Learning to dehaze with polarization

Anonymous Author(s)
Affiliation
Address
email

Abstract

Haze, a common kind of bad weather caused by atmospheric scattering, decreases the visibility of scenes and degenerates the performance of computer vision algorithms. Single-image dehazing methods have shown their effectiveness in a large variety of scenes, however, they are based on handcrafted priors or learned features, which do not generalize well to real-world images. Polarization information can be used to relieve its ill-posedness, however, real-world images are still challenging since existing polarization-based methods usually assume that the transmitted light is not significantly polarized, and they require specific clues to estimate necessary physical parameters. In this paper, we propose a generalized physical formation model of hazy images and a robust polarization-based dehazing pipeline without the above assumption or requirement, along with a neural network tailored to the pipeline. Experimental results show that our approach achieves state-of-the-art performance on both synthetic data and real-world hazy images.

1 Introduction

When taking photos in hazy environments, the visibility and color fidelity of recorded scenes are usually contaminated, because the captured images often contain a superposition of two unknown components: the transmitted light (an attenuated fraction of original scene radiance), and the airlight (ambient light scattered towards the viewer). It is highly ill-posed to separate them in a single hazy image as it requires estimating multiple unknowns from a single observation. Handcrafted priors from natural image statistics [21, 15, 1] have been wildly used to solve this problem. With the development of deep neural networks, learning-based methods (*e.g.*, CNN-based [24, 65, 7, 82] and GAN-based [79, 5]) have also been adopted to recover the haze-free images by extracting image features from a large amount of training data. However, these methods do not generalize well to real-world images, because they depend strongly on the image features extracted from training data and do not explicitly consider useful constraints from physical image formation models.

For better generalization, multi-image dehazing methods have been proposed. They capture multiple images from different viewpoints [38, 52, 84, 63], weather conditions [58, 55, 56, 57], or polarization angles [77, 78, 53, 81, 76, 25, 54]. Although all of these multi-image dehazing methods can relieve the ill-posedness, however, polarization-based ones have their unique advantages, since they directly utilize the physical image formation model with less dependency on image features extracted from training data. Nowadays, multiple polarized images can be conveniently captured in a single shot using a polarization camera such as Lucid Vision Phoenix polarization camera[1]. However, these polarization-based methods are not robust due to several issues: (1) They are largely based on a strong assumption that the transmitted light is not significantly polarized, while it is not the case for real-world images since both transmitted light and airlight contribute to the polarization [13, 29]. (2) They usually require specific clues (*e.g.*, sky regions [13, 77, 53], similar objects [54, 50], known

[1] https://thinklucid.com/product/phoenix-5-0-mp-polarized-model/

Submitted to 35th Conference on Neural Information Processing Systems (NeurIPS 2021). Do not distribute.

depth [54]) to estimate the infinite airlight and the degree of polarization (DoP), which significantly reduces their applicability since these requirements are not always met. (3) They are optimization-based methods which do not make full use of semantic and contextual information in image features to handle the spatially-variant real-world scattering.

In this paper, for enabling the polarization-based dehazing methods to handle images captured in the wild more robustly, we propose a generalized physical formation model of hazy images, without assuming that the transmitted light is not significantly polarized, while considering the spatially-variant real-world scattering. Based on the physical model, we propose a robust polarization-based dehazing pipeline to extend their applicability by adopting deep learning to estimate the infinite airlight and the DoP of both transmitted light and airlight without the requirement of specific clues like sky regions, similar objects, *etc*. According to our dehazing pipeline, we design a neural network to perform the dehazing process: It first estimates the DoP of both transmitted light and airlight to solve the transmitted light, then predicts the infinite airlight to reconstruct the original scene radiance. Thanks to our learning-based pipeline, our method extracts image features from training data and use semantic and contextual information to refine the results, which is suitable for handling the spatially-variant real-world scattering.

To summarize, this paper makes contributions by demonstrating: (1) A generalized physical formation model of hazy images, taking into account the polarization effects of both transmitted light and airlight, along with the spatially-variant real-world scattering. (2) A robust polarization-based dehazing pipeline without the requirement of specific clues, by adopting deep learning to estimate necessary physical parameters (infinite airlight, DoP of both transmitted light and airlight). (3) A neural network making full use of semantic and contextual information to handle the spatially-variant real-world scattering to improve the clarity of original scene radiance recovery.

Experimental results show that our approach achieves state-of-the-art performance on both synthetic data and real-world hazy images.

2 Related Work

Single-image dehazing. Single-image dehazing is a highly ill-posed problem because it requires estimating multiple unknowns (the transmitted light and airlight) from a single observation. Park *et al*. [64] estimated haze from the difference among the RGB channels. Some methods adopted an adaptive contrast enhancement strategy to maximize the local contrast of restored images [87, 17, 47]. Some works proposed several assumptions (*e.g*., the surface shading and transmission are locally uncorrelated [14], the scene albedo and depth are independent [59], both the scene albedo and transmission are constant inside each patch [86]) or image priors (*e.g*., dark channel prior [21, 89, 49, 37], color attenuation prior [15], non-local prior [1, 43], ellipsoid prior [19], color-lines [15]) to handle this problem. Recently, with the development of deep neural networks, learning-based methods have also been adopted to recover haze-free images by extracting image features from a large amount of training data. These learning-based methods could be divided into two groups: direct methods, which dehaze in an end-to-end manner using convolutional neural networks (CNN) [2, 68, 30, 31, 62, 6, 18, 45, 95, 44, 36, 24, 65, 7, 8, 70, 94, 51, 4, 48, 3] or generative adversarial networks (GAN) [93, 69, 35, 92, 66, 33, 34, 79, 5, 9, 12, 85, 10, 82], and indirect methods, which estimate image priors or their variants [88, 74, 91, 46, 20] first and then use them to dehaze. Although these methods have shown their effectiveness in a large variety of scenes, their generalization ability is still limited, since image priors are not always observed in the input and the image features extracted from synthetic training data often have a large domain gap with real-world ones.

Multi-image dehazing. For better generalization and less ill-posedness, multi-image dehazing methods have been proposed. They use computational photography techniques to capture multiple images with conventional or unconventional cameras for acquiring extra information. Some methods used separately measured range data [60] and georeferenced digital terrain with urban models [28] to facilitate dehazing. Some works captured multiple images under different unknown weather conditions to recover scene depth maps for dehazing [58, 55, 56, 57]. Some approaches take advantage of stereo vision to remove the effects of haze [38, 52, 84, 63] by taking multiple photos from different viewpoints. Some methods fuse RGB images with NIR (near-infrared) ones to help dehaze [75, 16, 42, 11, 83], because the scattering is significantly smaller in NIR than in visible

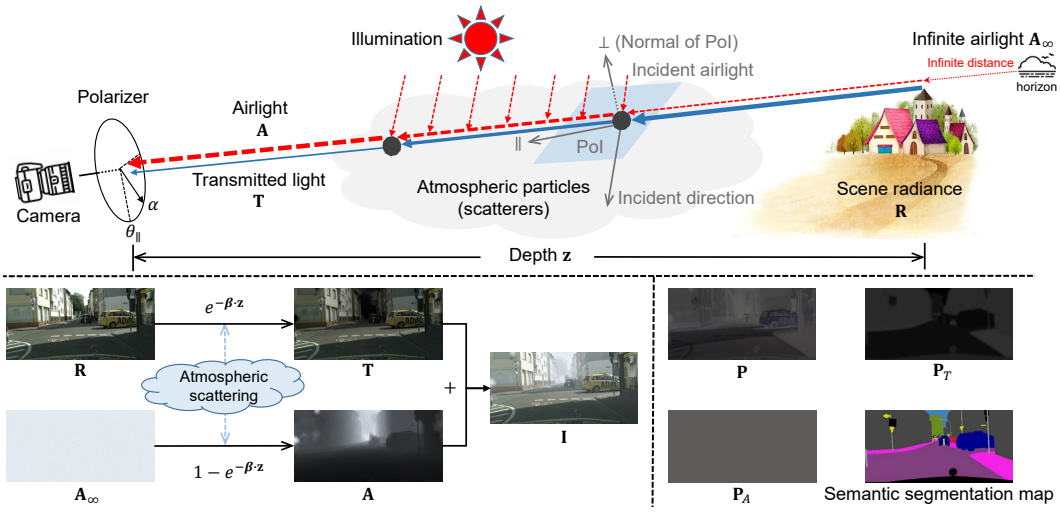

Figure 1: Top row: An illustration of the atmospheric scattering and polarization; transmitted light (blue solid line) **T** is an attenuated fraction of original scene radiance **R** that decreases with the scene depth **z**; airlight **A** (red dashed line) is the ambient light scattered towards the viewer that increases with **z**; when placing a linear polarizer with polarization angle α in front of the camera, the polarization component parallel to the plane of incidence (PoI) is best transmitted through the polarizer at $\alpha = \theta_\parallel$. Bottom left: A synthetic example for visualizing the formation of a hazy image (see Equation (1) for details). Bottom right: A synthetic example for visualizing **P**, $\mathbf{P}_T$, and $\mathbf{P}_A$ (the DoP of **I**, **T**, and **A**, see Equation (2) for details), along with the semantic segmentation map.

light since NIR wavelengths are longer. Although these methods have better generalization ability, capturing such data is not easy since they require multiple shots and/or complicated imaging systems.

