

UCL-Dehaze: Toward Real-World Image Dehazing via Unsupervised Contrastive Learning

Yongzhen Wang^{ID}, Xuefeng Yan^{ID}, Fu Lee Wang^{ID}, Senior Member, IEEE, Haoran Xie^{ID}, Senior Member, IEEE, Wenhan Yang^{ID}, Member, IEEE, Xiao-Ping Zhang^{ID}, Fellow, IEEE, Jing Qin^{ID}, Senior Member, IEEE, and Mingqiang Wei^{ID}, Senior Member, IEEE

Abstract—While the wisdom of training an image dehazing model on synthetic hazy data can alleviate the difficulty of collecting real-world hazy/clean image pairs, it brings the well-known domain shift problem. From a different yet new perspective, this paper explores contrastive learning with an adversarial training effort to leverage unpaired real-world hazy and clean images, thus alleviating the domain shift problem and enhancing the network’s generalization ability in real-world scenarios. We propose an effective unsupervised contrastive learning paradigm for image dehazing, dubbed UCL-Dehaze. Unpaired real-world clean and hazy images are easily captured, and will serve as the important positive and negative samples respectively when training our UCL-Dehaze network. To train the network more effectively, we formulate a new self-contrastive perceptual loss function, which encourages the restored images to approach the positive samples and keep away from the negative

Manuscript received 8 October 2022; revised 20 June 2023 and 15 November 2023; accepted 19 January 2024. Date of publication 9 February 2024; date of current version 22 February 2024. This work was supported in part by the National Natural Science Foundation of China under Grant T2322012, Grant 62172218, and Grant 62032011; in part by the National Defense Basic Scientific Research Program of China under Grant JCKY2020605C003; in part by the Shenzhen Science and Technology Program under Grant JCYJ20220818103401003 and Grant JCYJ20220530172403007; in part by the Guangdong Basic and Applied Basic Research Foundation under Grant 2022A1515010170; in part by the General Research Fund of Hong Kong Research Grants Council under Grant 15218521; in part by the Lam Woo Research Fund under Grant LWP20019; in part by the Direct Grant of Lingnan University, Hong Kong under Grant DR23B2; and in part by the Faculty Research Grants of Lingnan University, Hong Kong, under Grant DB23A3 and Grant DB23B2. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Siwei Ma. (*Corresponding author: Xuefeng Yan*)

Yongzhen Wang is with the School of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China, and also with the School of Computer Science and Technology, Anhui University of Technology, Ma’anshan 243032, China (e-mail: wangyz@nuaa.edu.cn).

Xuefeng Yan and Mingqiang Wei are with the School of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China, and also with the Shenzhen Institute of Research, Nanjing University of Aeronautics and Astronautics, Shenzhen 518038, China (e-mail: yxf@nuaa.edu.cn; mingqiang.wei@gmail.com).

Fu Lee Wang is with the School of Science and Technology, Hong Kong Metropolitan University, Hong Kong, China (e-mail: pwang@hkmu.edu.hk).

Haoran Xie is with the Department of Computing and Decision Sciences, Lingnan University, Hong Kong, China (e-mail: hrchie2@gmail.com).

Wenhan Yang is with the Peng Cheng Laboratory, Shenzhen 518000, China (e-mail: yangwh@pcl.ac.cn).

Xiao-Ping Zhang is with the Tsinghua-Berkeley Shenzhen Institute, Shenzhen 518055, China, and also with the Department of Electrical, Computer and Biomedical Engineering, Toronto Metropolitan University, Toronto, ON M5B 2K3, Canada (e-mail: xpzhang@ieee.org).

Jing Qin is with the School of Nursing, The Hong Kong Polytechnic University, Hong Kong, China (e-mail: harry.qin@polyu.edu.hk).

Digital Object Identifier 10.1109/TIP.2024.3362153

samples in the embedding space. Besides the overall network architecture of UCL-Dehaze, adversarial training is utilized to align the distributions between the positive samples and the dehazed images. Compared with recent image dehazing works, UCL-Dehaze does not require paired data during training and utilizes unpaired positive/negative data to better enhance the dehazing performance. We conduct comprehensive experiments to evaluate our UCL-Dehaze and demonstrate its superiority over the state-of-the-arts, even only 1,800 unpaired real-world images are used to train our network. *Source code is publicly available at <https://github.com/yz-wang/UCL-Dehaze>.*

Index Terms—UCL-Dehaze, image dehazing, unsupervised contrastive learning, unpaired data, adversarial training.

I. INTRODUCTION

IMAGES captured by outdoor vision systems often suffer from noticeable degradation of visibility and contrast due to the existence of haze. Such hazy images inevitably deteriorate the performance of various high-level vision tasks, e.g., traffic monitoring, object detection, and outdoor surveillance [1], [2], [3].

Image dehazing aims to recover sharp images from their hazy counterparts, which is a typical ill-posed problem. To make this problem well-posed, conventional efforts usually exploit hand-crafted priors with empirical observations, such as the dark channel prior (DCP) [7], color attenuation prior (CAP) [8], and non-local color prior (NCP) [9], etc. Although these methods improve the overall visibility of hazy images, making use of a particular prior assumption to dehaze arbitrary real-world images may not always produce satisfactory results. In addition, even professional users often have to carefully tweak various parameters in the formulas for dehazing different input images.

To overcome the aforementioned problems, numerous learning-based methods have been proposed [4], [10], [11], [12], [13], [14], [15]. They commonly employ CNNs or GANs to restore clean images from the corresponding hazy inputs under full supervision or even semi-supervision. Theoretically, if fed with enough paired data, these (semi-)supervised paradigms may generate very promising dehazing results. However, from a practical view, such paired data in the real world are difficult or even impossible to obtain. This explains why existing approaches resort to synthetic hazy data for training. But the gap between synthetic and real-world hazy images inevitably degrades their dehazing abilities to deal with real-world scenarios. Additionally, most of these



Fig. 1. Our UCL-Dehaze avoids bridging the gap between synthetic and real-world haze. It exploits unsupervised contrastive learning with an adversarial training effort to leverage unpaired real-world hazy (negatives) and clean (positives) images. Thus, when inputting an unpredictable real-world hazy image, UCL-Dehaze often generates a haze-free yet perceptually more pleasing result (f), compared to (b) AOD-Net [4], (c) EPDN [5], (d) FD-GAN [6], and (e) UCL-Dehaze w/o the self-contrastive perceptual (SCP) loss.

learning-based methods only exploit clean images as positive samples to guide the network's training, while ignoring the fact that the unexplored information in hazy images is valuable as negative samples. That is, these negative samples, like positive ones, can also provide beneficial information to improve the performance of cutting-edge dehazing models. As exhibited in Fig. 1, compared with the state-of-the-art dehazing approaches and our partial scheme that all only adopt the positive samples, the proposed UCL-Dehaze with the additional negative samples produces a much clearer and perceptually more pleasing dehazing result.

Recently, unsupervised learning-based [16], [17], [18], [19], [20] techniques of image dehazing have become a trend, as they can alleviate the problem of domain shifts between synthetic and real-world data. They commonly employ the framework of CycleGAN [21] or exploit physical priors as loss functions to constrain the network's training, both of which can generalize well to real-world scenes. However, the CycleGAN-based techniques mainly consider how to transform one image from the source domain (hazy) to the target domain (haze-free); during the transformation, the image's intrinsic details can be easily lost, leading to the low quality of restored images. The physical prior-based techniques are not always reliable, since the unseen yet real-world scenes (e.g., thick haze) may violate the prefixed priors; the restored images will be visually unnatural. Moreover, these unsupervised dehazing techniques only employ clean images to guide the network's training, ignoring the fact that the hazy images also provide meaningful guidance information to facilitate the network's training. Differently, we aim to propose an unsupervised contrastive learning paradigm that solves the problem of unpaired image training and leverages contrastive learning to benefit from real-world clean (positives) and hazy (negatives) images.

In this paper, we propose a novel unsupervised contrastive learning paradigm, which regards the real-world image dehazing as an image-to-image translation task (termed UCL-Dehaze). UCL-Dehaze builds itself on the contrastive learning framework and benefits from adversarial training efforts, thus can generalize well to real-world hazy scenarios. UCL-Dehaze consists of two types of contrastive loss functions, namely the patch-wise contrastive loss [22] and a new pixel-wise contrastive (self-contrastive perceptual) loss, both of which play an important role in dehazing. The patch-wise contrastive loss is employed to maximize the mutual information between the corresponding patches of the hazy and restored images, which enables a new unsupervised framework to collect

practical unpaired samples of real-world hazy images for training. Also, to better restore the clean images from their hazy counterparts, we formulate a new self-contrastive perceptual (SCP) loss to encourage the restored images and the clean images (positive samples) to pull together in the representation space while pushing them away from the hazy ones (negative samples). In this way, these two contrastive losses can collaborate with and contribute to each other. Furthermore, the adversarial training strategy is employed to further align the feature distributions between the positive samples and the restored images. Extensive experimental results prove that our UCL-Dehaze performs favorably against the state-of-the-art dehazing approaches, even only 1,800 unpaired real-world training images are used.

In summary, the contributions of this work are three-fold:

- We propose an unsupervised image dehazing network via contrastive learning and adversarial training (call UCL-Dehaze). UCL-Dehaze leverages patch-wise contrastive learning to maximize the mutual information between the input hazy and restored images, providing a new unsupervised training fashion. It can effectively address the restoration of unseen real-world hazy scenarios.
- We formulate a novel pixel-wise self-contrastive perceptual (SCP) loss to train UCL-Dehaze. Specifically, we employ SCP to learn a representation that pulls the restored images and real-world clean images (positives) together while pushing them away from the hazy ones (negatives). Moreover, SCP can be regarded as a universal module to enhance the performance of any other unsupervised dehazing approaches.
- UCL-Dehaze is compared with 22 representative state-of-the-art dehazing approaches via comprehensive experiments. The results are evaluated in terms of full-, reduced- and no-referenced image quality assessments, visual quality, and human subjective surveys. Consistently and substantially, UCL-Dehaze performs favorably against its competitors.

II. RELATED WORK

Image dehazing can be roughly divided into two categories: prior-based and learning-based approaches. In this section, we briefly introduce these two categories, followed by the introduction of contrastive learning.

