N is the total number of training batches, $f(t)$ is the learned transmission mapping function, t_{gt} is the corresponding truth filtered transmission map by guide filtering.

For the AL-Net, we also minimize the MSE loss function L_{AL} between the estimated value $f(A)$ and corresponding ground truth of the atmospheric light obtained from the true structural layer J_{struct} , which is expressed as:

$$L_{AL} = \frac{1}{NHW} \sum_{i=1}^N \|f(A)_i - A_{gt}\|^2, \quad (10)$$

where

$$A_{gt} = (I_{struct} - J_{struct} \times t_{gt}) / (1 - t_{gt}). \quad (11)$$

3.4 The Final Fusion

When obtaining the dehazing structural layer and the denoising textural layer, the next step is to combine them. However, since most enhancement functions f are not linear, then $f(I_S + I_T) \neq f(I_S) + f(I_T)$, we cannot sum them directly. As a consequence, we have to approximate the enhancement function by setting a scale multiplication factor β , which can obey the following condition: $f(I) = f(I_S) + \beta D(I_T)$. Where, $D(I_T)$ is the denoising operation. Thus, we rewrite the Eq. (1) as:

$$J(x) = \frac{I(x) - A(x)}{t(x)} + A(x). \quad (12)$$

The dehazing operation on the right side of the Eq.(12) can be regarded as an operator $f(\bullet)$, according to the Taylor's expansion method: $f(t + \nabla t) \approx f(t) + f'(t) \nabla t$, the Eq.(12) is rewritten as:

$$\begin{aligned} J &= f(I) = f(I_S + D(I_T)) \approx f(I_S) + f'(I_S)D(I_T) \\ &= J_S + \frac{1}{t}D(I_T). \end{aligned} \quad (13)$$

According to Eq.(13), the denoising textural layer $D(I_T)$ is added to the dehazing structural layer J_S to generate the dehazing, artifacts-free and denoising image J .

4 Experimental Results

In this section, we demonstrate the superiority of the proposed DPDP-Net on the Synthetic and Real-World Image datasets against several state-of-art single image dehazing methods, including DCP [He *et al.*, 2009], BCCR [Meng *et al.*, 2014], MSCNN [Ren *et al.*, 2016], DehazeNet [Cai *et al.*, 2016], AOD-Net [Li *et al.*, 2017], Gated-Net [Ren *et al.*, 2018], GMAN [Liu *et al.*, 2018], PAD-Net [Liu and Zhao, 2018], Bilinear-Net [Yang *et al.*, 2017] and DFCRN [Xi *et al.*, 2018].

4.1 Experimental Settings

Datasets

The training dataset is composed of 400 synthesized hazy images, which are generated with Eq.(1). Among them, 300 indoor images are from the NYU2 depth dataset [Silberman

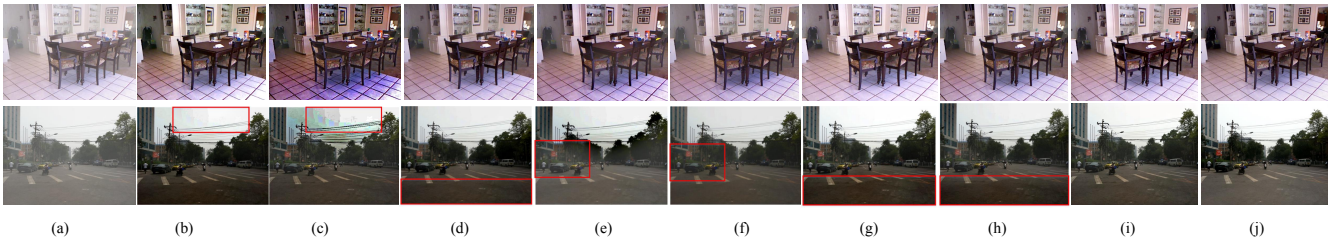


Figure 8: Dehazing results on the Indoor D-Hazy dataset (top) and Outdoor SOTS dataset (bottom). (a) Hazy image. (b) DCP. (c) BCCR. (d) DehazeNet. (e) MSCNN. (f) AOD-Net. (g) Bilinear-Net. (h) DFCRN. (i) DPDP-Net (ours). (j) Ground truth.

