

Semi-Supervised Domain Alignment Learning for Single Image Dehazing

Yu Dong^{ID}, Yunan Li^{ID}, Member, IEEE, Qian Dong, He Zhang^{ID}, and Shifeng Chen^{ID}

Abstract—Convolutional neural networks (CNNs) have attracted much research attention and achieved great improvements in single-image dehazing. However, previous learning-based dehazing methods are mainly trained on synthetic data, which greatly degrades their generalization capability on natural hazy images. To address this issue, this article proposes a semi-supervised learning approach for single-image dehazing, where both synthetic and realistic images are leveraged during training. Considering the situation that it is hard to obtain the realistic pairs of hazy and haze-free images, how to utilize the realistic data is not a trivial work. In this article, a domain alignment module is introduced to narrow the distribution distance between synthetic data and realistic hazy images in a latent feature space. Meanwhile, a haze-aware attention module is designed to describe haze densities of different regions in the image, thus adaptively responds for different hazy areas. Furthermore, the dark channel prior is introduced to the framework to improve the quality of the unsupervised learning results by considering the statistical characters of haze-free images. Such a semi-supervised design can significantly address the domain shift issue between the synthetic and realistic data, and improve generalization performance in the real world. Experiments indicate that the proposed method obtains state-of-the-art performance on both public synthetic and realistic hazy images with better visual results.

Index Terms—Dehazing, domain alignment, haze-aware attention, semi-supervised.

Manuscript received 19 May 2022; revised 15 July 2022 and 26 September 2022; accepted 3 November 2022. Date of publication 29 November 2022; date of current version 17 October 2023. This work was supported in part by the Key-Area Research and Development Program of Guangdong Province under Grant 2019B010155003; in part by the Shenzhen Science and Technology Innovation Commission under Grant JCYJ20200109114835623 and Grant JS20220831105002004; in part by the National Natural Science Foundation of China under Grant 62002271; and in part by the Fundamental Research Funds for the Central Universities under Grant XJS210310. This article was recommended by Associate Editor S. Senatore. (*Corresponding author: Shifeng Chen*.)

Yu Dong and Shifeng Chen are with the Shenzhen Key Laboratory of Computer Vision and Pattern Recognition, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China, and also with the Shenzhen Institute of Advanced Technology, University of Chinese Academy of Sciences, Beijing 101408, China (e-mail: yu.dong@siat.ac.cn; shifeng.chen@siat.ac.cn).

Yunan Li is with the School of Computer Science and Technology and the Xi'an Key Laboratory of Big Data and Intelligent Vision, Xidian University, Xi'an 710071, China (e-mail: yunanli@xidian.edu.cn).

Qian Dong is with the School of Mechanical Engineering and Automation, Harbin Institute of Technology (Shenzhen), Shenzhen 518055, China (e-mail: 190320418@stu.hit.edu.cn).

He Zhang is with Adobe Inc., San Jose, CA 95051 USA (e-mail: hezhan@adobe.com).

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/TCYB.2022.3221544>.

Digital Object Identifier 10.1109/TCYB.2022.3221544

I. INTRODUCTION

AS A COMMON atmospheric phenomenon, haze creates unwanted effects on imaging systems due to the scattering of atmospheric aerosol particles. However, for many high-level computer vision tasks such as object detection, image quality is critically important. There is a classical physical scattering model [3] widely used to describe the formation of a hazy image

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

where $I(x)$ is the observed hazy image, $J(x)$ is the real scene radiance to be recovered, $t(x)$ is the medium transmission map, x indicates the pixel coordinate, and A is the global atmospheric light. If the global atmosphere is homogenous, the transmission map $t(x)$ can be modeled by

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

where $d(x)$ represents the scene depth and β is the scattering coefficient of the atmosphere. Single image dehazing aims to elicit the haze-free scene $J(x)$ from a hazy input $I(x)$ without any other known parameters, which makes the task highly ill-posed.

Existing methods [1], [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], [16], [17], [18], [19], [20], [21], [22] can be divided into two trends: 1) prior-based methods and 2) learning-based methods. In prior-based methods [10], [11], [12], [17], [23], [24], [25], A and $t(x)$ that mentioned in (1) are evaluated under some assumptions and priors. Although, in some cases, these methods can improve the contrast and saturation of hazy images, they fail to recover clear images and even lead to some artifacts, such as color distortion, over-enhancement, and halo effect, particularly when the assumptions and the priors are invalid.

The rapid development of deep neural-networks boosts learning-based dehazing methods [1], [4], [5], [6], [7], [8], [9], [18], [19], [20], [21], [22], [26], [27], [28]. Such data-driven methods considerably alleviate the dependence on handcrafted priors by establishing mapping relations between hazy images and corresponding haze-free ones, resulting in fewer artifacts and clearer restorations. In an earlier processes, most learning-based methods [4], [13], [14], [15] leveraged $t(x)$ and A to recover a haze-free image according to (1). In recent years, some methods attempted to remove haze directly from the hazy inputs end-to-end.

Although the existing learning-based methods perform well in dehazing, they still have limitations as follows.

- 1) *Inadequacies in Common Training Datasets:* The training datasets for current learning-based methods mostly have drawbacks more or less. For example, NYU dataset [29] that is used in the early dehazing networks [4], [6], [13], [14] contains only indoor scenes that cannot simulate real hazy scenes; Make3D dataset [30] lacks a great diversity of scenes; RESIDE dataset [31] has unrealistic hazy conditions due to inaccurate outdoor depth estimation.
- 2) *Limited Generalization on Real-World Domain:* First, the quality of synthetic haze always suffers from the limitation of accurate outdoor depth estimation and the idealized atmosphere β . Second, the statistical properties of the real haze are hardly introduced in training strategies. Therefore, though most learning-based methods can perform relatively well on synthesized data, they still fail to generalize well toward arbitrary real-world hazy images.

To circumvent the above issues, a semi-supervised domain alignment learning strategy, called SDA-GAN, is proposed for image dehazing that uses both synthetic and realistic data simultaneously. The training process is divided into two stages. In the first stage, the latent features of synthetic and realistic haze are extracted by a shared encoder, then those features are aligned via adversarial learning and Kullback–Leibler divergence. We denote this distribution-consistency strategy as domain alignment. Meanwhile, an attention mechanism called haze-aware attention module, is proposed for adaptively responding to haze regions in images. In the second stage, the synthetic-paired branch is used to learn robustly supervised mapping via labeled data, while the real-unpaired branch is used to learn the statistical properties between unlabeled hazy images and clean images with the help of the dark channel prior (DCP) [10]. This joint learning method takes full advantage of the two domains, and thus significantly addresses the domain-shift problem and promotes better generalization on natural images. Fig. 1 shows our dehazing result compared with several latest methods.

The main contributions of this work are summarized as follows.

- 1) A semi-supervised network for single image dehazing is proposed, which is simultaneously trained on synthesized data and real-world data in a supervised and unsupervised way, respectively. To encourage the model generalization, a domain alignment strategy is adapted to narrow the distance between the extracted latent features of the two domains rather than directly share the network weights. This strategy effectively alleviates the domain-shift problem, thus making the model generalize better.
- 2) Based on the local entropy theory, a haze-aware attention module is designed to perceive haze information of different regions in the image, which further contributes to the domain alignment between synthetic and real haze.
- 3) Extensive experiments prove that the integration of our designs can ultimately improve the performance of restoration both qualitatively and quantitatively.

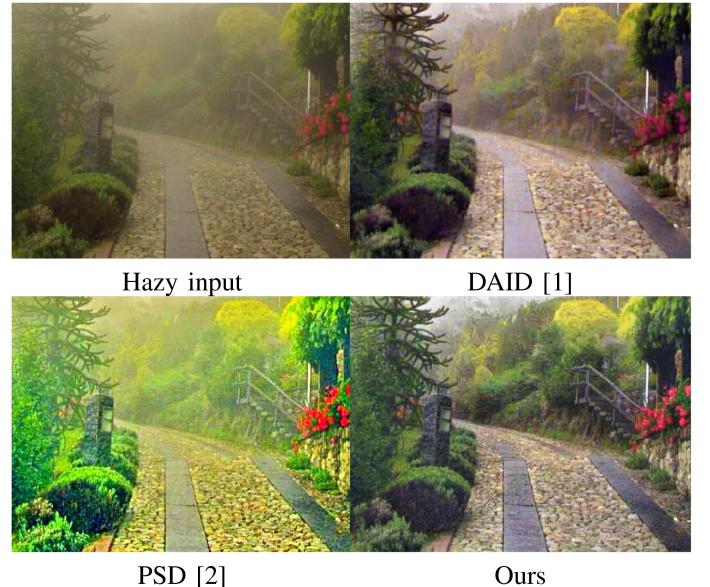


Fig. 1. Dehazing example on a real-world hazy image. Comparing with the latest state-of-the-art methods, our model can remove haze more thoroughly and generate more natural and visually pleasing results with less color distortion.

II. RELATED WORK

In recent decades, numerous dehazing methods have been proposed. These dehazing methods can be roughly divided into two categories: 1) prior-based methods and 2) learning-based methods. Besides, domain adaptation methods are related to our domain alignment strategy. In this section, we will give a brief review for them.