A. Single Image Dehazing

1) *Prior-based*: Conventional dehazing methods commonly explore hand-crafted priors to restore haze-free images based

on the ill-posed atmospheric scattering model [23]. Reference [24] develop a dehazing approach via compensating the local contrast of the hazy images. Reference [7] propose the well-known dark channel prior (DCP) for single image dehazing, which achieves impressive dehazing results. Recently, [9] observe that the colors of a haze-free image can be well approximated by several hundred distinct colors, and then exploit the non-local color prior (NCP)-based dehazing method. Although these methods have achieved promising results, their performances are limited by the accuracy of the hand-crafted priors adopted in the various real-world scenarios.

2) *Learning-based*: With the advances in deep learning, various learning-based models have been exploited for image dehazing [4], [10], [25], [26], [27], [28], [29], [30]. Early efforts focus on employing CNNs to estimate the transmission map and global atmospheric light in the atmospheric scattering model and then generate haze-free images. For instance, the MSCNN proposed by [10] is one of the early approaches that adopts CNNs for single image dehazing, where the model is trained to estimate the transmission map and then restore the clean result. Recently, some other learning-based approaches have tried to directly produce haze-free images in an end-to-end manner. Reference [4] develop a novel all-in-one dehazing network termed AOD-Net, which is the first model to directly learn the hazy-to-clean image translation. Since then, numerous end-to-end dehazing methods have sprung up. Reference [13] exploit a novel feature fusion network to directly restore haze-free images from the hazy inputs, which achieves remarkable performance on several benchmark datasets. Although we have witnessed promising dehazing results on synthetic datasets, these learning-based efforts trained on synthetic images cannot perform well under real-world scenarios due to the obvious domain gap.

Recently, this issue has been picked up by several semi-supervised-based studies [31], [32]. These approaches explore training their models on both synthetic data and real-world images. Although they alleviate the problem of domain shifts to some extent, their dehazing abilities still depend on the quality and quantity of synthetic data.

Motivated by CycleGAN in unpaired image-to-image translation [21], a handful of approaches attempt to exploit unsupervised frameworks for image dehazing in order to solve this domain shift issue (e.g., Cycle-Dehaze [16], Dehaze-GLCGAN [33]). Since such models can leverage unpaired images for network training, they commonly generalize well to real-world scenes. However, most learning-based efforts only employ clean images as positive samples to guide network training, while ignoring the fact that the negative samples can also provide additional beneficial information for the network's training, thus limiting the dehazing performance of the model.

B. Contrastive Learning

Contrastive learning is a kind of self-supervised learning framework, and it is widely used in the representation learning field [34], [35], [36], [37], [38]. These approaches aim to learn an embedding that brings the associated features close to each other, while the irrelevant samples are pushed away.

Existing efforts mainly focus on applying the contrastive learning paradigm on high-level vision tasks, since the data augmentation method is very suitable for modeling the contrast between positive and negative samples. Inspired by the success of contrastive learning in high-level vision tasks, several studies [39], [40], [41] have attempted to apply contrastive learning to low-level vision tasks, especially in image restoration tasks. Most existing efforts typically adopt the clean images (ground-truth) as positive samples while using the corresponding degraded images as negative samples to guide the training of the restoration network via contrastive learning. For instance, [39] develop a contrastive regularization term to leverage the information of both hazy and clean images for image dehazing. Reference [40] propose a segmentation-aware progressive network for image deraining, which utilizes contrastive learning to help the deraining network generate results closer to the positive images (ground-truth) and away from the negatives (rainy input).

Recently, a small minority of works [42], [43], [44] have attempted to explore the application of contrastive learning in unpaired image restoration networks to improve their performance. Reference [42] propose a dual contrastive learning framework (DCD-GAN) for single image deraining, which is a typical CycleGAN-based approach. Different from the aforementioned contrastive learning framework, DCD-GAN takes the input images and the reconstructed images as positive and negative samples, respectively, and then encourages the generated images to be close to the positives while keeping away from the negatives. Reference [43] develop a non-local contrastive learning (NLCL) method for unpaired image deraining. NLCL leverages the clean image patches as positives while employing rain layer patches as the negatives to better restore the clean images. However, these approaches only construct positive and negative samples from input images or images generated by their models, ignoring that numerous unpaired clean/degraded images in the real world can also be adopted as positive/negative samples to further enhance the generalization ability of the network in real-world scenarios.

Beyond the previous image dehazing wisdom, we do not plan to bridge the gap between synthetic and real-world haze. Thus, we explore unsupervised contrastive learning from an adversarial training perspective to leverage unpaired real-world hazy and clean images. Our proposed network does not require paired data during training. By training the network both pixel-wisely and patch-wisely in an unsupervised yet adversarial manner, we can better utilize unpaired positive/negative data to enhance its dehazing performance.

III. UCL-DEHAZE

In this section, we first describe the overview of UCL-Dehaze, and then detail its architecture. After that, we introduce how the patch-wise and pixel-wise contrastive learning paradigm with adversarial training works in UCL-Dehaze.

A. Overview

In our design, we cast image dehazing as an image-to-image translation task, and simplify it by seeking a mapping function

from hazy images to haze-free images. However, considering that the gap between synthetic and real-world data may cause a significant dehazing performance drop in real-world scenarios, we focus on developing an unsupervised learning framework to train our network via unpaired real-world images rather than paired synthetic images with ‘fake’ haze. In this way, abundant practical real-world images can contribute to the network’s training, thus boosting the generalization capacity of our network on real-world hazy images.

Towards real-world image dehazing, we propose an unsupervised contrastive learning paradigm, called UCL-Dehaze. As exhibited in Fig. 2, the overall architecture of UCL-Dehaze is an UNet-like generator [45] with nine residual blocks [46]. Given an input real-world hazy image x , we aim to employ the generator G to map x to the haze-free image y in an unsupervised training manner. To this end, we leverage contrastive learning with an adversarial training strategy for the network’s training. Specifically, we first feed x to G to produce the preliminary dehazing result y . Then, we employ the discriminator D to judge whether y is a real clean image or a fake image produced by G , which can further improve the quality of y . Finally, we leverage both patch-wise and pixel-wise contrastive losses to train the network in an unsupervised manner, so as to handle unpredictable real-world hazy scenes.

B. Network Architecture

We employ a UNet-based network as the generator module. Although using a more complex network structure would improve the performance of the model, we adopt a simple ResNet-based generator to achieve a better parameter-performance trade-off. As known, the training process of GANs is unstable, and problems such as model collapse and non-convergence often occur. We employ the spectral normalization strategy [47] in the design of residual blocks, which is a weight normalization technique commonly used in GANs, to enhance the training stability. This technique can ensure that the model is continuous without discontinuities, which can make the function smoother and prevent gradient explosion. Here, we adopt the default fast approximation strategy of spectral normalization described in [47]. To further boost the dehazing performance of UCL-Dehaze, an up-to-date multi-scale feature enhancement module (self-calibrated convolutions [48], i.e., SC Conv) is introduced into our network. Moreover, the skip connection mechanism is introduced in our generator to avoid the problem of gradient vanishing.

1) Generator: As shown in Fig. 2, the generator G can directly map an input image x to the haze-free image y in an end-to-end manner. To achieve it, G is supposed to preserve both image structures and details when removing the haze. The entire process of the generator can be described as follows: given a hazy image, we first employ a $4 \times$ down-sampling operation to encode the input hazy image into a low-resolution feature map. Then, nine residual blocks are adopted to extract more complex and deeper features in the low-resolution space and remove the haze simultaneously. After that, we employ the corresponding $4 \times$ up-sampling operation and a 7×7 convolutional layer to output the final restored image.

2) Self-calibrated Convolutions: One dehazing network should have a superior feature transformation capability, providing powerful representations to facilitate the subsequent restoration task. To attain this objective, we incorporate self-calibrated convolutions to build long-distance spatial and inter-channel dependencies around each spatial location. Self-calibrated convolutions can enlarge the receptive field of each convolutional layer to enrich the output features through internal multi-scale communications. Herein, we exploit self-calibrated convolution as a multi-scale feature enhancement module to further improve the performance of UCL-Dehaze.

As depicted in Fig. 3, given an input feature map X , we first split it into two feature maps X_1 and X_2 . Then, the self-calibrated convolution module leverages four different convolution operations (i.e., K_1 , K_2 , K_3 and K_4) to extract and fuse multi-scale features from X_1 and X_2 , thus enriching their feature representations. After that, we can obtain the output features Y_1 and Y_2 from the two branches of self-calibrated convolutions. Finally, we concatenate Y_1 and Y_2 to produce the final output Y . In the design, we incorporate the self-calibrated convolution module (thirteen in total) following the ReLU layer within each convolution operation of the generator. This intentional incorporation is geared towards expanding the receptive field of the convolutional layers and extracting features across multiple scales.

3) Discriminator: For adversarial training, we employ the well-known PatchGAN [49] as the discriminator, which can reduce the network’s parameters and perform faster than other conventional discriminators. The function of the discriminator is to judge whether a given image is a real clean image or a fake image produced by the generator, thus guiding the generator to produce more realistic images. The Least-Square GAN (LSGAN) loss [50] has demonstrated superior effectiveness compared to the vanilla GAN loss, ensuring a more stable training process. Hence, we employ the LSGAN loss in training our network, with the adversarial loss defined as follows:

$$L_{adv}(G) = E_{G(x) \sim P_{fake}}[(D(G(x)) - 1)^2], \quad (1)$$

$$\begin{aligned} L_{adv}(D) = & E_{y \sim P_{real}}[(D(y) - 1)^2] \\ & + E_{G(x) \sim P_{fake}}[(D(G(x)))^2], \end{aligned} \quad (2)$$

where y refers to the real-world clean images and $G(x)$ represents the restored haze-free images.

C. Unsupervised Contrastive Learning

Contrastive learning endeavors to learn a feature embedding that pushes the positive samples close to each other and pushes apart the embedding between negative samples. To enhance the generalization capacity of UCL-Dehaze on real-world hazy scenarios, we explore the joint use of patch- and pixel-wise contrastive learning paradigms for image dehazing tasks.

