

Domain Adaptation for Image Dehazing

Yuanjie Shao¹, Lerenhan Li¹, Wenqi Ren², Changxin Gao¹, Nong Sang^{1*}

¹National Key Laboratory of Science and Technology on Multispectral Information Processing,
School of Artificial Intelligence and Automation,
Huazhong University of Science and Technology, Wuhan, China

²Institute of Information Engineering, Chinese Academy of Sciences, Beijing, China
{shaoyuanjie, lrhli, cgao, nsang}@hust.edu.cn, rwq.renwenqi@gmail.com

Abstract

Image dehazing using learning-based methods has achieved state-of-the-art performance in recent years. However, most existing methods train a dehazing model on synthetic hazy images, which are less able to generalize well to real hazy images due to domain shift. To address this issue, we propose a domain adaptation paradigm, which consists of an image translation module and two image dehazing modules. Specifically, we first apply a bidirectional translation network to bridge the gap between the synthetic and real domains by translating images from one domain to another. And then, we use images before and after translation to train the proposed two image dehazing networks with a consistency constraint. In this phase, we incorporate the real hazy image into the dehazing training via exploiting the properties of the clear image (e.g., dark channel prior and image gradient smoothing) to further improve the domain adaptivity. By training image translation and dehazing network in an end-to-end manner, we can obtain better effects of both image translation and dehazing. Experimental results on both synthetic and real-world images demonstrate that our model performs favorably against the state-of-the-art dehazing algorithms.

1. Introduction

Single image dehazing aims to recover the clean image from a hazy input, which is essential for subsequent high-level tasks, such as object recognition and scene understanding. Thus, it has received significant attention in the vision community over the past few years. According to the physical scattering models [21, 23, 18], the hazing process is usually formulated as

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

*Corresponding author.



Figure 1. Dehazed results on a real world hazy image. Our method generates a cleaner image with fewer color distortion and brighter details. (a) Hazy image. (b) The result of EPDN [25]. (c) Our result.

where $I(x)$ and $J(x)$ denote the hazy image and the clean image, A is the global atmospheric light, and $t(x)$ is the transmission map. The transmission map can be represented as $t(x) = e^{-\beta d(x)}$, where $d(x)$ and β denote the scene depth and the atmosphere scattering parameter, respectively. Given a hazy image $I(x)$, most dehazing algorithms try to estimate $t(x)$ and A .

However, estimating the transmission map from a hazy image is an ill-posed problem generally. Early prior-based methods try to estimate the transmission map by exploiting the statistical properties of clear images, such as dark channel prior [9] and color-line prior [8]. Unfortunately, these image priors are easily inconsistent with the practice, which may lead to inaccurate transmission approximations. Thus, the quality of the restored image is undesirable.

To deal with this problem, convolutional neural networks (CNNs) have been employed to estimate transmission maps [4, 26, 35] or predict clear images directly [12, 27, 16, 25]. These methods are valid and superior to the prior-based algorithms with significant performance improvements. However, deep learning-based approaches need to rely on a large amount of real hazy images and their haze-free counterparts for training. In general, it is impractical to acquire large quantities of ground-truth images in the real world. Therefore, most dehazing models resort to training on synthetic hazy dataset. However, due to the domain shift problem, the models learned from synthetic data often fail

to generalize well to real data.

To address this issue, we propose a domain adaptation framework for single image dehazing. The proposed framework includes two parts, namely an image translation module and two domain-related dehazing modules (one for synthetic domain and another for real domain). To reduce the discrepancy between domains, our method first employs the bidirectional image translation network to translate images from one domain to another. Since image haze is a kind of noise and nonuniform highly depending on the scene depth, we incorporate the depth information into the translation network to guide the translation of synthetic to real hazy images. Then, the domain-related dehazing network takes images of this domain, including the original and translated images, as inputs to perform image dehazing. Moreover, we use a consistency loss to ensure that the two dehazing networks generate consistent results. In this training phase, to further improve the generalization of the network in the real domain, we incorporate the real hazy images into the training. We hope that the dehazing results of the real hazy image can have some properties of the clear images, such as dark channel prior and image gradient smoothing. We train the image translation network and dehazing networks in an end-to-end manner so that they can improve each other. As shown in Figure 1, our model produces a cleaner image when compared with recent dehazing work of EPDN [25].

We summarize the contributions of our work as follows:

- We propose an end-to-end domain adaptation framework for image dehazing, which effectively bridges the gap between the synthetic and real-world hazy images.
- We show that incorporating real hazy images into the training process can improve the dehazing performance.
- We conduct extensive experiments on both synthetic datasets and real-world hazy images, which demonstrate that the proposed method performs favorably against the state-of-the-art dehazing approaches.

2. Related Work

In this section, we briefly discuss the single image dehazing approaches and domain adaptation methods, which are related to our work.

2.1. Single Image Dehazing

Prior-based methods. Prior-based methods estimate the transmission maps and atmospheric light intensity based on the statistics of clear images. Representative works in this regard include [29, 9, 39, 8, 2]. Specifically, Tan [29] proposes a contrast maximization method for image dehazing since it observes that clear images tend to have higher contrast than their hazy counterparts. He et al. [9] make use of dark channel prior (DCP) to estimate the transmission

map, which is based on the assumption that pixels in haze-free patches are close to zero in at least one color channel. Follow-up works have improved the efficiency and performance of the DCP method [30, 22, 17, 24, 34]. Besides, the attenuation prior is adopted in [39] for recovering depth information of the hazy images. Fattal [8] uses the color-line assumption to recover the scene transmission, which asserts that pixels of small image patches exhibit a one-dimensional distribution. Similarly, Berman et al. [2] assume that several hundreds of distinct colors can well approximate colors of a clear image, and then perform image dehazing based on this prior. Though these methods have been shown effective for image dehazing, their performances are inherently limited since the assumed priors are not suited for all real-world images.