Polarization-based dehazing. Recently, polarization-based methods have been proposed to solve the dehazing problem by capturing multiple polarized images at the same view with different polarization angles. These methods have their unique advantages: they directly utilize the physical image formation model without dependence on image features extracted from training data, and multiple polarized images can be captured in a single shot using a polarization camera. However, most of them are based on a strong assumption that the transmitted light is not significantly polarized [77, 78, 53, 81, 76, 25, 54, 50, 41, 40, 39, 67, 80], while it is not the case for real-world images since both transmitted light and airlight contribute to the polarization [13, 29]. Fang *et al*. [13] takes the polarization of transmitted light into consideration, however, it supposes that the depth of sky regions is approximately infinite and require sky regions to estimate the infinite airlight and the DoP, just like [77, 53], which significantly reduces the applicability since sky regions are not always available.

3 Method

In this section, we show the physical formation model of hazy images in Section 3.1, demonstrate our polarization-based dehazing pipeline in Section 3.2, and introduce our neural network in Section 3.3.

3.1 Physical image formation model

As shown in Figure 1 (top row), when taking photos in hazy environments, caused by atmospheric scattering, the captured image $\mathbf{I} = \{I(\mathbf{x}, c)\}$ ($\mathbf{x}$ and c are pixel coordinates and channel respectively) is composed of two components: the transmitted light $\mathbf{T} = \{T(\mathbf{x}, c)\}$ (an attenuated fraction of original scene radiance $\mathbf{R} = R(\mathbf{x}, c)$), which decreases with the scene depth $\mathbf{z} = \{z(\mathbf{x})\}$, and the airlight $\mathbf{A} = \{A(\mathbf{x}, c)\}$ (ambient light scattered towards the viewer), which increases with $\mathbf{z}$. According to [77], the formation of a hazy image can be described as

$$\mathbf{I} = \mathbf{T} + \mathbf{A} = \mathbf{R} \cdot e^{-\boldsymbol{\beta} \cdot \mathbf{z}} + \mathbf{A}_\infty \cdot (1 - e^{-\boldsymbol{\beta} \cdot \mathbf{z}}), \tag{1}$$

where $\boldsymbol{\beta} = \{\beta(c)\}$ is the scattering coefficient, $\mathbf{A}_\infty = \{A_\infty(c)\}$ denotes the infinite airlight (the airlight radiance corresponding to an object at an infinite distance, *e.g.*, the horizon), and $\cdot$ stands

for element-wise multiplication. A synthetic example of its visualization can be found in Figure 1 (bottom left). However, real-word scattering does not always satisfy such an ideal model, which means that both $\boldsymbol{\beta}$ and $\mathbf{A}_\infty$ are not only dependent on wavelength, but also on the the size of the scattering particles [27, 22] and angular scattering coefficient [78]. To encode such variations, we replace them with $\boldsymbol{\beta} = \{\bar{\beta}(c) + N(\mathbf{x}, c)\}$ and $\mathbf{A}_\infty = \{\bar{A}_\infty(c) + N(\mathbf{x}, c)\}$ respectively, where $\bar{\cdot}$ marks the mean value and $N(\mathbf{x}, c)$ denotes the spatially-variant turbulence. Assume for a moment that the illumination of any scattering particle comes from one direction, the light ray from the source to a scatterer and the line-of-sight from the camera to the scatterer define a plane of incidence (PoI) [77], as shown in Figure 1 (top row). We divide $\mathbf{I}$, $\mathbf{T}$, and $\mathbf{A}$ into two components respectively: $\mathbf{I}^\parallel$ and $\mathbf{I}^\perp$, $\mathbf{T}^\parallel$ and $\mathbf{T}^\perp$, $\mathbf{A}^\parallel$ and $\mathbf{A}^\perp$, where the subscript $\parallel$ ($\perp$) means the component is parallel (perpendicular) to the PoI. The degrees of polarization (DoP) of $\mathbf{I}$, $\mathbf{T}$, and $\mathbf{A}$ are defined as

$$\mathbf{P} \triangleq \frac{\mathbf{I}^\perp - \mathbf{I}^\parallel}{\mathbf{I}}, \quad \mathbf{P}_T \triangleq \frac{\mathbf{T}^\perp - \mathbf{T}^\parallel}{\mathbf{T}}, \quad \text{and} \quad \mathbf{P}_A \triangleq \frac{\mathbf{A}^\perp - \mathbf{A}^\parallel}{\mathbf{A}}, \tag{2}$$

respectively, where

$$\mathbf{I} = \mathbf{I}^\perp + \mathbf{I}^\parallel, \quad \mathbf{T} = \mathbf{T}^\perp + \mathbf{T}^\parallel, \quad \text{and} \quad \mathbf{A} = \mathbf{A}^\perp + \mathbf{A}^\parallel. \tag{3}$$

Since the scattered light is partially polarized perpendicular to the PoI [77, 27, 22, 13], $\mathbf{P}$, $\mathbf{P}_T$, and $\mathbf{P}_A$ are not less than zero. Besides, although $\mathbf{P}$ is spatially-variant (*i.e.*, $\{\mathbf{P} = P(\mathbf{x}, c)\}$), the distribution of $\mathbf{P}_T$ and $\mathbf{P}_A$ are not irregular: the values of $\mathbf{P}_T$ are approximately uniform in the same semantic segment[2], while $\mathbf{P}_A$ can be regarded as spatially-uniform, *i.e.*, $\{\mathbf{P}_A = P_A(c)\}$, according to [13]. A synthetic example of their visualization can be found in Figure 1 (bottom right).

When we place a polarizer with polarization angle α in front of the camera, according to Malus' law [22], the captured polarized image $\mathbf{I}_\alpha$ can be calculated as

$$\mathbf{I}_\alpha = \frac{\mathbf{I} \cdot (1 - \mathbf{P} \cdot \cos(2(\alpha - \boldsymbol{\theta}_\parallel)))}{2}, \tag{4}$$

where $\boldsymbol{\theta}_\parallel = \{\bar{\theta}_\parallel + N(\mathbf{x}, c)\}$ denotes the orientation of the polarizer for best transmission of the component parallel to the PoI. Similarly, the two components $\mathbf{T}$ and $\mathbf{A}$ at angle alpha can be calculated as

$$\mathbf{T}_\alpha = \frac{\mathbf{T} \cdot (1 - \mathbf{P}_T \cdot \cos(2(\alpha - \boldsymbol{\theta}_\parallel)))}{2} \quad \text{and} \quad \mathbf{A}_\alpha = \frac{\mathbf{A} \cdot (1 - \mathbf{P}_A \cdot \cos(2(\alpha - \boldsymbol{\theta}_\parallel)))}{2}, \tag{5}$$

which satisfy $\mathbf{I}_\alpha = \mathbf{T}_\alpha + \mathbf{A}_\alpha$. Note that both $\mathbf{T}$ and $\mathbf{A}$ contribute to the polarization, and the polarization of $\mathbf{T}$ should not be ignored [13, 29]. From Equation (4) and Equation (5), we can derive the following equation

$$\mathbf{I} \cdot \mathbf{P} = \mathbf{T} \cdot \mathbf{P}_T + \mathbf{A} \cdot \mathbf{P}_A, \tag{6}$$

which reveals that the relationship among $\mathbf{I}$, $\mathbf{T}$, and $\mathbf{A}$ are determined by $\mathbf{P}$, $\mathbf{P}_T$, and $\mathbf{P}_A$.

