Learning to Dehaze From Realistic Scene with A Fast Physics Based Dehazing Network

Ruoteng Li · Xiaoyi Zhang · Shaodi You · Yu Li

Abstract Dehaze is one of the popular computer vision research topics for long. A realtime method with reliable performance is highly desired for a lot of applications such as autonomous driving. In recent years, while learning based methods require datasets containing pairs of hazy images and clean ground truth references, it is generally impossible to capture this kind of data in real. Many existing researches compromise this difficulty to generate hazy images by rendering the haze from depth on common RGBD datasets using the haze imaging model. However, there is still a gap between the synthetic datasets and real hazy images as large datasets with high quality depth are mostly indoor and depth maps for outdoor are imprecise. In this paper, we complement the exiting datasets with a new, large, and diverse dehazing dataset containing real outdoor scenes from HD 3D videos. We select large number of high quality frames of real outdoor scenes and render haze on them using depth from stereo. Our dataset is more realistic than existing ones and we demonstrate that using this dataset greatly improves the dehazing performance on real scenes. In addition to the dataset, inspired by the physics model, we also propose a light and reliable dehaze network. Our approach outperforms other methods by a large margin and becomes the new state-of-the-art method. Moreover, the light design of the network enables our methods to run at realtime speed that is much faster than other methods. We have made our code publicly available.

Keywords Single Image Dehaze · Realistic Dataset · Deep Learning, Recourse Constraint Applications

1 Introduction

Fog and haze are common phenomena caused by floating atmospheric particles usually degrade the contrast and visibility of images captured outside. In recent years, computer vision in outdoor environment and also in

Fig. 1: The proposed method achieves the best performance on our new realistic dehazing dataset, while it remains the fastest processing speed among all the baseline methods. The speed is measured on input resolution 460 × 620 (See Sect. 5 for more details).
wild environment is considered as a core challenge. Therefore, to overcome visibility degradation from haze and fog will significantly boost the performance of a various important computer vision tasks.

Unlike traditional model based dehazing algorithm, in recent years, data driven methods are more popular. In this paper, we argue two important problems of data driven dehazing:

- A large, high quality and realistic dataset. Especially, a dataset with a great variety of real in the wild images and with physics realistic ground truth. Also, a dataset with high quality and high resolution images is desire.

- A network that can work at realtime and with efficient power and memory usage. So that such network can be widely applied to a lot of resource constraint and dynamic outdoor scenarios, such as autonomous cars, UAV and underwater robots.

First, existing datasets suffer from the lack of realistic dynamic outdoor ground truth as an example shown in Fig.2. Unlike machine learning for high level computer vision tasks, machine learning for low level tasks requires pixel-wise accurate data. Obtaining such data for dehazing task is extremely challenging.

Second, while we find most of the existing methods demonstrates a satisfying PSNR as shown in Fig.1, the network size is rather huge and cannot be properly applied to an outdoor mobile vehicles, such as UAV, small size robot and low end autonomous cars. It can also be found that most of the existing networks work at 3 or lower FPS, which is not satisfying for an realtime application.

We aim to tackle both of the above mentioned problems through physics inspiration. Physically, the degradation in hazy and foggy images can be attributed to floating particles in the atmosphere that absorb and scatter light in the environment [16]. This scattering and absorption reduces the direct transmission from the scene to the camera and adds another layer of the scattered light, known as airlight [31]. The attenuated direct transmission causes the intensity from the scene to be weaker, while the airlight causes the appearance of the scene to be washed out. We provide a comprehensive review of the existing benchmark datasets as well as dehazing algorithms in Section 2 and introduce the physics-based haze modeling in Section 3.

Based on the physics modeling, we first propose a large, high quality, highly various dataset. The dataset selects 2000 high quality ground truth images from Blue-Ray videos with a 1920*1080 resolution. The images are captured by high end cameras where the color, exposure and sensor noise and optimized. We select the 2000 ground truth images from 20 different videos with 40 hours in total to insure the diversity. More importantly, all those videos are captured by multi-view stereo cameras, so that we can obtain high quality depth and enable us to render the high quality and physics realistic hazy images. Fig. 2 shows examples and comparisons with existing dataset. The details of the dataset creation are described in Section 4.

Second, inspired by the physics model, we also propose a light and reliable neural network. Particularly, the proposed network contains two stages. First stage incorporates the physics haze model to estimate the transmission map and atmospheric light. With the guidance of the transmission and atmospheric light, the second stage learns to recover the clean background image by taking in the hazy image and the previously estimated transmission and atmospheric light. The proposed network is simple, fast and accurate as shown in the experiment section. Fig. 1 shows the performance of our proposed network. Details of the network structure and implementation is described in Section 5.

Finally, this paper proposes a systematical and detailed benchmark using the proposed dataset and existing dataset, which we believe will enable more exciting and novel topics in dehazing. More importantly, we also demonstrate the state-of-the-art performance of the proposed network.

The main contribution of this paper can be summarized as:

- We create a large scale, high quality, outdoor dehazing dataset of real scenes. It is built from massive 3D images that contains large variety of outdoor scenes.
- We propose an efficient and high performance dehaze network base on physics inspiration. Which can be applied to various realtime dynamic and resource constraint outdoor scenes.
- We provide a systematical benchmark of the existing methods using the proposed dataset as well as existing dataset.
- We demonstrate state-of-the-art efficiency and accuracy of the proposed network through rigorous experiments.

2 Related Work

Haze removal from a single image or other inputs has attracted wide attention. Some works on dehazing require multiple images (e.g. [25,31]) or additional information acquired from other sources (e.g. [15,12]). A comprehensive survey for dehazing can be found in [24]. In this work, we only focus on single image dehazing.
Learning to Dehaze From Realistic Scene with A Fast Physics Based Dehazing Network

Fig. 2: A visualization of the existing datasets including Frida [50], Li et al. [24], Fattal [7], RESIDE [18], and our proposed dataset. ITS means indoor training set and OTS means outdoor training set.

