Pattern Recognition 165 (2025) 111596

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

Real-world nighttime image dehazing using contrastive and adversarial learning

Jingwen Deng [a], Patrick P.K. Chan [b,*], Daniel S. Yeung [c]

[a] School of Future Technology, South China University of Technology, Guangzhou, 511442, China
[b] Shien-Ming Wu School of Intelligent Engineering, South China University of Technology, Guangzhou, 511442, China
[c] Hong Kong Special Administrative Region of China

ARTICLE INFO

Keywords:
Nighttime dehaze
Contrastive learning
Adversarial learning

ABSTRACT

Nighttime image dehazing is a challenging task due to the scarcity of real hazy images and the domain gap between synthetic and real data. To address these challenges, we propose a novel deep learning framework that integrates contrastive and adversarial learning. In the initial training phase, the dehazing generator is trained on synthetic data to produce dehazed images that closely match the ground truths while maintaining a significant distance from the original hazy images through contrastive learning. Simultaneously, the contrastive learning encoder is updated to enhance its ability to distinguish between the dehazed images and ground truths, thereby increasing the difficulty of the dehazing task and pushing the generator to fully exploit feature information for improved results. To bridge the gap between synthetic and real data, the model is fine-tuned using a small set of real hazy images. To mitigate bias from the limited amount of real data, an additional constraint is applied to regulate model adjustments during fine-tuning. Empirical evaluation on multiple benchmark datasets demonstrates that our model outperforms state-of-the-art methods, providing an effective solution for improving visibility in hazy nighttime images by effectively leveraging both synthetic and real data.

1. Introduction

Nighttime hazy images, characterized by a misty and obscured appearance, often encounter various issues that degrade their quality, including inadequate lighting, reduced contrast, and limited visibility, as shown in Fig. 1. These challenges significantly impact computer vision tasks like semantic segmentation [1], and object detection [2]. Consequently, the restoration of hazy images plays a crucial role in enabling autonomous driving, surveillance systems, remote sensing in foggy conditions. Most dehazing methods focus on daytime images, making them inadequate for nighttime scenarios due to their inability to address the complexities introduced by multiple artificial light sources [3]. In particular, the uneven distribution and diverse color of nighttime lighting, as well as the inherent glow and glare effects present in such scenes, cannot be represented and processed accurately.

Several models [4,5] have been proposed specifically for nighttime hazy images, introducing new terms to capture the diverse illumination characteristics and glow effects. These techniques are designed to decompose a blurry image into several components, including the glow, atmospheric light, transmission map, and background layers. However, due to the lack of a definitive and stable solution, achieving a clear and haze-free background presents a significant challenge in this highly ill-posed problem. Moreover, the priors utilized for dehazing in these methods, such as DCP [6] and MRP [7], were originally designed to accommodate the characteristics of haze-free images during the day. Consequently, when applied to nighttime images, this discrepancy frequently results in distorted outputs [3]. Although deep learning-based methods [8] are able to overcome the issues associated with manually designed priors, collecting nighttime hazy images for training is a complex task as it requires long exposures, static scenes, and precise illumination calibration [9]. As a result, these methods often rely on synthetic data for training [10]. Consequently, the domain gap between synthetic and real data often leads to dehazing models trained on synthetic images exhibiting poor generalization when applied to real-world hazy images [9]. Unsupervised and self-supervised methods [11,12] have been proposed to eliminate the reliance on paired training data. However, the absence of supervision reduces their performance, leaving room for enhancement in both dehazing quality and the realism of the resulting images.

Real nighttime hazy images contain valuable information but are often too limited in quantity for effective training. In contrast, while

* Corresponding author.
E-mail address: patrickchan@ieee.org (P.P.K. Chan).

https://doi.org/10.1016/j.patcog.2025.111596
Received 9 August 2024; Received in revised form 12 February 2025; Accepted 10 March 2025
Available online 18 March 2025

(a) Low light (b) limited visibility (c) multiple artificial light source

Fig. 1. Examples of nighttime hazy images which suffer from multiple problems.

a large number of synthetic nighttime hazy images can be generated, a gap exists between synthetic and real data. To address this, a two-stage deep learning framework that integrates contrastive and adversarial learning is proposed. In the first stage, a novel contrastive and adversarial learning approach is used to train the dehazing model on synthetic nighttime hazy image pairs. The model aims to produce dehazed images that closely match their ground truths while maintaining a greater distance from the original hazy images in the feature space. During training, the contrastive learning encoder is updated continuously alongside the dehazing model, increasing the difficulty of the task by enhancing the difference between the dehazed images and their corresponding ground truths. This forces the generator to fully leverage the available information, resulting in improved dehazing quality. The model is then fine-tuned using a small set of real hazy images in the second stage, with histogram equalized images serving as pseudo ground truths. To mitigate bias from the limited real data and pseudo ground truths, an additional constraint is applied to control the adjustment size. This approach effectively bridges the gap between synthetic and real data, leveraging the strengths of both to improve dehazing performance.

In summary, the contributions of this work include:

- A novel deep learning framework combining contrastive and adversarial learning is proposed to address challenges related to limited real nighttime hazy images and the gap between synthetic and real data. The model is trained to generate dehazed images that closely match their ground truths while maintaining distinct separation from the original hazy images in the feature space.
- A key innovation of our approach is the alternating update of the contrastive learning encoder and the dehazing generator during training. This dynamic interaction increases the challenge of the dehazing task, compelling the generator to make full use of the available information and ultimately enhancing dehazing quality.
- Fine-tuning with real nighttime hazy images, processed using histogram equalization as pseudo ground truth, bridges the synthetic-real data gap. An additional constraint is applied to mitigate bias from the limited real data, ensuring better generalization during the fine-tuning process.
- Comprehensive evaluation on both real-world nighttime and daytime datasets demonstrates the model's superiority over state-of-the-art methods, validating its effectiveness in nighttime image dehazing.