3.2 Polarization-based dehazing pipeline

We aim to restore the original scene radiance $\mathbf{R}$ by using three polarized images $\mathbf{I}_{\alpha^{(i)}} (i = 1, 2, 3)$ captured at the same view with different polarization angles $\alpha^{(i)} (i = 1, 2, 3)$. Eliminating $\mathbf{A}$ from Equation (1) and Equation (6), $\mathbf{T}$ and $\mathbf{R}$ could be computed by the following two equations:

$$\text{(a)} \quad \mathbf{T} = \frac{\mathbf{P} \cdot \mathbf{I} - \mathbf{I} \cdot \mathbf{P}_A}{\mathbf{P}_T - \mathbf{P}_A} \quad \text{and} \quad \text{(b)} \quad \mathbf{R} = \frac{\mathbf{T} \cdot \mathbf{A}_\infty}{\mathbf{A}_\infty - (\mathbf{I} - \mathbf{T})}, \tag{7}$$

where $\mathbf{P}_T$, $\mathbf{P}_A$, and $\mathbf{A}_\infty$ are required to be estimated, $\mathbf{I}$ and $\mathbf{P}$ can be calculated by $\mathbf{I}_{\alpha^{(i)}} (i = 1, 2, 3)$.

Now we first explain how to calculate $\mathbf{I}$ and $\mathbf{P}$ using $\mathbf{I}_{\alpha^{(i)}} (i = 1, 2, 3)$. Expanding Equation (4), we obtain ($\langle\rangle$ denotes inner product)

$$\begin{aligned} &\mathbf{I}_\alpha = \left\langle \begin{bmatrix} \frac{1}{2} & \frac{-\cos(2\alpha)}{2} & \frac{-\sin(2\alpha)}{2} \end{bmatrix}, [\mathbf{D}_1 \quad \mathbf{D}_2 \quad \mathbf{D}_3] \right\rangle, \\ &\text{where} \quad \mathbf{D}_1 = \mathbf{I}, \quad \mathbf{D}_2 = \mathbf{I} \cdot \mathbf{P} \cdot \cos(2\theta_\parallel) \quad \text{and} \quad \mathbf{D}_3 = \mathbf{I} \cdot \mathbf{P} \cdot \sin(2\theta_\parallel). \end{aligned} \tag{8}$$

[2]The polarization properties of transmitted light depend on material properties of scene objects (*e.g.*, surface texture) [29], and objects in the same semantic segment often have similar material properties.

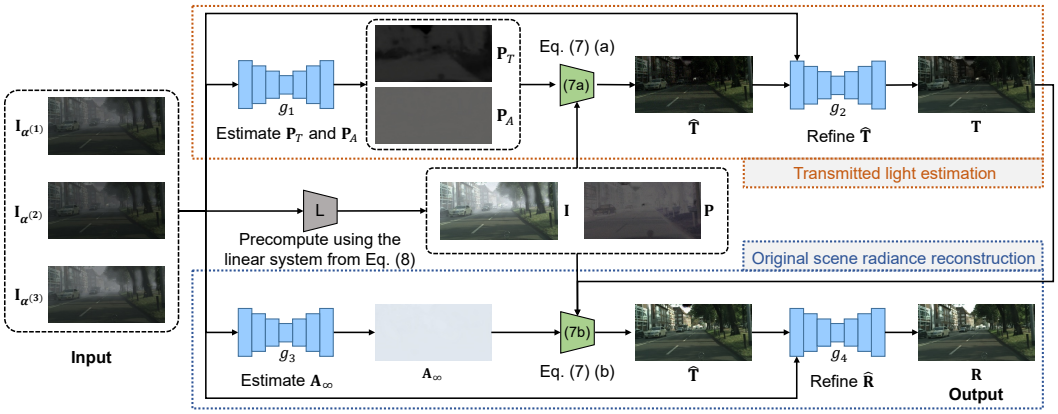

Figure 2: We design a network tailored to our polarization-based dehazing pipeline (in Section 3.2), which takes three polarized images $\mathbf{I}_{\alpha^{(i)}}(i = 1, 2, 3)$ captured at the same view with different polarization angles $\alpha^{(i)}(i = 1, 2, 3)$ as the input (along with the precomputed hazy image $\mathbf{I}$ and its DoP $\mathbf{P}$ using the linear system from Equation (8)) and outputs the reconstructed original scene radiance $\mathbf{R}$. It consists of two stages: transmitted light estimation and original scene radiance reconstruction. The first stage includes two subnetworks for estimating $\mathbf{P}_T$, $\mathbf{P}_A$, and refining $\widehat{\mathbf{T}}$. The second stage also includes two subnetworks for estimating $\mathbf{A}_\infty$ and refining $\widehat{\mathbf{R}}$. ($\widehat{\cdot}$ denotes the coarse value calculated from Equation (7)).

Since Equation (8) has three unknowns ($\mathbf{D}_i(i = 1, 2, 3)$), we can use $\mathbf{I}_{\alpha^{(i)}}(i = 1, 2, 3)$ to obtain a linear system that allows to compute them. Then, we could calculate $\mathbf{I}$ and $\mathbf{P}$ by $\mathbf{I} = \mathbf{D}_1$ and $\mathbf{P} = \frac{\sqrt{(\mathbf{D}_2^2+\mathbf{D}_3^2)}}{\mathbf{D}_1}$.

Next, we only need to estimate three parameters $\mathbf{P}_T$, $\mathbf{P}_A$, and $\mathbf{A}_\infty$ to reconstruct $\mathbf{R}$. To alleviate the dependency on specific clues (such as sky regions [13, 77, 53] or similar objects [54, 50], which are required by other polarization-based methods) for estimating these parameters, we choose to design a deep neural network that comprehensively explores physics and semantic features.

3.3 Polarization-based dehazing network

As shown in Figure 2, our network consists of two stages: transmitted light estimation and original scene radiance reconstruction.

Transmitted light estimation. As shown in the first stage of Figure 2, it aims to estimate the DoP of both transmitted light and airlight ($\mathbf{P}_T$ and $\mathbf{P}_A$) for solving the transmitted light $\mathbf{T}$. So, it adopts a subnetwork g_1 to estimate $\mathbf{P}_T$ and $\mathbf{P}_A$, then uses Equation (7) (a) to calculate $\widehat{\mathbf{T}}$ (the coarse value of $\mathbf{T}$). However, we cannot directly feed $\widehat{\mathbf{T}}$ into the second stage since the numerical problem will occur when the denominator of Equation (7) (a) approaches zero, which often happens in pixels where $\mathbf{P}_T \approx \mathbf{P}_A$ (the DoP of transmitted light and airlight are approximately the same). Besides, the estimated $\mathbf{P}_T$ and $\mathbf{P}_A$ by g_1 are prone to be noisy which distorts the calculated $\widehat{\mathbf{T}}$, because the spatially-variant turbulence is hard to learn due to its irregularities. So, we adopt another subnetwork g_2 to refine $\widehat{\mathbf{T}}$ using semantic and contextual information extracted from $\mathbf{I}_{\alpha^{(i)}}(i = 1, 2, 3)$. In practice, we construct g_1 using the U-Net architecture [71] since it works well on per-pixel estimation tasks such as semantic segmentation [71, 61]. As for g_2, we choose the autoencoder architecture [23], by virtue of its excellent context generalization ability for refining image contents.

Original scene radiance reconstruction. As shown in the second stage of Figure 2, it aims to estimate the infinite airlight $\mathbf{A}_\infty$ to reconstruct the original scene radiance $\mathbf{R}$. So, it first adopts a subnetwork g_3 to estimate $\mathbf{A}_\infty$, then uses Equation (7) (b) to calculate $\widehat{\mathbf{R}}$ (the coarse value of $\mathbf{R}$). However, $\widehat{\mathbf{R}}$ also needs to be refined, because when the haze in some pixels is very thick and leaves little information of the transmitted light ($\mathbf{T} \approx \mathbf{0}$), the numerator of Equation (7) (b) approaches zero, which leads to a wrong result that $\mathbf{R} \approx \mathbf{0}$. So, similar to the first stage, we also adopt a subnetwork g_4 to refine $\widehat{\mathbf{R}}$. We also choose the U-Net architecture [71] for g_3 and the autoencoder architecture [23] for g_4.

4 Data preparation and network training

In this section, we first detail our synthetic dataset generation pipeline in Section 4.1, then show our loss function and training strategy in Section 4.2.