Single image dehazing methods Early works on single image dehazing are mostly prior based. These works propose different priors in estimating the transmission map and the air light and then a clean image can be recovered by reverse the hazy image model in Eq. 1. Researches in this line include Specifically, [48] proposes to use local contract maximization and [7] use the idea that the shading and transmission functions are locally and statistically uncorrelated. [10,11] observes an interesting phenomenon of outdoor natural scenes with clear visibility and formulate as dark channel prior (DCP) which becomes one of the most successful priors to dehaze. [22] extends the idea of the dark channel prior in determining the initial values of transmission by the boundary constrain and refine the map by a contextual regularization. [49] gathers multiple and multiscale features from existing priors and uses a random forest regressor to learn to predict the transmission from the features. [49] introduces a color attenuation prior and create a linear model for scene depth, where the parameters of the model are learned in a supervised learning fashion. Unlike previous patch based priors, the work of [2] develops a non-local prior which relies on the assumption that colors of a haze-free image form tight clusters in RGB space. [11] propose to use optimization to solve the dehazing and in the meanwhile minimizing the visual artifacts. [20] links the traditional Retinex methods to the dehaze problem. Some researches address more specific problems, e.g. atmospheric light estimation [49] and nighttime dehazing [24,53].

With the advance of deep learning technologies and their success in many computer vision applications, convolution neural network (CNN) based methods for image dehazing gain increasing popularity in recent years. [3,9,17,19,21,29,30,36,37,38,53,54]. Unlike prior based methods that rely on hand-crafted features, the CNN based method try to learn the dehazing features directly from the training data. DehazeNet [3] trains a CNN to predict the transmission value at local patches. [37] proposes to use multi-scale CNN (MSCNN) to first generate a coarse transmission map and gradually refine it. The All-in-One Dehazing Network (AODNet) [17] re-writes the scattering model to enable a network to generate clean image directly without explicitly estimating the transmission and atmosphere light separately. [54] trains a Densely Connected Pyramid Dehazing Network (DCPDN) with an edge-preserving loss and introduces a discriminator in dehazing. Methods [38,54] incorporate traditional dehazing priors and color processing operations into the deep network. Generative adversarial network (GAN) based dehazing methods have been proposed in [19,30]. The recent state-of-the-art method [29] applies a novel attention-based multi-scale estimation on a grid network.

Dehazing datasets Quantitative evaluation for dehazing need the hazy image and the corresponding clean
background as reference. These pairs are generally difficult to capture in real as it is usually impossible to control the outdoor environments. Researchers in this field find other ways for generating the datasets. The most common approach is to render haze on clean images with its depth information using Eq. 1 [8,18,50]. An exception is the dataset of Li et al. [24] that applies physically based rendering technique (i.e., ray tracing) [35] to generate the haze effect on synthetic scenes. Fatal [8] provides a dehazing dataset rendered from real images and their depth maps. FRIDA [50] dataset focuses on driving scenario and uses synthetic road backgrounds from graphics models. These early datasets are all limited in sizes and can serve only as small benchmarking. To meet the need for training a neural network for dehazing, large scale datasets are then created. [40] generated a foggy Cityscapes dataset using images from the Cityscapes dataset [6]. [18] introduces a general dehazing dataset named RESIDE. Its indoor training set (ITS) is built on existing indoor RGBD datasets NYUv2 [43] and Middleburry stereo [41] and outdoor training set (OTS) collects real outdoor images with their depth estimated using [27]. Other kind of datasets like subjective evaluation datasets and task-driven datasets are not discussed here.

3 Physical Model Revisit

In this section, we first introduce the overall formulation of the imagery model. Following that, we review the widely used particle scattering modeling in Subsection 3.1. We also introduce the wavelength dependent module in order to demonstrate the way we render haze based on the 3D images.

Fig. 3 is the overall illustration of the hazy imaging environment. We name the reflected light from object as scene reflection $R$. Then scene reflection which travels though the scattering media will be attenuated depending on the wavelength. In addition to the direct transmission, the scattering media (micro water particles in this setting) will also scatter light from the environment. The scattering has two components, the scattering of the lighting from the object to the camera, which is usually called forward scattering. And the second, the scattering of the lighting from the environment to the camera, which is usually called back scattering.

3.1 Single Wavelength Image Formulation

In this subsection, we briefly review the derivation of the optical model for particle scattering, which is known as Koschmieder’s law [16]. The content is mainly based on Narasimhan and Nayar [33] and McCartney [31].

As illustrated in Fig. 3(a) and (b), for a specific wavelength $\lambda$, the light luminance $I$ captured at a certain pixel $x = (x, y)$ can be expressed as:

$$I(x, \lambda) = D(x, \lambda) + A(x, \lambda) = t(x, \lambda) R(x, \lambda) + L_{\infty}(1 - t(x, \lambda))$$

$$= e^{-\beta(\lambda)d(x)} R(x, \lambda) + L_{\infty}(1 - e^{-\beta(\lambda)d(x)})$$

where $D(x, \lambda)$ is the direct transmission, $A(x, \lambda)$ is the airlight, $t(x, \lambda)$ is the transmittance, $R(x, \lambda)$ is the ambient reflectance, $L_{\infty}$ is the infinite illumination, and $\beta(\lambda)$ is the scattering coefficient.
posed to be captured by the camera but however be scattered by the media and captured by the camera. The model is known as Koschmieder’s law \[10\]. The term \( p \) is the reflectance of an object (also called the albedo), \( \beta(\lambda) \) is the atmospheric attenuation coefficient, and \( d \) is the distance between an object and the camera.

Eq. 4 can be rewritten in a more commonly used denotations in dehazing literature. Where the first term, the scene reflection is rewritten as:

\[
R(x, \lambda) = L(\lambda)p(x, \lambda). \tag{2}
\]

The estimation of Eq. 2 terms is the goal of dehazing and/or visibility enhancement algorithms. Putting the scattering attenuation into the term, we have:

\[
D(x, \lambda) = R(x, \lambda)t(x, \lambda). \tag{3}
\]

The second term is commonly named as the airlight, denoted as \( A(x, \lambda) \):

\[
A(x, \lambda) = L(\lambda)(1-e^{-\beta d(x)}). \tag{4}
\]

The function \( t(x) \) represents the transmission, as

\[
t(x) = e^{-\beta d(x)}.
\]

Hence, the scattering model in Eq. (1) can be written as:

\[
I(x, \lambda) = D(x, \lambda) + A(x, \lambda)
= t(x, \lambda)R(x, \lambda) + L(\lambda)(1-t(x, \lambda))
= e^{-\beta d(x)}R(x, \lambda) + L(\lambda)(1-e^{-\beta d(x)}).
\tag{5}
\]

This equation formulates the dehazing task in a monochromatic setting. We now need to continue formulation it in a multi-wavelength setting, because in most of the scenarios, we capture color images.