1) *Prior-Based Methods:* As mentioned in Section I, assumptions and priors are required to estimate A and $t(x)$ for further restoring the clear image. Based on the observation that clear images usually have higher contrast, Tan [23] proposed a patch-based contrast maximization method. Fattal [17] estimated the albedo of the scene, with the assumption that the transmission and surface shading are locally uncorrelated. He et al. [10] demonstrated a classical statistical observation that in majority of nonsky regions, at least one color channel has extremely low intensities in some pixels, and it is the famous DCP. Based on such a prior, they estimated the transmission map and recovered the haze-free image. In [25], Fattal brought up color-line prior based on the observation that small image patches typically exhibit 1-D distribution in the RGB space. Then, Zhu et al. [11] estimated the scene depth of the hazy images and removed the haze based on a color attenuation prior. Based on this prior, the transmission map and atmospheric light can be estimated. Choi et al. [24] described a prediction model of perceptual fog density and a perceptual image-defogging algorithm, both based on fog-aware statistical features.

However, these priors may not hold for some complex and varied scenes, and involve extraordinary computation as well as limited performance.

2) *Learning-Based Methods:* As the technique of deep neural networks develops fast in past years, learning-based

methods have emerged in many image restoration tasks [32], [33], [34], [35]. In the early time, most dehazing methods effectively generate clean results by supervised learning. Recently, more training strategies, such as semi-supervised, weakly supervised, and unsupervised learning methods are gradually raised.

Supervised Learning Methods dominate in current learning-based dehazing methods. This kind of method utilizes paired synthetic hazy and haze-free images to train convolutional neural networks (CNNs) for dehazing.

Ren et al. [14] first learned the mapping between hazy images and transmission maps using a coarse-to-fine strategy. Cai et al. [13] introduced a simple but effective network for estimating the transmission map and then restored the hazy image. Differently, Li et al. [6] packed up the transmission map and atmospheric light into a parameter and used an ensemble network to directly yield dehazed results. Zhang et al. [4] built two densely connected pyramid branches for estimating the transmission map and atmospheric light, respectively. Qu et al. [8] exploited the pix2pix network and enhanced it for dehazing. Li et al. [15] presented a progressive dehazing network, which restored regions with different haze level by different stages, and adaptively fused them to derive the final result. Liu et al. [19] proposed a GridDehazeNet that combined multilevel features for fine-scale results in dehazing. Dong et al. [5] developed a novel Fusion-discriminator that integrates the frequency information as additional priors and constraints into the learning procedure. Liu et al. [36] formulated the image dehazing problem as the minimization of a variational model based on the classical gradient descent method for estimating intermediate parameters in the physical scattering model. Dong et al. [7] proposed a multiscale network to incorporate the boosting strategy and the back projection technique neatly for image dehazing. In [1], a domain adaptation framework like CycleGAN [37] was proposed by Shao et al., which translates the paired synthetic hazy image to the real haze style, then still proceeded dehazing learning in a supervised way. Zhang et al. [20] utilized haze density information to guide clean image reconstruction and alleviates the constraint of the atmospheric model. Zhu et al. [21] proposed DehazeGAN for haze removal, which uses an adversarial composition network to learn discriminative physical parameters for clean image recovery. Zhang et al. [22] learned the semantic prior for a single image dehazing task, which models the dehazing problem as a maximizing the probability of color conditioned on the semantic information.

However, although these methods outperform most prior-based methods, there are still some limitations due to the domain shift problem.

Semi-Supervised Learning Methods are trained on both synthetic and real data. As the overfitting problem is inevitable for most supervised learning methods, the semi-supervised scheme is proposed to jointly learn from both synthetic and real data. This scheme is used in many low-level vision tasks, such as generating flow warp [38], recovering the clearness, and chroma of old photos [39], and image deraining [40].

Taking the advantage of semi-supervised learning, Li et al. [16] developed a set of network architectures for

dehazing. The network contained a supervised learning branch trained on the paired synthetic data and an unsupervised learning branch trained on unpaired real data using DCP. Recently, Chen et al. [2] proposed a synthetic-to-real dehazing framework pretrained on synthetic data, then leveraged several well-grounded physical priors for fine-tuning.

Among these semi-supervised methods, [16] is the closest one to ours. However, our method remarkably differs from that one mainly from two aspects. First, to alleviate the domain gap, we adopt a domain alignment scheme to close the distributions of two domains features, which are extracted into latent high-level space, instead of simply using two reconstruction branches separately trained with paired/unpaired data. Second, we use DCP as guidance in adversarial learning rather than the pixel-wise ground truth for training. Compared with Li et al. [16] used \mathcal{L}_1 regularization to minimize values in the dark channel map of the dehazed image to zero. Our method can circumvent the limitations on special cases, where DCP fails to apply, avoiding distortion in sky regions and over bright objects.

Weakly Supervised and Unsupervised Learning Methods rely less or even none on paired training data. In the dehazing task, most methods use the DCP and GANs [41] to address the problem of lacking paired training images. Yang et al. [42] introduced a disentanglement and reconstruction mechanism based on the physical model, which disentangles the hazy image into three hidden factors constrained by adversarial loss and regularization, then recovers the input using the physical model (1). Similarly, Li et al. [43] proposed such an idea of layer disentanglement by viewing a hazy image as the entanglement of several hidden factors. However, instead of using adversarial learning, the DCP prior and empirical regularization are involved in this method. Zhao et al. [44] merged the merits of prior-based and learning-based approaches introduced a two-stage weakly supervised dehazing network, which adopts the DCP to restore visibility at first, and then refines results via adversarial learning with unpaired hazy and clean images. Golts et al. [45] presented a completely unsupervised method of training via minimization of the DCP energy function, which can be regarded as a fast and simple feed-forward approximator of the DCP by stopping the optimization early.

Although these methods achieve particularly effective performance in dehazing, due to lacking enough reliable priors and constraints, there may be some imperfect cases like artifacts, color distortion, and low resolution in their results.

3) *Domain Adaptation and Generalization*: Most existing methods trained under the independent identically distributed assumption of training datasets suffer from the bad generalization capability of testing on unseen datasets due to the interdomain gaps. Domain adaptation tasks [46], [47], [48], [49], [50] aim to effectively learn domain generic feature representations, on which the distribution discrepancy between source and target domains can be reduced. Existing methods can be roughly classified into two types: 1) feature-level or 2) pixel-level adaptation. Feature-level adaptation methods aim to align the feature distributions between the source and target domains either by minimizing the maximum mean discrepancy [51], or applying adversarial learning (GANs) [41]

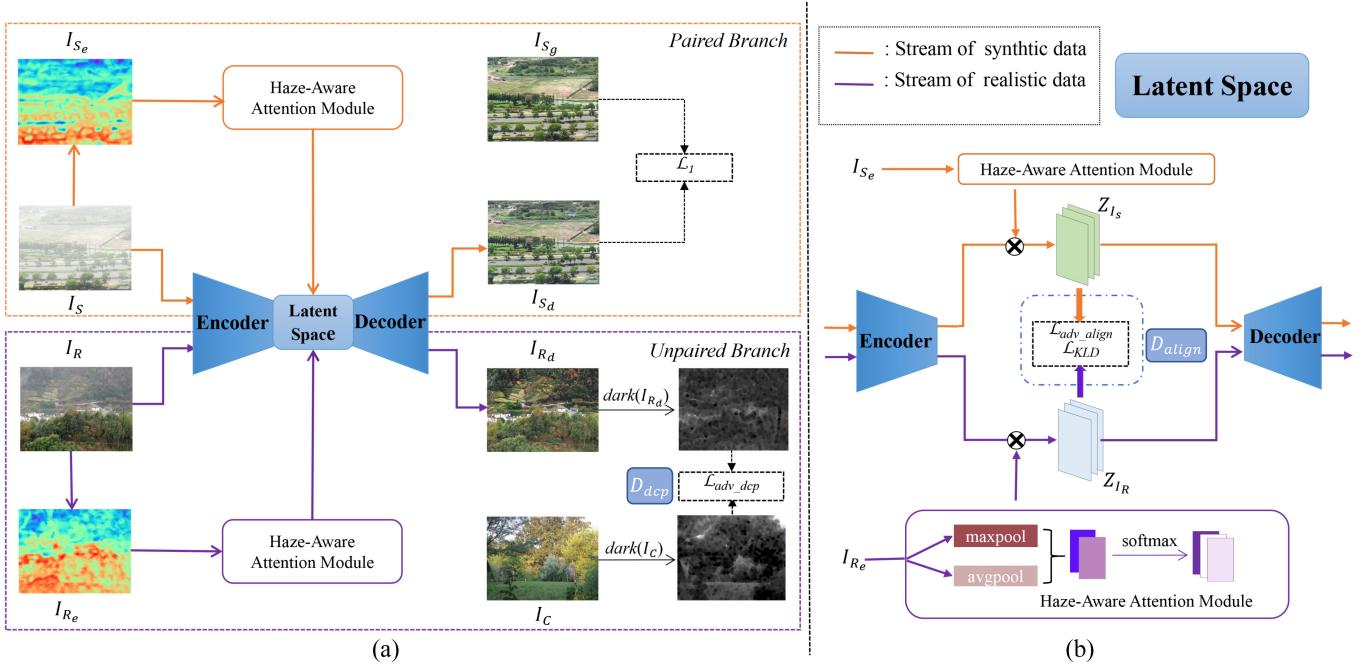


Fig. 2. Architecture of the proposed SDA-GAN. (a) Flow chart of the training method. Our design aims to alleviate the domain gap by constraining the high-level haze features in the latent space from two types of training data with the adversarial learning and divergence regularization. (b) Detailed operations in the latent space, including the domain alignment strategy and haze-aware attention module, where \otimes , D_{dcp} , and D_{align} denote the element-wise production operation and two discriminators, respectively. I_S and I_R represent the synthetic image and real-world image. Z_{I_S} and Z_{I_R} refer to high-level feature of I_S and I_R in the latent space, respectively. I_{R_e} and I_{S_e} are local entropy maps of I_R and I_S . And I_{S_g} , I_{S_d} , I_{R_d} , and I_C denote the ground truth and dehazed result of I_S , dehazed result of I_R , and the unpaired clean image, respectively.

strategies to confuse the discriminator so as to ensure that the feature distributions between two domains are similar [52], [53], [54]. Pixel-level adaptation [55], [56], [57] methods regard domain-shift problems as style transfer learning [10] or image-to-image transfer [55], [57] for augmenting the data in the target domain. In this article, to circumvent poor generalization of the well-trained model in the real world, we model the restoration of the unlabeled natural hazy images as a feature-level adaptation task with the labeled synthetic images and introduce adversarial learning and Kullback–Leibler divergence techniques to align the domain of synthetic and real data.