Learning-based Methods. With the advance in deep convolutional neural networks (CNNs) and the availability of large-scale synthetic datasets, data-driven approaches for image dehazing have received significant attention in recent years. Many methods [4, 26, 35, 12] directly utilize deep CNNs to estimate the transmissions and atmospheric light, and then restore the clean image according to the degradation model (1). Cai et al. [4] propose an end-to-end dehazing model, DehazeNet, to estimate the transmission map from hazy images. Ren et al. [26] utilize a coarse-to-fine strategy to learn the mapping between hazy inputs and transmission maps. Zhang and Patel [35] propose a densely connected pyramid network to estimate the transmission maps. Li et al. [12] propose an AOD-Net to estimate the parameter of the reformulated physical scattering model, which integrates the transmissions and atmospheric light. Also, some end-to-end methods [27, 16, 25] have been proposed to recover the clean image directly instead of estimating the transmission map and atmospheric light. Ren et al. [27] adopt a gated fusion network to recover the clean image from a hazy input directly. Qu et al. [25] transform the problem of image dehazing to an image-to-image translation problem, and propose an enhanced pix2pix dehazing network.

However, due to the domain gap between synthetic and real data, the CNN-based models trained on synthetic images tend to have a significant performance drop when applied to the real domain. To this end, Li et al. [14] propose a semi-supervised dehazing model, which is trained on both synthetic and real haze image, and thus enjoys domain adaptivity between synthetic and real-world hazy images. However, only applying the real haze image for training does not really solve the problem of domain shift. Different from the above methods, our model first applies the image translation network to translate images from one domain to another, and then performs image dehazing on both synthetic and real domain using the translated images and their original images (synthetic or real). The proposed ap-

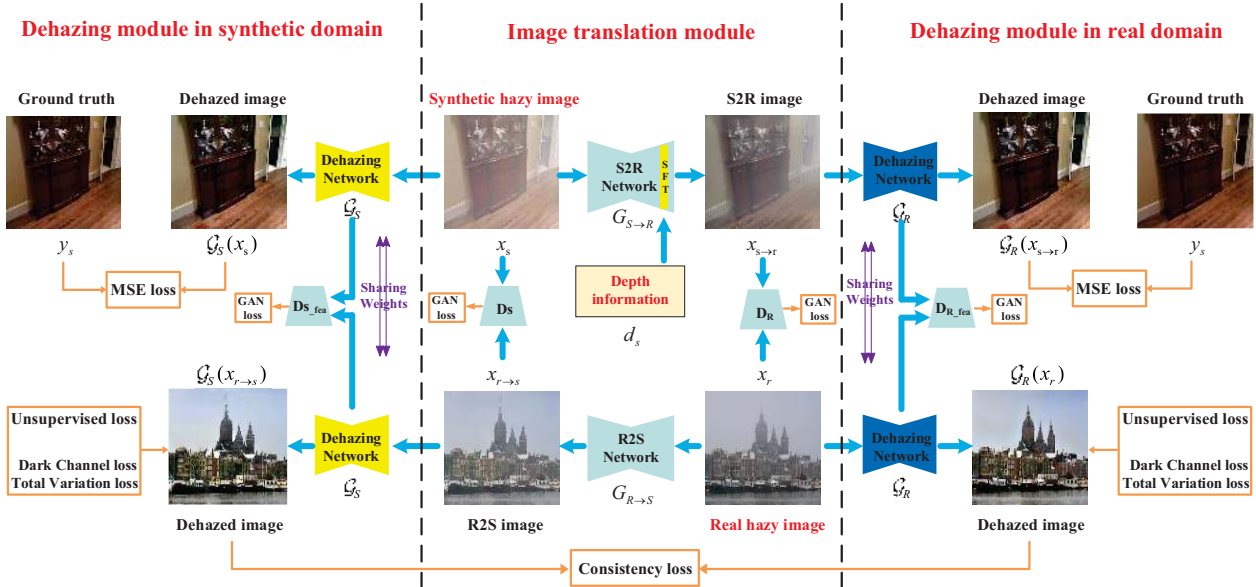


Figure 2. Architecture of the proposed domain adaptation framework for image dehazing. The framework consists of two parts, an image translation module and two image dehazing modules. The image translation module translates images from one domain to another to reduce the domain discrepancy. The image dehazing modules perform image dehazing on both synthetic and real domain.

proach can effectively solve the domain shift problem.

2.2. Domain Adaptation

Domain adaptation aims to reduce the discrepancy between different domains [1, 6, 20]. Existing work either to perform feature-level or pixel-level adaptation. Feature-level adaptation methods aim at aligning the feature distributions between the source and target domains through minimizing the maximum mean discrepancy [19], or applying adversarial learning strategies [32, 31] on the feature space. Another line of research focuses on pixel-level adaptation [3, 28, 7]. These approaches deal with the domain shift problem by applying image-to-image translation [3, 28] learning, or style transfer [7] methods to increase the data in the target domain.

Most recently, many methods perform feature-level and pixel-level adaptation jointly in many visual tasks, e.g., image classification [10], semantic segmentation [5], and depth prediction [37]. These methods [5, 37] translate images from one domain to another with pixel-level adaptation via image-to-image translation networks, e.g., the CycleGAN [38]. The translated images are then inputted to the task network with feature-level alignment. In this work, we take advantage of CycleGAN to adapt the real hazy images to our dehazing model trained on synthetic data. Moreover, since the depth information is closely related to the formulation of image haze, we incorporate the depth information into the translating network to better guide the real hazy image translation.

3. Proposed Method

This section presents the details of our domain adaptation framework. First, we provide an overview of our method. Then we describe the details of the image translation module and image dehazing module. Finally, we give the loss functions that are applied to train the proposed networks.

3.1. Method Overview

Given a synthetic dataset $X_S = \{x_s, y_s\}_{s=1}^{N_l}$ and a real hazy image set $X_R = \{x_r\}_{r=1}^{N_u}$, where N_l and N_u denote the number of the synthetic and real hazy images, respectively. We aim to learn a single image dehazing model which can accurately predict the clear image from real hazy image. Due to the domain shift, the dehazing model trained only on the synthetic data can not generalize well to the real hazy image.

To deal with this problem, we present a domain adaptation framework, which consists of two main parts: the image translation network $G_{S \rightarrow R}$ and $G_{R \rightarrow S}$, and two dehazing networks \mathcal{G}_S and \mathcal{G}_R . The image translation network translates images from one domain to another to bridge the gap between them. Then the dehazing networks perform image dehazing using both translated images and source images (e.g., synthetic or real).