The rest of the paper is organized as follows. Section 2 provides the background of hazy image models and dehazing methods. Section 3 introduces the rationale and procedure of our proposed method. Section 4 describes our experiments and results. Finally, the conclusion and future work are given in Section 5.

2. Related work

2.1. Hazy image model

The physical scattering model [13], shown in Eq. (1), is commonly used to represent a hazy image in most model-based dehazing approaches:

$$I(x) = J(x)t(x) + A(1 - t(x)), \tag{1}$$

where x represents an image coordinate, $I(x)$ denotes a hazy image, $J(x)$ represents the haze-free image of $I(x)$, $t(x)$ is the transmission map indicating the light reaching the camera without scattering, and A stands for the global atmospheric light. The model is ineffective for nighttime scenes because it relies on a constant A, which fails to capture the variability of artificial lighting at night. To address this limitation, several techniques have been proposed to improve the representation of nighttime hazy images. One such approach estimates environmental illumination at a point-wise level using a spatially varying atmospheric light term [4]. This method is crucial for accurately modeling the variations in lighting conditions across the scene. Additionally, [5] emphasized the importance of considering factors such as light shapes, illumination directions, and regions of light. To address this, their method processes the locations of glow sources separately using a binary mask, while intensity maps are handled independently to capture the varying light intensities. Another approach focuses on quantifying nighttime glow through the use of the atmospheric point spread function (APSF) [14]. This technique models the optical characteristics of nighttime glow, providing a more nuanced representation of lighting effects in hazy nighttime conditions. These methods represent significant progress in addressing the challenges of modeling nighttime haze, offering valuable insights into how to capture the complex lighting effects inherent in hazy nighttime scenes.

2.2. Single image dehazing methods

2.2.1. Prior-based methods

Early dehazing methods relied on statistical analysis or empirical observation to estimate atmospheric light (A) and transmission map (t). These methods included Dark Channel Prior [6], Maximum Reflectance Prior [7], Multiple Color Space Prior [15] and so on. However, these priors are often based on specific assumptions and may not generalize well in different scenarios. For instance, Dark Channel Prior struggles with spatially varying colors of atmospheric light, especially in sky regions. Other approaches [16] have applied the Retinex theory [17] to estimate the illumination component and detail-revealed reflectance component of hazy images. However, these methods may encounter difficulties when dealing with non-homogeneous and thick hazy conditions.

2.2.2. Deep learning-based methods

With the emergence of deep learning, an increasing number of methods have leveraged deep neural networks for estimating the atmospheric light term A and transmission map $t(x)$ [18]. These estimates, based on the physical scattering model, can be used to calculate a clear image $J(x)$. On the other hand, some methods, such as [19], focus on estimating t not to directly obtain $J(x)$, but rather to guide the image recovery process, as the transmission map provides insight into the haze distribution within the scene. Beyond the scattering model, other approaches decompose nighttime hazy images into structure, detail, and noise layers using hybrid regularization terms [20]. For example, [21] proposed extracting high- and low-frequency components of the hazy image to capture scene texture and structural information, respectively. However, many of these methods rely on a gradient-based layer separation technique, which assumes that noise layers exhibit strong gradients, while fog layers have minimal gradients. This assumption can be problematic, as it may inadvertently filter out high-frequency details, removing important image features in the process. Additionally, it may strip away light during fog removal, leading to dark or grayish outputs. To address the issue of detail loss, a two-subnetwork approach has been proposed, where one network handles dehazing and the other refines the details [22]. While this improves detail recovery, it also increases model complexity and may amplify artifacts, as the second network relies on the initial output, potentially exacerbating errors. End-to-end learning is usually adopted to generate haze-free images using models like GANs [23] and diffusion models [24]. For example, DW-GAN [25] employs a 2D discrete wavelet transform to capture multi-scale features and fully exploit image information. Diffusion models have gained attention for producing high-quality and diverse samples. However, their controllability remains a challenge, particularly in preserving structural details when removing fog. Some approaches, such as [26], condition the diffusion model on the original hazy image to guide dehazing, while RSHazeDiff [27] integrates frequency domain information to impose constraints during the intermediate process and employs a patch-level sampling strategy to accelerate inference.

However, these deep learning methods typically rely on large training datasets, which can be challenging to obtain for real-world hazy images. While training on synthetic datasets may yield satisfactory results on synthetic data, these methods may not generalize well to real data. In the case of synthetic hazy images, the generated fog typically exhibits a uniform color, unlike real hazy images where the fog may display varying colors due to the presence of artificial light sources. Additionally, synthetic hazy images face challenges in accurately simulating the distribution of fog, especially in capturing the varying concentrations of fog at different distances.

2.2.3. Self-supervised and unsupervised learning methods

Recent advances in self-supervised and unsupervised image enhancement have achieved promising results for image enhancement tasks without the need for paired training data. These methods often leverage contrastive learning to promote effective feature learning. For instance, UCL-Dehaze [11] addresses the dehazing problem using unpaired real-world hazy and clean images. The model employs patch-wise and pixel-wise contrastive learning loss to train the dehazing network. Similarly, the CDD-GAN model [28] uses contrastive disentanglement learning to separate task-relevant (clean image) and task-irrelevant (haze-related) factors within the feature space, isolating clean images from the hazy inputs.

To address the domain gap between training and testing data, Shyam et al. [12] proposed an augmentation technique generates hazy images across multiple domains, while adversarially varying the degradation factors of the generated images. By exposing the model to images from diverse domains, this approach enhances the robustness of the dehazing model. Additionally, domain adaptation techniques such as CycleGAN and other generative adversarial networks (GANs) have been employed to bridge the domain gap. These methods [29] typically involve min–max games between the generator and discriminator, aiming to align images from the source domain with those from the target domain. These frameworks offer valuable insights into addressing the challenge of insufficient real paired data. However, the absence of ground-truth supervision in these methods can result in visually unnatural outcomes, including color distortion and loss of fine details.