III. PROPOSED METHODS

In this section, the proposed SDA-GAN network is introduced, in which not only supervised synthesized hazy images but also unsupervised real hazy images are applied into the network training process for better generalization. We first describe the pipeline of our network and the design of the proposed domain alignment learning strategy. After that, the structure of the haze-aware attention module is introduced. The detail of the training process and loss functions will be illustrated at last.

A. Overview

In our task, the synthetic haze and real haze domain can be regarded as two different probability distributions and cause different effects on the image scenes. These differences represent different semantic information for learning-based tasks

and further significantly degrade the practical performance of models trained on synthetic images. Therefore, for better model generalization, an image alignment learning strategy is proposed to reduce the interdomain distance for closer semantic similarity.

The structure of the proposed network is illustrated in Fig. 2. The detailed configurations of each subnetwork are shown in Appendix A. The network has two branches of inputs, which are paired synthetic data I_S and unpaired real-world data I_R , respectively. The latent high-level features of these two branches are extracted by a shared encoder, and then aligned via adversarial learning and KL-divergence regularization for distribution consistency. Meanwhile, for adaptively responding to haze regions in images, a haze-aware attention module is designed according to the theory of local entropy. Next, the latent features are mapped into restored outputs via the shared decoder, and optimized in the supervised and unsupervised way, respectively.

B. Alignment Between Two Domains for Generalization

As in Fig. 2(a), the proposed model takes a real hazy image I_R and a synthetic image I_S as inputs, and extracts the high-level features, namely, Z_{I_R} and Z_{I_S} via a shared encoder. However, the shared encoder is not sufficient to make full use of the two kinds of data. For semantic similarity, the domain alignment of Z_{I_R} and Z_{I_S} is modeled as a closing process on the distribution probability of the two features. Specifically, a discriminator D_{align} in Fig. 3(b) is trained to distinguish Z_{I_R}

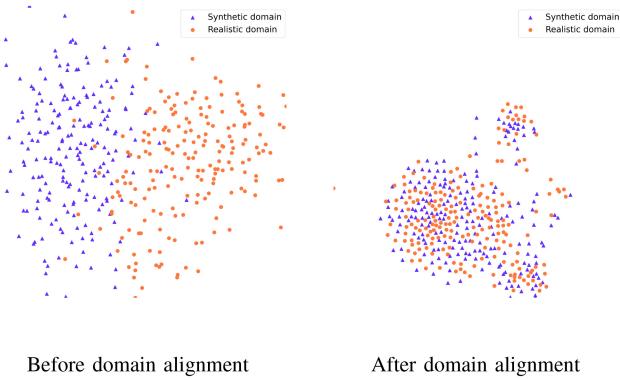


Fig. 3. Visualization of t -SNE distribution on the synthetic and realistic hazy datasets for the domain alignment strategy.

and Z_{I_S} and the adversarial loss can be formulated as follows:

$$\mathcal{L}_{\text{adv_align}} = \log(1 - D_{\text{align}}(Z_{I_S})) + \log D_{\text{align}}(Z_{I_R}). \quad (3)$$

Meanwhile, as Z_{I_R} and Z_{I_S} should be semantically similar, to further align them, the Kullback–Leibler divergence (KLD) loss is used to keep distribution consistency. The definition of the KLD loss is

$$\mathcal{L}_{\text{KLD}} = \sum_{i=0}^N [Z_{I_R} \log Z_{I_R} - Z_{I_R} \log Z_{I_S}] \quad (4)$$

where N indicates the number of samples in one mini-batch and log means the log-probabilities taken to base 2. In Fig. 3, the two domain distributions in the latent space before and after the proposed alignment are visualized. It can be observed that the distributions of synthetic and realistic data get closer, well meeting the agreement with our motivation.

C. Haze-Aware Attention Module for Haze Density

To facilitate haze features capture, the haze-aware attention module is introduced in feature representation based on the information entropy. For an image, the entropy values are positively correlated with the clearness. Accordingly, the local entropy can be further used to reflect the distribution of the haze densities of images [15], [24]. As illustrated in Fig. 4, with the density of haze increasing, the clearness of the image is significantly decreased, and the entropy of corresponding regions gets lower. This property of the local entropy is utilized into a haze-aware attention module. Therefore, the shared encoder is motivated by two local entropy maps I_{S_e} and I_{R_e} to extract the haze features, rather than different content information between synthetic and real images, which further contributes the effectiveness of domain alignment as verified in Section IV-D. The detail of the haze-aware attention module is shown in Fig. 2(b). Inspired by [58], a simple but effective subnetwork is designed to generate the attention feature from the local entropy map, and this subnetwork is called as haze-aware attention module. As the local entropy maps I_{S_e} and I_{R_e} are obtained for the synthetic- and real-branch, respectively, max-pooling and average-pooling are used to obtain the local details of heavy haze regions and the global haze distribution information, respectively. They encode the information

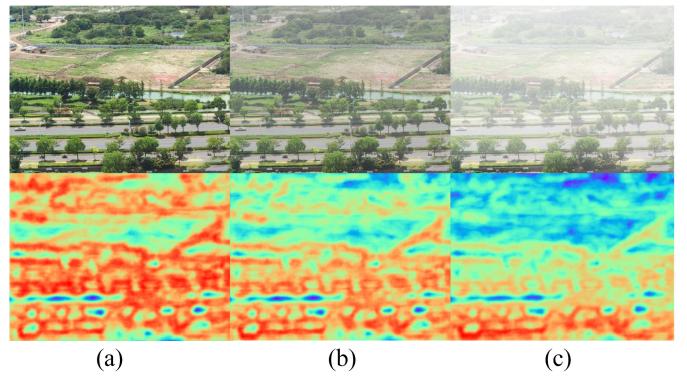


Fig. 4. From the top to the bottom: Hazy images with various haze density and their corresponding entropy information. The variations of the image entropy are relative to the haze density. Specifically, entropy values are negatively correlated with the increase of the haze density. In this illustration, the entropy map shows the regional haze density and it is colored from red to blue indicating the density from high to low. The denser the haze is, the lower (more blue) the entropy map is. (Best viewed in color.) (a) Cleaner image. (b) Mild haze. (c) Heavy haze.

of haze densities and thus help the network to handle areas with different haze density. Then two feature maps are normalized via softmax and concatenated as the attention map. The haze-aware attention module further contributes to the domain alignment as well as the model performance.

D. Restoration and Joint Optimization

As the features of synthetic and realistic branch are aligned with the guidance of the haze-aware attention module in the latent space, the mapping from labeled hazy images to corresponding haze-free ones can be much closer to learn in the real world. Then, the synthetic and realistic restoration results can be obtained via the decoding process. Similar to the encoder, a shared decoder is used for above two branches. For restoration in two branches of I_{S_d} and I_{R_d} , the supervised and unsupervised learning strategies are adopted, respectively.

For the Synthetic Branch, as the haze-free ground truth I_{S_g} is given, content loss \mathcal{L}_1 is used for optimizing. The content loss is defined as the Manhattan Distance $\|\cdot\|_1$ between the prediction I_{S_d} and ground truth I_{S_g}

$$\mathcal{L}_1 = \sum_{i=1}^N \|I_{S_d} - I_{S_g}\|_1. \quad (5)$$

For the Realistic Branch, as there is no haze-free ground truth for supervising, the classical prior, DCP is introduced in adversarial learning.

According to the DCP prior, the dark channel of a given image I_C can be calculated as follows:

$$\text{dark}(I_C(x)) = \min_{y \in N(x)} \left[\min_{c \in r, g, b} I_C^c(y) \right] \quad (6)$$

where I_C is from the haze-free part of the paired training data, x and y are pixel coordinates, I^c denotes the c th color channel, and $N(x)$ is an image patch centered at x . In Fig. 2, $\text{dark}(I_{R_d})$ and $\text{dark}(I_C)$ are denoted as the dark channel of the generated output and the unpaired haze-free image, respectively.