3.2 Modeling of Wavelength Dependent Attenuation

We now continue to study the image formulation with the consideration of wavelength. There are a few aspects which are determined on the wavelength, the first is the light source, the second is the scattering media and the third is the camera sensor. We first model the wavelength attenuation in physics and later develop it into the imagery model.

**Physical Model of Wavelength Attenuation**

Fig. 4 illustrates the wavelength attenuation from different material. The absorption has been well measured decades ago \[12,13\]. It is interesting to find out that airborne attenuation is mostly not wavelength dependent in visible light range, therefore, it is safe to drop the wavelength dependency for hazy and foggy scattering in our study. However, it also requires to pay attention that the pure air attenuation is wavelength dependent.

While underwater imagery is not the focus of this paper, we leave the information here and wish to encourage further research.

**Imagery Model of Wavelength Attenuation: RGB camera**

The color imagery can be developed from Eq. 5 by considering the sensor wave length response. For example, the red channel is developed as:

\[
I_r(x) = \int \gamma_r(\lambda)I_r(x, \lambda)d\lambda \tag{6}
\]

\[
= \int \gamma_r(\lambda)e^{-\beta(\lambda)d(x)}R(x, \lambda) \tag{7}
\]

\[
+ \int \gamma_r(\lambda)L(\lambda)(1-e^{-\beta(\lambda)d(x)})d\lambda, \tag{8}
\]

where \( \gamma_r(\lambda) \) is the red channel sensor response of at wave length \( \lambda \). The imagery of the green and blue channel can be defined similarly.

For degraded setting in most single image dehazing tasks, we no longer have the spectral response value. And the equations are usually degraded to three \((R, G, B)\) channels. And an example is given on the red channel:

\[
I_r(x) = D_r(x) + A_r(x)
= t(x)R(x) + L(1-t(x))
= e^{-\beta d(x)}R(x) + L(1-e^{-\beta d(x)}), \tag{9}
\]

where the scene radiance and the airlight is depending on three \((R, G, B)\) channels and the rest on channel independent.

The vectorized form can be similarly written as:

\[
I(x) = D(x) + A(x)
= t(x)R(x) + L(1-t(x))
= e^{-\beta d(x)}R(x) + L(1-e^{-\beta d(x)}), \tag{10}
\]

![Fig. 4: Wavelength attenuation in air(a) and water(b).](Image 297x637 to 535x752)
where $I(x) = (I_r(x), I_g(x), I_b(x))^\top$ and similarly for the rest.

4 Large and Physically Real Outdoor Dataset

In this section, we introduce the details of our large scale, high quality depth and real outdoor dataset for single image dehaze.

In order to advance the line of research of single image dehazing field, many public benchmarking datasets [13] have been made for evaluating image dehazing algorithms. However, most of the dehazing benchmark datasets contain non-negligible gap between the synthetic datasets and real hazy images due to the use of inaccurate depth map in the haze rendering process. We propose a new large-scale single image dehazing dataset for dehazing algorithm evaluation.

4.1 Dataset Overview

The proposed dehazing dataset contains in total 10000 synthetic hazy images, out of which 8000 images comprise the training set and the rest 2000 images belong to the test set. The hazy images are rendered on 2000 high-quality clean images, extracted from a series of High-Definition (HD) 3D videos. In order to keep high diversity of the background scenes, we extract around 100 key frame-pairs from each of the 22 videos and render 5 haze images on each clean background frame with various haze density. For the training set, each sample contains four items including the rendered hazy image $I$, the transmission map $T$, the atmospheric light value $A$ and the corresponding clean background image $C$.

In the following sections, we will explain clean images selection, image depth generation and hazy image rendering process in details.

4.2 Outdoor Image Data

We have listed the comparison of all the existing general dehazing datasets in Table 1. Note Sakaridis [13] dataset focuses specifically on semantic scene understanding for driving scene, which is not listed. From the table, one can see that most of the early datasets contain limited number of samples, which are not suitable for deep networks training. Existing large-scale outdoor dehazing dataset, i.e. the ”RESIDE-$\beta$” dataset [18], is rendered based on monocular depth estimation from single image [28]. However, monocular depth estimation is still an open problem. The depth estimation method used to generate the transmission map does not perform reliably at many texture-less and ambiguous areas [29]. To overcome the limitation of single image depth estimation, we consider to utilize stereo algorithm to provide robust and accurate outdoor depth. 3D videos provide the largest known source of stereo pairs, presenting the possibility of capturing millions of outdoor scenes with stereo information available. Thus, we propose to extract depth data from 3D videos. In recent time, high-quality 3D videos are usually captured by top-notch stereo cameras and feature diverse dynamic environments that range from human-centric imagery (such as Hollywood films) to nature scenes with landscapes and animals in scientific documentaries. To that end, we have selected a diverse set of 22 recent 3D videos with stereo image-pairs available.

4.3 3D Videos Pre-processing

Outdoor Scene Selection In order to obtain rich and diverse outdoor scene data with stereo camera settings, we carefully select a series of 3D videos that provide large number of high-quality stereo pairs in controlled conditions. In a typical 2-hour modern video, the video may contain up to hundreds of different clips including both indoor and outdoor scenes. The first step of extracting useful background images from the videos is to split the long video into individual segments separated by scene transitions. We apply the popular scene detection algorithm to split each video into independent clips, each of which contains a video segment recorded under the same scene. For each segment, we only extract up to 2 frames (image pairs) as our background clean image data, located at the front and the back of the segment. All the extracted frames form the image set $S_0$. From set $S_0$, we then apply the state of the art semantic segmentation methods [56] to select outdoor images (with proper scene) to form image set $S_1$. In order to make sure the selected scenes are mostly from outdoor environment, we only select the images that contain the common outdoor semantic objects including grassland, road, building, landscape, etc.. Not all the images from $S_1$ are useful for haze rendering because of various reasons. Some of the scenes are recorded under foggy/hazy/smoky conditions. Others may contain strong motion blur or de-focus blur. Hence, we manually scan over the image set $S_1$ to exclude blurred images, low-light dark images, texture-less images, foggy/smoky/hazy backgrounds, science-fiction unnatural images, etc. to guarantee the high quality of the selected background as the ground truth of our proposed dataset. We name the selected new image set as set $S_2$. 