3. Proposed method

In this section, a method is devised to overcome the challenges associated with limited volume in real nighttime hazy images for training and the disparity between synthetic and real data in nighttime image dehazing. The overall framework of the proposed method is illustrated in Fig. 2. Our model aims to train a generator $\mathcal{G}$ to produce a dehazed image from a given hazy input. An autoencoder-like dehazing network [30] is chosen as the backbone of $\mathcal{G}$. This network includes six Feature Attention (FA) blocks [31] and two Dynamic Feature Enhancement (DFE) modules [32], which together enable the network to maintain a large, adaptive receptive field. These components are effective in preserving the structural information of the images. The training process is composed of two phases: the training phase, represented by solid lines, and the fine-tuning phase, represented by dotted lines in Fig. 2.

During *the training phase*, synthetic nighttime hazy image pairs are utilized to train both the generator ($\mathcal{G}$) in the contrastive learning framework and the encoder ($\mathcal{F}$) through adversarial learning. $\mathcal{G}$ is optimized to ensure that its dehazed outputs closely align with the ground truth in both pixel and feature spaces, while maintaining distinguishability from the original hazy images in feature space. The total loss for optimizing $\mathcal{G}$ is as shown in Eq. (2).

$$\mathcal{L}_{G}^{Train} = \mathcal{L}_{rec} + \mathcal{L}_{rank}, \tag{2}$$

where $\mathcal{L}_{rec}$ and $\mathcal{L}_{rank}$ represent the reconstruction loss and pairwise ranking loss, which are explained in Sections 3.2 and 3.1, respectively.

Simultaneously, $\mathcal{F}$ is trained to enhance its discriminative ability between generated images and ground truths. The rationale for updating $\mathcal{F}$ is that as the distance between the generated images and ground truths increases, it becomes more challenging for $\mathcal{G}$ to generate images that are close with the ground truths in the feature space. This forces $\mathcal{G}$ to fully utilize the available information to update, leading to improved dehazing quality. $\mathcal{F}$ is updated by minimizing the loss $\mathcal{L}_{adv}$ of classifying generated images and ground truths. $\mathcal{L}_{adv}$ will be introduced in Section 3.3.

Finally, the model is further updated using a limited number of real nighttime hazy images during *the fine-tuning phase*. This phase is designed to align the feature distributions learned from synthetic data with those in real data, enhancing the model's performance on real-world tasks. In this phase, we assume that a small set of real-world hazy images, without corresponding ground truth, is available. Because collecting real nighttime hazy image pairs is costly, and the absence of corresponding real data and ground truth presents significant challenges for learning.

To generate a pseudo-ground truth for a real nighttime hazy image, our model employs Contrast Limited Adaptive Histogram Equalization (CLAHE) [33]. Nighttime hazy images typically exhibit a narrow intensity range, predominantly concentrated in the lower intensity regions, and the RGB channels often have highly overlapping distributions, resulting in a dark and blurry appearance. Fig. 3(a) visually demonstrates examples of nighttime haze images and their corresponding histograms for each color channel (RGB). To address the issues of uneven brightness and background noise, our model applies histogram equalization. CLAHE divides the image into patches and calculates a mapping relationship between the original and adjusted pixel values, while also imposing a contrast constraint during the mapping process. This technique effectively stretches the pixel distribution, improving brightness uniformity and reducing noise. CLAHE is particularly effective for nighttime dehazing scenarios.

Additionally, the results of applying CLAHE to each RGB, HSV, and CIELab channel independently, as well as to the three RGB channels jointly, are shown in Fig. 3. The application of CLAHE to the RGB channels jointly results in brighter images with higher saturation compared to the other color spaces. Although there may be slight

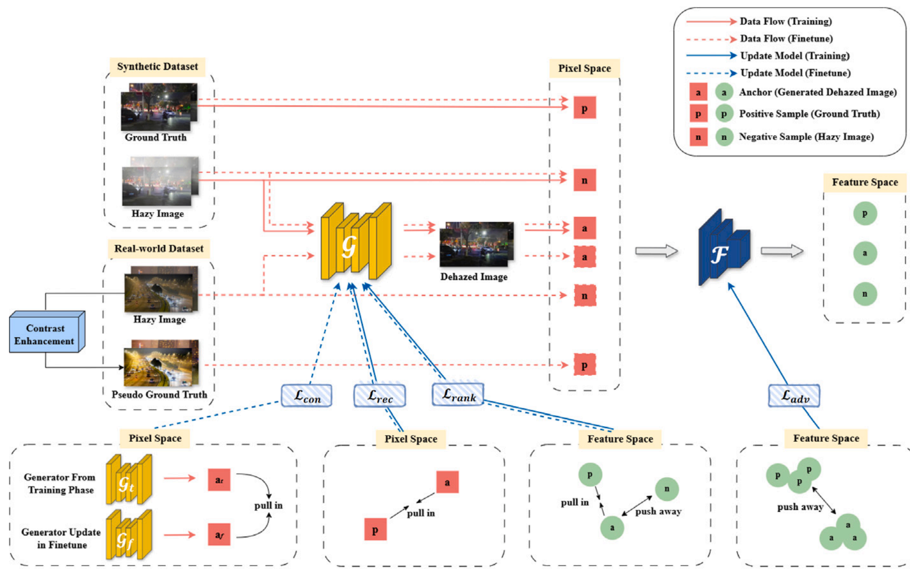
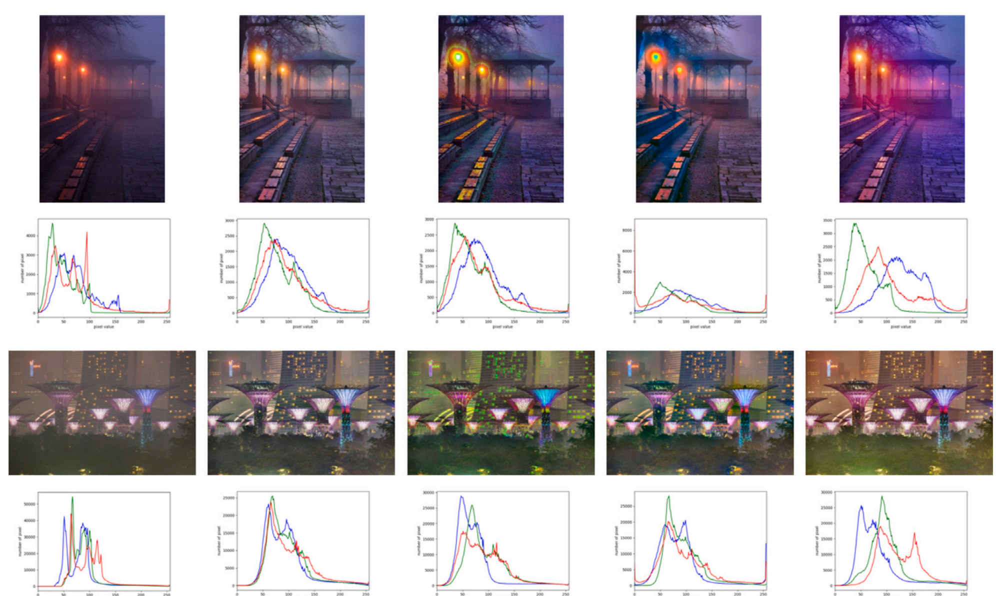