4.1 Synthetic dataset generation pipeline

It is difficult to obtain pairwise hazy and clear images with three polarized observations at a large scale. Besides, getting the ground truth values of the DoP or infinite airlight is not feasible. So, we propose to generate a synthetic dataset for training our network. Since we require spatially-variant $\boldsymbol{\beta}$ and $\mathbf{A}_\infty$ to simulate real-world scattering, and need the semantic segmentation map $\mathbf{S}$ for generating reasonable $\mathbf{P}_T$ (see Section 3.1 for details about the properties of $\mathbf{P}_T$), we cannot directly generate the polarized images from the hazy images in existing dehazing benchmarks [32, 97, 96, 73, 72]. The desired data source for generating our dataset should provide: (1) clear image $\mathbf{R}$ with depth map $\mathbf{z}$, from which we can calculate $\mathbf{I}$ using Equation (1) by generating spatially-variant $\boldsymbol{\beta}$ and $\mathbf{A}_\infty$; (2) semantic segmentation map $\mathbf{S}$, from which we can generate reasonable $\mathbf{P}_T$ using $\mathbf{P}_T = f(\mathbf{S})$, where f denotes a function which randomly maps each semantic segment to a value of $\mathbf{P}_T$.

The Foggy Cityscapes-DBF dataset [72] meets the above two requirements[3], so we use the provided $\mathbf{z}$, $\mathbf{R}$, and $\mathbf{S}$ to generate our synthetic dataset. In short, with $\mathbf{z}$, $\mathbf{R}$, and $\mathbf{S}$ available, our synthetic dataset generation pipeline could be described as: (1) Randomly generate $\boldsymbol{\beta}$ (in $[0.01, 0.02]$), $\mathbf{A}_\infty$ (in $[0.85, 0.95]$), and $\mathbf{P}_A$ (in $[0.05, 0.4]$) to calculate $\mathbf{T}$, $\mathbf{A}$, and $\mathbf{I}$ using Equation (1); (2) generate $\mathbf{P}_T$ from $\mathbf{S}$ using $\mathbf{P}_T = f(\mathbf{S})$ (in $[0.025, 0.2]$), then calculate $\mathbf{P}$ using Equation (6); (3) randomly generate $\theta_\parallel$ (in $[-45°, 45°]$), then use Equation (4) to calculate $\mathbf{I}_{\alpha^{(i)}} (i = 1, 2, 3)$ ($\alpha^{(i)} (i = 1, 2, 3)$ are set to be $0°$, $45°$, and $90°$ respectively).

The visualization of above mentioned parameters can be found in the bottom row of Figure 1. Note that for $\boldsymbol{\beta} = \{\bar{\beta}(c) + N(\mathbf{x}, c)\}$ and $\mathbf{A}_\infty = \{\bar{A}_\infty(c) + N(\mathbf{x}, c)\}$, we first generate their mean values $\bar{\beta}(c)$ and $\bar{A}_\infty(c)$ for each channel, then add 5% Gaussian noise to make them spatially-variant. Besides, we also add 2% Gaussian noise to $\mathbf{I}_{\alpha^{(i)}} (i = 1, 2, 3)$. To conform to the real-world scattering [77], we ensure that $\bar{\beta}(r) < \bar{\beta}(g) < \bar{\beta}(b)$ and $P_A(r) > P_A(g) > P_A(b)$. The images are resized and randomly cropped to 240×240 patches during the training process, and cropped to 496×240 patches for test[4].

4.2 Loss function and training strategy

Loss function. The total loss function of our network is $\mathcal{L} = \lambda_1 \cdot \mathcal{L}_{g_1} + \lambda_2 \cdot \mathcal{L}_{g_2} + \lambda_3 \cdot \mathcal{L}_{g_3} + \lambda_4 \cdot \mathcal{L}_{g_4}$, where $\mathcal{L}_{g_i} (i = 1, 2, 3, 4)$ define the loss of our four subnetworks. Each of them could be described as $\mathcal{L}_{g_i} = 2 \cdot \mathcal{L}_1 + \mathcal{L}_2$, where $\mathcal{L}_1$ and $\mathcal{L}_2$ denote the L_1 and L_2 loss respectively. $\lambda_i (i = 1, 2, 3, 4)$ are empirically set to be 1.0, 1.0, 2.0, 2.0 respectively.

Training strategy. We implement our network using PyTorch on an NVIDIA 2080Ti GPU and apply a two-phase training strategy. First, to ensure a stable initialization of the training process, we train our two network stages independently for 400 epochs. ADAM optimizer [26] is used with an initial learning rate 5×10^{-4} for the first 300 epochs, and a linear decay to 2.5×10^{-4} in the next 100 epochs. Then, we finetune the entire network in an end-to-end manner for another 300 epochs, keeping the learning rate to 5×10^{-4}. Instance normalization [90] are added during training.

5 Experiments

5.1 Evaluation on synthetic data

We compare our results to a representative polarization-based dehazing algorithm SPCVE [54] which also takes three polarized images as the input and five state-of-the-art learning-based dehazing methods including GDN [44], BPP [82], FFA [65], HardGAN [5], and MSBDN [7] which take a

[3] Although it does not directly provide $\mathbf{z}$, it offers the transmittance ($e^{-\boldsymbol{\beta} \cdot \mathbf{z}}$) with known spatially-uniform scattering coefficient $\boldsymbol{\beta}$, so that we can can compute $\mathbf{z}$ by ourselves.

[4] Our training (test) images are generated from the training (test) images of the Foggy Cityscapes-DBF dataset [72].

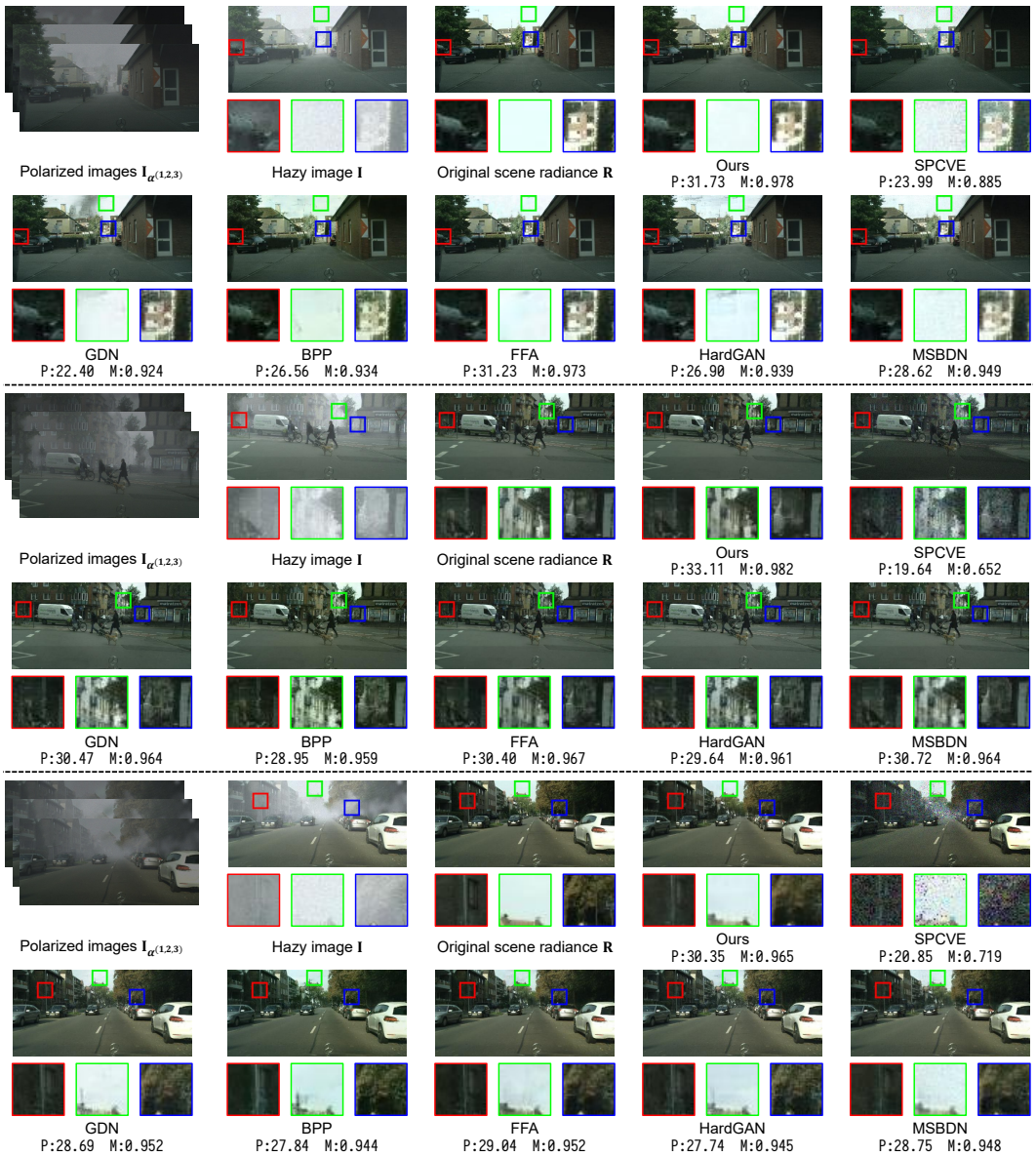

Figure 3: Qualitative comparisons on synthetic data among our method, a representative polarization-based dehazing algorithm SPCVE [54] which also takes three polarized images as the input, and five state-of-the-art learning-based dehazing methods including GDN [44], BPP [82], FFA [65], HardGAN [5], and MSBDN [7] which take a single hazy image as the input. Quantitative results evaluated using PSNR (P) and MS-SSIM (M) are displayed below each image.