As mentioned in Section II, using DCP as a pixel-wise regularization to minimize values in the dark channel map of the dehazed image to zero [1], [16], [43] may lead to color-shift in white and bright areas, because these areas do not follow the DCP prior, in other words, the dark channels of these areas are not close to zero. Instead, we use D_{dcp} to capture the statistical properties of natural haze-free images, which can circumvent the limitations on special cases and avoid distortion in sky regions and over bright objects. Therefore, an adversarial learning procedure between unpaired haze-free real images and generated output images are established, where the proposed SDA-GAN network serves as a generator to generate more natural haze-free outputs.

A discriminator D_{dcp} is trained to distinguish dark(I_{R_d}) and dark(I_C). Thus, the loss function can be expressed as follows:

$$\begin{aligned} \mathcal{L}_{\text{adv_dcp}} = & \log(1 - D_{\text{dcp}}(\text{dark}(I_{R_d}))) \\ & + \log D_{\text{dcp}}(\text{dark}(I_C)). \end{aligned} \quad (7)$$

Overall Loss: By using the above dehazing scheme, the encoder E , decoder D , discriminator D_{align} , and D_{dcp} can be jointly optimized with an overall loss as follows:

$$\begin{aligned} \min_{E, D} \max_{D_{\text{align}}, D_{\text{dcp}}} (E, D, D_{\text{align}}, D_{\text{dcp}}) = & \mathcal{L}_1 + \lambda_1 \mathcal{L}_{\text{adv_align}} \\ & + \lambda_2 \mathcal{L}_{\text{KLD}} + \lambda_3 \mathcal{L}_{\text{adv_dcp}} \end{aligned} \quad (8)$$

where λ_1 , λ_2 , and λ_3 are weights to control the importance of different loss terms.

IV. EXPERIMENTS

In this section, the detailed experimental settings are introduced, and then the qualitative and quantitative comparisons on synthetic and real-world datasets with eight state-of-the-art methods are exhibited. After that, ablation studies are conducted to analyze the effect of different modules, including semi-supervised learning, domain alignment, haze-aware attention strategy, and DCP-based adversarial learning modules.

A. Experimental Settings

1) *Training Dataset:* For supervised training, a high-quality paired dataset is desired, which includes a sufficient amount of paired clear and hazy images of multiple scenes, preferably with depth information. Therefore, we randomly select 8000 haze-free images derived from the Microsoft COCO dataset [59] (4000), NYU dataset (2000), and OTS of RESIDE [31] (2000), respectively, to construct a new synthetic dataset. As depth information is contained in the NYU dataset, and synthetic haze is contained in OTS, we only estimate the scene depth of the Microsoft COCO dataset according to the method in [60]. Then similar ways in [14] are adopted to synthesize hazy images for Microsoft COCO and NYU. The atmospheric light A is randomly sampled in $[0.5, 1]$, and the scattering coefficient β is randomly set within $[1.2, 2.0]$. For unsupervised training, a real-world hazy dataset consisting of 8000 images (augmented from 1044 realistic hazy images) is collected from Internet and the LIVE Image Defogging

TABLE I
QUANTITATIVE COMPARISONS WITH OTHER STATE-OF-THE-ART METHODS ON PUBLIC SYNTHETIC TESTING DATASETS. IT CAN BE SEEN THAT OUR METHOD REACHES THE HIGHEST VALUES IN BOTH PSNR AND SSIM BY A LARGE MARGIN

Method	Training Type	RESIDE-SOTS PSNR/SSIM	HazeRD PSNR/SSIM
DCP [10]	Hand-crafted	18.34 / 0.868	14.38 / 0.746
DCPDN [4]	Supervised	17.97 / 0.856	15.83 / 0.769
LAP [15]	Supervised	20.54 / 0.862	15.63 / 0.812
DAID [1]	Supervised	26.37 / 0.937	16.96 / 0.809
ZID [43]	Unsupervised	14.84 / 0.687	13.72 / 0.639
RefineD [44]	Weakly	20.59 / 0.873	17.39 / 0.806
SSID [16]	Semi	24.04 / 0.913	16.30 / 0.808
PSD [2]	Semi	14.39 / 0.726	13.56 / 0.722
Ours	Semi	27.44 / 0.944	17.88 / 0.834

Database [24], in which images are mostly with low resolution, overcomplicated degeneration, and too heavy smoke.

2) *Implementation Details:* In our experiments, the loss weights λ_1 , λ_2 , and λ_3 are set to 0.1, 10, and 0.05, respectively. Our model is implemented with the Pytorch platform on an Nvidia RTX 2080Ti GPU. In the training phase, the size of the input image is fixed to $256 \times 320 \times 3$. The initial learning rate is set to 2×10^{-3} . The Adam optimizer is used with the parameters β_1 and β_2 set to 0.5 and 0.99, respectively. The training process is stopped after 50 epochs.

B. Comparison With State-of-the-Art Methods

To evaluate the performance of our method, we compared it with eight state-of-the-art methods, which are representative prior-based methods (DCP [10]), CNN-based supervised (DCPDN [4], LAP [15], DAID [1]), semi-supervised (SSID [16] and PSD [2]), weakly supervised (RefineD [44]), and un-supervised (ZID [43]).

1) *Quantitative Evaluation on Synthetic Datasets:* For synthetic testing, two public datasets are used: synthetic objective testing set (SOTS) (1000 pairs of images) of RESIDE and HazeRD (75 pairs of images) [61]. Considering the input size limit for some methods, the test images of HazeRD are resized to 768×768 .

Peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) are used for quantitative evaluation. The results are shown in Table I. It can be seen that our method reaches the highest values in both PSNR and SSIM by a large margin. (All methods are tested by the codes released by their authors, and then measured by the unified templates).¹ Several examples are illustrated in Fig. 5, Our method can generate better-dehazed results, which are shown to be more visually faithful to the ground truth.

2) *Visual Quality on Nonreference Real Datasets:* To verify the generalization performance of our model on natural hazy images, we also compare it with other methods on two real-world datasets unannotated real hazy images (URHI) [31] and DHQ [62]. URHI is a real-world part of RESIDE including

¹The adopted templates for qualitative measurement is distributed by NTIRE 2020 challenge on image dehazing.

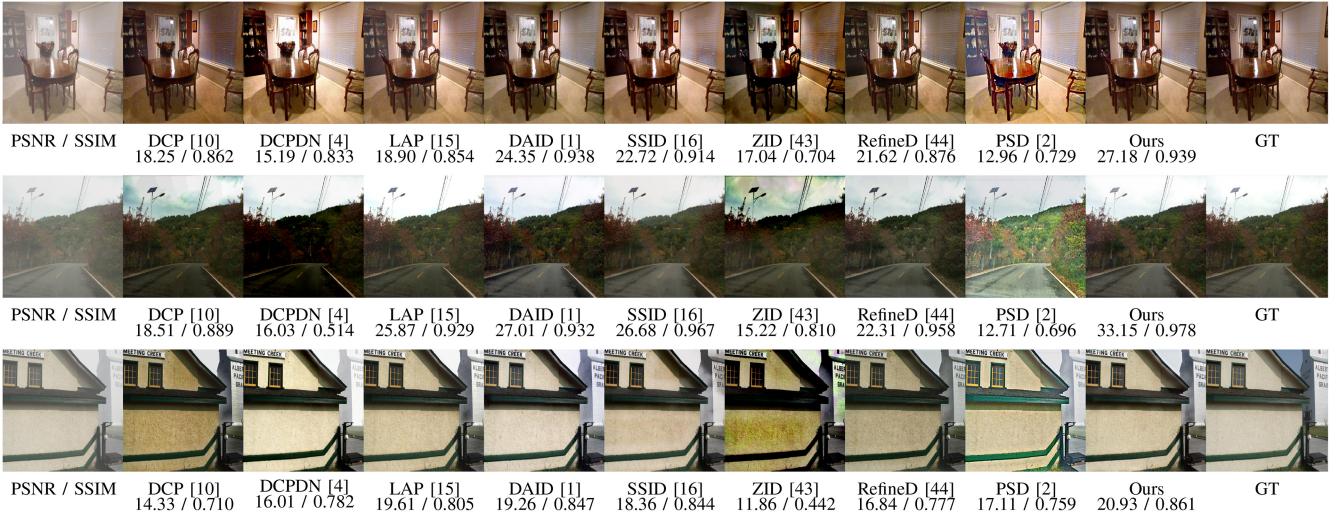


Fig. 5. Visual comparisons on synthetic datasets. The top row: Sample from SOTS-indoor; the middle row: Sample from SOTS-outdoor; and the bottom row: Sample from HazeRD dataset. Our method yields the highest PSNR and SSIM on public synthetic datasets.

4809 hazy images, and DHQ includes 1750 dehazed images generated from 250 hazy images of various haze densities.

The samples are shown in Fig. 6. It can be observed that DCP [10] appears dimmer than others and fails in the sky area. LAP [15] and SSID [16] leave lots of haze residuals. RefineDNet [44] is unable to remove haze thoroughly in the heavily hazy area. The results of DAID [1] are over-smoothing thus look a little blurry. ZID [43] causes serious color distortion, making the dehazed results unrealistic.² Similarly, these cases also exist in DCPDN [4]. PSD [2] transforms the images with over-high brightness and saturation. Comparing with those methods, our model can not only remove haze much more thoroughly, whatever in the mild or heavy haze area, but also generate much sharper details with less color distortion and few artifacts.