Fig. 2. The framework of our proposed method consists of a two-stage training strategy. In the *training phase*, the dehazed image generator $\mathcal{G}$ is initially trained using synthetic images, while real images are utilized in the *fine-tuning phase*. Throughout the entire training process, the encoder $\mathcal{F}$ is updated simultaneously to increase the difficulty of the dehazing task, thereby enhancing the training of $\mathcal{G}$. The loss functions used during the training process are outlined below.



(a) Avg FADE: 0.79 (b) Avg FADE: 0.52 (c) Avg FADE: 0.40 (d) Avg FADE: 0.48 (e) Avg FADE: 0.42

Fig. 3. Visual comparisons between different histogram equalization processing methods and corresponding histograms. (a) original hazy image; (b) CLAHE applied to each RGB channel independently; (c) CLAHE applied to each HSV channel independently; (d) CLAHE applied to each CIELab channel independently; (e) CLAHE applied to three RGB channels jointly.

color distortion in the darker regions, the overall output provides a satisfactory pseudo-ground truth for real nighttime haze images.

Similar to the initial training phase, both $\mathcal{L}_{rank}$ and $\mathcal{L}_{rec}$ are incorporated during the fine-tuning phase. In this phase, the pseudo-ground truth, generated by applying CLAHE to the three RGB channels jointly, serves as p. To address potential bias from the limited number of real images, synthetic images are also included alongside the real ones. The encoder $\mathcal{F}$ is not updated during fine-tuning, as it has already demonstrated strong discriminative capability between hazy and clean images, and the small data size during this phase could lead to inaccurate updates. An additional loss term, $\mathcal{L}_{con}$ (discussed in Section 3.4), is introduced to regulate the extent of model adjustments. Thus, the total loss for training $\mathcal{G}$ is as shown in Eq. (3), with the newly introduced consistency loss $\mathcal{L}con$ detailed in Section 3.4.

$$\mathcal{L}_{G}^{Tuning} = \mathcal{L}_{rec} + \mathcal{L}_{rank} + \mathcal{L}_{con}. \tag{3}$$

3.1. Pairwise ranking loss ($\mathcal{L}_{rank}$)

By maintaining proximity between the dehazed and clean images, our framework follows the principles of *contrastive learning* to improve the accuracy of the dehazing process. Specifically, $\mathcal{G}$ is optimized by maximizing the distance between a hazy image (n) and its generated dehazed image ($a = \mathcal{G}(n)$), while minimizing the distance between generated dehazed image (a) and the ground truth (p) in the space of the encoder ($\mathcal{F}$). The pairwise ranking loss is defined as Eq. (4).

$$\mathcal{L}_{rank} = \sum_{i \in \mathcal{A}} w_i \times \frac{D(\mathcal{F}_i(a), \mathcal{F}_i(p))}{D(\mathcal{F}_i(a), \mathcal{F}_i(n)) + \delta}, \tag{4}$$

where $D(s,t)$ denotes the $L1$ distance between s and t, $\mathcal{F}_i$ represents the latent feature of the i_{th} convolution layer in $\mathcal{F}$, and $\mathcal{A}$ is a set containing all considered layers. δ is a small value to prevent the occurrence of division by 0. Following the empirical settings of perceptual loss, the output features from the layers conv1_2, conv2_2, conv3_4, conv4_4, and conv5_4 are selected, and their corresponding weights w_i are set to 1/32, 1/16, 1/8, 1/4, and 1, respectively, for the calculation of $\mathcal{L}_{rank}$.

3.2. Reconstruction loss ($\mathcal{L}_{rec}$)

While $\mathcal{L}_{rank}$ effectively reduces the distance between the generated dehazed output (a) and its corresponding clean image (p) in feature space of $\mathcal{F}$, an additional constraint is introduced to minimize their distance in the pixel space as well. This reconstruction loss, defined in Eq. (5), encourages the generated image to closely approximate the clean image.

$$\mathcal{L}_{rec} = \|a - p\|_1 . \tag{5}$$

3.3. Adversarial loss ($\mathcal{L}_{adv}$)

The quality of the generated image is further enhanced through *adversarial learning*. Unlike previous contrastive learning-based approaches, $\mathcal{F}$ in our model is updated simultaneously with the $\mathcal{G}$, which increases the difficulty of the dehazing task. This joint optimization encourages the generator to produce more realistic and high-quality dehazed images. The objective of $\mathcal{F}$ is to minimize the loss $\mathcal{L}_{adv}$, defined as Eq. (6), which measures its ability to distinguish between the generated images and the ground truths.

$$\mathcal{L}_{adv} = y log(\mathcal{F}(x)) + (1-y) log(1-\mathcal{F}(x)), \tag{6}$$

where x represents an image, which can either be a generated dehazed image a or the ground truth p, and y denotes the label of x. $y = 0$ when x is a, and $y = 1$ when x is p.