...
Table 1: Comparison of the proposed dataset with existing dehaze datasets. One may find our dataset contains the most diverse and accurate depths.

|                | Indoor | Outdoor | Background | Diversity | Haze effect       | Depth accuracy |
|----------------|--------|---------|------------|-----------|-------------------|----------------|
| Fattal [8]     | 4      | 8       | Real       | Low       | Using depth       | High           |
| FRIDA [50]     | -      | 480     | Synthetic  | Low       | Using depth       | High           |
| Li et al. [24] | -      | 5       | Synthetic  | Low       | Ray tracing       | -              |
| RESIDE-ITS     | 1399   | -       | Real       | Low       | Using depth       | High           |
| RESIDE-OTS     | -      | 2061    | Real       | High      | Using depth       | Low            |
| Ours           | -      | 2000    | Real       | High      | Using depth       | High           |

Fig. 5: Some Samples of the proposed dataset. The atmospheric light RGB values from top to bottom is (252,252,252), (251, 251, 251), (254, 254, 251) and (244,244,224), and the airlight are generated in consistent with the color tone. (Zoom in to see details.)

**Depth Map** Accurate depth map of a scene is the key to make a high-quality dehazing dataset useful. However, the depth map of the outdoor scene is extremely difficult to obtain. Existing dehazing datasets use either [28] depth estimation algorithm [18] to predict or semantic annotations [40] to obtain coarse depth maps. Due to the limitation of these annotation methods, the generated depth maps demonstrate compromised accuracy as shown in [6]. To that end, one of the key contributions in the proposed dataset is to leverage stereo cameras to produce more accurate depth map for haze rendering. Although 3D videos provide stereo image pairs for every frame, the video data comes with its own challenges as well. First, some 3D videos are shot with monocular cameras and the stereo effects are manually added by post-processing. In this case, we only select videos that were shot using physical stereo cameras. Second, we only select Blu-ray format high definition videos, which allow us to make high-resolution images. Third, focal lengths, stereo camera baselines and camera intrinsic parameters are usually unknown and vary from one video to another. It is difficult to directly compute the stereo disparity solely from the image pairs using current state of the art stereo estimation algorithms. In addition, most of the stereo algorithms are designed and trained to estimate disparity in positive ranges only. However, the 3D video data may contain negative disparity value because of the camera rotation in the stereo rig configuration. Instead of applying stereo algorithms, we apply a state of the art optical
flow estimation method [47] to handle the positive and negative disparity values. We only retain the horizontal component of the flow field as the disparity. **Depth Map Refinement** Since PWCNet [47] produces flow fields in a quarter of the original image size, a bi-linear up-sampling post-processing is usually used to up-sample the flow fields to the full size for optical flow accuracy evaluation. In our case of haze rendering, the bi-linear up-sampling usually makes the object’s boundary blur (shown in Fig. 6), which also leads to halo effect on the rendered object’s boundary. To solve this problem, we adopt FGI [22] algorithm to up-sample the flow fields based on the boundary, contours and edges of the input image. Fig. 6 compares the depth maps of these two up-sampling methods. One can observe that the depth discontinuities are more aligned with the actual objects’ boundaries in the input images. Hence, the halo effect of the FGI up-sampled result is significantly reduced.

### 4.4 Haze Rendering

We follow the widely used haze model [3,9,17,19]:

\[
I(x) = J(x)t(x) + A(1 - t(x))
\]

where \(I(x)\) is the fog image at location \(x\) and \(J\) is the fog-free scene radiance. While the transmission map \(t\) is defined as:

\[
t(x) = \exp(-\beta d(x))
\]

where \(\beta\) indicates the scattering coefficient of the atmosphere, \(d(x)\) is the distance between the object and the camera. We in total collect 2000 high resolution clean images from 3D videos and use these clean images as background for haze rendering. The haze rendering principally follows the haze model Eq. (11). The detailed rendering process is described in the Algorithm 1. For each clean background image, we synthesize 5 hazy images using different \(\beta\) values uniformly sampled from \([1.0, 3.0]\), resulting in 10000 hazy images. 8000 hazy images are split into training set and the rest 2000 images are for the test set.

The atmospheric light value \(A\) is determined using maximum color method described by Eq. (13):

\[
A_C = \max_{x \in I} C(x), C = R, G, B,
\]

which is a widely used approach in white balance. We need to determine the airlight because airlight need to be consistent with the environment lighting in the haze free image. Since we are using video images for the haze free image, the lighting and color tone are usually adjusted from their natural exposure. Therefore, we need to adjust the airlight accordingly.

### 5 Physics Inspired Efficient Dehaze Network

Along with our propose large scale realistic dataset, we also propose a two-stage light and fast CNN-based dehazing method to show that the proposed dataset has better generalised than existing dehazing dataset. We also argue that the physics modeling allows us to develop a network which is fast and light, which is essential for on-line visibility enhancement. Because online dehaze are widely used in UAV, autonomous cars and
Learning to Dehaze From Realistic Scene with A Fast Physics Based Dehazing Network

\[ J = \frac{I - (1 - T) \odot A}{T} \]  

The overall architecture of the proposed method is shown in Fig. 7. Before describing the proposed 2-stage network, we first discuss the overall input and output of the network, as well as the intermediate output by the first stage. Referring to Fig. 7, we call the first stage as the physics-based network. It takes in a single haze image as input and extracts the physical parameters of haze, including the transmission map \( T \) and the atmospheric light \( A \). The output of this first stage is the clean background image \( J \), which is reconstructed by the following equation:

\[ J = \frac{I - (1 - T) \odot A}{T}. \]  

In the second stage, we constructed a basic conditional generative adversarial network (cGAN) to refine the output image \( J \). The generative network takes in the reconstructed image \( J \) as well as the predicted transmission map \( T \) and predicted atmospheric light \( A \), and produce the refined clean image \( C \). The discriminative network is used to determine whether this is a real clean image. Hence, the refinement network directly learns to map the reconstructed image \( J \) to clean image \( C \).