1) Patch-wise Contrastive Learning: In the design, our goal is to maximize the mutual information between corresponding patches of the input and output images to capture the structure and content feature correspondences between two image domains. To this end, we leverage path-wise contrastive

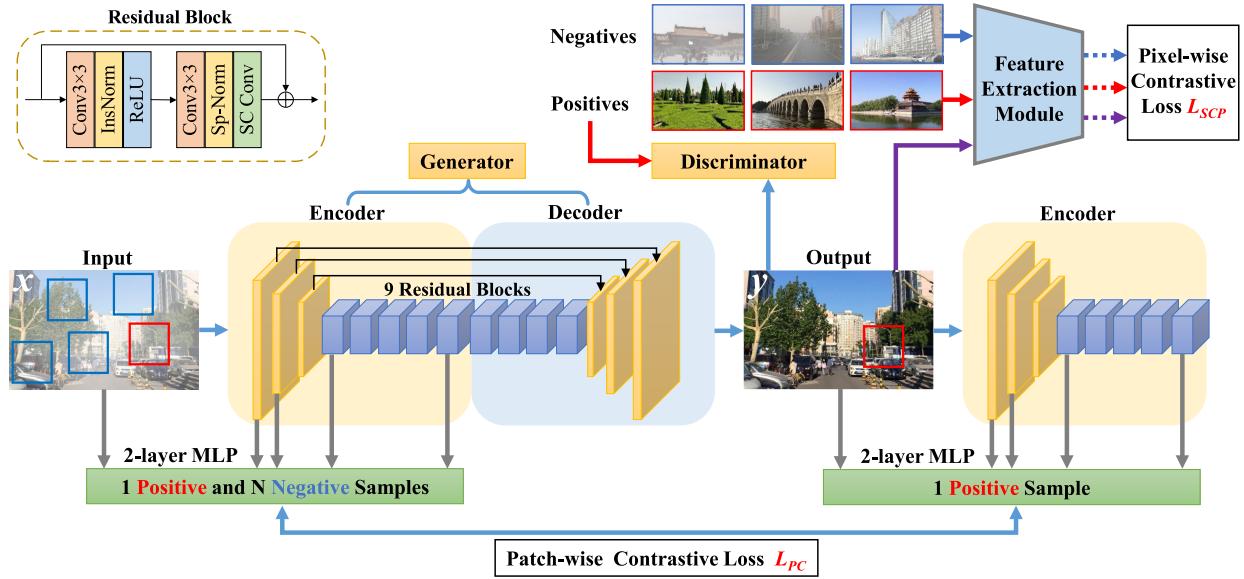


Fig. 2. Overview of UCL-Dehaze. UCL-Dehaze aims to learn a mapping function F , which can map the input hazy image x to the corresponding haze-free image y in an unsupervised manner. It leverages both patch-wise and pixel-wise contrastive losses for the network's training, thus enhancing the dehazing capacity of UCL-Dehaze, especially in real-world scenarios. L_{SCP} represents self-contrastive perceptual loss, Sp-Norm refers to spectral Normalization, and SC Conv refers to self-calibrated convolutions.

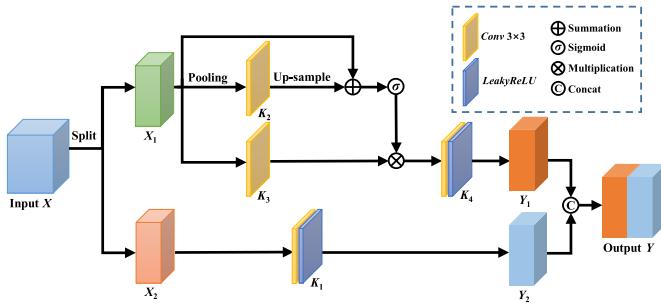


Fig. 3. Architecture of Self-calibrated Convolutions.

learning to train our network in an unsupervised manner. The first thing we need to consider is how to construct the positive and negative samples. Inspired by CUT [22] and CWR [51], we randomly choose $N + 1$ patches from the input image x and one corresponding patch from the restored image y (see Fig. 2). We denote the two corresponding patches as the positive sample, with the remaining N patches in x serving as the negatives. Subsequently, we employ a noisy contrastive estimation module to maximize the mutual information between positive samples. Specifically, we first map the anchor (the patch in y), the positive and the N negatives to K -dimensional vectors and denote them as v , v^+ , and v^- , respectively. After that, we establish an $(N + 1)$ -way classification problem and compute the probability of v^+ being selected over v^- , which is formulated as a cross-entropy loss:

$$\ell(v, v^+, v^-) = -\log \times \left(\frac{\exp(\text{sim}(v, v^+)/\tau)}{\exp(\text{sim}(v, v^+)/\tau) + \sum_{n=1}^N \exp(\text{sim}(v, v_n^-)/\tau)} \right), \quad (3)$$

where $\text{sim}(u, v)$ refers to the cosine similarity between u and v . τ is a temperature parameter to adjust the distance between the anchor and other samples and is set to 0.07. v_n^- denotes the n -th negative sample.

We employ the encoder G_{enc} and a two-layer MLP network H to extract features from the input image x (see Fig. 2), and embed them to a stack of features $\{z_l\}_L = \{H^l(G_{enc}^l(x))\}_L$, where L is the number of layers choose from G_{enc} , and l is the specific l -th layers. These stack of features represent different patches from the image, and we denote the spatial locations in each selected layer as $s \in \{1, \dots, S_l\}$, where S_l refers to the number of spatial locations in each layer. We select an anchor each time and denote its feature as $\hat{z}_l^s \in \mathbb{R}^{C_l}$, where C_l is the number of channels in each layer. Additionally, we refer to the corresponding feature (positive) as $z_l^s \in \mathbb{R}^{C_l}$ and the other features (negatives) as $z_l^{S_l \setminus s} \in \mathbb{R}^{(S_l-1) \times C_l}$.

Ultimately, the patch-wise, multi-layer contrastive loss for the mapping $X \rightarrow Y$ (hazy \rightarrow clean) can be formulated as:

$$L_{PC}(G, H, X) = \mathbb{E}_{x \sim X} \sum_{l=1}^L \sum_{s=1}^{S_l} \ell(\hat{z}_l^s, z_l^s, z_l^{S_l \setminus s}). \quad (4)$$

2) *Pixel-wise Contrastive Learning*: The aforementioned patch-wise contrastive learning introduces a new unsupervised training paradigm by maximizing mutual information between the corresponding patches in input and output images. To help UCL-Dehaze better restore the clean images, we designate real-world clean and hazy images as positive and negative samples, respectively, facilitating the reconstruction of sharp images from their corresponding hazy counterparts. Note that all these positive/negative samples are randomly chosen from the real-world images and are unpaired from each other. In the design, our goal is to advance the dehazing capacity of UCL-Dehaze by exploiting both the information from negative and positive images via a contrastive learning paradigm.

We devise a novel pixel-wise self-contrastive perceptual (SCP) loss to guide UCL-Dehaze in generating results in proximity to the positive samples (real-world clean images) while diverging from the negative ones. Inspired by AECR-Net [39], we denote the group of a real-world clean image R_c and a preliminary restored image $G(x)$ as the positive pair. Similarly, the negative pair is generated by the group of a real-world hazy image R_h and the restored image $G(x)$. In addition to constructing the positive and negative pairs, we need to find a latent feature space of these pairs for contrast. Here, we leverage a pre-trained VGG-16 network to extract the feature maps of different samples. Therefore, the pixel-wise contrastive loss can be expressed as:

$$L_{SCP} = \sum_{i=1}^n \omega_i \cdot \frac{\|\psi_i(R_c) - \psi_i(G(x))\|_1}{\|\psi_i(R_h) - \psi_i(G(x))\|_1}, \quad (5)$$

where $\psi_i(\cdot)$, $i = 1, 2, \dots, n$, refer to extracting the i -th hidden features from the VGG-16 network pre-trained on ImageNet. Here we choose the 2nd, 3rd and 5th max-pooling layers. The coefficients ω_i denote the weight coefficients, and we empirically set $\omega_1 = 0.4$, $\omega_2 = 0.6$, and $\omega_3 = 1$. In this way, the patch-wise and pixel-wise contrastive learning paradigms synergistically collaborate and contribute to each other, thereby enhancing the dehazing capacity of UCL-Dehaze.

Besides these two contrastive loss functions, we adopt the identity loss to keep the structure identical after dehazing as:

$$L_{ide} = E_{y \sim P_{data(Y)}} [\|(G(y) - y)\|_1]. \quad (6)$$

Such an identity loss can encourage the output image to have the same color composition and structure as the input image, thus enhancing the quality of the generated image. The total loss function can be formulated as:

$$L_{Total} = \lambda_1 L_{adv}(G) + \lambda_2 L_{PC} + \lambda_3 L_{SCP} + \lambda_4 L_{ide}, \quad (7)$$

where λ_i , $i = 1, 2, \dots, 4$, are hyperparameters, and we set $\lambda_1 = 1$, $\lambda_2 = 1$, $\lambda_3 = 0.0002$ and $\lambda_4 = 5$ in our experiments.

IV. EXPERIMENTS

In this section, comprehensive experiments are conducted to evaluate the dehazing performance of UCL-Dehaze and other methods. All the experiments are implemented by PyTorch 1.7 on a system with an Intel(R) Core(TM) i9-10920X CPU and an NVIDIA GeForce RTX 3090 GPU.

A. Implementation Details

1) *Dataset*: Since UCL-Dehaze is trained in an unsupervised manner, real-world hazy images can contribute to the network's training for real-world scenarios. We randomly choose the real-world hazy and clean images from the most accessible publicly dataset RESIDE [52] as our training set. RESIDE is a widely used benchmark dataset for image dehazing, which consists of six subsets, i.e., OTS (Outdoor Training Set), ITS (Indoor Training Set), SOTS (Synthetic Object Testing Set), RTTS (Real-world Task-driven Testing Set), HSTS (Hybrid Subjective Testing Set), and URHI (Unannotated Real Hazy Images). In our experiments, the training set is composed of 1,800 real-world hazy images chosen from RTTS and

URHI. For haze-free images, we randomly choose 1,800 clean images from ITS. Note that all the hazy and clean images in our training set are real-world images and are unpaired from each other. Although increasing the number of images in the training set can further improve the dehazing performance of UCL-Dehaze, even in this case, our approach surpasses various state-of-the-art dehazing approaches trained on the entire ITS. To more comprehensively evaluate the dehazing performance of UCL-Dehaze, we select SOTS, HSTS, and the Middlebury portion of the D-HAZY dataset [53] as the testing set.