As illustrated in Figure 2, the proposed model takes a real hazy image x_r and a synthetic image x_s along with its corresponding depth images d_s as input. We first obtain the corresponding translated images $x_{s \rightarrow r} = G_{S \rightarrow R}(x_s, d_s)$

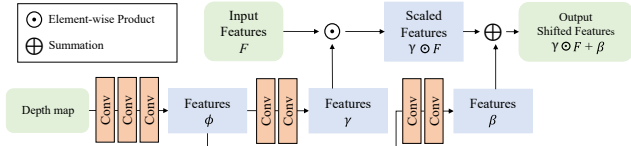


Figure 3. Structure of the SFT layer. In the translator $G_{S \rightarrow R}$, we consider the depth map as the guidance to assist the image translation.

and $x_{r \rightarrow s} = G_{R \rightarrow S}(x_r)$ using two image translators. And then, we pass x_s and $x_{r \rightarrow s}$ to \mathcal{G}_S , x_r and $x_{r \rightarrow s}$ to \mathcal{G}_R to perform image dehazing.

3.2. Image Translation Module

The image translation module includes two translators: synthetic to real network $G_{S \rightarrow R}$ and real to synthetic network $G_{R \rightarrow S}$. The $G_{S \rightarrow R}$ network takes (X_s, D_s) as inputs, and generates translated images $G_{S \rightarrow R}(X_s, D_s)$ with similar style to the real hazy images. Another translator $G_{R \rightarrow S}$ performs image translation inversely. Since the depth information is highly correlated to the hazing formulation, we incorporate it into the generator $G_{S \rightarrow R}$ to produce images with similar haze distribution in real cases.

We adopt the spatial feature transform (SFT) layer [33, 15] to incorporate the depth information into the translation network, which can fuse features from depth map and synthetic image effectively. As shown in Fig. 3, the SFT layer first applies three convolution layers to extract conditional maps ϕ from the depth map. The conditional maps are then fed to the other two convolution layers to predict the modulation parameters, γ and β , respectively. Finally, we can obtain the output shifted features by:

$$\text{SFT}(F|\gamma, \beta) = \gamma \odot F + \beta. \quad (2)$$

where \odot is the element-wise multiplication. In the translator $G_{S \rightarrow R}$, we treat the depth map as the guidance and use the SFT layer to transform the features of the penultimate convolution layer. As shown in Fig. 4, the synthetic images are relative closer to the real-world hazy image after the translation.

We show the detailed configurations of the translator $G_{S \rightarrow R}$ in Table 1. We also employ the architectures, provided by CycleGAN [38], for the generator $G_{R \rightarrow S}$ and discriminators (D_R^{img} and D_S^{img}).

3.3. Dehazing Module

Our method includes two dehazing modules \mathcal{G}_S and \mathcal{G}_R , which perform image dehazing on synthetic and real domains, respectively. \mathcal{G}_S takes the synthetic image x_s and the translated image $x_{r \rightarrow s}$ as inputs to perform image dehazing. And \mathcal{G}_R is trained on x_r and $x_{s \rightarrow r}$. For these two image dehazing networks, we both utilize a standard encoder-decoder architecture with skip connections and side outputs



(a) Synthetic hazy image (b) Translated image (c) Real hazy image

Figure 4. Translated results on two synthetic hazy images.

as [37]. The dehazing network in each domain shares the same network architecture but with different learned parameters.

3.4. Training Losses

In the domain adaptation framework, we adopt the following losses to train the network.

Image translation Losses. The aim of our translate module is to learn the translators $G_{S \rightarrow R}$ and $G_{R \rightarrow S}$ to reduce the discrepancy between the synthetic domain X_S and the real domain X_R . For translators $G_{S \rightarrow R}$, we expect the $G_{S \rightarrow R}(x_s, d_s)$ to be indistinguishable from the real hazy image x_r . Thus, we employ an image-level discriminators D_R^{img} and a feature-level discriminators D_R^{feat} , to perform a minmax game via an adversarial learning manner. The D_R^{img} aims at aligning the distributions between the real image x_r and the translated image $G_{S \rightarrow R}(x_s, d_s)$. The discriminator D_R^{feat} helps align the distributions between the feature map of x_r and $G_{S \rightarrow R}(x_s, d_s)$. The adversarial losses are defined as:

$$\begin{aligned} L_{Gan}^{img}(X_R, (X_S, D_S), D_R^{img}, G_{S \rightarrow R}) \\ = \mathbb{E}_{x_s \sim X_S, d_s \sim D_S} [D_R^{img}(G_{S \rightarrow R}(x_s, d_s))] \\ + \mathbb{E}_{x_r \sim X_R} [D_R^{img}(x_r) - 1]. \end{aligned} \quad (3)$$

$$\begin{aligned} L_{Gan}^{feat}(X_R, (X_S, D_S), D_R^{feat}, G_{S \rightarrow R}, \mathcal{G}_R) \\ = \mathbb{E}_{x_s \sim X_S, d_s \sim D_S} [D_R^{feat}(\mathcal{G}_R(G_{S \rightarrow R}(x_s, d_s)))] \\ + \mathbb{E}_{x_r \sim X_R} [D_R^{feat}(\mathcal{G}_R(x_r)) - 1]. \end{aligned} \quad (4)$$

Similar to $G_{S \rightarrow R}$, the translator $G_{R \rightarrow S}$ has another image-level adversarial loss and feature-level adversarial loss, which are denoted as $L_{Gan}^{img}(X_S, X_R, D_S^{img}, G_{R \rightarrow S})$, $L_{Gan}^{feat}(X_S, X_R, D_S^{feat}, G_{R \rightarrow S}, \mathcal{G}_S)$, respectively.

In addition, we utilize the cycle-consistency loss [38] to regularize the training of translation network. Specifically,

Table 1. Configurations of image translation module. ‘‘Conv’’ denotes the convolution layer, ‘‘Res’’ denotes the residual block, ‘‘Upconv’’ denotes the up-sample layer by transposed convolution operator and ‘‘Tanh’’ denotes the non-linear Tanh layer.