3.4. Consistency loss ($\mathcal{L}_{con}$)

$\mathcal{L}_{con}$, defined in Eq. (7), is considered exclusively during the fine-tuning phase to alleviate the bias caused by the limited number of real images. The extent of model adjustments is controlled by minimizing the changes during the fine-tuning process.

$$\mathcal{L}_{con} = \left\| \mathcal{G}_t(n) - \mathcal{G}_f(n) \right\|_1 , \tag{7}$$

where $\mathcal{G}_t$ denotes the generator obtained from the training phase, while $\mathcal{G}_f$ refers to the generator that is being updated during the fine-tuning phase.

4. Experiments

4.1. Experimental settings

4.1.1. Implementation on our proposed method

The implementation of our method utilizes PyTorch on an NVIDIA RTX 3090 GPU. The batch size is configured to be 24, and the training process is divided into two phases, each consisting of 100 and 50 epochs, respectively.

The synthetic nighttime dataset [34] is used in our experiments. It consists of 3649 pairs of nighttime clean and hazy images. In our experiments, 1839 and 1810 pairs are used for training and testing respectively. The clean images are extracted from 100 video clips and the hazy images are synthesized based on physical scattering model [13]. The raw images at original 4K resolution are resized into 256 × 256 for training. On the other hand, the real nighttime hazy images, collected by [14] from the Internet, consist of 440 images and are used in the fine-tuning phase.

Table 1
Quantitative comparisons on real-world dataset NHRW. The best and second best results are indicated by bold and underlined respectively.

Method	FADE(↓)	NIQE(↓)	BRISQUE(↓)	PaQ-2-PiQ(↑)	UNIQUE(↑)
Original image	0.78	12.39	26.05	67.38	0.02
MRP [7]	0.44	16.23	24.07	66.75	−0.12
OSFD [9]	0.38	14.60	**20.92**	67.29	−0.05
FFANet [31]	0.60	**12.43**	24.32	67.42	0.03
AECRNet [30]	0.57	13.93	31.80	67.21	0.10
Zheng [34]	0.79	<u>13.80</u>	28.74	<u>67.69</u>	<u>0.18</u>
Jin [14]	<u>0.29</u>	16.79	27.97	67.23	−0.13
UCL-Dehaze [11]	0.43	14.50	30.65	66.19	−0.37
FC-DM [26]	0.52	15.23	30.56	66.58	−0.21
Ours	**0.21**	14.44	<u>22.23</u>	**70.21**	**0.32**

4.1.2. Evaluation metric

The visibility of a real-world hazy image without ground truth is estimated using the Fog Aware Density Evaluator (FADE) [35]. FADE is based on statistical observations from a large dataset of natural foggy and fog-free images. A smaller FADE value indicates a better quality of dehazing for the image. To evaluate the quality of the generated images from different perspectives, four non-reference image quality assessment metrics are employed. These metrics include two traditional statistical-based and two deep learning-based image quality assessments. For traditional statistical-based metrics, such as Natural Image Quality Evaluator (NIQE) [36] and Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE) [37], a lower score from both metrics indicates better perceptual quality of the generated images. For deep learning-based models, such as From Patches to Pictures Deep Quality Prediction Models (PaQ-2-PiQ) [38] and Unified No-reference Image Quality and Uncertainty Evaluator (UNIQUE) [39], a model with a higher score indicates better performance. For the paired clean and hazy images, two well-known metrics, Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM), are used to evaluate the image quality.

4.2. Results on real-world datasets

Our proposed method is evaluated and compared with the existing state-of-the-art (SOTA) methods of image dehazing, including UCL-Dehaze [11], Jin [14], Zheng [34], AECRNet [30], FFANet [31], OSFD [9], MRP [7] and FC-DM [26], in the real-world hazy image dataset, NHRW [9]. It is important to note that none of these images were included in the training phase of the methods.

Table 1 presents the experimental results in terms of the metrics. The dehazed images produced by the methods, except Zheng's, exhibit improved visibility as evidenced by their lower FADE values compared to the original images. Among these methods, our approach achieves the lowest FADE value of 0.21, indicating superior performance. In terms of NIQE, which evaluates the naturalness of images, FFANet performs the best, while our method achieves a competitive score of 14.44. Regarding the BRISQUE metric, only OSFD, our method, and MRP generate higher quality images than the original ones. Our method achieves the second lowest BRISQUE value. Additionally, on the latest no-reference IQA metrics, PaQ-2-PiQ and UNIQUE, our method outperforms all others, achieving the highest scores of 70.21 for PaQ-2-PiQ and 0.32 for UNIQUE. Notably, the diffusion-based method FC-DM does not demonstrate a clear advantage over other approaches in these evaluations. In conclusion, the dehazed images produced by our method demonstrate good quality across various aspects and generally outperform those generated by state-of-the-art (SOTA) methods.

The dehazing methods are also evaluated through subjective visualization. Fig. 4 illustrates visual examples of each method. In the images processed by Zheng, FFANet, OSFD, and MRP, some residual haze is still noticeable, particularly in the background areas. When the sky region occupies a substantial portion of the image, as seen in the example (a) and (c), the results from OSFD and MRP exhibit a halo effect in the sky region. This effect is primarily caused by these methods estimating the ambient illumination and transmission map based on patch information, leading to discontinuous colors.