2) *Training Details*: UCL-Dehaze is trained using the Adam optimizer [54] with a batch size of 1, where the momentum parameters β_1 and β_2 are set to 0.5 and 0.999, respectively. The initial learning rate l for both generator and discriminator is set to 2×10^{-4} . We empirically set the total number of epochs to 100 and adopt a linear decay strategy to adjust l after 50 epochs. Additionally, the patch-wise contrastive loss L_{PC} is computed by the features from five layers of encoder G_{enc} , i.e., the input RGB image, the 1-st and 2-nd down-sampling convolutional layers, as well as the 1-st and 5-th residual blocks. For each layer's features, we randomly sample 256 locations and apply a 2-layer MLP module to produce the final 256-dim features.

Moreover, we are surprised to find that if we employ L_{PC} for mapping both $X \rightarrow Y$ and $Y \rightarrow X$ (i.e., clean images \rightarrow clean images), the restored images will be much clearer and more realistic. This is because using L_{PC} for images from domain Y prevents the generator from making unnecessary changes. Therefore, we adopt a dual-direction L_{PC} to train our UCL-Dehaze, and the two hyperparameters λ_2 are both set to 1.

3) *Evaluation Settings*: UCL-Dehaze is compared quantitatively and qualitatively with various dehazing approaches. They can be classified into three categories: 1) prior-based DCP [7], BCCR [55] and NCP [9] supervised-based AOD-Net [4], GFN [26], EPDN [5], GCANet [57], MSCNN-HE [58], FD-GAN [6], GFN-IJCV [59], and MSFNet [60]; and 3) unsupervised-based CycleGAN [21], Cycle-Dehaze [16], Deep DCP [17], Dehaze-GLCGAN [33], YOLY [19], LIGHT-Net [18], and Cycle-SNSPGAN [20]. Besides, we also compared UCL-Dehaze with two recent semi-supervised image dehazing frameworks (Semi-dehazing [31] and PSD (PSD-FFANet) [30]), a prior + supervised dehazing network (Dual-ScaleNet [56]), and a self-supervised dehazing method (SLAdehazing [61]). We employ the average Peak Signal to Noise Ratio (PSNR), Structural Similarity index (SSIM), and CIEDE2000 [62] for quantitative evaluation of the recovered images, which are the most widely used image full-reference evaluation indexes. Furthermore, to assess the quality of the dehazed images more comprehensively, four reduced-reference indicators are employed to evaluate the contrast (Contrast gain [63]), visibility (e, \bar{r}) [64] and saturation (σ) [64] of the restored images.

1) Contrast gain refers to the mean contrast difference between the dehazed image and its hazy counterpart, which is formulated as:

$$C_{gain} = \bar{C}_R - \bar{C}_H, \quad (8)$$

TABLE I
QUANTITATIVE PSNR AND SSIM VALUES OF THE PROPOSED UCL-DEHAZE AND 22 STATE-OF-THE-ART DEHAZING APPROACHES ON SYNTHETIC DATASETS. **RED** AND **BLUE** COLORS ARE USED TO INDICATE THE 1st AND 2nd RANKS, RESPECTIVELY

Method	Publication	Type	SOTS outdoor		HSTS	
			PSNR↑	SSIM↑	PSNR↑	SSIM↑
DCP [7]	TPAMI'10	Prior	18.38	0.819	17.01	0.803
BCCR [55]	ICCV'13	Prior	15.71	0.769	15.21	0.747
NCP [9]	CVPR'16	Prior	18.07	0.802	17.62	0.798
Dual-ScaleNet [56]	TIP'22	Prior + Supervised	21.76	0.909	24.94	0.912
AOD-Net [4]	ICCV'17	Supervised	20.08	0.861	19.68	0.835
GFN [26]	CVPR'18	Supervised	21.49	0.838	22.94	0.894
EPDN [5]	CVPR'19	Supervised	22.57	0.863	20.37	0.877
GCANet [57]	WACV'19	Supervised	21.66	0.867	21.37	0.874
MSCNN-HE [58]	IJCV'20	Supervised	22.72	0.871	21.23	0.851
FD-GAN [6]	AAAI'20	Supervised	23.76	0.926	23.28	0.914
GFN-IJCV [59]	IJCV'20	Supervised	24.21	0.849	23.17	0.829
MSFNet [60]	TIP'21	Supervised	30.07	0.939	31.03	0.931
Semi-dehazing [31]	TIP'20	Semi-supervised	24.79	0.892	24.36	0.889
PSD [30]	CVPR'21	Semi-supervised	20.49	0.844	19.37	0.824
SLAdehazing [61]	IJCAI'22	Self-supervised	24.33	0.932	24.18	0.893
CycleGAN [21]	ICCV'17	Unsupervised	17.32	0.706	16.05	0.703
Cycle-Dehaze [16]	CVPRW'18	Unsupervised	18.60	0.797	17.96	0.777
Deep DCP [17]	TIP'20	Unsupervised	20.99	0.893	21.21	0.871
Dehaze-GLCGAN [33]	arXiv'20	Unsupervised	23.03	0.917	-	-
YOLY [19]	IJCV'21	Unsupervised	20.39	0.889	21.02	0.905
LIGHT-Net [18]	TETCI'22	Unsupervised	23.11	0.917	22.27	0.906
Cycle-SNPGAN [20]	TITS'22	Unsupervised	23.91	0.911	25.04	0.927
UCL-Dehaze	Ours	Unsupervised	25.21	0.927	26.87	0.933

where \bar{C}_R and \bar{C}_H are the mean contrast of the restored image and hazy image respectively. Given an image with the size of $N_x \times N_y$, its mean contrast can be expressed by:

$$\bar{C} = \frac{1}{N_x N_y} \sum_{y=1}^{N_y} \sum_{x=1}^{N_x} C(x, y), \quad (9)$$

where C represents the contrast of the image in a small window and can be calculated by:

$$C(x, y) = \frac{S(x, y)}{m(x, y)}, \quad (10)$$

where $S(x, y) = \frac{1}{(2r+1)^2} \sum_{j=-r}^r \sum_{i=-r}^r (I(x+i, y+j) - m(x, y))^2$, $m(x, y) = \frac{1}{(2r+1)^2} \sum_{j=-r}^r \sum_{i=-r}^r I(x+i, y+j)$. $I(x, y)$ refers to the original hazy image with the size of $N_x \times N_y$. r is the radius of the local region. A larger value of Contrast gain indicates a better result.

2) The indicators (e, \bar{r}) evaluate image visibility by measuring the enhanced degree of image edges [64]. The first indicator e represents the restoration rate of visible edges after image dehazing and can be expressed as:

$$e = \frac{n_r - n_o}{n_o}, \quad (11)$$

where n_r and n_o refer to the cardinal numbers of the set of visible edges in the dehazing image I_r and the original image I_o . The second indicator \bar{r} is employed to assess the restoration degree of the image edge and texture information. It takes into account both invisible and visible image edges in I_o , which is formulated as:

$$\bar{r} = \exp \left[\frac{1}{n_r} \sum_{i \in \varphi_r} \log r_i \right], \quad (12)$$

where $r_i = \Delta I_i^r / \Delta I_i^o$, ΔI_i^r and ΔI_i^o denote the gradient of the dehazing image and original hazy image, respectively. φ_r refers to the set of visible edges of the restored image. Similar to e , a larger \bar{r} means better results.

3) The indicator σ is adopted to evaluate the color restoration performance of dehazing algorithms [64]. σ represents the rate of the saturated pixels (black or white) after image dehazing and can be expressed as:

$$\sigma = \frac{n_s}{N_x \times N_y}, \quad (13)$$

where n_s represents the number of pixels that are saturated after applying the image restoration but were not before. A smaller value of σ usually indicates a better result.

B. Comparison With State-of-the-Arts

1) *Results on Synthetic Dataset:* We report the average PSNR and SSIM of 22 state-of-the-art dehazing methods on SOTS outdoor and HSTS datasets in Table I. Since haze affects only outdoor vision systems, we mainly focus on evaluating these methods on outdoor datasets. For all compared supervised-based approaches, we either retrain their models on the ITS dataset or directly use the pre-trained models provided by the authors for evaluation. To further demonstrate the effectiveness of our UCL-Dehaze, all compared unsupervised-based methods (except YOLY [19]) are retrained on the URHI datasets (containing 4809 real-world hazy images). For YOLY, since it is an untrained dehazing network, it does not require training data and only uses the given hazy images to produce dehazing results. For Semi-dehazing [31] and PSD [30] (semi-supervised), we retrain the models using both synthetic hazy images and real images (from the URHI dataset) according to the settings in their paper. As discerned, UCL-Dehaze

TABLE II
QUANTITATIVE COMPARISONS (CIEDE2000/CONTRAST GAIN/ $(e, \bar{r})/\sigma$) WITH STATE-OF-THE-ART DEHAZING ALGORITHMS ON SYNTHETIC DATASETS