single hazy image as the input[5]. SPCVE [54] assumes that the transmitted light is not significantly polarized ($\mathbf{P}_T = \mathbf{0}$), and uses optimization to estimate $\mathbf{P}_A$ and $\mathbf{A}_\infty$, while our method takes into account the polarization effects of transmitted light and adopts deep learning to estimate $\mathbf{P}_T$, $\mathbf{P}_A$, and $\mathbf{A}_\infty$. All of these learning-based methods are re-trained on our dataset using $\mathbf{R}$ and $\hat{\mathbf{I}}$ (the calculated hazy image from $\mathbf{I}_{\alpha^{(i)}}\,(i = 1, 2, 3)$ using the linear system from Equation (8))[6]. Note that comparing

[5]Note that the code of SPCVE [54] is not available and the demonstrated results are based on our own implementation. We directly provide the ground truth $\mathbf{A}_\infty$ to our implementation as its upper bound performance, also owing to that SPCVE [54] requires similar objects or known depth to estimate $\mathbf{A}_\infty$, which are not always available in our scenes.

[6]We should not use the ground truth $\mathbf{I}$ as the input of these learning-based methods since we could only get $\hat{\mathbf{I}}$ during the inferring phase of real data; and if we use $\mathbf{I}$ to re-train them, their results will be degenerated due to the large domain gap between $\mathbf{I}$ and $\hat{\mathbf{I}}$ (caused by the noise in $\mathbf{I}_{\alpha^{(i)}}\,(i = 1, 2, 3)$).

Table 1: Quantitative evaluation results on synthetic data among our method, a representative polarization-based dehazing algorithm SPCVE [54] (taking three polarized images as the input), and five state-of-the-art learning-based dehazing methods including GDN [44], BPP [82], FFA [65], HardGAN [5], and MSBDN [7] (taking a single hazy image as the input).

	Ours	SPCVE [54]	GDN [44]	BPP [82]	FFA [65]	HardGAN [5]	MSBDN [7]
PSNR	**28.32**	15.94	26.54	24.93	26.84	26.22	26.94
MS-SSIM	**0.951**	0.521	0.928	0.915	0.934	0.928	0.932

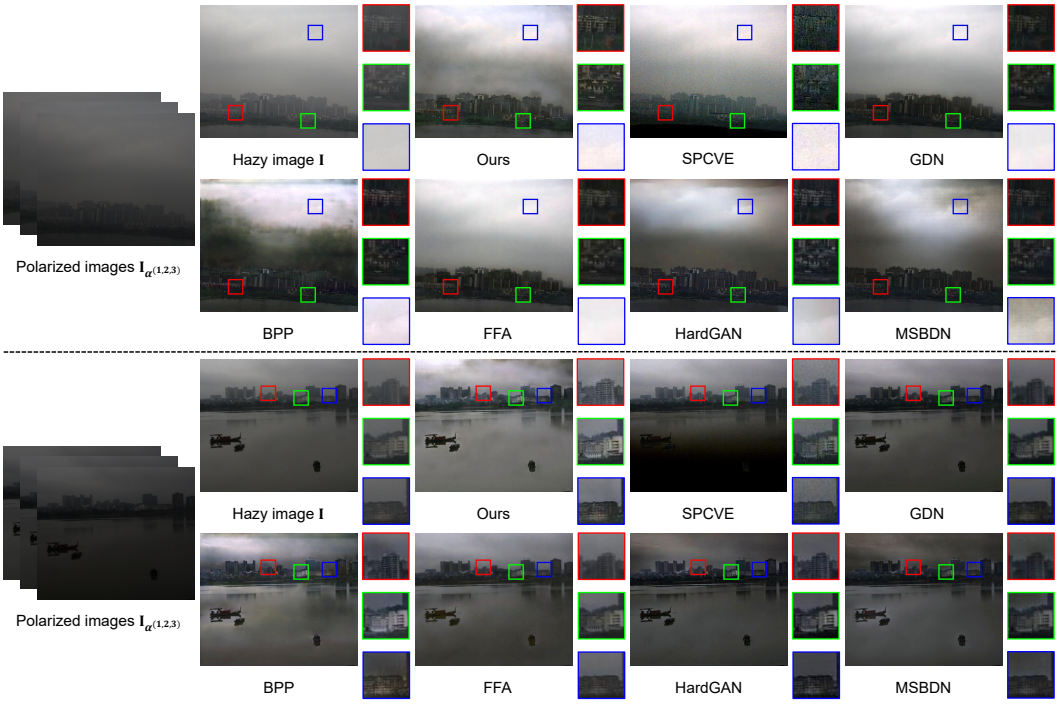

Figure 4: Qualitative comparisons on real data. See the caption of Figure 3 for explanation. All dehazing results are white-balanced to the similar color appearance and multiplied by a factor of 1.25 for better visualization. Please zoom-in for better details.

with learning-based dehazing methods might be a bit unfair because of the difference in types of input data (ordinary image *vs*. polarized image), and we conduct such a comparison to show the advantage of using polarized images over image-only approaches. Visual quality comparisons of dehazed results are shown in Figure 3. Compared to the polarization-based dehazing algorithm SPCVE [54], our method can dehaze robustly with fewer artifacts; compared to the learning-based methods, our method performs better in recovering details. Taking the sky region (green box) in the first group of Figure 3 as an example, SPCVE [54] suffers severely from noise, while the learning-based methods yields bad pixels (shown as black streaks in the sky). This is because the scattering is spatially-variant, and the polarization-based methods such as SPCVE [54] does not consider semantic and contextual information to refine the results, while the learning-based methods [44, 82, 65, 5, 7] are prone to artifacts for the pixels with large spatially-variant turbulence. To evaluate the results quantitatively, we adopt two frequently-used image quality metrics including PSNR and MS-SSIM (multi-scale SSIM). Results are shown in Table 1 (also below corresponding examples in Figure 3). Our model consistently outperforms the polarization-based and learning-based dehazing methods on all metrics.

5.2 Evaluation on real data

We use the Lucid Vision Phoenix polarization camera (RGB) to capture real data. The polarization camera can take four images with different polarization angles ($0°$, $45°$, $90°$, and $135°$) at a single shot. We use three of them ($0°$, $45°$, and $90°$) as the input to our method and SPCVE [54], and calculate the hazy image $\hat{\mathbf{I}}$ from the polarized images $\mathbf{I}_{\alpha(i)}\,(i = 1, 2, 3)$ using the linear system from

Table 2: Quantitative evaluation results of ablation study.

	PSNR	MS-SSIM
Ignoring the polarization effects of the transmitted light	27.63	0.943
Neglecting the spatially-variant real-world scattering	27.86	0.948
Directly estimating **T** and **R**	27.27	0.945
Removing g_2 (g_4)	21.55 (21.28)	0.740 (0.903)
Removing both g_2 and g_4	17.04	0.662
Our final model	**28.32**	**0.951**

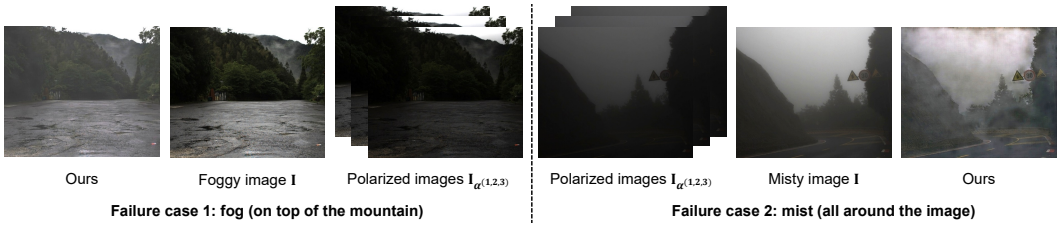

Figure 5: Two failure cases (fog and mist) in which our method shows degenerate performance.