et al., 2012] and 100 outdoor images (most of them have large areas of the sky) are from ImageNet dataset [Deng *et al.*, 2009]. Because there is no depth information available in ImageNet dataset, we extract it for outdoor images based on the method in [Liu *et al.*, 2015]. Following [Ren *et al.*, 2016], we randomly set different atmospheric light $A(x)$ by choosing each channel uniformly between $[0.7, 1.0]$. As for the atmospheric attenuation coefficient, we set it from 0.5 to 1.5 for the indoor images since they contain thin haze, and from 1.5 to 3 for outdoor images to improve the generalization capability of the DPDP-Net. The training is conducted on patches with size 64×64 samples from synthesized hazy images. In total, we sample more than 30,000 patches to train the network. In addition, a small validation dataset of some additional hazy and ground truth pairs are selected randomly for tracking model performance and empirically determining the parameters of the proposed model.

To keep the error curve smoothly, the learning rate is set to 0.001 for the first 60 epochs, and 0.0001 for the remaining epochs. Stochastic Gradient Descent (SGD) is employed for learning transmission map and atmospheric light with 0.9 momentum and 0.05 decay parameter for training. The batch size is 64. The whole experiment is conducted on a PC with an Intel(R) Xeon(R) CPU E5-1607 v3@3.1GHz and an Nvidia GeForce GTX 1080 Ti GPU.

4.2 Comparisons on Synthetic Test Datasets

To demonstrate the effectiveness of the proposed DPDP-Net, we compare it with several state-of-the-art methods on four synthetic test datasets. To fully evaluate these methods, we employ the popular PSNR (peak signal-to-noise ratio), SSIM (structural similarity) and FSIM (feature similarity) [Zhang *et al.*, 2011] as the image dehazing evaluation criteria.

First, we synthesize hazy images by using stereo images



Figure 7: Synthetic image examples on Middlebury Stereo Dataset and their corresponding DPDP-Net results.

from Middlebury stereo dataset (2003-2007) [Scharstein and Szeliski, 2003; Scharstein and Pal, 2007] as test images to report the comparison results in Table 1. We can observe that the proposed DPDP-Net performs the best among the competitors. Figure 7 illustrates some dehazing results, from which we can find that the DPDP-Net achieves good visual contrast and restores vivid edge information.

Then, we demonstrate the comparisons on the indoor images from SOTS released by RESIDE [Li *et al.*, 2019] dataset in Table 2. Since there are no FSIM provided in the competing approaches, we do not apply it in Table 2. It is observed that our DPDP-Net achieves the highest SSIM as well as maintaining a good PSNR.

Next, we show the comparison on indoor hazy images using 100 hazy images from D-Hazy dataset [Ancuti *et al.*, 2016], as shown in table 3. We can observe that the DPDP-Net has the highest performances on PSNR and FSIM, and competing performance on SSIM. Several dehazing examples are shown the top line in Figure 8. It can be observed that although the haze can be removed by DCP and BCCR, there still exist serious color distortion in the dehazing images since some priors could be invalid for certain images. On the other hand, DehazeNet often makes some parts of image darker, whereas MSCNN and Bilinear-Net often significantly leads to be oversaturation in images. In addition, AOD-Net and DFCRN can only remove haze to a small extent. In contrast, the results from DPDP-Net are clear and the details are better enhanced with perceptual qualities due to the accurate estimation of the atmospheric light and transmission map.

Finally, we show the comparison result on outdoor images from the SOTS released by RESIDE as shown in the bottom line in Figure 8. The dehazing results of DCP and BCCR have some color distortions and blocking artifacts when the scene objects are similar to the atmospheric light (e.g., the part enclosed in red box on the image from the bottom row). The deep learning-based methods, such as DehazeNet, Bilinear-Net and DFCRN, tend to result in color oversaturation on the ground. The result of MSCNN and ADO-Net have color oversaturation to a small extent, and also have some hazy residuals. In contrast, our proposed DPDP-Net is able to remove haze from the image while preserving the image details well. What's more, from the bottom line of Figure 8, we can see that our results have better color characteristics in the sky than the competing methods.