Method	Type	SOTS outdoor						HSTS					
		CIDED2000↓	Contrast gain↑	$e\uparrow$	$\bar{r}\uparrow$	$\sigma\downarrow$	CIDED2000↓	Contrast gain↑	$e\uparrow$	$\bar{r}\uparrow$	$\sigma\downarrow$		
DCP	Prior	10.199	0.284	12.285	1.439	0.0037	9.186	0.294	11.108	1.473	0.0025		
BCCR	Prior	12.381	0.285	10.583	1.963	0.0022	13.022	0.304	12.206	2.103	0.0022		
NCP	Prior	10.420	0.354	9.312	1.947	0.0136	10.722	0.325	9.049	1.762	0.0077		
Dual-ScaleNet	Prior + Supervised	6.394	0.260	10.562	1.553	0.0078	5.187	0.241	10.382	1.552	0.0021		
AOD-Net	Supervised	7.287	0.230	7.755	1.673	0.0021	7.646	0.202	7.824	1.711	0.0018		
GFN	Supervised	6.879	0.276	8.829	1.544	0.0017	6.287	0.304	9.196	1.544	0.0013		
EPDN	Supervised	8.447	0.287	9.531	1.647	0.0010	6.936	0.326	7.821	1.744	0.0003		
GCANet	Supervised	7.314	0.210	7.274	1.568	0.0121	8.107	0.234	8.999	1.471	0.0195		
MSCNN-HE	Supervised	9.971	0.281	6.505	1.454	0.0039	10.271	0.293	7.427	1.507	0.0042		
FD-GAN	Supervised	6.537	0.262	8.736	1.520	0.0204	7.122	0.297	7.247	1.892	0.0170		
GFN-IJCV	Supervised	4.841	0.140	3.864	1.212	0.0014	4.677	0.166	4.535	1.272	0.0004		
MSFNet	Supervised	2.914	0.262	7.129	1.578	0.0128	2.593	0.315	7.223	1.637	0.0041		
Semi-dehazing	Semi-supervised	4.856	0.205	8.642	1.521	0.0124	5.312	0.191	7.270	1.628	0.0190		
PSD	Semi-supervised	14.292	0.281	8.721	3.199	0.0185	14.820	0.272	7.866	3.230	0.0021		
SLAdehazing	Self-supervised	5.753	0.239	10.354	1.647	0.0011	5.423	0.232	10.088	1.703	0.0003		
CycleGAN	Unsupervised	13.394	0.280	9.984	1.526	0.0059	15.074	0.317	7.780	1.810	0.0023		
Cycle-Dehaze	Unsupervised	13.967	0.287	9.288	1.390	0.0024	13.535	0.324	8.417	1.375	0.0034		
Deep DCP	Unsupervised	8.042	0.284	9.120	1.684	0.0037	8.723	0.115	5.136	1.028	0.0016		
YOLY	Unsupervised	8.557	0.286	3.298	1.220	0.0384	7.910	0.323	3.377	1.279	0.0287		
LIGHT-Net	Unsupervised	5.874	0.277	10.503	1.629	0.0015	6.074	0.312	9.972	1.724	0.0011		
Cycle-SNSPGAN	Unsupervised	6.589	0.132	5.643	1.241	0.0011	6.163	0.188	5.867	1.387	0.0009		
UCL-Dehaze	Unsupervised	4.784	0.289	10.685	1.685	0.0008	4.612	0.330	10.481	1.797	0.0002		

achieves impressive performance in terms of PSNR and SSIM, and outperforms other unsupervised dehazing algorithms by a large margin on both datasets. Moreover, juxtaposed against (semi-)supervised-based methodologies, our approach exhibits distinctive merits.

In addition to PSNR and SSIM, we report the average CIEDE2000, Contrast gain, (e, \bar{r}) , and σ for a comprehensive evaluation of the different dehazing algorithms from color difference, contrast, visibility, and saturation, respectively. As exhibited in Table II, the proposed UCL-Dehaze achieves encouraging performance in terms of CIEDE2000, Contrast gain, and σ , which indicates that the restored images by our method have more realistic colors and higher contrast. Furthermore, we are surprised to find that while the prior-based image dehazing algorithms (including PSD, which employs physical priors as constraints to train the model) perform poorly on the full-reference metrics (PSNR/SSIM), they achieve promising results in terms of visibility enhancement (e, \bar{r}) . We argue it is possible that the prior-based approaches usually focus only on improving the overall visibility of the image, while ignoring the problems of color distortion and loss of details that often occur during image restoration. For instance, DCP cannot dehaze the sky regions well and introduce artifacts. Despite not claiming the pinnacle in visibility compared to these methodologies, UCL-Dehaze still achieves remarkable performance, securing a position among the top five among the 22 dehazing algorithms.

Moreover, we exhibit qualitative comparisons of the dehazing results in Fig. 4 and Fig. 5. It can be observed that MSFNet achieves very promising results in both synthetic datasets, and most of the haze in the images is removed. The dehazing results of Semi-dehazing are pretty good but still cannot completely remove the haze in some regions. Although Cycle-Dehaze and Deep DCP succeed in removing the haze to

some extent, they also cause color degradation and make the dehazed images look darker. Cycle-SNSPGAN can improve the overall visibility of the hazy images, but there is still some haze left in the images. Similar to MSFNet, our UCL-Dehaze can produce much clearer and more natural dehazing results.

Furthermore, Table III illustrates the quantitative evaluation results of different dehazing approaches on the D-HAZY dataset. As observed, our UCL-Dehaze achieves the best performance in both PSNR and SSIM, fully demonstrating the superiority of our method in image dehazing tasks.

2) *Results on Real-world Hazy Images:* To evaluate the performance of UCL-Dehaze in real-world hazy conditions, we compare our method with SOTAs on real-world hazy images. Fig. 6 exhibits five real-world hazy samples and the dehazing results by different approaches. As observed, the dehazing results of MSFNet still have a large number of haze residuals, which further verifies that the supervised-based algorithms generalize poorly on real-world hazy images. Semi-dehazing cannot achieve promising results due to color distortion and halo artifacts. Similar to the results in Fig. 4 and Fig. 5, Cycle-Dehaze and Deep DCP tend to darken the images and cannot remove the haze completely. Although Cycle-SNSPGAN improves the overall visibility of the images, there is still some remaining haze. Compared with these SOTAs, our UCL-Dehaze produces the most natural haze-free images with perceptually pleasing and consistent quality.

For the quantitative comparison, we employ four well-known no-reference image quality assessment metrics: NIQE [65], BRISQUE [66], SSEQ [67], and PI [68]. All these indicators are evaluated on the 50 images prepared for the user study. Evaluation results are illustrated in Table IV. NIQE and BRISQUE are used to evaluate the overall quality of the images, and lower values indicate better results. As a display, our UCL-Dehaze achieves the best performance in these two

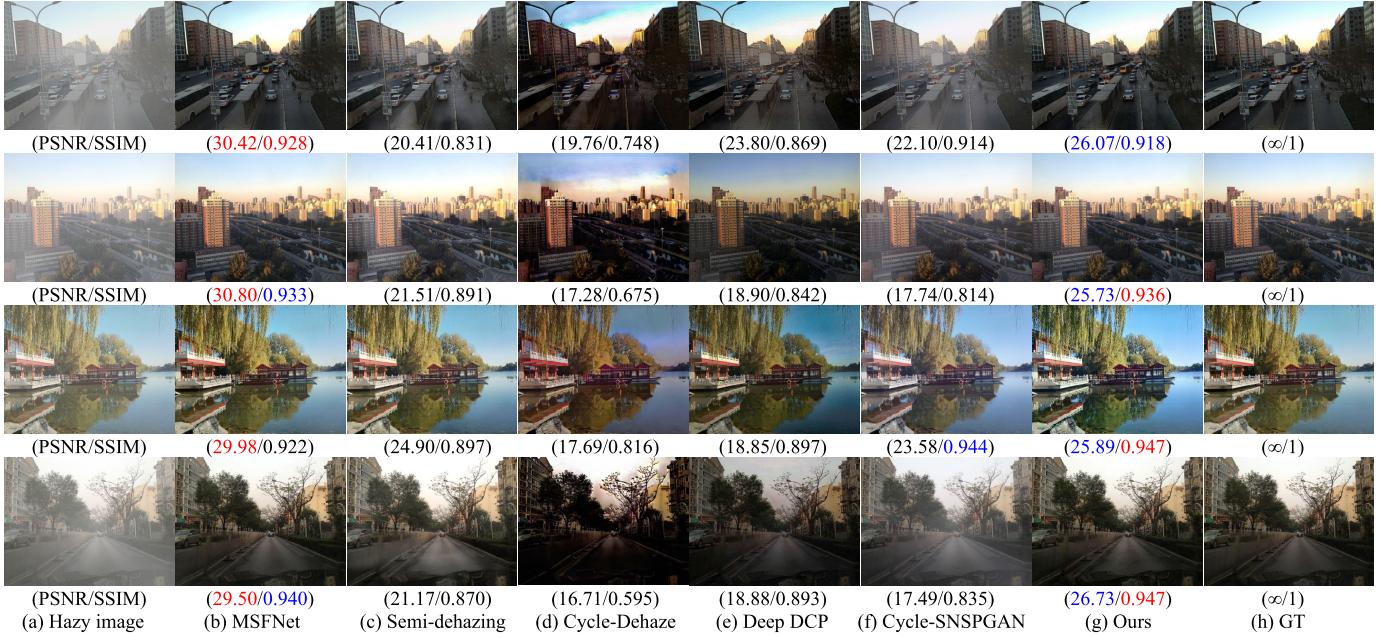


Fig. 4. Image dehazing results on the SOTS outdoor dataset. From (a) to (h): (a) the hazy image, and the dehazing results of (b) MSFNet [60], (c) Semi-dehazing [31], (d) Cycle-Dehaze [16], (e) Deep DCP [17], (f) Cycle-SNSPGAN [20], (g) our UCL-Dehaze, respectively, and (h) the ground-truth image. Our UCL-Dehaze can produce much clearer dehazing images with well-preserved details.



Fig. 5. Image dehazing results on the HSTS dataset. From (a) to (h): (a) the hazy image, and the dehazing results of (b) MSFNet [60], (c) Semi-dehazing [31], (d) Cycle-Dehaze [16], (e) Deep DCP [17], (f) Cycle-SNSPGAN [20], (g) our UCL-Dehaze, respectively, and (h) the ground-truth image. UCL-Dehaze can produce much clearer results with perceptually pleasing.

indicators. SSEQ evaluates image quality by counting the entropy in the spatial and frequency domains of image patches, and UCL-Dehaze wins first place again, which indicates that the images restored by UCL-Dehaze are clean and perceptually pleasing. PI is a criterion that bridges the visual effect with computable index and has been widely used in the field of image super-resolution. Clearly, our UCL-Dehaze also achieves impressive performance in terms of PI. In general, UCL-Dehaze wins three of the four indicators, which further verifies the superiority of our method on real-world dehazing tasks.

In addition to these four no-reference image quality assessment indexes, we also report the average (e , \bar{r}) and σ metrics

for a more comprehensive evaluation of different dehazing algorithms in real-world experiments. As illustrated in Table V, UCL-Dehaze achieves first place in e and σ , as well as an impressive performance in \bar{r} , fully demonstrating the superiority of our network in real-world image dehazing tasks.

C. Ablation Study

1) *Effect of Different Components in UCL-Dehaze:* The proposed network shows superior dehazing performance compared to SOTAs. To further study the effectiveness of UCL-Dehaze, we implement extensive ablation studies to analyze the effectiveness of its components.



Fig. 6. Image dehazing results on the real-world hazy images. From (a) to (g): (a) the real-world hazy image, and the dehazing results of (b) MSFNet [60], (c) Semi-dehazing [31], (d) Cycle-Dehaze [16], (e) Deep DCP [17], (f) Cycle-SNSPGAN [20], and (g) our UCL-Dehaze, respectively. Our UCL-Dehaze can produce both haze-free and more natural images.