The reason of proposing a 2-stage network is that the physics model (as indicated by Eq. 11) is an approximated representation of real fog scenes, and thus can provide constraints to our Network, such as. Most of the existing network does not generalize very well when trained on a particular dataset and tested on other data from a different domain. (i.e. train a network on synthetic data and test it on real haze images will result in poor performance). As mentioned in the introduction, the damages introduced by haze cannot be fully expressed by the coarse model Eq. 11. For this reason, we should add another refinement network, the physical-model free network, which does not assume any hand-crafted model. Hence, unlike the first stage, the refinement network is not guided by the proposed equation, and directly learns the transfer function from the hazy image to clean images.

5.1 Physics-based Stage

The physics-based stage of the proposed network contains two branches. The first branch is an atmospheric light estimation branch, called AtmNet (shown in Fig. 7 top). Another branch is a transmission estimation sub-network called TransNet (shown in Fig. 7 bottom). The details of the structure and processes are discussed in these subsequent sections.

Learning Atmospheric Light
The atmospheric light sub-network learns to first predict the global atmospheric light from the input hazy image. This sub-network is composed of 5 Conv+ReLU blocks appended with 2 fully-connected layers. The output scalar value \( A \) from the second fully-connected layer is then up-sampled to the size of the input image. The loss function for learning atmospheric light \( A \) is defined as:

\[ \mathcal{L}_A = \mathcal{L}_{MSE}(A, A_{gt}) \]  

where \( A_{gt} \) is the ground truth of the atmospheric light and \( A \) is the predicted atmospheric light map. The loss function \( \mathcal{L}_{MSE} \) computes the mean-square-error between the predicted \( A \) and the ground truth \( A_{gt} \).

Learning Transmission Map
The transmission map encodes the global depth information of a scene \( [29] \). To learn the global transmission map well, we use an auto-encoder with skip connection with dilated convolutions to provide the network larger receptive fields. Instead of using batch normalization, we adopt the instance normalization \( [51] \) in the first two convolutional layers in our experiments because the batch normalization usually performs poorly when the testing data is from a different domain. The loss function for learning the
5.2 Refinement Stage

The model-free refinement stage contains a conditional generative adversarial network. The generative network takes in the estimated image $\hat{I}$ and rain image $I$ The goal of the generative network is to generate a refined clean version $\hat{I}$ from the estimated image $\hat{I}$ and the artefacts produced by previous stage. The input of this generator is the concatenated results of reconstructed image $\hat{I}$ predicted transmission map $T$ and the predicted atmospheric light $A$. We also add MSE and perceptual losses [13] for training the generative network. They are given by the first and second terms in the following loss functions:

$$L_T = L_{MSE}(T, T^\text{gt}),$$

where $\lambda_p$ represents the weighting term between MSE loss and perceptual loss. In this experiment, we set it equal to 1. $VGG$ represents the VGG16 [44] network pretrained on ImageNet dataset [39].

Overall, the loss function for the generative network is as follows:

$$L_G = L_c + \lambda_{GAN} L_{GAN}(C),$$

where $L_{GAN}$ is the adversarial loss and hence $L_{GAN}(C) = \log(1 - D(C))$ and the weighting parameter $\lambda_{GAN}$ is set to 0.01 in our experiments.

5.3 Implementation

The proposed network is firstly trained in the stage-wise manner and then fine-tuned on an end-to-end basis. The first stage, physics-based stage is trained the proposed dataset, to learn transmission and atmospheric light estimation. We use Adam [14] optimizer with weight decay of $10^{-4}$ and only supervise the loss $L_T$ and $L_A$. The learning rate is set to 0.001 initially and is divided by 2 after every 10 epochs until the 50 epoch. When training the second stage, we fix the learnable parameters of the first stage and only update the refinement subnetwork with a comparatively smaller learning rate, starting from 0.0005. The optimizer for training the second network is the same. The model free network is trained from 60th epoch to 100th epoch. The entire network is implemented in PyTorch framework and is processed on NVIDIA GTX1080 Ti GPU. The entire training time is approximately 24 hours.

6 Benchmark

In this section, We evaluate our algorithm compared with popular conventional methods and recent CNN based methods. We will mainly use the proposed dataset for the benchmark. The existing methods we choose to compare are including Boundary Constrained Context Regularization (BCCR) [32], Color Attenuation Prior (CAP) [57], Dark-Channel Prior (DCP) [11], Artifact Suppression via Gradient Residual Minimization (GRM) [4], and Non-local Image Dehazing (non-local) [2]. For CNN-based methods, we compared with DehazeNet [3], Multi-scale CNN (MSCNN) [57], All-in-One Dehazing (AODNet) [17], Densely Connected Pyramid Dehazing Network (DCPDN) [34], and grid [29]. We perform the testing on our proposed dataset on Table 2 and the most recent dehazing dataset Reside [18] on Table 3.

6.1 Results on the Proposed Dataset

We first evaluated the classic dehazing algorithms and CNN-based methods on the proposed datasets as shown in Table 2. We use PSNR [12] and SSIM [52] metrics to evaluate all the methods. The conventional methods are directly applied on the proposed dataset test set. For CNN-based solutions, we directly used the author provided weights to run on the proposed datasets. From the table, one can see that the proposed method outperforms other methods for PSNR and SSIM evaluations. Since the proposed method contains comparatively denser haze, the CNN-based baseline methods could not perform well since most of the methods are single-stage, which are not designed for solving dense haze conditions. The conventional dehazing algorithms tend to make the background darker and over saturated as shown in Fig. [9, 10, and 11]. From the figures, one may also find that baseline CNN methods tend to predict wrong depth at some objects as indicated in the red boxes. However, our proposed method is able...
6.2 Results on RESIDE-β dataset

We finetune our proposed network on the training set of RESIDE[19] and test it on the public test set. The results are shown in Fig. 12 and Fig. 13. Since the RESIDE dataset renders comparatively light haze and the transmission maps do not align with the actual background depths well. One may observe that most of the CNN methods can perform well on this test data. The atmospheric light estimation makes an important role for the conventional baseline methods. GRM [4] and AODNet[17] result in poor performance because of inaccurate estimation (implicit) on atmospheric light. Therefore the results’ color deviates from the ground truth.