The method of Zheng cannot effectively remove the fog in the images, especially in example (b), the concentration of the fog in the background even increases. On the other hand, Jin's method demonstrates better performance by successfully removing a significant amount of fog and producing visually improved outcomes compared to the aforementioned methods. However, it tends to over-suppress the light-effects present in hazy images, resulting in color distortion in regions near light sources, as observed in examples (b) and (c). AECRNet exhibits similar performance, though, as shown in example (a), it tends to darken the images. Images generated by UCL-Dehaze, on the other hand, often show noticeable artifacts, likely due to its unsupervised nature, which may compromise its ability to preserve image naturalness and quality. FC-DM generates images with fewer artifacts but struggles to preserve lighting and glow regions, as seen in example (b). Much of the lighting from the train and station is removed, resulting in a loss of realism and failing to accurately represent the original scene. In contrast, our method excels in significantly reducing fog while simultaneously enhancing low-light regions. This results in improved visibility and visually appealing outcomes. Unlike the mentioned methods, our approach strikes a balance by effectively reducing fog without over-suppressing the light-effects. As a result, color distortion near light sources is minimized, leading to a more accurate and visually pleasing representation of the scene.

4.3. Results on synthetic datasets

Our proposed method is compared with the SOTA methods of nighttime image dehazing on the well-known NHR dataset [9] and the Nighttime with Haze 4K dataset [34]. Table 2 presents quantitative comparisons of our method and the other methods in terms of PSNR and SSIM.

Among the NHR dataset, the Jin and our methods achieve the highest and second-highest PSNR values respectively. Furthermore, when it comes to SSIM, our method achieves the highest score of 0.91, indicating superior structural similarity and better preservation of image details. Regarding the Nighttime with Haze 4K dataset, our method achieves the highest PSNR and SSIM values, followed by the Jin method.

The synthesis methods in the Nighttime with Haze 4K and NHR datasets differ significantly. The Nighttime with Haze 4K dataset is representative of scenarios with white fog that is unevenly distributed. In contrast, the NHR dataset includes images with varied color designed to simulate nighttime, dusk, and other low-light scenarios. The consistent performance of our method across these datasets demonstrates its robustness and generalization capability. While the Jin method performs well in terms of PSNR on the NHR dataset, our method strikes a balance between PSNR and SSIM, resulting in the best overall image quality. On the other hand, similar to its performance on real-world images, the diffusion-based method FC-DM delivers less satisfying results on synthetic images compared to other approaches.

4.4. Ablation study

In this ablation study, the impact of each module in our model is investigated, with four modules being considered. The modules are denoted as R, A, F, and C, which respectively represent the usage of $\mathcal{L}_{rank}$, $\mathcal{L}_{adv}$, Fine-Tuning, and $\mathcal{L}_{con}$. It should be noted that C relies on F, meaning that if a model does not consider fine-tuning, $\mathcal{L}_{con}$ must not be used. As a result, different variations of our model are considered, including *Ours-R*, *Ours-RA*, *Ours-RFC*, *Ours-RAF*, and *Ours-RAFC*, which is our full model. Additionally, a baseline model (*Base*) that only contains the generator trained by minimizing $\mathcal{L}_{rec}$ is also included for comparison purposes. Since real images lack corresponding ground truths, we use no-reference metrics instead of PSNR and SSIM for evaluation. The detailed settings and the quantified results on the NHRW dataset in terms of FADE, NIQE, BRISQUE, PaQ-2-PiQ, and UNIQUE are presented in Table 3.

Table 2
Quantitative comparisons on synthetic dataset NHR [9] and Nighttime with Haze 4K dataset [34]. The best and second best results are indicated by bold and underlined respectively. The results of MRP, OSFD, and Jin tested on NHR dataset are borrowed from [14].

Method	NHR		Nighttime with haze 4K dataset	
	PSNR(↑)	SSIM(↑)	PSNR(↑)	SSIM(↑)
MRP [7]	19.93	0.78	10.48	0.50
OSFD [9]	21.32	0.80	8.51	0.47
FFANet [31]	10.78	0.66	7.94	0.52
AECRNet [30]	9.07	0.44	<u>15.12</u>	<u>0.69</u>
Zheng [34]	12.80	0.74	12.48	0.61
Jin [14]	**26.56**	<u>0.89</u>	13.50	0.67
UCL-Dehaze [11]	11.73	0.55	11.25	0.49
FC-DM [26]	12.69	0.75	14.31	0.61
Ours	<u>21.49</u>	**0.91**	**16.39**	**0.83**

The results clearly demonstrate the contribution of the each component in our proposed method towards image dehazing. The FADE values consistently decrease with the addition of each new module, indicating a reduction in fog concentration and improved visibility in the dehazed images. Furthermore, the perceptual quality of the generated images is enhanced, as evidenced by the decreasing BRISQUE, PaQ-2-PiQ, and UNIQUE values. This suggests that the addition of modules such as ranking loss, adversarial loss, and consistency loss leads to improved image quality in terms of naturalness and visual appeal. However, it is worth noting that the performance of Ours-RAF is significantly lower compared to the full model, highlighting the critical role of the consistency loss ($\mathcal{L}_{con}$) in our framework. Without $\mathcal{L}_{con}$, the model becomes overly reliant on the pseudo ground truths during retraining. Since pseudo ground truth images inherently lack the full detail and quality of real haze-free images, this dependency restricts the model's ability to generalize effectively, especially given the limited amount of real-world data in the dataset. Unlike other metrics, the NIQE value shows a slight increase after fine-tuning. This may be attributed to the model making larger modifications to the image during this phase, resulting in a higher score in terms of perceived image quality. Despite this slight increase, the overall visual improvement achieved by fine-tuning is evident from the other metrics.

A visual example from NHRW shown in Fig. 5 provide a more intuitive understanding. As depicted in (a), (b), and (c), the generated images progressively become clearer as the ranking loss regularization and the design of adversarial architecture are incorporated. This clearly demonstrates the superiority of these components in improving the quality of the dehazed images. However, it is evident that without the fine-tuning module, residual fog still persists in the generated images, as seen in the first row of images. We observe that the fine-tuned model produces images with significantly improved visibility in the second row of images. However, it is worth noting that in the absence of the consistency loss module, which constrains the model to retain the knowledge learned before, the resulting images appear overly brightened. This observation aligns with our previous analysis.