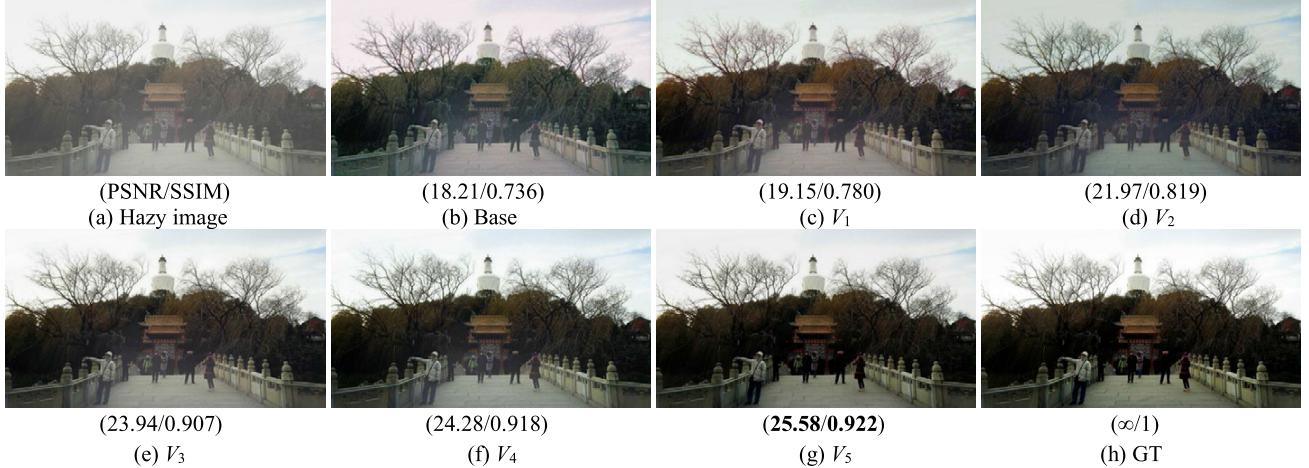


Fig. 7. Visual results of ablation studies. From (a) to (h): (a) the input hazy image, and the results of (b) Base, (c) V_1 , (d) V_2 , (e) V_3 , (f) V_4 , (g) V_5 , respectively, and (h) the ground-truth image.

We first construct our base network with the original ResNet-based generator, and then we train this model through the LSGAN loss and unidirectional L_{PC} . Subsequently, we progressively add different components to the base network as outlined below:

- 1) base network + $L_{ide} \rightarrow V_1$,
- 2) V_1 + dual-direction $L_{PC} \rightarrow V_2$,
- 3) V_2 + self-contrastive perceptual loss $L_{SCP} \rightarrow V_3$,
- 4) V_3 + spectral normalization $\rightarrow V_4$,
- 5) V_4 + self-calibrated convolutions $\rightarrow V_5$ (full model).

All these variants are retrained in the same way as before and tested on the SOTS outdoor dataset and the HSTS dataset. The

performances of these variants are summarized in Table VI and Fig. 7.

As shown, each component of our UCL-Dehaze contributes to image dehazing, especially the proposed self-contrastive perceptual loss L_{SCP} , which achieves 3.12dB PSNR gains over variant V_2 . If the implementation details in this work are fully adopted, we will achieve satisfactory dehazing results.

2) Effect of Different Training Scales: To better understand the effectiveness of UCL-Dehaze under different training scales, we train the proposed network on a different number of training samples and evaluate it on the SOTS outdoor dataset. Note that all the implementation details here are the

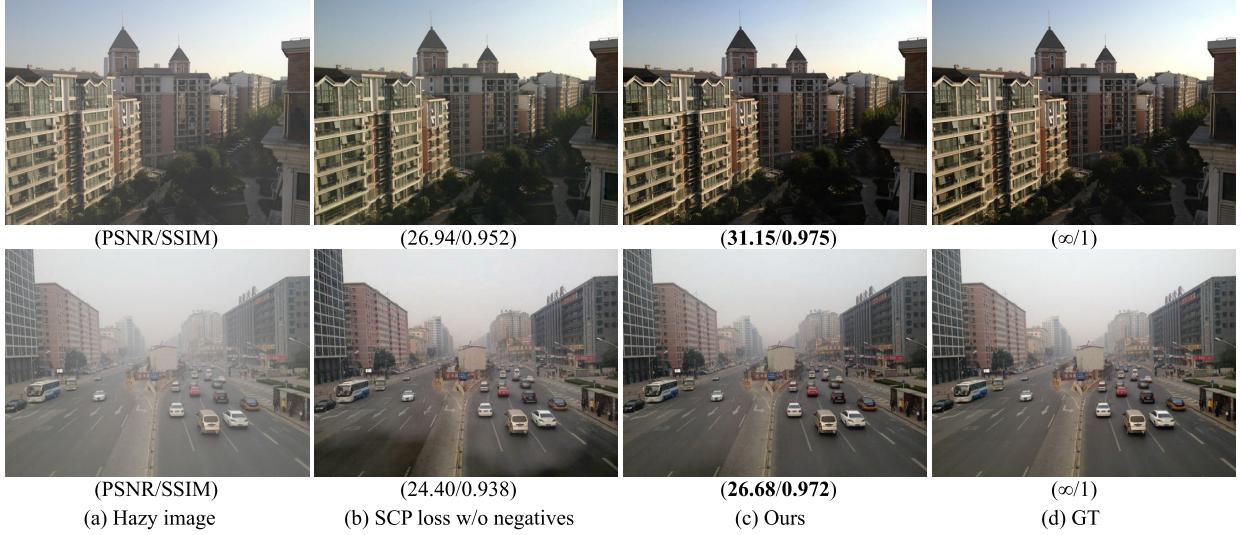


Fig. 8. Visual results of ablation studies on negative samples.

TABLE III

QUANTITATIVE COMPARISONS (AVERAGE PSNR/SSIM) WITH SOTA DEHAZING APPROACHES ON THE D-HAZY DATASET

Method	Type	PSNR↑	SSIM↑
DCP [7]	Prior	15.09	0.830
BCCR [55]	Prior	15.02	0.821
MSCNN [10]	Supervised	13.57	0.798
AOD-Net [4]	Supervised	13.13	0.795
GFN [26]	Supervised	13.15	0.796
EPDN [5]	Supervised	14.44	0.819
MSFNet [60]	Supervised	15.93	0.826
Cycle-Dehaze [16]	Unsupervised	13.28	0.714
Deep DCP [17]	Unsupervised	15.31	0.817
UCL-Dehaze	Unsupervised	16.14	0.833

TABLE IV

QUANTITATIVE COMPARISONS (NIQE/BRISQUE/SSEQ/PI) WITH SOTAs ON 50 REAL-WORLD IMAGES

Method	NIQE↓	BRISQUE↓	SSEQ↓	PI↓
Hazy	4.390	31.515	29.665	4.097
DCP [7]	3.743	26.831	26.146	3.418
AOD-Net [4]	4.117	28.019	27.076	3.435
FD-GAN [6]	3.825	25.886	27.718	3.463
MSFNet [60]	4.173	28.351	26.774	3.510
Semi-dehazing [31]	4.062	24.768	29.321	3.457
PSD [30]	3.775	25.265	28.549	3.342
UCL-Dehaze	3.736	24.658	26.028	3.412

TABLE V

QUANTITATIVE (e , \bar{r}) AND σ METRICS WITH SOTAs ON 50 REAL-WORLD IMAGES

Method	$e \uparrow$	$\bar{r} \uparrow$	$\sigma \downarrow$
DCP [7]	11.724	2.002	0.0043
AOD-Net [4]	9.523	1.550	0.0024
FD-GAN [6]	10.540	1.197	0.0042
MSFNet [60]	5.582	1.263	0.0019
Semi-dehazing [31]	11.005	1.508	0.0016
PSD [30]	10.321	3.875	0.0239
UCL-Dehaze	12.047	2.160	0.0003

same as before. As depicted in Table VII, employing more training samples in the training phase leads to better performance. However, more training samples may incur additional

TABLE VI

ABLATION ANALYSIS OF DIFFERENT VARIANTS ON THE SOTS OUTDOOR DATASET. OUR FULL MODEL OUTPERFORMS ITS ALTERNATIVES

Variants	Base	V_1	V_2	V_3	V_4	V_5
L_{ide}	w/o	✓	✓	✓	✓	✓
Dual- L_{PC}	w/o	w/o	✓	✓	✓	✓
L_{SCP}	w/o	w/o	w/o	✓	✓	✓
Sp-Norm	w/o	w/o	w/o	w/o	✓	✓
SC Conv	w/o	w/o	w/o	w/o	w/o	✓
PSNR	18.14	19.57	21.07	24.19	24.74	25.21
SSIM	0.764	0.804	0.837	0.909	0.916	0.927

TABLE VII

ABLATION STUDY ON DIFFERENT AMOUNT OF TRAINING SAMPLES (N)

	$N = 500$	$N = 1000$	$N = 1500$	$N = 1800$	$N = 2200$
PSNR	21.94	22.77	23.81	25.21	25.07
SSIM	0.884	0.901	0.914	0.927	0.934

computational overhead and training time, which is not preferable in resource-limited applications. Therefore, to achieve a better trade-off between performance and computational overhead, we choose 1800 unpaired images to train our UCL-Dehaze.

3) *Effect of Negative Samples*: To evaluate the effectiveness of negative samples in the proposed pixel-wise contrastive learning loss, we train UCL-Dehaze with/without negative samples and test it on the SOTS outdoor dataset. Note that without negative samples mean that only positive samples in the SCP loss are used to train our network. All the implementation details are the same as before. As illustrated in Table VIII and Fig. 8, employing both negative and positive samples can help the network achieve better performance, verifying that negative samples can also provide beneficial supervision information for network training.