3) *Visual Quality on Referenced Real Datasets*: For a more objective comparison, we also evaluate our method with other five dehazing methods, whose training datasets include real hazy images (SSID [16], ZID [43], DAID [1], RefineD [44], and PSD [2]), on a new benchmark real-world outdoor dataset BeDDE [63]. This dataset contains 208 clear and foggy image pairs from 23 cities, and is used to evaluate dehazing methods via full reference image quality assessment (FR-IQA) metrics. Several visual comparisons are shown in Fig. 7. As we can see, the similar problems of the previous methods introduced in Section IV-B2 still exist. By comparison, our method preserves better detail while removing haze and simultaneously gets much more closer to the reference images.

4) *No Reference Image Quality Assessments*: We test four well-known no-reference image quality assessment (NR-IQA) indicators: 1) BRISQUE [64]; 2) NIQE [65]; 3) fog aware density evaluator (FADE) [24]; and 4) dehazing quality index (DHQI) [62]. As shown in Table II, compared with other not fully supervised methods DC, DAID, ZID, RefineD, and PSD on DHQ dataset, our proposed method obtains competitive

²We run the released codes provided by the author under default parameter settings.

TABLE II
QUANTITATIVE RESULTS USING NR-IQA METRICS ON DHQ. THE ORIENTATIONS OF THE ARROWS REPRESENT THE ORDER OF BETTER VALUES. RED AND PINK INDICATE THE BEST AND THE SECOND BEST. IT CAN BE OBSERVED THAT OUR PROPOSED METHOD OBTAINS COMPETITIVE PERFORMANCE

Method	DHQI \uparrow	NIQE \downarrow	FADE \downarrow	BRISQUE \downarrow
Input	50.4423	3.1029	0.8992	27.8548
DCP	51.2065	4.3644	0.4232	26.6585
DAID	51.9686	3.7220	0.5240	28.1121
ZID	62.2235	3.2021	0.2908	27.3342
RefineD	53.2287	4.7891	0.4983	32.2328
PSD	54.8741	3.6788	0.3167	26.7894
Ours	53.8966	4.2122	0.3031	26.4901

performance. But whether NR-IQA metrics do usually match well with human perception? We introduce a discussion on several popular NR-IQA metrics in Appendix B.

C. Running Time

We evaluate the inference time of the proposed model and the state-of-the-art dehazing methods on the same CPU device (56 Intel Xeon CPU E5-2683 v3 @ 2.00 GHz), and the 500 test images size of 512×512 are selected from the RESIDE-SOTS dataset. As shown in Table III,³ with fewer parameters and smaller computing complexity, our model has high efficiency in practice.

D. Ablation Study

To confirm our analysis that the model trained on synthetic datasets greatly degrades its generalization capability on real-world images, and demonstrate the effectiveness of each module in our proposed method, an ablation study is

³ZID generates per dehazed result by 500 training iterations, so the calculation takes longer than others.

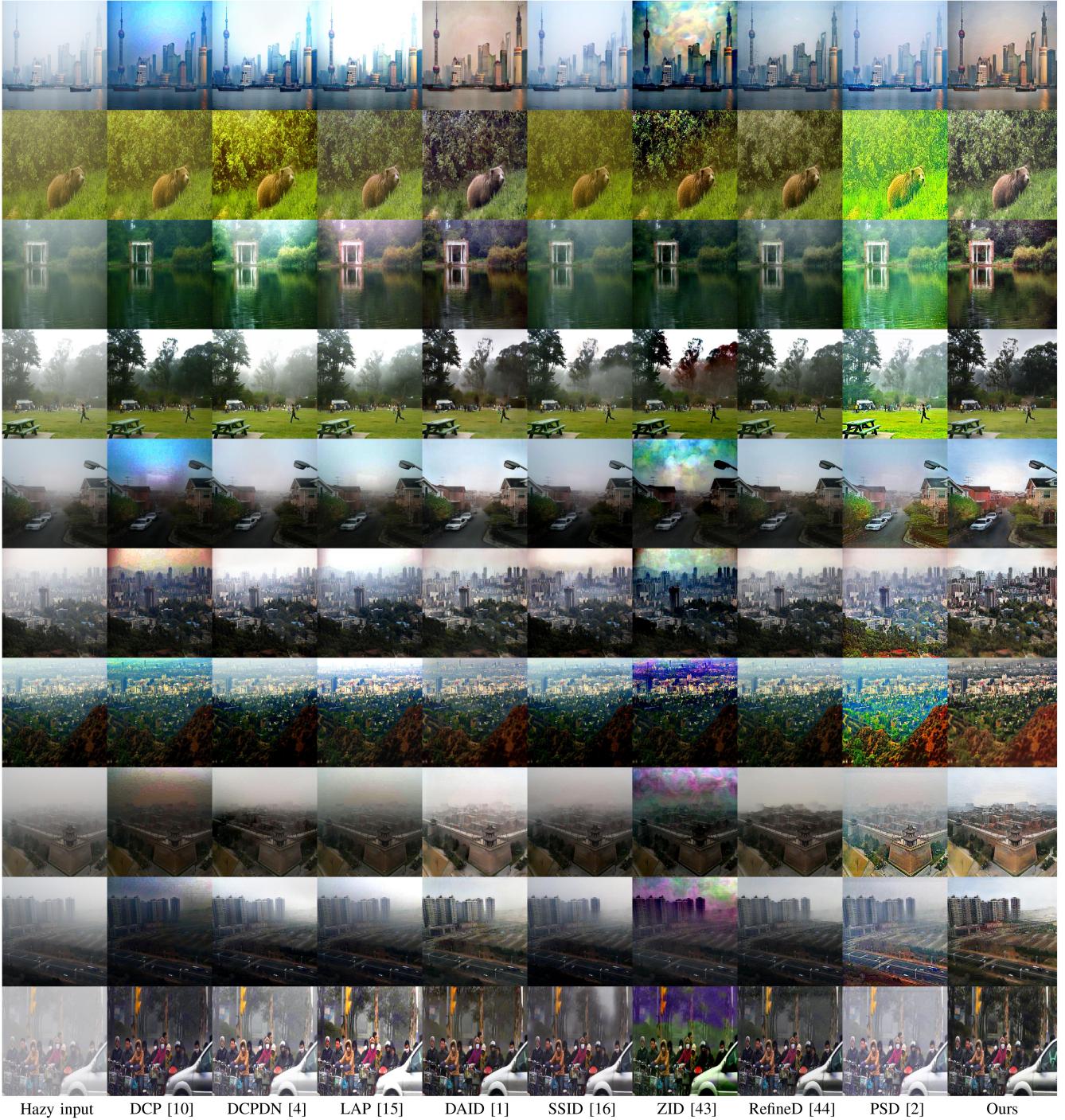


Fig. 6. Visual comparisons on nonreference real datasets. The top five rows: Samples from DHQ dataset; the bottom five rows: Samples from URHI dataset. Our model can generate more natural and visual pleasing dehazed results with less color distortion. Due to the layout limitations, tens of thousands of comparisons with more methods, such as NLD [12], AOD [6], EPDN [8], and FDGan [5] on the two datasets are made publicly available on the website: <https://drive.google.com/drive/folders/1VUfepuqsI99QXMHDvxF4shvUmSMRKA55?usp=sharing>.

conducted. Specifically, five different settings are defined as shown in Table IV.

- 1) *Supervised*: The model trained with labeled synthetic data only.
- 2) *Supervised w/ Haze-Aware Attention*: Adding haze-aware attention module to *Supervised*.
- 3) *Ours Semi-Supervised w/o DCP*: Removing the DCP loss $\mathcal{L}_{\text{adv_dcp}}$ based on our model.

4) *Ours Semi-Supervised w/o Domain Alignment*: Removing the domain alignment strategy $\mathcal{L}_{\text{adv_align}}$ and \mathcal{L}_{KL}) based on our model.

5) *Ours*: To make a fair comparison, the same network architecture and training settings are kept for all settings above.

We compare these settings on both synthetic (SOTS-Outdoor) and real hazy images from the perspective of

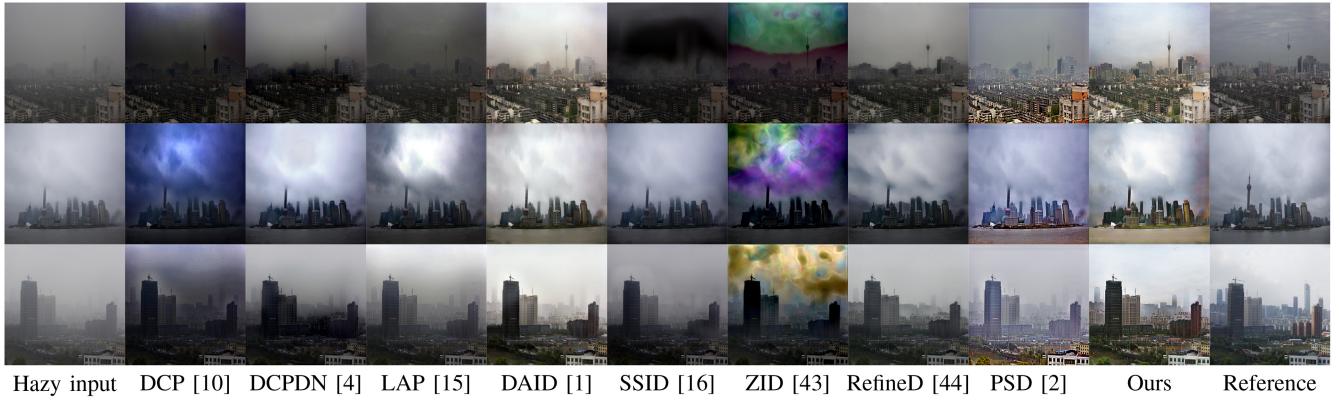


Fig. 7. Visual comparisons on referenced real datasets: BeDDE. Our results are much more closer to the reference images. Please zoom in for best view.