Layer	Conv1	Conv2	Conv3	Res4-12	Upconv13	Upconv14	SFT layer	Conv15	Tanh
In_channels	3	64	128	256	256	128	64	64	3
Out_channels	64	128	256	256	128	64	64	3	3
Kernel size	7	3	3	-	3	3	-	7	-
Stride	1	2	2	-	2	2	-	1	-
Pad	-	1	1	-	1	1	-	-	-

when passing an image x_s to $G_{S \rightarrow R}$ and $G_{R \rightarrow S}$ sequentially, we expect the output should be the same image, and vice versa for x_r . Namely, $G_{R \rightarrow S}(G_{S \rightarrow R}(x_s, d_s)) \approx x_s$ and $G_{S \rightarrow R}(G_{R \rightarrow S}(x_r, d_r)) \approx x_r$. The cycle consistency loss can be expressed as:

$$L_c = \mathbb{E}_{x_s \sim X_S, d_s \sim D_S} [\|G_{R \rightarrow S}(G_{S \rightarrow R}(x_s, d_s) - x_s\|_1] + \mathbb{E}_{x_r \sim X_R, d_r \sim D_R} [\|G_{S \rightarrow R}(G_{R \rightarrow S}(x_r, d_r) - x_r\|_1]. \quad (5)$$

Finally, to encourage the generators to preserve content information between the input and output, we also utilize an identity mapping loss [38], which is denoted as:

$$L_{idt} = \mathbb{E}_{x_s \sim X_S} [\|G_{R \rightarrow S}(x_s) - x_s\|_1] + \mathbb{E}_{x_r \sim X_R, d_r \sim D_R} [\|G_{S \rightarrow R}(x_r, d_r) - x_r\|_1]. \quad (6)$$

The full loss function for the the translating module is as follow:

$$L_{tran} = L_{Gan}^{img}(X_R, (X_S, D_S), D_R^{img}, G_{S \rightarrow R}) + L_{Gan}^{feat}(X_R, (X_S, D_S), D_R^{feat}, G_{S \rightarrow R}, \mathcal{G}_R) + L_{Gan}^{img}(X_S, X_R, D_S^{img}, G_{R \rightarrow S}) + L_{Gan}^{feat}(X_S, X_R, D_S^{feat}, G_{R \rightarrow S}, \mathcal{G}_S) + \lambda_1 L_c + \lambda_2 L_{idt}. \quad (7)$$

Image dehazing Losses. We can now transfer the synthetic images X_S and the corresponding depth images D_S to the generator $G_{S \rightarrow R}$, and obtain a new dataset $X_{S \rightarrow R} = G_{S \rightarrow R}(X_S, D_S)$, which has a similar style with real hazy images. And then, we train a image dehazing network \mathcal{G}_R on $X_{S \rightarrow R}$ and X_R in a semi-supervised manner. For supervised branch, we apply the mean squared loss to ensure the predicted images $J_{S \rightarrow R}$ is close to clean images Y_S , which can be defined as:

$$L_{rm} = \|J_{S \rightarrow R} - Y_S\|_2^2. \quad (8)$$

In the unsupervised branch, we introduce the total variation and dark channel losses, which regularize the dehazing network to produce images with similar statistical characteristics of the clear images. The total variation loss is an ℓ_1 -regularization gradient prior on the predicted images J_R :

$$L_{rt} = \|\partial_h J_R\|_1 + \|\partial_v J_R\|_1. \quad (9)$$

where ∂_h denotes the horizontal gradient operators, and ∂_v represents the vertical gradient operators.

Furthermore, the recent work [9] has proposed the concept of the dark channel, which can be expressed as:

$$D(I) = \min_{y \in N(x)} \left[\min_{c \in \{r, g, b\}} I^c(y) \right]. \quad (10)$$

where x and y are pixel coordinates of image I , I^c represents c -th color channel of I , and $N(x)$ denotes the local neighborhood centered at x . He et al. [9] have also shown that most intensity of the dark channel image are zero or close to zero. Therefore, we apply the following dark channel (DC) loss to ensure that the dark channel of the predicted images are in consistence with that of clean image:

$$L_{rd} = \|D(J_R)\|_1. \quad (11)$$

In addition, we also train a complementary image dehazing network \mathcal{G}_S on X_S and $X_{R \rightarrow S}$. Similarly, we apply the same supervised loss and unsupervised loss to train the dehazing network \mathcal{G}_S , which are as follows:

$$L_{sm} = \|J_S - Y_S\|_2^2, \quad (12)$$

$$L_{st} = \|\partial_h J_{R \rightarrow S}\|_1 + \|\partial_v J_{R \rightarrow S}\|_1, \quad (13)$$

$$L_{sd} = \|D(J_{R \rightarrow S})\|_1. \quad (14)$$

Finally, considering that the outputs of the two dehazing networks should be consistency for real hazy images, i.e., $\mathcal{G}_R(X_R) \approx \mathcal{G}_S(G_{R \rightarrow S}(X_R))$, we introduce following consistency loss:

$$L_c = \|\mathcal{G}_R(X_R) - \mathcal{G}_S(G_{R \rightarrow S}(X_R))\|_1. \quad (15)$$

Overall Loss Function. The overall loss function are defined as follow:

$$L = L_{tran} + \lambda_m(L_{rm} + L_{sm}) + \lambda_d(L_{rd} + L_{sd}) + \lambda_t(L_{rt} + L_{st}) + \lambda_c L_c. \quad (16)$$

where λ_m , λ_d , λ_t and λ_c are trade-off weights.

4. Experimental Results

In this part, we first present the implementation details of our framework. Then, we evaluate our domain adaptation method on synthetic datasets and the real images, respectively. Finally, ablation studies are conducted to analyze the proposed approach.