4.5. Generalization to daytime hazy images

To evaluate the generalization capability of our model to daytime hazy images, we conducted experiments on several real-word daytime datasets, including NH-HAZE [40], O-HAZE [41] and 23 well-known real-world hazy images (23 Daytime Images) [42], which are widely used in previous studies. The first two datasets have paired images, while the last one does not.

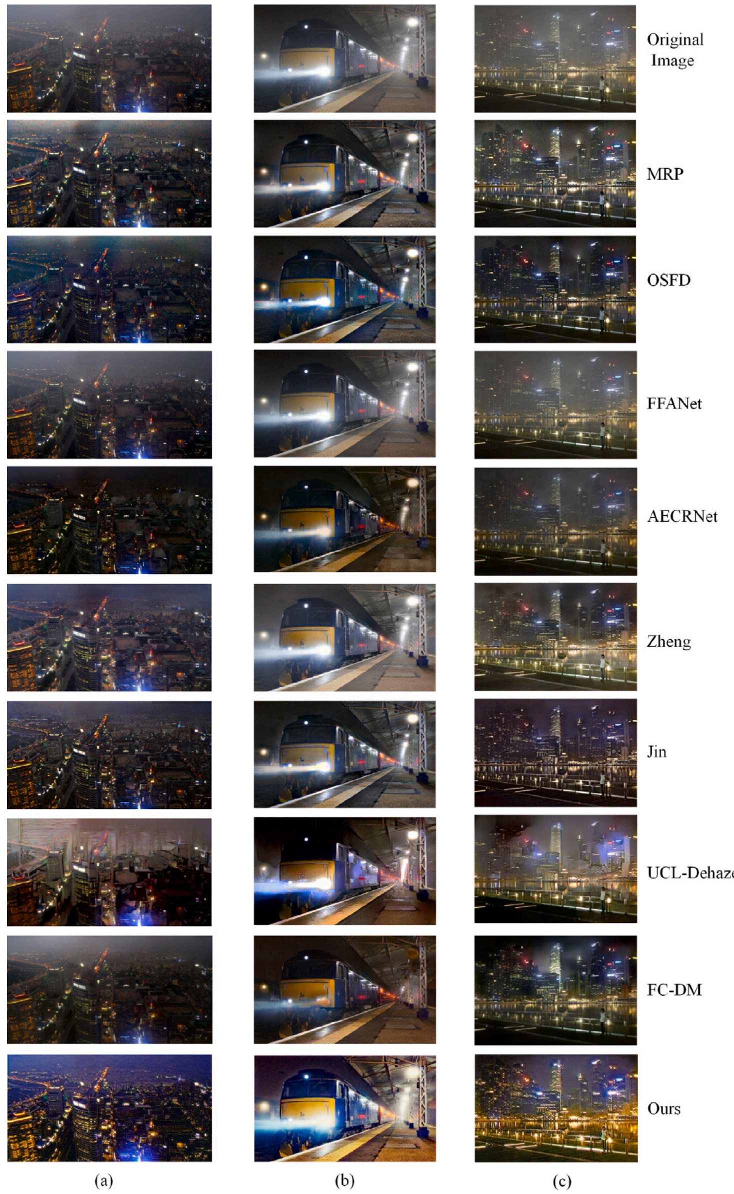


Fig. 4. Three visual examples of NHRW handled by MRP [7], OSFD [9], FFANet [31], AECRNet [30], Zheng [34], Jin [14], UCL-Dehaze [11], FD-DM [26] and our method.

Table 3
Ablation experiment results on real-world dataset NHRW. The best and second best results are indicated by bold and underlined respectively.

Model	$\mathcal{L}_{rank}$	$\mathcal{L}_{adv}$	FT	$\mathcal{L}_{con}$	FADE(↓)	NIQE(↓)	BRISQUE(↓)	PaQ-2-PiQ(↑)	UNIQUE(↑)
Base	×	×	×	×	0.99	14.72	49.57	55.82	−1.23
Ours-R	✓	×	×	×	0.74	14.57	39.68	65.21	−0.69
Ours-RA	✓	✓	×	×	0.52	**14.12**	36.36	65.85	−0.39
Ours-RFC	✓	×	✓	✓	0.32	15.93	22.33	69.65	−0.18
Ours-RAF	✓	✓	✓	×	0.34	15.22	27.87	55.82	−1.22
Ours-RAFC (full)	✓	✓	✓	✓	**0.21**	14.44	**22.23**	**70.21**	**0.32**

In addition to the previously mentioned nighttime defogging methods, we compared our method with SOTA daytime dehazing techniques, including Dark Channel Prior (DCP) [6], Zheng et al. (Zheng) [34], and Liu et al. (Liu) [43]. By comparing our method with these SOTA approaches, we can assess how well our model performs in the context of daytime dehazing.

The quantitative comparisons are shown in Table 4. The results demonstrate that our method achieves competitive performance across these datasets. Be noted that PSNR and SSIM only be applied to NH-HAZE and O-HAZE but not 23 Daytime Images due to the availability of image pairs. In terms of overall performance, Liu's daytime dehaze method [43] stands out as the best performer across the three datasets. It takes the first rank in five out of thirteen metrics and secures the second rank in two metrics. Although our model was primarily designed for nighttime hazy images and trained on a nighttime hazy dataset, it demonstrates impressive generalization ability in removing fog. It achieves the highest FADE values on the O-HAZE dataset and the second-highest values on the other two datasets, indicating its effectiveness in fog removal across different scenarios. Furthermore, the quality of the output images generated by our model is highly satisfying. For