D. Runtime Analysis

Efficiency is essential for a computer vision system [69]. We evaluate the computational performance of various

TABLE VIII
ABLATION STUDY ON NEGATIVE SAMPLES IN SCP LOSS

	SCP w/o negatives	Ours
PSNR	24.63	25.21
SSIM	0.918	0.927

TABLE IX
AVERAGE RUNNING TIMES (SECONDS) OF DIFFERENT METHODS
TESTED ON THE HSTS DATASET

Method	Platform	Average time
DCP	Python (CPU)	1.41
AOD-Net	PyTorch (GPU)	0.11
GPN	PyTorch (GPU)	0.44
Cycle-Dehaze	TensorFlow (GPU)	1.97
GCANet	PyTorch (GPU)	0.21
Deep DCP	TensorFlow (GPU)	0.65
PSD	PyTorch (GPU)	0.39
UCL-Dehaze (ours)	PyTorch (GPU)	0.08

state-of-the-art dehazing methods and report their average running times in Table IX. All the approaches are implemented on a system with an Intel(R) Core(TM) i9-10920X CPU, 32 GB RAM, and an NVIDIA GeForce RTX 3090 GPU. The model size of UCL-Dehaze is 19.451M, and it takes about 0.08s to process a single hazy image from the HSTS dataset on average. This inference speed is faster and more efficient compared to other dehazing algorithms.

V. CONCLUSION

In this paper, we avoid bridging the gap between synthetic and real-world haze. We explore unsupervised contrastive learning from an adversarial training perspective to leverage unpaired real-world hazy and clean images. Accordingly, we propose an effective unsupervised contrastive learning paradigm for image dehazing, termed UCL-Dehaze. Unlike most existing image dehazing works, UCL-Dehaze does not require paired data during training and utilizes unpaired positive/negative data to better enhance the dehazing performance. It leverages an adversarial training effort and benefits from unpaired real-world training data, thus can generalize smoothly to real-world hazy scenarios. In addition, to effectively train the network in an unsupervised manner, we formulate a new pixel-wise contrastive loss function, i.e., the self-contrastive perceptual loss, which encourages the restored images to approach the clean images while keeping away from the hazy ones in the embedding space. Finally, comprehensive evaluations demonstrate that our method performs favorably against the state-of-the-arts, even only 1,800 unpaired real-world images are consumed to train our UCL-Dehaze.

REFERENCES

- [1] D.-Y. Choi, J.-H. Choi, J. Choi, and B. C. Song, "Sharpness enhancement and super-resolution of around-view monitor images," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 8, pp. 2650–2662, Aug. 2018.
- [2] 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/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2018, pp. 3339–3348.
- [3] M. Nasir, K. Muhammad, J. Lloret, A. K. Sangaiah, and M. Sajjad, "Fog computing enabled cost-effective distributed summarization of surveillance videos for smart cities," *J. Parallel Distrib. Comput.*, vol. 126, pp. 161–170, Apr. 2019.
- [4] B. Li, X. Peng, Z. Wang, J. Xu, and D. Feng, "AOD-Net: All-in-one dehazing network," in *Proc. IEEE Int. Conf. Comput. Vis.*, Jun. 2017, pp. 4770–4778.
- [5] Y. Qu, Y. Chen, J. Huang, and Y. Xie, "Enhanced pix2pix dehazing network," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 8160–8168.
- [6] Y. Dong, Y. Liu, H. Zhang, S. Chen, and Y. Qiao, "FD-GAN: Generative adversarial networks with fusion-discriminator for single image dehazing," in *Proc. AAAI Conf. Artif. Intell.*, 2020, pp. 10729–10736.
- [7] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 12, pp. 2341–2353, Dec. 2010.
- [8] Q. Zhu, J. Mai, and L. Shao, "A fast single image haze removal algorithm using color attenuation prior," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3522–3533, Nov. 2015.
- [9] D. Berman, T. Treibitz, and S. Avidan, "Non-local image dehazing," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1674–1682.
- [10] 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.
- [11] Y. Pang, J. Xie, and X. Li, "Visual haze removal by a unified generative adversarial network," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 11, pp. 3211–3221, Nov. 2019.
- [12] C. Li, C. Guo, J. Guo, P. Han, H. Fu, and R. Cong, "PDR-Net: Perception-inspired single image dehazing network with refinement," *IEEE Trans. Multimedia*, vol. 22, no. 3, pp. 704–716, Mar. 2020.
- [13] X. Qin, Z. Wang, Y. Bai, X. Xie, and H. Jia, "FFA-Net: Feature fusion attention network for single image dehazing," in *Proc. AAAI Conf. Artif. Intell.*, vol. 34, no. 7, 2020, pp. 11908–11915.
- [14] T. Song, Y. Kim, C. Oh, H. Jang, N. Ha, and K. Sohn, "Simultaneous deep stereo matching and dehazing with feature attention," *Int. J. Comput. Vis.*, vol. 128, no. 4, pp. 799–817, Apr. 2020.
- [15] Y. Li, Y. Liu, Q. Yan, and K. Zhang, "Deep dehazing network with latent ensembling architecture and adversarial learning," *IEEE Trans. Image Process.*, vol. 30, pp. 1354–1368, 2021.
- [16] D. Engin, A. Genc, and H. K. Ekenel, "Cycle-Dehaze: Enhanced CycleGAN for single image dehazing," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2018, pp. 825–833.
- [17] 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, 2020.
- [18] A. Dudhane, P. W. Patil, and S. Murala, "An end-to-end network for image de-hazing and beyond," *IEEE Trans. Emerg. Topics Comput. Intell.*, vol. 6, no. 1, pp. 159–170, Feb. 2022.
- [19] B. Li, Y. Gou, S. Gu, J. Z. Liu, J. T. Zhou, and X. Peng, "You only look yourself: Unsupervised and untrained single image dehazing neural network," *Int. J. Comput. Vis.*, vol. 129, no. 5, pp. 1754–1767, 2021.
- [20] Y. Wang et al., "Cycle-SNSPGAN: Towards real-world image dehazing via cycle spectral normalized soft likelihood estimation patch GAN," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 11, pp. 20368–20382, Nov. 2022.
- [21] J. 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. (ICCV)*, Oct. 2017, pp. 2223–2232.
- [22] T. Park, A. A. Efros, R. Zhang, and J.-Y. Zhu, "Contrastive learning for unpaired image-to-image translation," in *Proc. Eur. Conf. Comput. Vis.*, Cham, Switzerland: Springer, 2020, pp. 319–345.
- [23] S. G. Narasimhan and S. K. Nayar, "Chromatic framework for vision in bad weather," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2000, pp. 598–605.
- [24] R. T. Tan, "Visibility in bad weather from a single image," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2008, pp. 1–8.
- [25] J. L. Yin, Y. C. Huang, B. H. Chen, and S. Z. Ye, "Color transferred convolutional neural networks for image dehazing," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 11, pp. 3957–3967, Nov. 2020.
- [26] W. Ren et al., "Gated fusion network for single image dehazing," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 3253–3261.

- [27] 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. (ICCV)*, Oct. 2019, pp. 7314–7323.
- [28] H. Dong et al., "Multi-scale boosted dehazing network with dense feature fusion," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 2154–2164.
- [29] Q. Wu, W. Ren, and X. Cao, "Learning interleaved cascade of shrinkage fields for joint image dehazing and denoising," *IEEE Trans. Image Process.*, vol. 29, pp. 1788–1801, 2020.
- [30] Z. Chen, Y. Wang, Y. Yang, and D. Liu, "PSD: Principled synthetic-to-real dehazing guided by physical priors," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 7180–7189.
- [31] L. Li et al., "Semi-supervised image dehazing," *IEEE Trans. Image Process.*, vol. 29, pp. 2766–2779, 2020.
- [32] S. An, X. Huang, L. Wang, L. Wang, and Z. Zheng, "Semi-supervised image dehazing network," *Vis. Comput.*, vol. 38, no. 6, pp. 2041–2055, Jun. 2022.
- [33] Z. Anvari and V. Athitsos, "Dehaze-GLCGAN: Unpaired single image de-hazing via adversarial training," 2020, *arXiv:2008.06632*.
- [34] P. Sermanet et al., "Time-contrastive networks: Self-supervised learning from video," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, Jun. 2018, pp. 1134–1141.
- [35] A. van den Oord, Y. Li, and O. Vinyals, "Representation learning with contrastive predictive coding," 2018, *arXiv:1807.03748*.
- [36] K. He, H. Fan, Y. Wu, S. Xie, and R. Girshick, "Momentum contrast for unsupervised visual representation learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 9729–9738.
- [37] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," in *Proc. Int. Conf. Mach. Learn.*, 2020, pp. 1597–1607.
- [38] O. Henaff, "Data-efficient image recognition with contrastive predictive coding," in *Proc. Int. Conf. Mach. Learn.*, 2020, pp. 4182–4192.
- [39] H. Wu et al., "Contrastive learning for compact single image dehazing," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2021, pp. 10551–10560.
- [40] S. Zheng, C. Lu, Y. Wu, and G. Gupta, "SAPNet: Segmentation-aware progressive network for perceptual contrastive deraining," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. Workshops (WACVW)*, Jan. 2022, pp. 52–62.
- [41] A. Marathe, P. Jain, R. Walambe, and K. Kotecha, "RestoreX-AI: A contrastive approach towards guiding image restoration via explainable AI systems," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2022, pp. 3029–3038.
- [42] X. Chen et al., "Unpaired deep image deraining using dual contrastive learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 2017–2026.
- [43] Y. Ye et al., "Unsupervised deraining: Where contrastive learning meets self-similarity," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 5821–5830.
- [44] X. Chen et al., "Unpaired deep image dehazing using contrastive disentanglement learning," in *Proc. Eur. Conf. Comput. Vis.*, 2022, pp. 632–648.
- [45] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* in Lecture Notes in Computer Science, vol. 9351, 2015, pp. 234–241.
- [46] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
- [47] T. Miyato, T. Kataoka, M. Koyama, and Y. Yoshida, "Spectral normalization for generative adversarial networks," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2018, pp. 1–26.
- [48] J.-J. Liu, Q. Hou, M.-M. Cheng, C. Wang, and J. Feng, "Improving convolutional networks with self-calibrated convolutions," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 10096–10105.
- [49] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 1125–1134.
- [50] X. Mao, Q. Li, H. Xie, R. Y. K. Lau, Z. Wang, and S. P. Smolley, "Least squares generative adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2794–2802.
- [51] J. Han et al., "Underwater image restoration via contrastive learning and a real-world dataset," *Remote Sens.*, vol. 14, no. 17, p. 4297, Aug. 2022.
- [52] B. Li et al., "Benchmarking single-image dehazing and beyond," *IEEE Trans. Image Process.*, vol. 28, no. 1, pp. 492–505, Jan. 2019.
- [53] C. Ancuti, C. O. Ancuti, and C. De Vleeschouwer, "D-HAZY: A dataset to evaluate quantitatively dehazing algorithms," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2016, pp. 2226–2230.
- [54] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," in *Proc. 3rd Int. Conf. Learn. Represent.*, 2015, pp. 1–15.
- [55] G. Meng, Y. Wang, J. Duan, S. Xiang, and C. Pan, "Efficient image dehazing with boundary constraint and contextual regularization," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 617–624.
- [56] Z. Li, C. Zheng, H. Shu, and S. Wu, "Dual-scale single image dehazing via neural augmentation," *IEEE Trans. Image Process.*, vol. 31, pp. 6213–6223, 2022.
- [57] D. Chen et al., "Gated context aggregation network for image dehazing and deraining," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Jun. 2019, pp. 1375–1383.
- [58] W. Ren, J. Pan, H. Zhang, X. Cao, and M.-H. Yang, "Single image dehazing via multi-scale convolutional neural networks with holistic edges," *Int. J. Comput. Vis.*, vol. 128, no. 1, pp. 240–259, Jan. 2020.
- [59] X. Zhang, H. Dong, Z. Hu, W.-S. Lai, F. Wang, and M.-H. Yang, "Gated fusion network for degraded image super resolution," *Int. J. Comput. Vis.*, vol. 128, pp. 1699–1721, Jan. 2020.
- [60] X. Zhu, S. Li, Y. Gan, Y. Zhang, and B. Sun, "Multi-stream fusion network with generalized smooth L_1 loss for single image dehazing," *IEEE Trans. Image Process.*, vol. 30, pp. 7620–7635, 2021.
- [61] Y. Liang, B. Wang, W. Zuo, J. Liu, and W. Ren, "Self-supervised learning and adaptation for single image dehazing," in *Proc. 31st Int. Joint Conf. Artif. Intell.*, Jul. 2022, pp. 1137–1143.
- [62] G. Sharma, W. Wu, and E. N. Dalal, "The CIEDE2000 color-difference formula: Implementation notes, supplementary test data, and mathematical observations," *Color Res. Appl.*, vol. 30, no. 1, pp. 21–30, 2004.
- [63] T. L. Economopoulos, P. A. Asvestas, and G. K. Matsopoulos, "Contrast enhancement of images using partitioned iterated function systems," *Image Vis. Comput.*, vol. 28, no. 1, pp. 45–54, Jan. 2010.
- [64] N. Hautière, J.-P. Tarel, D. Aubert, and É. Dumont, "Blind contrast enhancement assessment by gradient ratioing at visible edges," *Image Anal. Stereol. J.*, vol. 27, no. 2, pp. 87–95, Jun. 2008.
- [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] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Trans. Image Process.*, vol. 21, no. 12, pp. 4695–4708, Dec. 2012.
- [67] L. Liu, B. Liu, H. Huang, and A. C. Bovik, "No-reference image quality assessment based on spatial and spectral entropies," *Signal Process., Image Commun.*, vol. 29, no. 8, pp. 856–863, Sep. 2014.
- [68] Y. Blau, R. Mechrez, R. Timofte, T. Michaeli, and L. Zelnik-Manor, "The 2018 PIRM challenge on perceptual image super-resolution," in *Proc. Eur. Conf. Comput. Vis. (ECCV) Workshops*, 2018, pp. 334–355.
- [69] Y. Pang, J. Cao, and X. Li, "Cascade learning by optimally partitioning," *IEEE Trans. Cybern.*, vol. 47, no. 12, pp. 4148–4161, Dec. 2017.