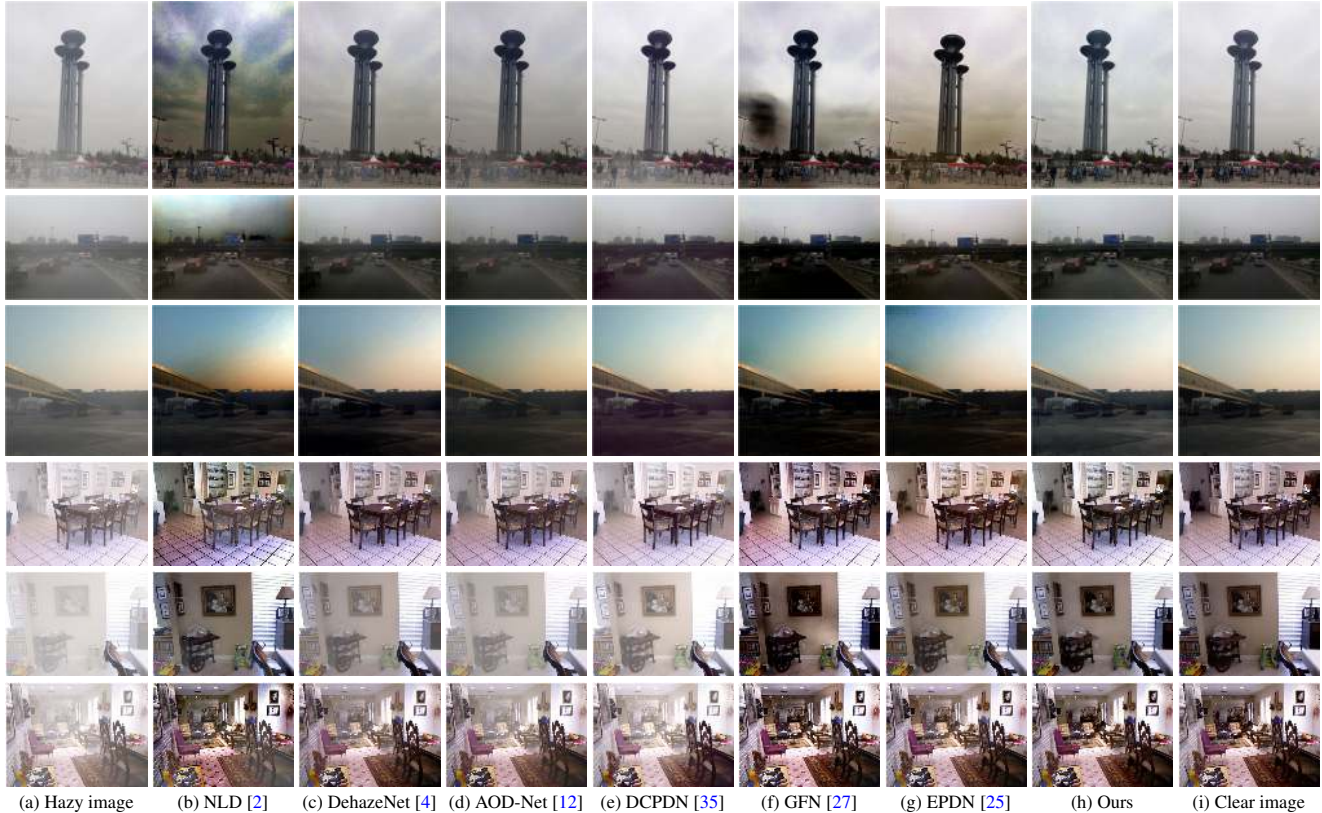


Figure 5. Visual comparisons on the SOTS [13] dataset.

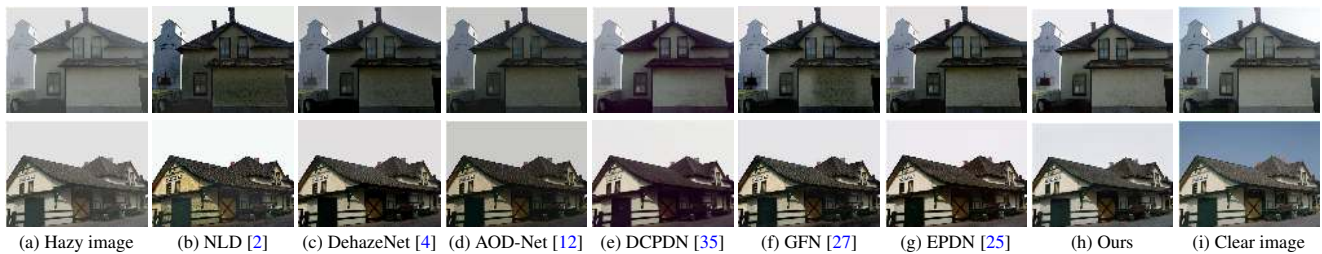


Figure 6. Visual comparisons on the HazeRD [36] dataset.

4.1. Implementation Details

Datasets. We randomly choose both synthetic and real-world hazy images from the RESIDE dataset [13] for training. The dataset is divided into five subsets, namely, ITS (Indoor Training Set), OTS (Outdoor Training Set), SOTS (Synthetic Object Testing Set), URHI (Unannotated real Hazy Images), and RTTS (real Task-driven Testing Set). For synthetic dataset, we choose 6000 synthetic hazy images for training, 3000 from the ITS and 3000 from the OTS. For real-world hazy images, we train the network by randomly selecting 1000 real hazy images from the URHI. In the training phase, we randomly crop all the images to 256×256 and normalize the pixel values to $[-1, 1]$.

Training details. We implement our framework in PyTorch [24] and utilize ADAM [11] optimizer with a batch

size 2 to train the network. First, we train the image translation network $G_{S \rightarrow R}$ and $G_{R \rightarrow S}$ for 90 epochs with the momentum $\beta_1 = 0.5$, $\beta_2 = 0.999$, and the learning rate is set as 5×10^{-5} . Then we train \mathcal{G}_R on $\{X_R, G_{S \rightarrow R}(X_s, D_s)\}$, and \mathcal{G}_S on $\{X_S, G_{R \rightarrow S}(X_R)\}$ for 90 epochs using the pre-trained $G_{S \rightarrow R}$ and $G_{R \rightarrow S}$ models. The momentum and the learning rate are set as: $\beta_1 = 0.95$, $\beta_2 = 0.999$, $lr = 10^{-4}$. Finally, we fine tune the whole network using the above pre-trained models. When computing the DC loss, we set the patch as 35×35 . The trade-off weights are set as: $\lambda_{tran} = 1$, $\lambda_m = 10$, $\lambda_d = 10^{-2}$, $\lambda_t = 10^{-3}$ and $\lambda_c = 10^{-1}$.