TABLE III
RUNNING TIME COMPARISON WITH OTHER
STATE-OF-THE-ART METHODS

Method	Parameters	Platform	Time
DCP [10]	-	Matlab	22.61 s
DCPDN [4]	66.89 M	Pytorch	24.49 s
LAP [15]	45.00 M	Caffe	6.58 s
DAID [1]	54.59 M	Pytorch	7.39 s
ZID [43]	41.40 M	Pytorch	366.79 s
RefineD [44]	65.80 M	Pytorch	7.71 s
SSID [16]	9.23 M	Pytorch	2.35 s
PSD [2]	6.20 M	Pytorch	25.28 s
Ours	9.96 M	Pytorch	1.61 s

TABLE IV
FIVE DIFFERENT SETTINGS FOR ABLATION STUDY AND QUANTITATIVE
RESULTS OF DIFFERENT SETTINGS ON THE SYNTHETIC TESTING SET

Setting	Synthetic Data	Real Data	Haze-Aware Attention	Domain Alignment	DCP Loss	PSNR / SSIM
1) Supervised	✓					25.12 / 0.936
2) Supervised w/ haze-aware attention	✓		✓			26.94 / 0.951
3) Ours semi-supervised w/o DCP	✓	✓	✓	✓		26.95 / 0.953
4) Ours semi-supervised w/o domain alignment	✓	✓	✓		✓	27.13 / 0.952
5) Ours	✓	✓	✓	✓	✓	27.91 / 0.960

quantitative indexes of PSNR and SSIM and visual results. Table IV shows the quantitative comparison of the effect with different modules. Compared with the *Supervised*, adding the haze-aware attention module can significantly improve both indexes. It means such an attention scheme can help to adaptively respond to the hazy regions and guide the network to remove haze more purposeful. Meanwhile, adding the domain alignment is more helpful to improve SSIM whereas adding the DCP loss to improve PSNR. The reason may be that the domain alignment benefits learning the structure information from the synthetic pairs, and the DCP loss helps to keep the color and texture details of haze-free regions. Fig. 9 can exhibit these improvements more apparently. As marked in the red box, *Supervised* fails to restore the distant regions. With the attention module, *Supervised w/ haze-aware attention* yields a clear view since it helps the network to focus on

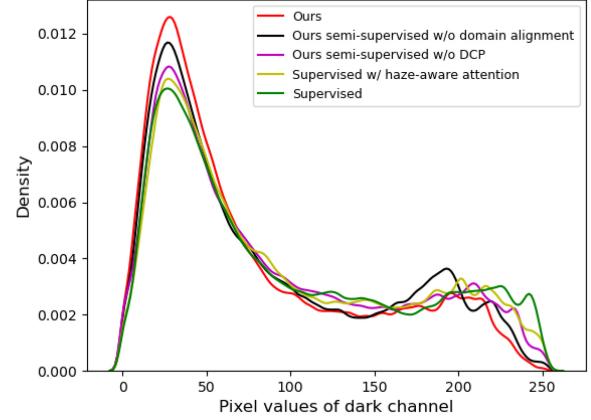


Fig. 8. Learned latent statistical regularities of the five different designs. Our methods learn a much more effective statistical regularity, which also agrees with the qualitative and visual results.

the hazy regions. *Ours semi-supervised w/o DCP* reveals more far details while *Ours semi-supervised w/o domain alignment* has better color vividness, which is consistent with the quantitative observation. Finally, combining all these settings, our model can not only remove haze much more completely, particularly in heavy haze regions, but also generate more visually pleasing results with clearer details.

To compare the learned statistical regularities of different methods, we further collect the dark channel statistics of these five settings on the 100 realistic images and plot the histogram of them, which are shown in Fig. 8. It can be observed that the dark channel values of our method are most concentrated in the range with lower values. Based on the statistical observation in [10] that most local patches in haze-free outdoor images contain some pixels which have very low intensities in at least one color channel, we can say our method performs well on haze removal.

E. User Study

In order to further evaluate the performance of our method from the perspective of human subjective perception, we perform a user study on 20 randomly selected live-action shooting images filmed in natural hazy conditions. For a fair comparison, these images are excluded from the realistic images in

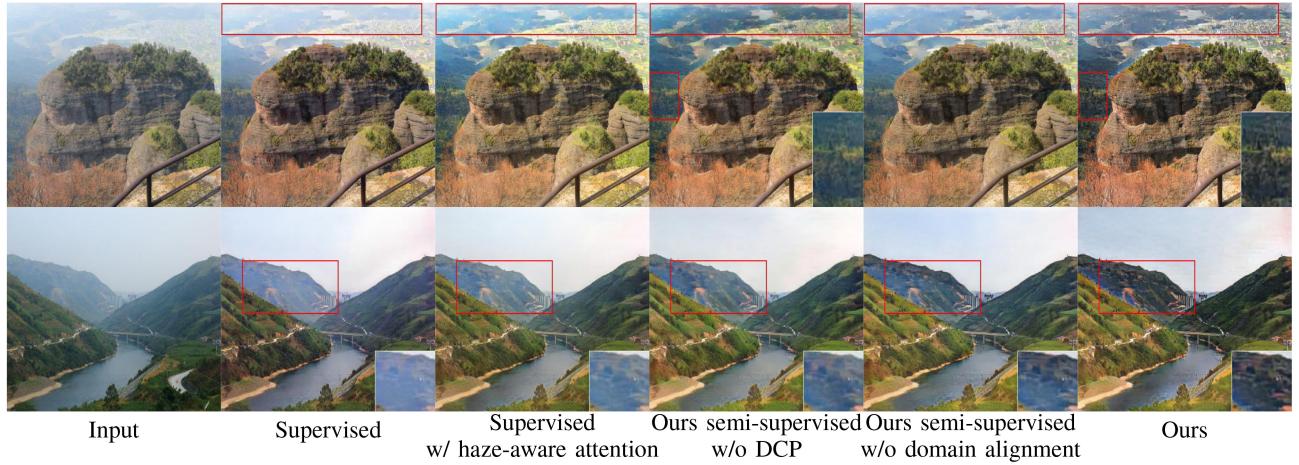


Fig. 9. By introducing the proposed designs, our model can generate more natural dehazed images with shaper details and better contrast. Please zoom in for better visualization and partly details are enlarged in lower right white rectangle.

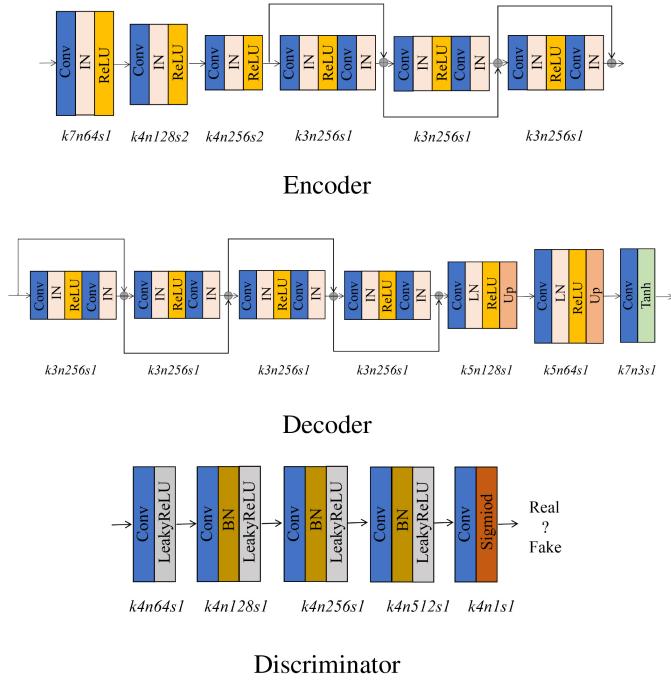


Fig. 10. Parameters and structures of the proposed networks, where k , n , and s are the number of kernels, channels, and stride, respectively.

the training stage. We invite 29 subjects to give their evaluations, among which 21 are males and 8 are females. They also have various occupations include student, teacher, marketing manager, computer engineer, etc. As we compare with the eight state-of-the-art methods with our method, the subjects are asked to score the restoration quality of results from 1 (worst) to 9 (best). The average score for each image and the average for all the 20 images are shown in Table V. It is shown that the subjects consider most of our results are better than the others, and it is consistent with the qualitative and quantitative analysis in Section IV.