4.3 Comparisons on Real-World Images

To prove the generalization ability of the proposed method, we evaluate DPDP-Net and other methods on some real-world hazy images. For space limitation, we only provide the results against the latest end-to-end dehazing methods, including Gated-Net [Ren *et al.*, 2018], AIP-Net [Wang *et al.*, 2019], and DCPDN [Zhang and Patel, 2018], as illustrated in Figure 9. It can be observed that dehazing images by Gated-Net have apparent color shift. AIP-Net has a significant color cast and low visual quality. The results by DCPDN also have significant color cast. In contrast, the results of our method not only restore similar visual effect to the original ones, but have better contrast, vivid color and visually pleasing visibility.

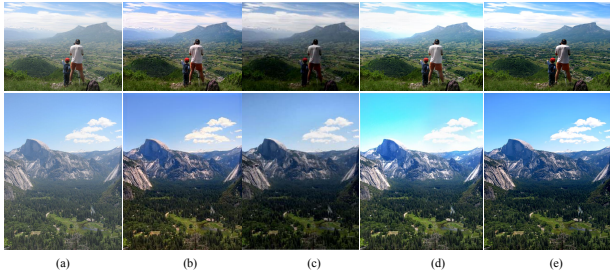


Figure 9: Qualitative Results against end-to-end approaches. (a) Hazy image. (b) Gated-Net. (c) AIP-Net. (d) DCPDN. (e) DPDP-Net (ours).

4.4 Robustness Analysis

Following [Cai *et al.*, 2016], we further demonstrate the robustness of our DPDP-Net. The Airlight Robustness Evaluation (ARE), Coefficient Robustness Evaluation (CRE), Scale Robustness Evaluation (SRE), and Noise Robustness Evaluation (NRE) are employed as criteria.

ARE is an important impact to the restoration quality of dehazing image. This paper generates different test images with different atmospheric light values to verify the robustness of atmospheric light. CRE reflects the haze concentration. An excellent dehazing algorithm should show a good performance to images with different haze concentrations. This paper generates test images with different scattering coefficients to demonstrate the robustness of the scattering coefficient. SRE is necessary for the adaptability of the algorithm. Image scaling and transformation in visual fields usually occur in real life. This paper generates different test images with different scales to verify the robustness of image scales. NRE reflects the noise robustness of the algorithm. It is unavoidable that hazy images containing noise. This paper generates test images with different types of noise and different noise levels to verify the noise robustness.

Metric	DCP	BCCR	MSCNN	DehazeNet	AOD-Net	Bilinear-Net	DPDP-Net (ours)
PSNR	18.03	17.26	16.46	18.76	18.74	18.74	18.76
SSIM	0.82	0.80	0.70	0.82	0.89	0.90	0.91
FSIM	0.92	0.90	0.91	0.93	0.92	0.74	0.94

Table 1: Quantitative comparison on Middlebury stereo dataset.

Metric	DCP	BCCR	MSCNN	DehazeNet	AOD-Net	Gated-Net	GMAN	PAD-Net	DPDP-Net (ours)
PSNR	16.62	16.88	17.57	21.14	19.06	22.30	20.53	20.68	20.18
SSIM	0.81	0.74	0.81	0.84	0.85	0.88	0.81	0.82	0.88

Table 2: Quantitative comparison on SOTS dataset from RESIDES.

Metric	DCP	BCCR	MSCNN	DehazeNet	AOD-Net	Bilinear-Net	DFCRN	DPDP-Net (ours)
PSNR	18.09	14.51	17.99	20.91	16.40	20.46	20.32	21.24
SSIM	0.82	0.80	0.86	0.81	0.75	0.92	0.90	0.86
FSIM	0.96	0.91	0.94	0.96	0.93	0.67	0.97	0.97

Table 3: Quantitative comparison on D-Hazy dataset.

In Table 4, we show the average performances of PSNR and SSIM on an additional 200 pictures from the NYU2 dataset except for training dataset. For ARE, we synthesize hazy images with $\beta=1$ and A (0.7,1.0). For CRE, hazy images are synthesized with A = 0.85 and β (0.5, 1.5). To analyze the influence of the scale variation, we select four scale coefficients, i.e., 1, 0.8, 0.6, 0.4, to generate multi-scale images with $\beta = 1$ and A = 0.85. Finally, Gaussian noise and salt-and-pepper noise are added to the hazy images with $\beta = 1$ and A = 0.85 for NRE. It can be found that our method achieves the best performance on 7 metrics among 8 indicators. The reason is that we optimize the atmospheric light and transmission map via the independent loss function that avoids interact effects. Moreover, due to the Denoising Network, the metrics of the NRE are significantly higher than the other methods.