Equation (8) as the input to learning-based methods (GDN [44], BPP [82], FFA [65], HardGAN [5], and MSBDN [7]). Visual quality comparisons of dehazed results are shown in Figure 4. As shown in Figure 4, our method is able to generate clearer and brighter images than those by the state-of-the-art polarization-based and learning-based methods. For example, the color of the buildings (red box) in the first group of Figure 4, is correctly recovered by our method, while other methods suffer from color distortion artifacts which dim the results.

5.3 Ablation study

To verify the validity of each model design choice, we conduct a series of ablation studies and show comparisons in Table 2. We first show the effectiveness of our physical image formation model by comparing with a model that ignores the polarization effects of the transmitted light (by taking $\mathbf{P}_T$ as zero) and a model that neglects the spatially-variant real-world scattering (by taking $\mathbf{P}_T$, $\mathbf{P}_A$, and $\mathbf{A}_\infty$ as spatially-uniform parameters). From the results we can see that our model is more generalized and reasonable. We further verify the contribution of our dehazing pipeline which estimates three parameters ($\mathbf{P}_T$, $\mathbf{P}_A$, and $\mathbf{A}_\infty$) to solve **T** and **R** by comparing with a model that directly estimates **T** and **R**. We find that our dehazing pipeline is better than directly estimating **T** and **R** since these parameters are easier to learn than **T** and **R**. Then, we demonstrate the necessity of the refinement subnetworks (g_2 and g_4) by removing g_2, g_4, and both of them. We could tell that without the refinement subnetworks, the performance degenerates rapidly, while it still outperforms the existing polarization-based (also optimization-based) dehazing algorithm SPCVE [54] (see Table 1 for the performance of SPCVE [54]) thanks to our generalized physical image formation model.

6 Conclusion

We presented a learning-based solution which leverages the properties of polarized light for image dehzaing. To handle the images captured in the wild, we proposed a generalized physical formation model of hazy images, introduced a robust polarization-based dehazing pipeline, and designed a neural network tailored to the pipeline, showing state-of-the-art performance. Our solution extended the applicability of polarization-based dehazing methods by adopting deep learning to estimate the infinite airlight and the DoP of both transmitted light and airlight without the requirement of specific clues (*e.g.*, sky regions, similar objects), while considering the spatially-variant real-world scattering.

Limitations. Since our method is based on the physical image formation model of hazy images, it may fail in situations which does not conform to the model, such as fog or mist. As shown in Figure 5, our method shows degenerate performance on those images, because fog and mist are caused by a suspension of water droplets, while haze is a suspension of extremely small particles (other than water droplets) in the air. As future work, we plan to extend our model to support other situations.

References

[1] Dana Berman, Tali Treibitz, and Shai Avidan. Non-local image dehazing. In *Proc. of Computer Vision and Pattern Recognition*, pages 1674–1682, 2016.

[2] Bolun Cai, Xiangmin Xu, Kui Jia, Chunmei Qing, and Dacheng Tao. DehazeNet: An end-to-end system for single image haze removal. *IEEE Transactions on Image Processing*, 25(11):5187–5198, 2016.

[3] Dongdong Chen, Mingming He, Qingnan Fan, Jing Liao, Liheng Zhang, Dongdong Hou, Lu Yuan, and Gang Hua. Gated context aggregation network for image dehazing and deraining. In *Proc. of Winter Conference on Applications of Computer Vision*, pages 1375–1383, 2019.

[4] Shuxin Chen, Yizi Chen, Yanyun Qu, Jingying Huang, and Ming Hong. Multi-scale adaptive dehazing network. In *Proc. of Computer Vision and Pattern Recognition Workshops*, 2019.

[5] Qili Deng, Ziling Huang, Chung-Chi Tsai, and Chia-Wen Lin. HardGAN: A haze-aware representation distillation GAN for single image dehazing. In *Proc. of European Conference on Computer Vision*, pages 722–738, 2020.

[6] Zijun Deng, Lei Zhu, Xiaowei Hu, Chi-Wing Fu, Xuemiao Xu, Qing Zhang, Jing Qin, and Pheng-Ann Heng. Deep multi-model fusion for single-image dehazing. In *Proc. of International Conference on Computer Vision*, pages 2453–2462, 2019.

[7] Hang Dong, Jinshan Pan, Lei Xiang, Zhe Hu, Xinyi Zhang, Fei Wang, and Ming-Hsuan Yang. Multi-scale boosted dehazing network with dense feature fusion. In *Proc. of Computer Vision and Pattern Recognition*, pages 2157–2167, 2020.

[8] Jiangxin Dong and Jinshan Pan. Physics-based feature dehazing networks. In *Proc. of European Conference on Computer Vision*, pages 188–204, 2020.

[9] Akshay Dudhane, Kuldeep M Biradar, Prashant W Patil, Praful Hambarde, and Subrahmanyam Murala. Varicolored image de-hazing. In *Proc. of Computer Vision and Pattern Recognition*, pages 4564–4573, 2020.

[10] Akshay Dudhane, Harshjeet Singh Aulakh, and Subrahmanyam Murala. RI-GAN: An end-to-end network for single image haze removal. In *Proc. of Computer Vision and Pattern Recognition Workshops*, 2019.

[11] Frederike Dümbgen, Majed El Helou, Natalija Gucevska, and Sabine Süsstrunk. Near-infrared fusion for photorealistic image dehazing. *Electronic Imaging*, 2018(16):321–1, 2018.

[12] Deniz Engin, Anil Genç, and Hazim Kemal Ekenel. Cycle-Dehaze: Enhanced CycleGAN for single image dehazing. In *Proc. of Computer Vision and Pattern Recognition Workshops*, pages 825–833, 2018.

[13] Shuai Fang, XiuShan Xia, Xing Huo, and ChangWen Chen. Image dehazing using polarization effects of objects and airlight. *Optics Express*, 22(16):19523–19537, 2014.

[14] Raanan Fattal. Single image dehazing. *ACM Transactions on Graphics (Proc. of ACM SIGGRAPH)*, 27(3):1–9, 2008.

[15] Raanan Fattal. Dehazing using color-lines. *ACM Transactions on Graphics (Proc. of ACM SIGGRAPH)*, 34(1):1–14, 2014.

[16] Chen Feng, Shaojie Zhuo, Xiaopeng Zhang, Liang Shen, and Sabine Süsstrunk. Near-infrared guided color image dehazing. In *Proc. of International Conference on Image Processing*, pages 2363–2367, 2013.

[17] Adrian Galdran, Javier Vazquez-Corral, David Pardo, and Marcelo Bertalmio. Enhanced variational image dehazing. *SIAM Journal on Imaging Sciences*, 8(3):1519–1546, 2015.

[18] Yosef Gandelsman, Assaf Shocher, and Michal Irani. "Double-DIP": Unsupervised image decomposition via coupled deep-image-priors. In *Proc. of Computer Vision and Pattern Recognition*, pages 11026–11035, 2019.

[19] Kristofor B Gibson and Truong Q Nguyen. An analysis of single image defogging methods using a color ellipsoid framework. *EURASIP Journal on Image and Video Processing*, 2013(1):1–14, 2013.

[20] Alona Golts, Daniel Freedman, and Michael Elad. Unsupervised single image dehazing using dark channel prior loss. *IEEE Transactions on Image Processing*, 29:2692–2701, 2019.

[21] Kaiming He, Jian Sun, and Xiaoou Tang. Single image haze removal using dark channel prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(12):2341–2353, 2010.

[22] Eugene Hecht et al. *Optics*, volume 5. Addison Wesley San Francisco, 2002.

[23] Geoffrey E Hinton and Ruslan R Salakhutdinov. Reducing the dimensionality of data with neural networks. *Science*, 313(5786):504–507, 2006.

[24] Ming Hong, Yuan Xie, Cuihua Li, and Yanyun Qu. Distilling image dehazing with heterogeneous task imitation. In *Proc. of Computer Vision and Pattern Recognition*, pages 3462–3471, 2020.

[25] Ran Kaftory, Yoav Y Schechner, and Yehoshua Y Zeevi. Variational distance-dependent image restoration. In *Proc. of Computer Vision and Pattern Recognition*, pages 1–8, 2007.

[26] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[27] GP Können. *Polarized light in nature*. CUP Archive, 1985.

[28] Johannes Kopf, Boris Neubert, Billy Chen, Michael Cohen, Daniel Cohen-Or, Oliver Deussen, Matt Uyttendaele, and Dani Lischinski. Deep Photo: Model-based photograph enhancement and viewing. *ACM Transactions on Graphics (Proc. of ACM SIGGRAPH)*, 27(5):1–10, 2008.