Comparison methods. We evaluate the proposed method against the following state-of-the-art approaches: DCP [9], MSCNN [26], DehazeNet [4], NLD [2], AOD-Net [12],

Table 2. Quantitative comparison (Average PSNR/SSIM) of the dehazing results on two synthetic datasets.

	DCP [9]	NLD [2]	MSCNN [26]	DehazeNet [4]	AOD-Net [12]	DCPDN [35]	GFN [27]	EPDN [25]	Ours
SOTS	15.49/0.64	17.27/0.75	17.57/0.81	21.14/0.85	19.06/0.85	19.39/0.65	22.30/0.88	23.82/0.89	27.76/0.93
HazeRD	14.01/0.39	16.16/0.58	15.57/0.42	15.54/0.41	15.63/0.45	16.12/0.34	13.98/0.37	17.37/0.56	18.07/0.63

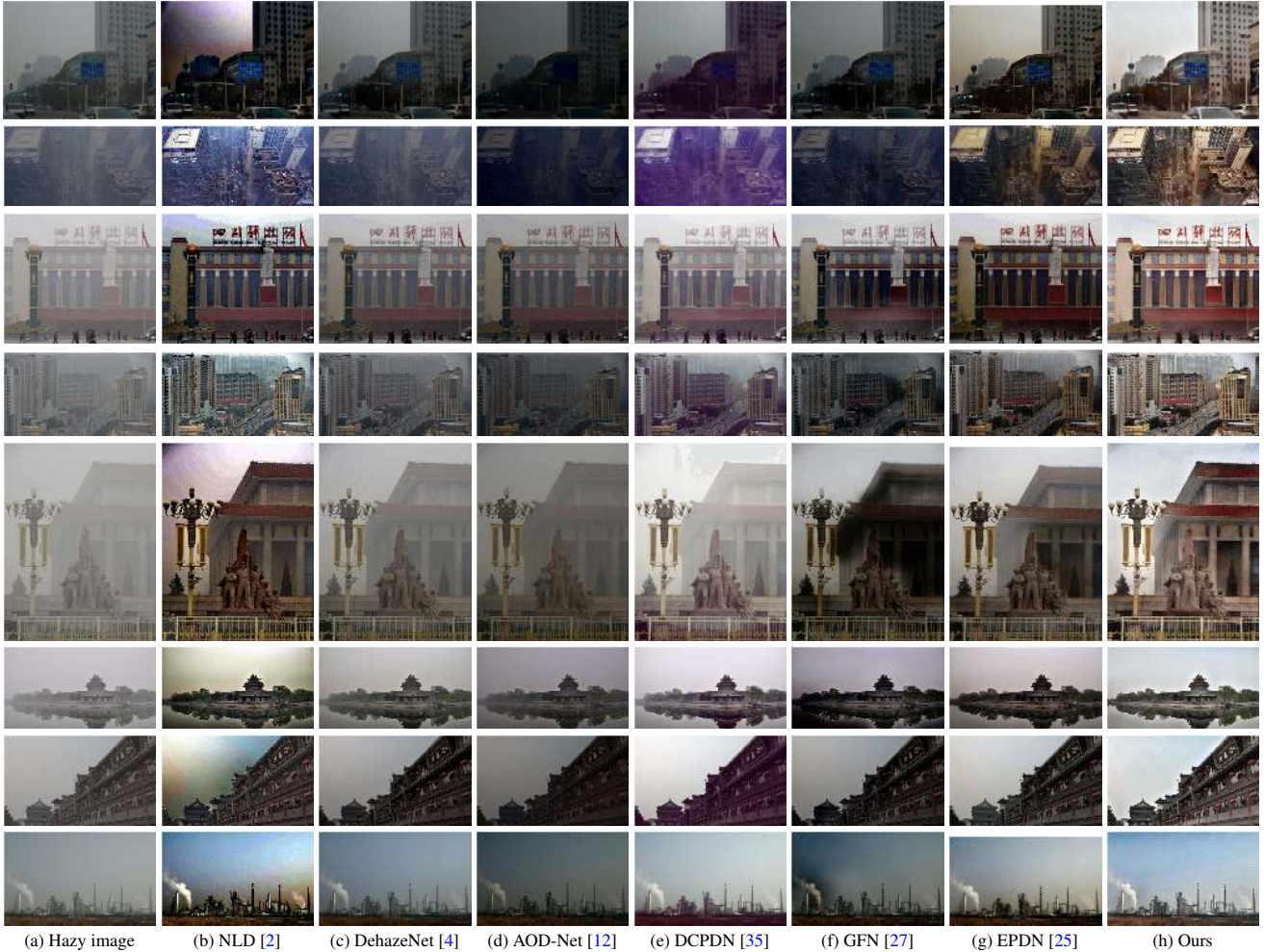


Figure 7. Visual comparisons on the real hazy images.

GFN [27], DCPDN [35], and EPDN [25]. More image dehazing results and comparisons against other dehazing approaches are included in the supplementary materials.

4.2. Experiments on Synthetic Datasets

We use two synthetic datasets, namely SOTS [13] and HazeRD [36], to evaluate the performance of our proposed method. The dehazed images of different methods on these two datasets are shown in Figure 5 and 6. From Figure 5 (b), we can observe that both NLD [2] and GFN [27] suffer from some color distortion where the result looks unrealistic. The dehazing results by EPDN [25] are also darker than the ground-truths in some cases, as shown in Figure 5 (g). Besides, there is still some remaining haze in the dehazed

images by DehazeNet [4], AOD-Net [12] and DCPDN [35]. Compared with these methods, our algorithm restores images with sharper structures and details, which are closer to the ground-truths. Similar results can be found in the dehazed results on the HazeRD dataset in Figure 6, our algorithm generates results with better visual effect.

We also give the quantitative comparison of dehazed results in Table 2. As shown, the proposed method obtains the highest PSNR and SSIM values on both datasets. Compared with the state-of-the-art EPDN [25], our approach achieves the gain with 3.94 dB and 0.04 in terms of PSNR and SSIM on the STOS dataset, respectively. For the HazeRD dataset, the PSNR and SSIM produced by our method are higher than EPDN [25] by up to 0.7dB and 0.07, respectively.