	Metric	DehazeNet	MSCNN	AOD-Net	Bilinear-Net	DFCRN	DPDP-Net (ours)
ARE	PSNR	21.01	17.75	19.26	22.60	20.43	21.67
	SSIM	0.89	0.84	0.86	0.90	0.89	0.91
CRE	PSNR	21.09	18.21	19.79	22.44	20.66	22.55
	SSIM	0.88	0.84	0.86	0.90	0.89	0.90
SRE	PSNR	21.76	18.46	20.31	21.82	20.4042	23.17
	SSIM	0.9013	0.82	0.87	0.91	0.87	0.91
NRE	PSNR	15.56	14.82	14.64	16.13	15.33	18.52
	SSIM	0.34	0.33	0.35	0.36	0.34	0.60

Table 4: Quantitative comparison on NYU2 dataset.

5 Conclusion

The Dual-Path in Dual-Path Network (DPDP-Net) proposed in this paper perform dehazing and denoising operations simultaneously and outperforms the state-of-the-art methods. The Dehazing Network estimates atmospheric light and transmission map with independent loss functions that ensures an accurate estimation. Particularly, the atmospheric light is set to a non-uniform value to restore illumination intensity of the ambient light to eliminate color imbalance.

Acknowledgements

The work was supported by the National Natural Science Foundation of China under Grant No.61632018, No.61771329 and No.61372145.