[29] Meredith K Kupinski, Christine L Bradley, David J Diner, Feng Xu, and Russell A Chipman. Angle of linear polarization images of outdoor scenes. *Optical Engineering*, 58(8):082419, 2019.

[30] Boyi Li, Xiulian Peng, Zhangyang Wang, Jizheng Xu, and Dan Feng. AOD-Net: All-in-one dehazing network. In *Proc. of International Conference on Computer Vision*, pages 4770–4778, 2017.

[31] Boyi Li, Xiulian Peng, Zhangyang Wang, Jizheng Xu, and Dan Feng. End-to-end united video dehazing and detection. In *Proc. of the AAAI Conference on Artificial Intelligence*, 2018.

[32] Boyi Li, Wenqi Ren, Dengpan Fu, Dacheng Tao, Dan Feng, Wenjun Zeng, and Zhangyang Wang. Benchmarking single-image dehazing and beyond. *IEEE Transactions on Image Processing*, 28(1):492–505, 2018.

[33] Chongyi Li, Chunle Guo, Jichang Guo, Ping Han, Huazhu Fu, and Runmin Cong. PDR-Net: Perception-inspired single image dehazing network with refinement. *IEEE Transactions on Multimedia*, 22(3):704–716, 2019.

[34] Lerenhan Li, Yunlong Dong, Wenqi Ren, Jinshan Pan, Changxin Gao, Nong Sang, and Ming-Hsuan Yang. Semi-supervised image dehazing. *IEEE Transactions on Image Processing*, 29:2766–2779, 2019.

[35] Runde Li, Jinshan Pan, Zechao Li, and Jinhui Tang. Single image dehazing via conditional generative adversarial network. In *Proc. of Computer Vision and Pattern Recognition*, pages 8202–8211, 2018.

[36] Yunan Li, Qiguang Miao, Wanli Ouyang, Zhenxin Ma, Huijuan Fang, Chao Dong, and Yining Quan. LAP-Net: Level-aware progressive network for image dehazing. In *Proc. of International Conference on Computer Vision*, pages 3276–3285, 2019.

[37] Yu Li, Robby T Tan, and Michael S Brown. Nighttime haze removal with glow and multiple light colors. In *Proc. of International Conference on Computer Vision*, pages 226–234, 2015.

[38] Zhuwen Li, Ping Tan, Robby T Tan, Danping Zou, Steven Zhiying Zhou, and Loong-Fah Cheong. Simultaneous video defogging and stereo reconstruction. In *Proc. of Computer Vision and Pattern Recognition*, pages 4988–4997, 2015.

[39] Jian Liang, Liyong Ren, Haijuan Ju, Wenfei Zhang, and Enshi Qu. Polarimetric dehazing method for dense haze removal based on distribution analysis of angle of polarization. *Optics Express*, 23(20):26146–26157, 2015.

[40] Jian Liang, Liyong Ren, Enshi Qu, Bingliang Hu, and Yingli Wang. Method for enhancing visibility of hazy images based on polarimetric imaging. *Photonics Research*, 2(1):38–44, 2014.

[41] Jian Liang, Li-Yong Ren, Hai-Juan Ju, En-Shi Qu, and Ying-Li Wang. Visibility enhancement of hazy images based on a universal polarimetric imaging method. *Journal of Applied Physics*, 116(17):173107, 2014.

[42] Jian Liang, Wenfei Zhang, Liyong Ren, Haijuan Ju, and Enshi Qu. Polarimetric dehazing method for visibility improvement based on visible and infrared image fusion. *Applied Optics*, 55(29):8221–8226, 2016.

[43] Qi Liu, Xinbo Gao, Lihuo He, and Wen Lu. Single image dehazing with depth-aware non-local total variation regularization. *IEEE Transactions on Image Processing*, 27(10):5178–5191, 2018.

[44] Xiaohong Liu, Yongrui Ma, Zhihao Shi, and Jun Chen. GridDehazeNet: Attention-based multi-scale network for image dehazing. In *Proc. of International Conference on Computer Vision*, pages 7314–7323, 2019.

[45] Xing Liu, Masanori Suganuma, Zhun Sun, and Takayuki Okatani. Dual residual networks leveraging the potential of paired operations for image restoration. In *Proc. of Computer Vision and Pattern Recognition*, pages 7007–7016, 2019.

[46] Yang Liu, Jinshan Pan, Jimmy Ren, and Zhixun Su. Learning deep priors for image dehazing. In *Proc. of International Conference on Computer Vision*, pages 2492–2500, 2019.

[47] Raúl Luzón-González, Juan L Nieves, and Javier Romero. Recovering of weather degraded images based on RGB response ratio constancy. *Applied Optics*, 54(4):B222–B231, 2015.

[48] Kangfu Mei, Aiwen Jiang, Juncheng Li, and Mingwen Wang. Progressive feature fusion network for realistic image dehazing. In *Proc. of Asian Conference on Computer Vision*, pages 203–215, 2018.

[49] Gaofeng Meng, Ying Wang, Jiangyong Duan, Shiming Xiang, and Chunhong Pan. Efficient image dehazing with boundary constraint and contextual regularization. In *Proc. of International Conference on Computer Vision*, pages 617–624, 2013.

[50] Daisuke Miyazaki, Daisuke Akiyama, Masashi Baba, Ryo Furukawa, Shinsaku Hiura, and Naoki Asada. Polarization-based dehazing using two reference objects. In *Proc. of International Conference on Computer Vision Workshops*, pages 852–859, 2013.

[51] Peter Morales, Tzofi Klinghoffer, and Seung Jae Lee. Feature forwarding for efficient single image dehazing. In *Proc. of Computer Vision and Pattern Recognition Workshops*, 2019.

[52] Jeong-Yun Na and Kuk-Jin Yoon. Stereo vision aided image dehazing using deep neural network. In *Proc. of the 1st Workshop and Challenge on Comprehensive Video Understanding in the Wild*, pages 15–19, 2018.

[53] Einav Namer and Yoav Y Schechner. Advanced visibility improvement based on polarization filtered images. In *Polarization Science and Remote Sensing II*, volume 5888, page 588805, 2005.

[54] Einav Namer, Sarit Shwartz, and Yoav Y Schechner. Skyless polarimetric calibration and visibility enhancement. *Optics Express*, 17(2):472–493, 2009.

[55] Srinivasa G Narasimhan and Shree K Nayar. Chromatic framework for vision in bad weather. In *Proc. of Computer Vision and Pattern Recognition*, volume 1, pages 598–605, 2000.

[56] Srinivasa G Narasimhan and Shree K Nayar. Vision and the atmosphere. *International Journal of Computer Vision*, 48(3):233–254, 2002.

[57] Srinivasa G. Narasimhan and Shree K. Nayar. Contrast restoration of weather degraded images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(6):713–724, 2003.

[58] Shree K Nayar and Srinivasa G Narasimhan. Vision in bad weather. In *Proc. of International Conference on Computer Vision*, volume 2, pages 820–827, 1999.

[59] Ko Nishino, Louis Kratz, and Stephen Lombardi. Bayesian defogging. *International Journal of Computer Vision*, 98(3):263–278, 2012.

[60] John P Oakley and Brenda L Satherley. Improving image quality in poor visibility conditions using a physical model for contrast degradation. *IEEE Transactions on Image Processing*, 7(2):167–179, 1998.

[61] Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y Hammerla, Bernhard Kainz, Ben Glocker, and Daniel Rueckert. Attention U-Net: Learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999*, 2018.

[62] Jinshan Pan, Sifei Liu, Deqing Sun, Jiawei Zhang, Yang Liu, Jimmy Ren, Zechao Li, Jinhui Tang, Huchuan Lu, Yu-Wing Tai, et al. Learning dual convolutional neural networks for low-level vision. In *Proc. of Computer Vision and Pattern Recognition*, pages 3070–3079, 2018.

[63] Yanwei Pang, Jing Nie, Jin Xie, Jungong Han, and Xuelong Li. BidNet: Binocular image dehazing without explicit disparity estimation. In *Proc. of Computer Vision and Pattern Recognition*, pages 5931–5940, 2020.

[64] Dubok Park, David K Han, Changwon Jeon, and Hanseok Ko. Fast single image de-hazing using characteristics of RGB channel of foggy image. *IEICE Transactions on Information and Systems*, 96(8):1793–1799, 2013.