6.3 Running time on test images

The running time comparison of our proposed network and other baseline methods are shown in the last column of Table 3. The running time testing setup is followed [18]. We averaged over the synthetic indoor images of size 620×460 in the synthetic outdoor testing sets. For CPU algorithms [32,57,11,4,2] and AOD-Net [17], we follow the reported results in [18]. For GPU algorithms, we rerun them on our machine with NVIDIA GTX1080 Ti GPU. As we have proposed a 2-stage simple network, from the table we can see that our algorithm shows a clear superiority over other GPU-based methods in efficiency due to the light weight design of our network. The two-stage network is able to maintain good performance while working in a interactive processing speed.

6.4 Ablation Study

Effectiveness of refinement In our proposed method, the dehazing network contains two stages: reconstruction and refinement. To verify the necessity and effectiveness, we conducted the ablation study to compare the restore results between the final output C and intermediate reconstructed result J. The quantitative results on the proposed dataset is shown in Fig[8]. From the table, one can find that the intermediate reconstructed image J is very sensitive to the estimation of atmospheric light, failing to estimate which will result in dark result of J. The second stage refinement C will work based on the reconstructed image J to recover the details of the dark regions in J.

Categorical Analysis of the proposed dataset In order to provide a deep analysis on the proposed dataset, we have classified the dataset into different categories based on the nature of the scenes, including Cityscape, Landscape, Natural rural scenes. Since this dataset is created with respect to the 3D videos, we have also classify the test sets into Human-centric and non-Human categories. We then test all the existing methods as well as the proposed method on these different categories. The results are shown in Table 5. From the table, one can see that our proposed method consistently outperforms others on all the categories. It is also interesting to study if the fog density makes a difference. To that end, we also classify the test sets based on the fog density β into 5 different categories. The testing performance is shown in Table 4. In this analysis, the

### Table 2: The quantitative comparison between our method and baseline methods on the proposed dataset test set.

| Methods   | PSNR  | SSIM  |
|-----------|-------|-------|
| BCCR [32] | 17.91 | 0.729 |
| DCP [11]  | 19.56 | 0.761 |
| CAP [57]  | 14.64 | 0.715 |
| Non-Local [2]<sup>2</sup> | 18.06 | 0.727 |
| GRM [4]<sup>3</sup> | 17.48 | 0.682 |
| MS-CNN [37] | 10.85 | 0.687 |
| AODnet [17] | 12.42 | 0.639 |
| AODnet-ft [17] | 18.33 | 0.728 |
| DehazeNet [2] | 19.76 | 0.788 |
| DCPDN [54] | 15.38 | 0.660 |
| GridDehazeNet [29] | 12.37 | 0.659 |
| GridDehazeNet-ft [29] | 22.19 | 0.837 |
| Ours      | 26.71 | 0.923 |

### Table 3: The quantitative comparison between our method and baseline methods on RESIDE [18] dataset.

| Methods   | PSNR  | SSIM  | Time (s) |
|-----------|-------|-------|----------|
| BCCR [32] | 16.88 | 0.7913| 3.85     |
| DCP [11]  | 16.62 | 16.62 | 1.62     |
| CAP [57]  | 19.05 | 0.8364| 0.95     |
| Non-Local [2]<sup>2</sup> | 17.29 | 0.7489 | 9.89 |
| GRM [4]<sup>3</sup> | 18.86 | 0.8553 | 83.96 |
| MS-CNN [37] | 17.57 | 0.8102 | 2.60 |
| AODnet [17] | 19.06 | 0.8504 | 0.65 |
| DehazeNet [2] | 21.14 | 0.8472 | 2.51 |
| GridDehazeNet [29] | 30.86 | 0.9819 | 0.26 |
| Ours      | 30.33 | 0.9473 | 0.03 |

To predict appropriate depth value and object depth boundaries.

---

1 Learning to Dehaze From Realistic Scene with A Fast Physics Based Dehazing Network

---
Table 4: The quantitative comparison between our method and baseline methods on the proposed dataset with respect to different haze/fog density $\beta$.

| Method       | $\beta = 1.5$ | $\beta = 1.4$ | $\beta = 1.3$ | $\beta = 1.2$ | $\beta = 1.1$ |
|--------------|--------------|--------------|--------------|--------------|--------------|
|              | PSNR | SSIM  | PSNR | SSIM  | PSNR | SSIM  | PSNR | SSIM  | PSNR | SSIM  |
| BCCR [32]    | 17.95| 0.7333| 17.85| 0.7299| 17.77| 0.7175| 17.99| 0.7279| 18.05| 0.7376|
| DCP [11]     | 19.80| 0.7704| 19.65| 0.7641| 19.59| 0.7507| 19.32| 0.7556| 19.53| 0.7654|
| CAP [57]     | 13.83| 0.7233| 14.17| 0.7017| 14.11| 0.6945| 16.08| 0.7401| 15.10| 0.7262|
| Non-Local [2]| 17.71| 0.7269| 18.01| 0.7206| 17.88| 0.7199| 18.27| 0.7302| 18.34| 0.7342|
| GRM [4]      | 17.03| 0.6872| 17.11| 0.6764| 17.19| 0.6723| 18.08| 0.6870| 17.98| 0.6872|
| AODnet [17]  | 18.27| 0.7926| 18.56| 0.7383| 18.11| 0.7150| 18.51| 0.7314| 18.30| 0.7290|
| GridDehazeNet [29] | 21.47| 0.8255| 22.30| 0.8458| 22.64| 0.8402| 22.92| 0.8480| 21.67| 0.8275|
| Ours         | 24.89| 0.9012| 25.36| 0.9173| 25.29| 0.9076| 25.45| 0.9148| 25.13| 0.9009|

Table 5: The quantitative comparison between our method and baseline methods on the proposed dataset.

| Method       | Human-centric | Non-Human | Cityscapes | Landscapes | Natural rural |
|--------------|--------------|-----------|------------|------------|--------------|
|              | PSNR | SSIM  | PSNR | SSIM  | PSNR | SSIM  | PSNR | SSIM  | PSNR | SSIM  |
| BCCR [32]    | 17.93| 0.7242| 18.04| 0.7403| 18.03| 0.7397| 17.88| 0.7369| 18.04| 0.7575|
| DCP [11]     | 19.62| 0.7546| 19.52| 0.7702| 19.18| 0.7616| 19.70| 0.7710| 19.83| 0.7948|
| CAP [57]     | 14.21| 0.6991| 15.07| 0.7300| 14.85| 0.7272| 14.59| 0.7185| 15.08| 0.73985|
| Non-Local [2]| 17.53| 0.7075| 18.38| 0.7430| 18.02| 0.7344| 18.38| 0.7405| 18.64| 0.7537|
| GRM [4]      | 17.28| 0.6671| 17.64| 0.6960| 17.63| 0.6899| 17.70| 0.7021| 17.76| 0.7075|
| AODnet [17]  | 18.49| 0.6997| 18.71| 0.7486| 18.53| 0.7302| 18.48| 0.7369| 18.68| 0.7772|
| GridDehazeNet [29] | 22.45| 0.8110| 22.58| 0.8539| 22.17| 0.8412| 22.44| 0.8501| 23.07| 0.8832|
| Ours         | 25.81| 0.9037| 25.91| 0.9126| 26.16| 0.9142| 26.47| 0.9168| 25.98| 0.9213|