References

- [Ancuti *et al.*, 2016] Cosmin Ancuti, Codruta O. Ancuti, and Christophe De Vleeschouwer. D-hazy: A dataset to evaluate quantitatively dehazing algorithms. In *ICIP*, pages 2226–2230, 2016.
- [Berman *et al.*, 2016] Dana Berman, Tali Treibitz, and Shai Avidan. Non-local image dehazing. In *CVPR*, pages 1674–1682, 2016.
- [Cai *et al.*, 2016] Bolun Cai, Xiangmin Xu, Kui Jia, Chunmei Qing, and Dacheng Tao. Dehazenet: An end-to-end system for single image haze removal. *TIP*, 25(11):5187–5198, 2016.
- [Deng *et al.*, 2009] Jia Deng, Wei Dong, Richard Socher, Li Jia Li, Kai Li, and Fei Fei Li. Imagenet: a large-scale hierarchical image database. In *CVPR*, pages 248–255, 2009.
- [He *et al.*, 2009] Kaiming He, Jian Sun, and Xiaoou Tang. Single image haze removal using dark channel prior. In *CVPR*, pages 1956–1963, 2009.
- [He *et al.*, 2010] Kaiming He, Sun Jian, and Xiaoou Tang. Guided image filtering. In *ECCV*, 2010.
- [Lai *et al.*, 2015] Yi Hsuan Lai, YiLei Chen, Chuan Ju Chiou, and Chiou Ting Hsu. Single image dehazing via a optimal transmission map under scene priors. *TCSVT*, 25(1):1–14, 2015.
- [Li *et al.*, 2014] Yu Li, Fangfang Guo, Robby T. Tan, and Michael S. Brown. A contrast enhancement framework with jpeg artifacts suppression. In *ECCV*, pages 174–188, 2014.
- [Li *et al.*, 2017] Boyi Li, Xiulian Peng, Zhangyang Wang, Jizheng Xu, and Dan Feng. Aod-net: All-in-one dehazing network. In *ICCV*, pages 4780–4788, 2017.
- [Li *et al.*, 2019] Boyi Li, Wenqi Ren, Dengpan Fu, Dacheng Tao, Dan Feng, Wenjun Zeng, and Zhangyang Wang. Benchmarking single-image dehazing and beyond. *TIP*, 28(1):492–505, 2019.
- [Liu and Zhao, 2018] Yu Liu and Guanlong Zhao. Pad-net: A perception-aided single image dehazing network. *CoRR*, abs/1805.03146, 2018.
- [Liu *et al.*, 2015] Fayao Liu, Chunhua Shen, and Guosheng Lin. Deep convolutional neural fields for depth estimation from a single image. In *CVPR*, pages 5162–5170, 2015.
- [Liu *et al.*, 2018] Zheng Liu, Botao Xiao, Muhammad Alrabeiah, Keyan Wang, and Jun Chen. Generic model-agnostic convolutional neural network for single image dehazing. *CoRR*, abs/1810.02862, 2018.
- [Meng *et al.*, 2014] Gaofeng Meng, Ying Wang, Jianguo Duan, Shiming Xiang, and Chunhong Pan. Efficient image dehazing with boundary constraint and contextual regularization. In *ICCV*, pages 617–624, 2014.
- [Narasimhan and Nayar, 2002] Srinivasa G. Narasimhan and Shree K. Nayar. Vision and the atmosphere. *IJCV*, 48(3):233–254, 2002.
- [Narasimhan and Nayar, 2003] S. G. Narasimhan and S. K. Nayar. Contrast restoration of weather degraded images. *TPAMI*, 25(6):713–724, 2003.
- [Nayar, 1999] Shree K. Nayar. Vision in bad weather. *ICCV*, 1999.
- [Qingsong *et al.*, 2015] Zhu Qingsong, Mai Jiaming, and Shao Ling. A fast single image haze removal algorithm using color attenuation prior. *TIP*, 24(11):3522–3533, 2015.
- [Ren *et al.*, 2016] Wenqi Ren, Si Liu, Hua Zhang, Jinshan Pan, Xiaochun Cao, and Ming-Hsuan Yang. Single image dehazing via multi-scale convolutional neural networks. In *ECCV*, pages 154–169, 2016.
- [Ren *et al.*, 2018] Wenqi Ren, Lin Ma, Jiawei Zhang, Jinshan Pan, Xiaochun Cao, Wei Liu, and Ming-Hsuan Yang. Gated fusion network for single image dehazing. In *CVPR*, 2018.
- [Scharstein and Pal, 2007] D. Scharstein and C. Pal. Learning conditional random fields for stereo. In *CVPR*, pages 1–8, 2007.
- [Scharstein and Szeliski, 2003] D. Scharstein and R. Szeliski. High-accuracy stereo depth maps using structured light. In *CVPR*, pages 195–202, 2003.
- [Silberman *et al.*, 2012] Nathan Silberman, Derek Hoiem, Pushmeet Kohli, and Rob Fergus. Indoor segmentation and support inference from rgb-d images. In *ECCV*, pages 746–760, 2012.
- [Tan, 2008] Robby T. Tan. Visibility in bad weather from a single image. In *CVPR*, pages 1–8, 2008.
- [Wang *et al.*, 2019] A. Wang, W. Wang, J. Liu, and N. Gu. Aipnet: Image-to-image single image dehazing with atmospheric illumination prior. *TIP*, 28(1):381–393, 2019.
- [Xi *et al.*, 2018] Zhao Xi, Keyan Wang, Yunsong Li, and Jiaojiao Li. Deep fully convolutional regression networks for single image haze removal. In *VCIP*, 2018.
- [Xiaogang *et al.*, 2014] Wang Xiaogang, Wang Meng, and Li Wei. Scene-specific pedestrian detection for static video surveillance. *TPAMI*, 36(2):361–374, 2014.
- [Yang *et al.*, 2017] Hui Yang, Jinshan Pan, Qiong Yan, Wenxiu Sun, Jimmy S. J. Ren, and Yu-Wing Tai. Image dehazing using bilinear composition loss function. *CoRR*, abs/1710.00279, 2017.
- [Zhang and Patel, 2018] H. Zhang and V. M. Patel. Densely connected pyramid dehazing network. In *CVPR*, pages 3194–3203, 2018.
- [Zhang *et al.*, 2011] Lin Zhang, Lei Zhang, Xuanqin Mou, and David Zhang. Fsim: A feature similarity index for image quality assessment. *TIP*, pages 2378 – 2386, 2011.
- [Zhang *et al.*, 2017] Kai Zhang, Wangmeng Zuo, and Lei Zhang. Ffdnet: Toward a fast and flexible solution for cnn based image denoising. *TIP*, 2017.
- [Zheng *et al.*, 2012] W. S. Zheng, S. Gong, and T. Xiang. Quantifying and transferring contextual information in object detection. *TPAMI*, 34(4):762–777, 2012.