V. CONCLUSION

In this article, a semi-supervised learning paradigm is proposed for image dehazing. This paradigm has one branch

TABLE V
SCORES OF 20 LIVE-ACTION SHOOTING IMAGES FOR USER STUDY

Image	DCP	DCPDN	LAP	SSID	RefineD	DAID	ZID	PSD	Ours
1	2.94	4.95	6.52	6.08	6.51	7.24	2.37	3.77	7.46
2	4.17	6.01	7.32	7.66	8.07	7.93	1.96	5.67	8.26
3	3.92	5.98	6.86	6.69	7.01	6.71	2.00	5.24	7.83
4	4.07	7.02	7.13	5.86	5.68	6.61	2.52	4.54	7.61
5	4.18	5.06	6.79	5.96	6.13	6.58	2.05	4.10	7.80
6	3.42	6.73	5.78	6.44	6.73	6.53	1.80	4.89	7.01
7	4.15	5.81	6.65	5.95	7.24	6.40	2.08	4.11	7.56
8	4.24	6.66	5.27	5.30	6.56	5.88	1.69	4.52	6.89
9	4.12	4.58	6.89	6.45	5.67	6.53	1.92	4.00	7.16
10	4.33	5.43	7.13	7.53	7.70	7.44	2.16	4.97	8.09
11	3.62	6.59	4.55	5.93	6.00	5.90	1.86	3.45	6.79
12	4.00	6.86	6.55	6.72	7.17	5.79	1.83	5.34	7.41
13	3.76	6.28	7.55	6.00	6.79	5.86	1.83	4.86	8.76
14	3.59	4.86	5.55	6.69	6.97	6.66	2.17	5.45	7.55
15	4.24	6.90	6.52	5.86	6.14	7.10	1.83	5.79	7.31
16	3.90	6.34	7.28	7.90	7.52	7.96	1.97	4.28	8.76
17	3.66	6.41	6.76	5.97	6.31	7.90	1.69	5.21	8.79
18	3.93	4.76	6.97	7.72	5.62	6.14	1.83	3.97	7.76
19	3.72	5.93	7.69	7.34	7.17	6.94	1.90	4.17	8.83
20	3.59	5.93	5.62	5.76	5.55	7.43	1.83	4.93	7.66
average	3.88	5.95	6.57	6.49	6.63	6.78	1.96	4.66	7.76

with synthesized hazy images for supervised learning and the other with real hazy images for unsupervised learning. For a better generalization on real-world images, a domain alignment strategy is adapted to shorten the distance between the high-level haze features of the two branches in the latent space rather than directly share the network weights. Moreover, we introduce a haze-aware attention module according to the local entropy theory to facilitate adaptive attention on hazy regions. Therefore, our semi-supervised design can powerfully alleviate the domain-shift problem and generate better-dehazed results with clearer details and more natural color. Extensive experiments have demonstrated that our method performs favorably against state-of-the-art methods on both synthetic datasets and real-world hazy images.

APPENDIX A

DETAILED CONFIGURATIONS OF THE SUBNETWORKS

As shown in Fig. 10, we use a symmetric structure to compose the encoder and decoder, except a conv-Tanh layer at the end of the decoder for ouputing the result. Take the encoder

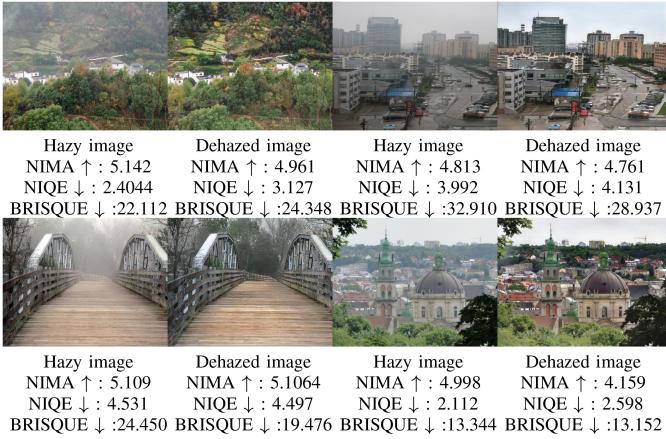
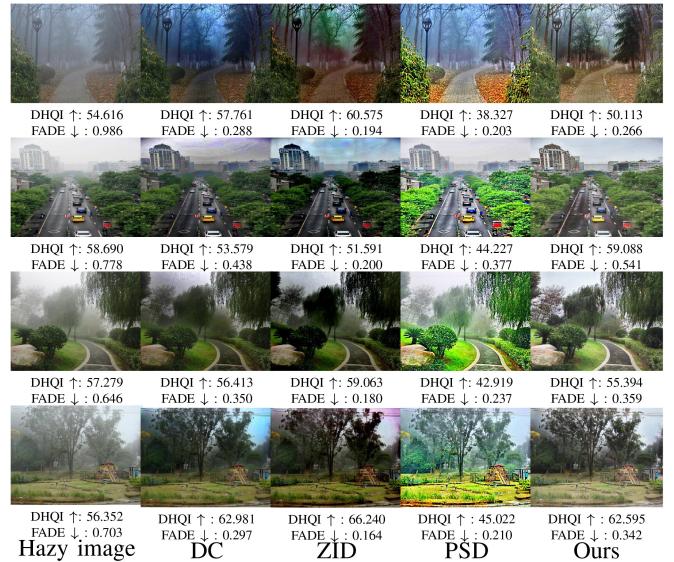


Fig. 11. Quantitative results using NIMA, NIQE, and BRISQUE metrics on real hazy images and their corresponding dehazed versions. The orientations of the arrows represent the order of better values. The higher values indicate the better image quality in NIMA. On the contrary, The lower values indicate the better image quality in NIQE and BRISQUE. Therefore, these three NR-IQA metrics can not well match with human perception.

for example, convolutional layer, instance normalization layer, and ReLU serve as the basic elements for each block. After the third group, we add an identity mapping to keep the information passing through the network. The discriminator starts with a conv-LeakyReLU layer and contains three blocks of conv-BN-LeakyReLU structure. The result is derived from a conv-Sigmoid group, and is used to discriminate whether it is real or fake image.

APPENDIX B DISCUSSION ON THE NR-IQA

There are several measures [24], [62], [64], [65], [66] commonly used in NR-IQAs. NIMA [66] predicts the distribution of human evaluation of images from both technical perspective and aesthetic perspective, which can be used for image quality assessment. But hazy images are not considered to be a special class of degradation in training, therefore, the prediction results of hazy and dehazed images can not be well correlated with human perception. Similar problems also exist with BRISQUE [64] and NIQE [65], in which some universal objective evaluation indexes closing to subjective quality universal are introduced. As shown in Fig. 11, we can see that these metrics are not usually match well with human perceptions. As specifical assessment algorithms to dehazing quality, Choi et al. [24] proposed another no-reference assessment method, called the FADE, which focused on the characteristics of hazy images, including low contrast, faint color, and shifted luminance. Min et al. [62] introduced a DHQI for evaluating dehazed images, in terms of haze-removing features, structure-preserving features, and over-enhancement features. However, they both limit themselves to certain kinds of distortions, whereas different dehazing methods may involve various types of distortions. We further illustrate this view by human vision and quantitative evaluation, as shown in Fig. 12. It can be observed that the two no-reference metrics are not also