[65] Xu Qin, Zhilin Wang, Yuanchao Bai, Xiaodong Xie, and Huizhu Jia. FFA-Net: Feature fusion attention network for single image dehazing. In *Proc. of the AAAI Conference on Artificial Intelligence*, pages 11908–11915, 2020.

[66] Yanyun Qu, Yizi Chen, Jingying Huang, and Yuan Xie. Enhanced pix2pix dehazing network. In *Proc. of Computer Vision and Pattern Recognition*, pages 8160–8168, 2019.

[67] Yufu Qu and Zhaofan Zou. Non-sky polarization-based dehazing algorithm for non-specular objects using polarization difference and global scene feature. *Optics Express*, 25(21):25004–25022, 2017.

[68] Wenqi Ren, Si Liu, Hua Zhang, Jinshan Pan, Xiaochun Cao, and Ming-Hsuan Yang. Single image dehazing via multi-scale convolutional neural networks. In *Proc. of European Conference on Computer Vision*, pages 154–169, 2016.

[69] Wenqi Ren, Lin Ma, Jiawei Zhang, Jinshan Pan, Xiaochun Cao, Wei Liu, and Ming-Hsuan Yang. Gated fusion network for single image dehazing. In *Proc. of Computer Vision and Pattern Recognition*, pages 3253–3261, 2018.

[70] Wenqi Ren, Jinshan Pan, Hua Zhang, Xiaochun Cao, and Ming-Hsuan Yang. Single image dehazing via multi-scale convolutional neural networks with holistic edges. *International Journal of Computer Vision*, 128(1):240–259, 2020.

[71] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional networks for biomedical image segmentation. In *Proc. of International Conference on Medical Image Computing and Computer Assisted Intervention*, pages 234–241, 2015.

[72] Christos Sakaridis, Dengxin Dai, Simon Hecker, and Luc Van Gool. Model adaptation with synthetic and real data for semantic dense foggy scene understanding. In *Proc. of European Conference on Computer Vision*, pages 687–704, 2018.

[73] Christos Sakaridis, Dengxin Dai, and Luc Van Gool. Semantic foggy scene understanding with synthetic data. *International Journal of Computer Vision*, 126(9):973–992, 2018.

[74] Sanchayan Santra, Ranjan Mondal, and Bhabatosh Chanda. Learning a patch quality comparator for single image dehazing. *IEEE Transactions on Image Processing*, 27(9):4598–4607, 2018.

[75] Lex Schaul, Clément Fredembach, and Sabine Süsstrunk. Color image dehazing using the near-infrared. In *Proc. of International Conference on Image Processing*, pages 1629–1632, 2009.

[76] Yoav Y Schechner and Yuval Averbuch. Regularized image recovery in scattering media. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(9):1655–1660, 2007.

[77] Yoav Y Schechner, Srinivasa G Narasimhan, and Shree K Nayar. Instant dehazing of images using polarization. In *Proc. of Computer Vision and Pattern Recognition*, volume 1, pages I–I, 2001.

[78] Yoav Y Schechner, Srinivasa G Narasimhan, and Shree K Nayar. Polarization-based vision through haze. *Applied Optics*, 42(3):511–525, 2003.

[79] Yuanjie Shao, Lerenhan Li, Wenqi Ren, Changxin Gao, and Nong Sang. Domain adaptation for image dehazing. In *Proc. of Computer Vision and Pattern Recognition*, pages 2808–2817, 2020.

[80] Linghao Shen, Yongqiang Zhao, Qunnie Peng, Jonathan Cheung-Wai Chan, and Seong G Kong. An iterative image dehazing method with polarization. *IEEE Transactions on Multimedia*, 21(5):1093–1107, 2018.

[81] Sarit Shwartz, Einav Namer, and Yoav Y Schechner. Blind haze separation. In *Proc. of Computer Vision and Pattern Recognition*, volume 2, pages 1984–1991, 2006.

[82] Ayush Singh, Ajay Bhave, and Dilip K Prasad. Single image dehazing for a variety of haze scenarios using back projected pyramid network. In *Proc. of European Conference on Computer Vision Workshops*, pages 166–181, 2020.

[83] Chang-Hwan Son and Xiao-Ping Zhang. Near-infrared fusion via color regularization for haze and color distortion removals. *IEEE Transactions on Circuits and Systems for Video Technology*, 28(11):3111–3126, 2017.

[84] Taeyong Song, Youngjung Kim, Changjae Oh, and Kwanghoon Sohn. Deep network for simultaneous stereo matching and dehazing. In *Proc. of British Machine Vision*, page 5, 2018.

[85] Patricia L Suárez, Angel D Sappa, Boris X Vintimilla, and Riad I Hammoud. Deep learning based single image dehazing. In *Proc. of Computer Vision and Pattern Recognition Workshops*, pages 1169–1176, 2018.

[86] Matan Sulami, Itamar Glatzer, Raanan Fattal, and Mike Werman. Automatic recovery of the atmospheric light in hazy images. In *Proc. of International Conference on Computational Photography*, pages 1–11, 2014.

[87] Robby T Tan. Visibility in bad weather from a single image. In *Proc. of Computer Vision and Pattern Recognition*, pages 1–8, 2008.

[88] Ketan Tang, Jianchao Yang, and Jue Wang. Investigating haze-relevant features in a learning framework for image dehazing. In *Proc. of Computer Vision and Pattern Recognition*, pages 2995–3000, 2014.

[89] Jean-Philippe Tarel and Nicolas Hautiere. Fast visibility restoration from a single color or gray level image. In *Proc. of International Conference on Computer Vision*, pages 2201–2208, 2009.

[90] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Instance Normalization: The missing ingredient for fast stylization. *arXiv preprint arXiv:1607.08022*, 2016.

[91] Dong Yang and Jian Sun. Proximal Dehaze-Net: A prior learning-based deep network for single image dehazing. In *Proc. of European Conference on Computer Vision*, pages 702–717, 2018.

[92] Xitong Yang, Zheng Xu, and Jiebo Luo. Towards perceptual image dehazing by physics-based disentanglement and adversarial training. In *Proc. of the AAAI Conference on Artificial Intelligence*, 2018.

[93] He Zhang and Vishal M Patel. Densely connected pyramid dehazing network. In *Proc. of Computer Vision and Pattern Recognition*, pages 3194–3203, 2018.

[94] He Zhang, Vishwanath Sindagi, and Vishal M Patel. Multi-scale single image dehazing using perceptual pyramid deep network. In *Proc. of Computer Vision and Pattern Recognition Workshops*, pages 902–911, 2018.

[95] Jing Zhang and Dacheng Tao. FAMED-Net: A fast and accurate multi-scale end-to-end dehazing network. *IEEE Transactions on Image Processing*, 29:72–84, 2019.

[96] Yanfu Zhang, Li Ding, and Gaurav Sharma. HAZERD: An outdoor scene dataset and benchmark for single image dehazing. In *Proc. of International Conference on Image Processing*, pages 3205–3209, 2017.

[97] Shiyu Zhao, Lin Zhang, Shuaiyi Huang, Ying Shen, and Shengjie Zhao. Dehazing Evaluation: Real-world benchmark datasets, criteria, and baselines. *IEEE Transactions on Image Processing*, 29:6947–6962, 2020.

Checklist

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes]
   (b) Did you describe the limitations of your work? [Yes] See Section 6.
   (c) Did you discuss any potential negative societal impacts of your work? [No] We do not think this work would raise any negative societal concerns.
   (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]
2. If you are including theoretical results...
   (a) Did you state the full set of assumptions of all theoretical results? [Yes] See Section 3.1.
   (b) Did you include complete proofs of all theoretical results? [Yes] See Section 3.2.
3. If you ran experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [No] No (during peer review). We will release the code once the paper is accepted.
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [Yes] See Section 4.1 and Section 4.2.

(c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [No] We think error bars are not needed in this task.

(d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [Yes] See Section 4.2.

4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...

(a) If your work uses existing assets, did you cite the creators? [Yes] We use the Foggy Cityscapes-DBF dataset [72] as the data source for generating our own dataset (see Section 4.1 for details).

(b) Did you mention the license of the assets? [Yes] See Section 4.1.

(c) Did you include any new assets either in the supplemental material or as a URL? [No] No (during peer review). We will release the implementation code of our dataset generation pipeline once the paper is accepted.

(d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [No] The Foggy Cityscapes-DBF dataset [72] we used is open-source.

(e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [No] Our data do not contain any personally identifiable information or offensive content.

5. If you used crowdsourcing or conducted research with human subjects...

(a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A]

(b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A]

(c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A]