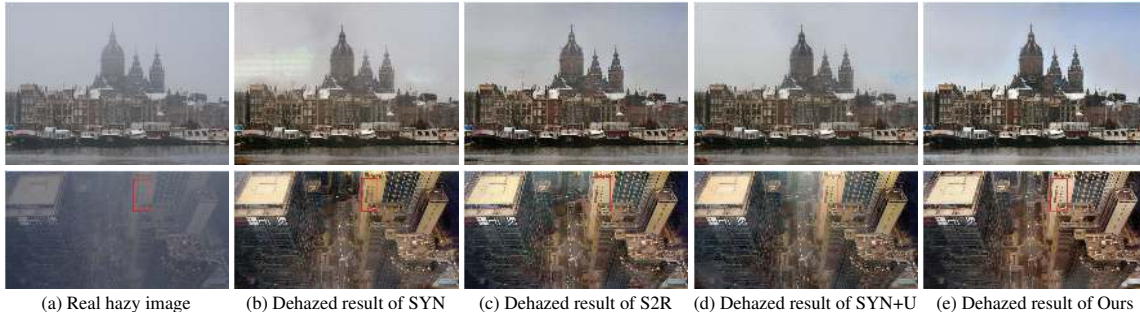


Figure 8. Comparison of dehazed results of several dehazing models on real hazy images.

4.3. Experiments on Real Images

To evaluate the generalization of our method on real images, we compare the visual results of different approaches on real hazy images from the URHI dataset. As shown in Figure 7, NLD [2] suffers from serious color distortions (see, e.g., the sky in Figure 7 (b)). From Fig 7 (f), we can figure out that the GFN [27] also suffers from color distortion in some cases, and the dehazed results look darker than our method. In addition, the dehazed results by DehazeNet [4], AOD-Net [12], and DCPDN [35] have some remaining haze artifacts as shown in the fifth row in Figure 7 (c-e)). Though EPDN [25] has achieved a better visual effect than the above methods, the brightness of the dehazing results is lower than our method in general. Overall, our proposed method restores more details and obtains visually pleasing images.

4.4. Ablation Study

To verify the effectiveness of the image translation network and the unsupervised loss, we conduct a series of ablations to analyze our method. We construct the following dehazing models for comparison: 1) **SYN**: \mathcal{G}_S is only trained on X_S ; 2) **SYN+U**: \mathcal{G}_S is trained on both X_S and X_R ; 3) **R2S+U**: \mathcal{G}_S is only trained on X_S and $G_{R \rightarrow S}(X_R)$; 4) **S2R**: \mathcal{G}_R is trained on $G_{S \rightarrow R}(X_S, D_s)$.

We compare the proposed domain adaptation method against these four dehazing models on both synthetic and real hazy images. The visual and quantitative results are shown in Table 3 and Figure 8, which demonstrate that our approach achieves the best performance of image dehazing in terms of PSNR and SSIM as well as visual effects. As shown in Figure 8 (b), due to the domain shift, the **SYN** method causes color distortion or darker artifacts (see, e.g., the sky part and the red rectangle). In contrast, the dehazing model \mathcal{G}_R trained on the translated images (**S2R**) achieves better quality image as shown in Figure 8 (c), which demonstrates that the translators effectively reduces the discrepancy between synthetic data and real images. Moreover, Figure 8 (b) and (d) show that the dehazing model with unsupervised loss (**SYN+U**) can produce better

Table 3. Quantitative results of different dehazing models on synthetic domain.

Method	PSNR	SSIM
SYN	25.67	0.8801
SYN+U	25.75	0.8699
R2S+U	25.91	0.8822
Ours	27.76	0.9284

results than **SYN**, which demonstrate the effectiveness of the unsupervised loss. Finally, we can observe that the proposed method with both translators and unsupervised loss generates cleaner and visually more pleasing results (e.g., the sky is brighter) in Figure 8 (e). The quantitative results in Table 3 by applying image translation and unsupervised loss also agree with the qualitative results in Figure 8.

As a conclusion, these ablations demonstrate that the image translation model and unsupervised loss are useful to reduce the domain gap between synthetic data and real-world images and improve the performance of image dehazing on both synthetic and real domains.

5. Conclusions

In this work, we propose a novel domain adaptation framework for single image dehazing, which contains an image translation module and two image dehazing modules. We first use the image translation network to translate images from one domain to another to reduce the domain discrepancy. And then, the image dehazing networks take the translated images and their original images as inputs to perform image dehazing. To further improve the generalization, we incorporate the real hazy image into the dehazing training by exploiting the properties of clean images. Extensive experimental results on both synthetic datasets and real-world images demonstrate that our algorithm performs favorably against state-of-the-arts.

Acknowledgements. This work is supported by the Project of the National Natural Science Foundation of China No.61433007 and No.61901184. W. Ren is supported in part by the Zhejiang Labs International Talent Fund for Young Professionals and the CCF-Tencent Open Fund.