Yongzhen Wang received the Ph.D. degree in computer science and technology from the Nanjing University of Aeronautics and Astronautics (NUAA), Nanjing, China, in 2023. He is currently a Lecturer with the Anhui University of Technology, Ma'anshan, China. He has published more than 20 research publications, including *IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS*, *IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING*, and *Computer Graphics Forum*. His research interests include deep learning, image processing, and computer vision.



Xuefeng Yan received the Ph.D. degree from the Beijing Institute of Technology in 2005. He was a Visiting Scholar with Georgia State University, Atlanta, GA, USA, in 2008 and 2012, respectively. He is currently a Professor with the School of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics (NUAA), China. His research interests include intelligent computing, MBSE/complex system modeling, simulation, and evaluation.



Fu Lee Wang (Senior Member, IEEE) received the B.Eng. degree in computer engineering and the M.Phil. degree in computer science and information systems from The University of Hong Kong, Hong Kong, and the Ph.D. degree in systems engineering and engineering management from The Chinese University of Hong Kong, Hong Kong. He is currently the Dean and a Professor with the School of Science and Technology, Hong Kong Metropolitan University, Hong Kong. He has over 300 publications in international journals and conferences and led more than 20 competitive grants with more than 20 million Hong Kong dollars. His current research interests include educational technology, information retrieval, computer graphics, and bioinformatics. He is a fellow of BCS, HKIE, and IET, and a Senior Member of ACM. He was the Chair of the IEEE Hong Kong Section Computer Chapter and the ACM Hong Kong Chapter.



Haoran Xie (Senior Member, IEEE) received the Ph.D. degree in computer science from the City University of Hong Kong, Hong Kong, SAR, China, and the Ed.D. degree in digital learning from the University of Bristol, U.K. He is currently the Department Head and an Associate Professor with the Department of Computing and Decision Sciences, Lingnan University, Hong Kong. He has published 400 research publications, including 226 journal articles, such as *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, *IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING*, *IEEE TRANSACTIONS ON AFFECTIVE COMPUTING*, and *IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY*. His research interests include artificial intelligence, big data, and educational technology. He is the Editor-in-Chief of *Natural Language Processing Journal*, *Computers & Education: Artificial Intelligence*, and *Computers & Education: X Reality*. He has been selected as the World's Top 2% Scientists by Stanford University.



He received the 2023 IEEE Multimedia Rising Star Runner-Up Award, the IEEE ICME-2020 Best Paper Award, the IFTC 2017 Best Paper Award, the IEEE CVPR-2018 UG2 Challenge First Runner-up Award, and the MSA-TC Best Paper Award of ISCAS 2022. He was the Candidate of CSIG Best Doctoral Dissertation Award in 2019. He served as the Area Chair for IEEE ICME-2021 and 2022, the Session Chair for IEEE ICME-2021, and the Organizer for IEEE CVPR-2019/2020/2021 UG2+ Challenge and Workshop.



Xiao-Ping Zhang (Fellow, IEEE) received the B.S. and Ph.D. degrees in electronic engineering from Tsinghua University, in 1992 and 1996, respectively, and the M.B.A. degree (Hons.) in finance, economics, and entrepreneurship from The University of Chicago Booth School of Business, Chicago, IL, USA.

He is the Chair Professor at Tsinghua-Berkeley Shenzhen Institute (TBSI). He was with the Department of Electrical, Computer and Biomedical Engineering, Toronto Metropolitan University (Formerly Ryerson University), Toronto, ON, Canada, as a Professor; and the Director of the Communication and Signal Processing Applications Laboratory. He served as the Program Director of Graduate Studies. He is cross-appointed at the Finance Department, Ted Rogers School of Management, Toronto Metropolitan University. His research interests include image and multimedia content analysis, sensor networks, the IoT, machine learning, statistical signal processing, and applications in big data, finance, and marketing.

Dr. Zhang is a fellow of the Canadian Academy of Engineering and the Engineering Institute of Canada. He is a registered Professional Engineer in Ontario, Canada, and a member of the Beta Gamma Sigma Honor Society. He is the General Co-Chair of the IEEE International Conference on Acoustics, Speech, and Signal Processing in 2021, the 2017 GlobalSIP Symposium on Signal and Information Processing for Finance and Business, and the 2019 GlobalSIP Symposium on Signal, Information Processing. He is an AI for Finance and Business. He was an Elected Member of the ICME Steering Committee. He is the Editor-in-Chief of the *IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING*. He is a Senior Area Editor of the *IEEE TRANSACTIONS ON IMAGE PROCESSING*. He served as a Senior Area Editor for the *IEEE TRANSACTIONS ON SIGNAL PROCESSING* and an Associate Editor for the *IEEE TRANSACTIONS ON IMAGE PROCESSING*, the *IEEE TRANSACTIONS ON MULTIMEDIA*, the *IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY*, the *IEEE TRANSACTIONS ON SIGNAL PROCESSING*, and the *IEEE SIGNAL PROCESSING LETTERS*. He was selected as the IEEE Distinguished Lecturer by the IEEE Signal Processing Society and by the IEEE Circuits and Systems Society.



Jing Qin (Senior Member, IEEE) is currently an Associate Professor with the School of Nursing, The Hong Kong Polytechnic University, and a Key Member of the Centre for Smart Health. His research interests include creatively leveraging advanced virtual reality (VR) and artificial intelligence (AI) techniques in healthcare and medicine applications and his achievements in relevant areas have been well recognized by the academic community. He won the Hong Kong Medical and Health Device Industries Association Student Research Award for the Ph.D. study on VR-based simulation systems for surgical training and planning. He won three best paper awards for his research on AI-driven medical image analysis and computer-assisted surgery, including one of the most prestigious awards in this field: the MIA-MICCAI Best Paper Award in 2017. He served as the Local Organization Chair for MICCAI 2019, a technical program committee (TPC) member for many academic conferences, a speaker for many invited talks, and a referee for many prestigious journals in relevant fields.



Mingqiang Wei (Senior Member, IEEE) received the Ph.D. degree in computer science and engineering from The Chinese University of Hong Kong (CUHK), Hong Kong, in 2014. He is currently a Professor with the School of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics (NUAA), Nanjing, China. Before joining NUAA, he was an Assistant Professor with the Hefei University of Technology and a Postdoctoral Fellow with CUHK. His research interests include 3D vision, computer graphics, and deep learning. He was a recipient of the CUHK Young Scholar Thesis Awards in 2014. He is also an Associate Editor of *ACM TOMM*, *The Visual Computer*, and *Journal of Electronic Imaging*, and the Guest Editor of *IEEE TRANSACTIONS ON MULTIMEDIA*. He has published more than 150 research papers, including *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, *SIGGRAPH*, *IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS*, *CVPR*, and *ICCV*.