Fig. 8: Comparison of the intermediate reconstructed result $J$ and the final output $C$. The reconstructed image $J$ usually contains darker results. However, the final output $C$ recovers the details of the darker regions.

The proposed method also outperforms other method significantly.

7 Conclusion

We propose a new, large, and diverse dehazing dataset containing real outdoor scenes from HD 3D videos. We select high quality frames of real outdoor scenes and render haze on them using depth from stereo pair. Comparing with previous datasets, our dataset is more realistic thanks to the high quality depth map. We demonstrate that using this dataset greatly improves the dehazing performance on real scenes. In addition, we design a simple physics inspired network model and train it on our proposed dataset. In this way, our approach outperforms other methods by a large margin and is the only one runs at a realtime speed. Our dataset is a good complement to existing dehazing dataset and our method provide a practical dehazing solution that are both efficient and effective.

References

1. Akkaynak, D., Treibitz, T.: A revised underwater image formation model. In: IEEE Conference on Computer Vision and Pattern Recognition(2018)
2. Berman, D., Treibitz, T., Avidan, S.: Non-local image dehazing. In: IEEE Conference on Computer Vision and Pattern Recognition(2016)
3. Cai, B., Xu, X., Jia, K., Qing, C., Tao, D.: Dehazenet: An end-to-end system for single image haze removal. IEEE Transactions on Image Processing 25(11), 5187–5198 (2016)
4. Chen, C., Do, M.N., Wang, J.: Robust image and video dehazing with visual artifact suppression via gradient residual minimization. In: European Conference on Computer Vision(2016)
5. Chen, C.C.: Attenuation of electromagnetic radiation by haze, fog, clouds, and rain. Tech. rep., RAND CORP SANTA MONICA CA (1975)
6. Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., Franke, U., Roth, S., Schiele,
Fig. 9: Results comparison on the proposed dataset.

B.: The cityscapes dataset for semantic urban scene understanding. In: IEEE Conference on Computer Vision and Pattern Recognition(2016)
7. Fattal, R.: Single image dehazing. ACM Transactions on Graphics27(3), 72 (2008)
8. Fattal, R.: Dehazing using color-lines. ACM Trans. Graph. 34(1), 13:1–13:14 (2014)
9. Galdran, A., Alvarez-Gilia, A., Bria, A., Vazquez-Corral, J., Bertalmío, M.: On the duality between retinex and image dehazing. In: IEEE Conference on Computer Vision and Pattern Recognition(2018)
10. He, K., Sun, J., Tang, X.: Single image haze removal using dark channel prior. In: IEEE Conference on Computer Vision and Pattern Recognition(2009)
11. He, K., Sun, J., Tang, X.: Single image haze removal using dark channel prior. IEEE Transactions on Pattern Analysis and Machine Intelligence33(12), 2341–2353 (2011)
12. Huynh-Thu, Q., Ghanbari, M.: Scope of validity of psnr in image/video quality assessment. Electronics Letters 44(13), 800–801 (2008). DOI 10.1049/el:20080522
13. Johnson, J., Alahi, A., Fei-Fei, L.: Perceptual losses for real-time style transfer and super-resolution. In: European Conference on Computer Vision(2016)
14. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. CoRR abs/1412.6980 (2014). URL http://arxiv.org/abs/1412.6980
Fig. 10: Results comparison on the proposed dataset.

15. Kopf, J., Neubert, B., Chen, B., Cohen, M.F., Cohen-Or, D., Deussen, O., Uyttendaele, M., Lischinski, D.: Deep photo: Model-based photograph enhancement and viewing. ACM Transactions on Graphics 27(5), 116:1–116:10 (2008)

16. Koschmieder, H.: Theorie der horizontalen Sichtweite: Kontrast und Sichtweite. Keim & Nemnich (1925)

17. Li, B., Peng, X., Wang, Z., Xu, J., Feng, D.: Aod-net: All-in-one dehazing network. In: IEEE International Conference on Computer Vision(2017)

18. Li, B., Ren, W., Fu, D., Tao, D., Feng, D., Zeng, W., Wang, Z.: Benchmarking single-image dehazing and beyond. IEEE Transactions on Image Processing 28(1), 492–505 (2019)

19. Li, R., Pan, J., Li, Z., Tang, J.: Single image dehazing via conditional generative adversarial network. In: IEEE Conference on Computer Vision and Pattern Recognition(2018)

20. Li, Y., Guo, F., Tan, R.T., Brown, M.S.: A contrast enhancement framework with jpeg artifacts suppression. In: European Conference on Computer Vision(2014)

21. Li, Y., Miao, Q., Ouyang, W., Ma, Z., Fang, H., Dong, C., Quan, Y.: Lap-net: Level-aware progressive network for image dehazing. In: IEEE International Conference on Computer Vision(2019)
Fig. 11: Results comparison on the proposed dataset.