References

- [1] Amir Atapour-Abarghouei and Toby P Breckon. Real-time monocular depth estimation using synthetic data with domain adaptation via image style transfer. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2018. 3
- [2] Dana Berman, Shai Avidan, et al. Non-local image dehazing. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2016. 2, 6, 7, 8
- [3] Konstantinos Bousmalis, Nathan Silberman, David Dohan, Dumitru Erhan, and Dilip Krishnan. Unsupervised pixel-level domain adaptation with generative adversarial networks. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2017. 3
- [4] Bolun Cai, Xiangmin Xu, Kui Jia, Qing Chunmei, and Dacheng Tao. Dehazenet: An end-to-end system for single image haze removal. *IEEE Transactions on Image Processing*, 2016. 1, 2, 6, 7, 8
- [5] Yuhua Chen, Wen Li, Xiaoran Chen, and Luc Van Gool. Learning semantic segmentation from synthetic data: A geometrically guided input-output adaptation approach. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2019. 3
- [6] Yuhua Chen, Wen Li, Christos Sakaridis, Dengxin Dai, and Luc Van Gool. Domain adaptive faster r-cnn for object detection in the wild. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2018. 3
- [7] Aysegul Dundar, Ming-Yu Liu, Ting-Chun Wang, John Zedlewski, and Jan Kautz. Domain stylization: A strong, simple baseline for synthetic to real image domain adaptation. *arXiv preprint arXiv:1807.09384*, 2018. 3
- [8] Raanan Fattal. Dehazing using color-lines. *ACM Transactions on Graphics*, 2014. 1, 2
- [9] Kaiming He, Jian Sun, and Xiaoou Tang. Single image haze removal using dark channel prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011. 1, 2, 5, 6, 7
- [10] Judy Hoffman, Eric Tzeng, Taesung Park, Jun-Yan Zhu, Phillip Isola, Kate Saenko, Alexei A Efros, and Trevor Darrell. Cycada: Cycle-consistent adversarial domain adaptation. In *International Conference on Machine Learning*, 2018. 3
- [11] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv*, 2014. 6
- [12] Boyi Li, Xiulian Peng, Zhangyang Wang, Jizheng Xu, and Dan Feng. Aod-net: All-in-one dehazing network. In *IEEE International Conference on Computer Vision*, 2017. 1, 2, 6, 7, 8
- [13] 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*, 2019. 6, 7
- [14] 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. 2
- [15] Lerenhan Li, Jinshan Pan, Wei-Sheng Lai, Changxin Gao, Nong Sang, and Ming-Hsuan Yang. Dynamic scene deblurring by depth guided model. *IEEE Transactions on Image Processing*, 2020. 4
- [16] Runde Li, Jinshan Pan, Zechao Li, and Jinhui Tang. Single image dehazing via conditional generative adversarial network. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2018. 1, 2
- [17] Yu Li, Robby T Tan, and Michael S Brown. Nighttime haze removal with glow and multiple light colors. In *IEEE International Conference on Computer Vision*, 2015. 2
- [18] Yu Li, Shaodi You, Michael S Brown, and Robby T Tan. Haze visibility enhancement: A survey and quantitative benchmarking. *Computer Vision and Image Understanding*, 165:1–16, 2017. 1
- [19] Mingsheng Long, Yue Cao, Jianmin Wang, and Michael I Jordan. Learning transferable features with deep adaptation networks. In *International Conference on Machine Learning*, 2015. 3
- [20] Mingsheng Long, Guiguang Ding, Jianmin Wang, Jianguang Sun, Yuchen Guo, and Philip S Yu. Transfer sparse coding for robust image representation. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2013. 3
- [21] Earl J McCartney. Optics of the atmosphere: scattering by molecules and particles. *New York, John Wiley and Sons, Inc., 1976. 421 p.*, 1976. 1
- [22] Gaofeng Meng, Ying Wang, Jianguo Duan, Shiming Xiang, and Chunhong Pan. Efficient image dehazing with boundary constraint and contextual regularization. In *IEEE International Conference on Computer Vision*, 2013. 2
- [23] Srinivasa G Narasimhan and Shree K Nayar. Vision and the atmosphere. *International Journal of Computer Vision*, 48(3):233–254, 2002. 1
- [24] Ko Nishino, Louis Kratz, and Stephen Lombardi. Bayesian defogging. *International Journal of Computer Vision*, 2012. 2, 6
- [25] Yanyun Qu, Yizi Chen, Jingying Huang, and Yuan Xie. Enhanced pix2pix dehazing network. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2019. 1, 2, 6, 7, 8
- [26] Wenqi Ren, Si Liu, Hua Zhang, Jinshan Pan, Xiaochun Cao, and Ming-Hsuan Yang. Single image dehazing via multi-scale convolutional neural networks. In *European Conference on Computer Vision*, 2016. 1, 2, 6, 7
- [27] Wenqi Ren, Lin Ma, Jiawei Zhang, Jinshan Pan, Xiaochun Cao, Wei Liu, and Ming-Hsuan Yang. Gated fusion network for single image dehazing. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2018. 1, 2, 6, 7, 8
- [28] Ashish Shrivastava, Tomas Pfister, Oncel Tuzel, Joshua Susskind, Wenda Wang, and Russell Webb. Learning from simulated and unsupervised images through adversarial training. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2017. 3
- [29] Robby T. Tan. Visibility in bad weather from a single image. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2008. 2
- [30] Jean-Philippe Tarel and Nicolas Hautiere. Fast visibility restoration from a single color or gray level image. In *IEEE International Conference on Computer Vision*. 2

- [31] Yi-Hsuan Tsai, Wei-Chih Hung, Samuel Schulter, Kihyuk Sohn, Ming-Hsuan Yang, and Manmohan Chandraker. Learning to adapt structured output space for semantic segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2018. [3](#)
- [32] Eric Tzeng, Judy Hoffman, Kate Saenko, and Trevor Darrell. Adversarial discriminative domain adaptation. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2017. [3](#)
- [33] Xintao Wang, Ke Yu, Chao Dong, and Chen Change Loy. Recovering realistic texture in image super-resolution by deep spatial feature transform. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2018. [4](#)
- [34] Dong Yang and Jian Sun. Proximal dehaze-net: A prior learning-based deep network for single image dehazing. In *European Conference on Computer Vision*, 2018. [2](#)
- [35] He Zhang and Vishal M. Patel. Densely connected pyramid dehazing network. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2018. [1, 2, 6, 7, 8](#)
- [36] Yanfu Zhang, Li Ding, and Gaurav Sharma. Hazerd: an outdoor scene dataset and benchmark for single image dehazing. In *IEEE International Conference on Image Processing*, 2017. [6, 7](#)
- [37] Chuanxia Zheng, Tat-Jen Cham, and Jianfei Cai. T2net: Synthetic-to-realistic translation for solving single-image depth estimation tasks. In *European Conference on Computer Vision*, 2018. [3, 4](#)
- [38] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *IEEE International Conference on Computer Vision*, 2017. [3, 4, 5](#)
- [39] Qingsong Zhu, Jiaming Mai, Ling Shao, et al. A fast single image haze removal algorithm using color attenuation prior. *IEEE Transactions on Image Processing*, 2015. [2](#)