- [14] W. Ren, S. Liu, H. Zhang, J. Pan, X. Cao, and M.-H. Yang, "Single image dehazing via multi-scale convolutional neural networks," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 154–169.
- [15] Y. Li et al., "LAP-net: Level-aware progressive network for image dehazing," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 3276–3285.
- [16] L. Li et al., "Semi-supervised image dehazing," *IEEE Trans. Image Process.*, vol. 29, pp. 2766–2779, 2020.
- [17] R. Fattal, "Single image dehazing," *ACM Trans. Graph.*, vol. 27, no. 3, pp. 1–9, 2008.
- [18] K. Zhang, W. Zuo, and L. Zhang, "FFDNet: Toward a fast and flexible solution for CNN-based image denoising," *IEEE Trans. Image Process.*, vol. 27, pp. 4608–4622, 2018.
- [19] X. Liu, Y. Ma, Z. Shi, and J. Chen, "GridDehazeNet: Attention-based multi-scale network for image dehazing," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 7314–7323.
- [20] J. Zhang et al., "Hierarchical density-aware Dehazing network," *IEEE Trans. Cybern.*, vol. 52, no. 10, pp. 11187–11199, Oct. 2022.
- [21] H. Zhu et al., "Single-image dehazing via compositional adversarial network," *IEEE Trans. Cybern.*, vol. 51, no. 2, pp. 829–838, Feb. 2021.
- [22] S. Zhang et al., "Semantic-aware Dehazing network with adaptive feature fusion," *IEEE Trans. Cybern.*, early access, Nov. 19, 2021, doi: [10.1109/TCYB.2021.3124231](https://doi.org/10.1109/TCYB.2021.3124231).
- [23] R. T. Tan, "Visibility in bad weather from a single image," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2008, pp. 1–8.
- [24] L. K. Choi, J. You, and A. C. Bovik, "Referenceless prediction of perceptual fog density and perceptual image defogging," *IEEE Trans. Image Process.*, vol. 24, pp. 3888–3901, 2015.
- [25] R. Fattal, "Dehazing using color-lines," *ACM Trans. Graph.*, vol. 34, no. 1, pp. 1–14, 2014.
- [26] K. Fukushima and S. Miyake, "Neocognitron: A self-organizing neural network model for a mechanism of visual pattern recognition," in *Competition and Cooperation in Neural Nets*. Berlin, Germany: Springer, 1982, pp. 267–285.
- [27] Y. LeCun et al., "Backpropagation applied to handwritten zip code recognition," *Neural Comput.*, vol. 1, no. 4, pp. 541–551, Dec. 1989.
- [28] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.
- [29] N. Silberman, D. Hoiem, P. Kohli, and R. Fergus, "Indoor segmentation and support inference from RGBD images," in *Proc. Eur. Conf. Comput. Vis.*, 2012, pp. 746–760.
- [30] A. Saxena, S. H. Chung, and A. Y. Ng, "Learning depth from single monocular images," in *Proc. NIPS*, vol. 18, 2005, pp. 1–8.
- [31] B. Li et al., "Benchmarking single-image dehazing and beyond," *IEEE Trans. Image Process.*, vol. 28, pp. 492–505, 2019.
- [32] Y. Zhou, X. Du, M. Wang, S. Huo, Y. Zhang, and S.-Y. Kung, "Cross-scale residual network: A general framework for image super-resolution, denoising, and deblocking," *IEEE Trans. Cybern.*, vol. 52, no. 7, pp. 5855–5867, Jul. 2022.
- [33] M. Zhang, Q. Wu, J. Zhang, X. Gao, J. Guo, and D. Tao, "Fluid micelle network for image super-resolution reconstruction," *IEEE Trans. Cybern.*, early access, Apr. 20, 2022, doi: [10.1109/TCYB.2022.3163294](https://doi.org/10.1109/TCYB.2022.3163294).
- [34] K. Panetta, S. K. K. M., S. P. Rao, and S. S. Agaian, "Deep perceptual image enhancement network for exposure restoration," *IEEE Trans. Cybern.*, early access, Jan. 25, 2022, doi: [10.1109/TCYB.2021.3140202](https://doi.org/10.1109/TCYB.2021.3140202).
- [35] R. Ma, S. Li, B. Zhang, L. Fang, and Z. Li, "Flexible and generalized real photograph denoising exploiting dual Meta attention," *IEEE Trans. Cybern.*, early access, May 17, 2022, doi: [10.1109/TCYB.2022.3170472](https://doi.org/10.1109/TCYB.2022.3170472).
- [36] Y. Liu, J. Pan, J. Ren, and Z. Su, "Learning deep priors for image dehazing," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 2492–2500.
- [37] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2223–2232.
- [38] W.-S. Lai, J.-B. Huang, and M.-H. Yang, "Semi-supervised learning for optical flow with generative adversarial networks," in *Proc. 31st Int. Conf. Neural Inf. Process. Syst.*, 2017, pp. 353–363.
- [39] Z. Wan et al., "Bringing old photos back to life," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 2747–2757.
- [40] W. Wei, D. Meng, Q. Zhao, Z. Xu, and Y. Wu, "Semi-supervised transfer learning for image rain removal," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 3877–3886.
- [41] I. Goodfellow et al., "Generative adversarial nets," in *Proc. Int. Conf. Adv. Neural Inf. Process. Syst.*, vol. 27, 2014, pp. 139–144.
- [42] X. Yang, Z. Xu, and J. Luo, "Towards perceptual image dehazing by physics-based disentanglement and adversarial training," in *Proc. AAAI Conf. Artif. Intell.*, vol. 32, 2018, pp. 7485–7492.
- [43] B. Li, Y. Gou, J. Z. Liu, H. Zhu, J. T. Zhou, and X. Peng, "Zero-shot image dehazing," *IEEE Trans. Image Process.*, vol. 29, pp. 8457–8466, 2020.
- [44] S. Zhao, L. Zhang, Y. Shen, and Y. Zhou, "RefineDNet: A weakly supervised refinement framework for single image dehazing," *IEEE Trans. Image Process.*, vol. 30, pp. 3391–3404, 2021.
- [45] A. Golts, D. Freedman, and M. Elad, "Unsupervised single image dehazing using dark channel prior loss," *IEEE Trans. Image Process.*, vol. 29, pp. 2692–2701, 2019.
- [46] Y. Chen et al., "D-BIN: A generalized disentangling batch instance normalization for domain adaptation," *IEEE Trans. Cybern.*, early access, Sep. 21, 2022, doi: [10.1109/TCYB.2021.3110128](https://doi.org/10.1109/TCYB.2021.3110128).
- [47] S. Yang, K. Yu, F. Cao, H. Wang, and X. Wu, "Dual-representation-based autoencoder for domain adaptation," *IEEE Trans. Cybern.*, vol. 52, no. 8, pp. 7464–7477, Aug. 2022.
- [48] A. Atapour-Abarghouei and T. P. Breckon, "Real-time monocular depth estimation using synthetic data with domain adaptation via image style transfer," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 2800–2810.
- [49] Y. Chen, W. Li, C. Sakaridis, D. Dai, and L. Van Gool, "Domain adaptive faster R-CNN for object detection in the wild," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 3339–3348.
- [50] M. Long, G. Ding, J. Wang, J. Sun, Y. Guo, and P. S. Yu, "Transfer sparse coding for robust image representation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2013, pp. 407–414.
- [51] M. Long, Y. Cao, J. Wang, and M. Jordan, "Learning transferable features with deep adaptation networks," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 97–105.
- [52] Y.-H. Tsai, W.-C. Hung, S. Schulter, K. Sohn, M.-H. Yang, and M. Chandraker, "Learning to adapt structured output space for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 7472–7481.
- [53] E. Tzeng, J. Hoffman, K. Saenko, and T. Darrell, "Adversarial discriminative domain adaptation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 7167–7176.
- [54] Y. Ganin and V. Lempitsky, "Unsupervised domain adaptation by back-propagation," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 1180–1189.
- [55] K. Bousmalis, N. Silberman, D. Dohan, D. Erhan, and D. Krishnan, "Unsupervised pixel-level domain adaptation with generative adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 3722–3731.
- [56] A. Dundar, M.-Y. Liu, T.-C. Wang, J. Zedlewski, and J. Kautz, "Domain stylization: A strong, simple baseline for synthetic to real image domain adaptation," 2018, [arXiv:1807.09384](https://arxiv.org/abs/1807.09384).
- [57] A. Shrivastava, T. Pfister, O. Tuzel, J. Susskind, W. Wang, and R. Webb, "Learning from simulated and unsupervised images through adversarial training," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 2107–2116.
- [58] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 3–19.
- [59] T.-Y. Lin et al., "Microsoft COCO: Common objects in context," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 740–755.
- [60] Z. Li and N. Snavely, "MegaDepth: Learning single-view depth prediction from Internet photos," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 2041–2050.
- [61] Y. Zhang, L. Ding, and G. Sharma, "HazeRD: An outdoor scene dataset and benchmark for single image dehazing," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, 2017, pp. 3205–3209.
- [62] X. Min, G. Zhai, K. Gu, X. Yang, and X. Guan, "Objective quality evaluation of dehazed images," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 8, pp. 2879–2892, Aug. 2019.
- [63] S. Zhao, L. Zhang, S. Huang, Y. Shen, and S. Zhao, "Dehazing evaluation: Real-world benchmark datasets, criteria, and baselines," *IEEE Trans. Image Process.*, vol. 29, pp. 6947–6962, 2020.
- [64] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Trans. Image Process.*, vol. 21, pp. 4695–4708, 2012.
- [65] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a ‘completely blind’ image quality analyzer," *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 209–212, Mar. 2013.
- [66] H. Talebi and P. Milanfar, "NIMA: Neural image assessment," *IEEE Trans. Image Process.*, vol. 27, pp. 3998–4011, 2018.



Yu Dong received the B.S. degree from the Chongqing University of Posts and Telecommunications, Chongqing, China, in 2016. She is currently pursuing the Ph.D. degree with the Shenzhen Key Laboratory of Computer Vision and Pattern Recognition, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China, and the University of Chinese Academy of Sciences, Beijing, China.

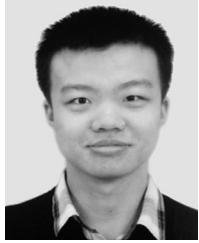
Her research interests include deep learning, computer vision, and image enhancement.



He Zhang received the Ph.D. degree from Rutgers, the State University of New Jersey, New Brunswick, NJ, USA, in 2018.

He is a Research Scientist with Adobe Inc., San Jose, CA, USA, where his main focus is on computer vision.

Dr. Zhang has received multiple awards including the TCSVT 2022 Best Paper Award.



Yunan Li (Member, IEEE) received the B.S. and Ph.D. degrees from the School of Computer Science and Technology, Xidian University, Xi'an, China, in 2014 and 2019, respectively.

He is currently a Huashan Elite Associate Professor with Xidian University. His research interests include computer vision and pattern recognition, especially their applications in image enhancement and action/gesture recognition.



Qian Dong is currently pursuing the B.S. degree with the School of Harbin Institute of Technology, Shenzhen, China.

His research interests include artificial intelligence, machine learning, computer vision, and automatic control.



Shifeng Chen received the B.E. degree from the University of Science and Technology of China, Hefei, China, in 2002, the M.Phil. degree from the City University of Hong Kong, Hong Kong, in 2005, and the Ph.D. degree from The Chinese University of Hong Kong, Hong Kong, in 2008.

He is currently an Associate Professor with the Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China. His current research interests include computer vision, machine learning, and artificial intelligence.