22. Li, Y., Min, D., Do, M.N., Lu, J.: Fast guided global interpolation for depth and motion. In: B. Leibe, J. Matas, N. Sebe, M. Welling (eds.) Computer Vision – ECCV 2016, pp. 717–733. Springer International Publishing, Cham (2016)

23. Li, Y., Tan, R.T., Brown, M.S.: Nighttime haze removal with glow and multiple light colors. In: IEEE International Conference on Computer Vision (2015)

24. Li, Y., You, S., Brown, M.S., Tan, R.T.: Haze visibility enhancement: A survey and quantitative benchmarking. Computer Vision and Image Understanding 165, 1–16 (2017)

25. Li, Z., Tan, P., Tan, R.T., Zhou, S.Z., Cheong, L.F.: Simultaneous video dehazing and stereo reconstruction. In: IEEE Conference on Computer Vision and Pattern Recognition (2015)

26. Liu, F., Shen, C., Lin, G.: Deep convolutional neural fields for depth estimation from a single image. In: IEEE Conference on Computer Vision and Pattern Recognition (2015)

27. Liu, F., Shen, C., Lin, G., Reid, I.: Learning depth from single monocular images using deep convolutional neural fields. IEEE Transactions on Pattern Analysis and Machine Intelligence 38(10), 2024–2039 (2015)

28. Liu, F., Shen, C., Lin, G., Reid, I.D.: Learning depth from single monocular images using deep convolutional neural fields. CoRR abs/1502.07411 (2015). URL [http://arxiv.org/abs/1502.07411](http://arxiv.org/abs/1502.07411)

29. Liu, X., Ma, Y., Shi, Z., Chen, J.: Griddehazenet: Attention-based multi-scale network for image dehazing. In: IEEE International Conference on Computer Vision (2019)
16 Ruoteng Li et al.

30. Liu, Y., Pan, J., Ren, J., Su, Z.: Learning deep priors for image dehazing. In: IEEE International Conference on Computer Vision(2019)

31. McCartney, E.J.: Optics of the atmosphere: scattering by molecules and particles. New York, John Wiley and Sons, Inc., 1976. 421 p. 1 (1976)

32. Meng, G., Wang, Y., Duan, J., Xiang, S., Pan, C.: Efficient image dehazing with boundary constraint and contextual regularization. In: IEEE International Conference on Computer Vision(2013)

33. Narasimhan, S.G., Nayar, S.K.: Chromatic framework for vision in bad weather. In: IEEE Conference on Computer Vision and Pattern Recognition(2000)

34. Narasimhan, S.G., Nayar, S.K.: Contrast restoration of weather degraded images. IEEE Transactions on Pattern Analysis and Machine Intelligence25(6), 713–724 (2003)

35. Pharr, M., Humphreys, G.: Physically based rendering: From theory to implementation. Morgan Kaufmann (2010)

36. Qi, Y., Chen, Y., Huang, J., Xie, Y.: Enhanced pix2pix dehazing network. In: IEEE Conference on Computer Vision and Pattern Recognition

37. Ren, W., Liu, S., Zhang, H., Pan, J., Cao, X., Yang, M.H.: Single image dehazing via multi-scale convolutional neural networks. In: European Conference on Computer Vision(2016)

38. Ren, W., Ma, L., Zhang, J., Pan, J., Cao, X., Liu, W., Yang, M.H.: Gated fusion network for single image dehazing. In: IEEE Conference on Computer Vision and Pattern Recognition(2018)

39. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A.C., Fei-Fei, L.: ImageNet Large Scale Visual Recognition Challenge. International Journal of Computer Vision115(3), 211–252 (2015). DOI 10.1007/s11263-015-0816-y

40. Sakaridis, C., Dai, D., Van Gool, L.: Semantic foggy scene understanding with synthetic data. International Journal of Computer Vision126(9), 973–992 (2018)

41. Scharstein, D., Szeliski, R.: High-accuracy stereo depth maps using structured light. In: IEEE Conference on Computer Vision and Pattern Recognition(2003)

42. Schechner, Y.Y., Narasimhan, S.G., Nayar, S.K.: Instant dehazing of images using polarization. In: IEEE Conference on Computer Vision and Pattern Recognition(2001)

43. Silberman, N., Hoiem, D., Kohli, P., Fergus, R.: Indoor segmentation and support inference from rgbd images. In: European Conference on Computer Vision(2012)

44. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. CoRR abs/1409.1556 (2014). URL http://arxiv.org/abs/1409.1556

45. Solonenko, M.G., Mobley, C.D.: Inherent optical properties of jerlov water types. Applied optics 54(17), 5392–5401 (2015)

46. Sulami, M., Geltzer, I., Fattal, R., Werman, M.: Automatic recovery of the atmospheric light in hazy images. In: IEEE Int’l Conf. Computational Photography (2014)

47. Sun, D., Yang, X., Liu, M.Y., Kautz, J.: PWC-Net: CNNs for optical flow using pyramid, warping, and cost volume. In: IEEE Conference on Computer Vision and Pattern Recognition(2018)

48. Tan, R.T.: Visibility in bad weather from a single image. In: IEEE Conference on Computer Vision and Pattern Recognition(2008)

49. Tang, K., Yang, J., Wang, J.: Investigating haze-relevant features in a learning framework for image dehazing. In: IEEE Conference on Computer Vision and Pattern Recognition(2014)

50. Tarel, J.P., Hautière, N., Caraffa, L., Cord, A., Halmaoui, H., Gruyer, D.: Vision enhancement in homogeneous and heterogeneous fog. IEEE Intelligent Transportation Systems Magazine 4(2), 6–20 (2012)

51. Ulyanov, D., Vedaldi, A., Lempitsky, V.S.: Instance normalization: The missing ingredient for fast stylization. CoRR abs/1607.08022 (2016). URL http://arxiv.org/abs/1607.08022

52. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing 13(4), 600–612 (2004). DOI 10.1109/TIP.2003.819861

53. Yang, D., Sun, J.: Proximal dehaze-net: a prior learning-based deep network for single image dehazing. In: European Conference on Computer Vision(2018)

54. Zhang, H., Patel, V.M.: Densely connected pyramid dehazing network. In: IEEE Conference on Computer Vision and Pattern Recognition(2018)

55. Zhang, J., Cao, Y., Fang, S., Kang, Y., Wen Chen, C.: Fast haze removal for nighttime image using maximum reflectance prior. In: IEEE Conference on Computer Vision and Pattern Recognition(2017)

56. Zhou, B., Zhao, H., Puig, X., Xiao, T., Fidler, S., Barriuso, A., Torralba, A.: Semantic understanding of scenes through the ade20k dataset. International Journal of Computer Vision(2018)

57. Zhu, Q., Mai, J., Shao, L.: A fast single image haze removal algorithm using color attenuation prior. IEEE Transactions on Image Processing24(11), 3522–3533 (2015)
Fig. 12: Results comparison on RESIDE-$\beta$ [18] dataset.
Fig. 13: Results comparison on RESIDE-β [18] dataset.