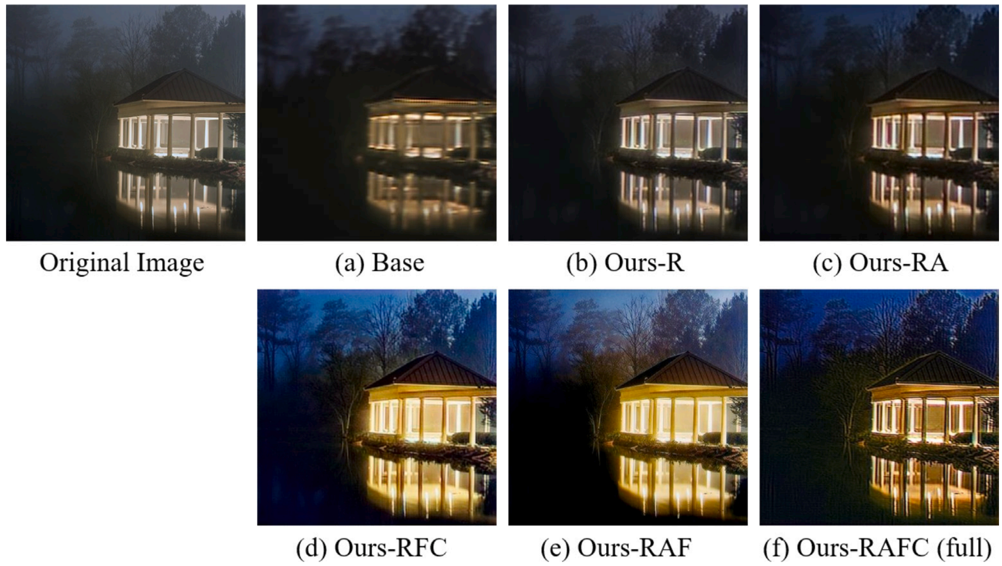


Fig. 5. Visual example of ablation experiment results in NHRW.

Table 4
Comparisons on real-world daytime hazing datasets: NH-HAZE [40], O-HAZE [41] and 23 daytime images [42]. The best and second best results are indicated by bold and underlined respectively.

Method	NH-HAZE [40]					O-HAZE [41]					23 daytime images [42]		
	PSNR	SSIM	FADE	NIQE	BRISQUE	PSNR	SSIM	FADE	NIQE	BRISQUE	FADE	NIQE	BRISQUE
DCP [6]	12.22	0.47	0.40	<u>10.18</u>	<u>17.81</u>	16.10	<u>0.65</u>	0.37	15.66	26.51	0.33	13.40	**19.35**
Zheng [34]	12.43	0.47	0.96	10.32	29.39	14.90	0.62	0.78	15.45	22.05	0.44	14.60	<u>20.36</u>
Liu [43]	11.59	0.45	**0.24**	**9.28**	**16.34**	14.36	0.55	<u>0.22</u>	<u>13.13</u>	26.88	**0.20**	**12.01**	23.22
MRP [7]	11.09	**0.61**	0.31	12.39	20.52	16.66	**0.70**	0.30	16.38	<u>21.79</u>	0.30	13.27	21.87
OSFD [9]	<u>12.82</u>	<u>0.55</u>	0.33	10.71	20.43	**17.30**	0.55	0.33	**10.71**	**20.43**	0.28	<u>13.05</u>	21.57
Jin [14]	**13.12**	0.40	0.83	16.62	52.25	<u>16.74</u>	0.61	0.52	18.76	41.50	0.40	20.02	32.37
Ours	12.31	0.49	<u>0.27</u>	12.07	25.44	16.32	0.58	**0.20**	14.55	31.23	<u>0.23</u>	15.64	34.01

instance, the PSNR values surpass those of all daytime dehaze methods on the O-HAZE datasets. This highlights the superior image quality achieved by our model in comparison to other methods specifically designed for daytime dehazing. Among the nighttime dehaze methods, MRP and OSFD also exhibit competitive performance. This can be attributed to the prior knowledge they utilize, which is statistically derived from a large number of haze-free daytime images. However, it is worth noting that these methods may underperform in nighttime dehazing tasks due to their reliance on daytime-specific priors.

Fig. 6 provides visual examples comparing the results generated by different dehazing methods. In the images generated by Liu and DCP, there is a noticeable reduction in light and color distortion. Zheng produces overall cleaner images, but there is still residual fog in the regions with thick haze. This suggests that their approach may not fully address the challenges posed by non-homogeneous haze. On the other hand, the nighttime dehazing methods, including MRP, OSFD and Jin, produce images with obvious artifacts. Given that our method was originally designed for nighttime hazy images, the dehazed images produced by our method for daytime hazy images appear brighter than the original input. However, they are not over-exposed. This is achieved through the use of histogram equalization, which limits contrast while enhancing the overall image quality. The majority of pixel values are concentrated in the center, resulting in improved visibility without sacrificing image details. Furthermore, we introduce the concept of consistency loss, which helps regulate the model to partially rely on the knowledge learned from real normal images. This regularization technique contributes to maintaining a balance between image enhancement and preserving the natural appearance of the scene.

The evaluations conducted on various daytime datasets demonstrate the effectiveness and versatility of our model in addressing the dehazing task. Our model performs well not only in the specific context of nighttime images but also in the more general case of daytime dehazing. This highlights the robustness of our approach and its ability to handle different lighting conditions and haze types.

4.6. Visualization on encoder $\mathcal{F}$

The encoder $\mathcal{F}$ used in contrastive learning is critical for training the dehazing generator $\mathcal{G}$, as the generated samples must be distinct from the original hazy images and closer to the ground truth in the space mapped by $\mathcal{F}$. This section evaluates the performance of the fixed $\mathcal{F}$, employed in previous contrastive learning-based models [30], and the dynamically updated $\mathcal{F}$ used in our method, in separating the generated samples from the ground truth during training. Improved separation increases the challenge for $\mathcal{G}$, leading to higher-quality dehazed images.

Let $\mathcal{F}_{fix}$ be the VGG19 model pre-trained on ImageNet, which will not be updated, while $\mathcal{F}_{dyn}$ be the same VGG19 model continuously updated in our model. The outputs of the second-to-last layer of the model on the dehazed images generated by our method, trained for 60 and 100 epochs, for 500 randomly selected samples and their corresponding ground truths, are mapped to a two-dimensional space using t-SNE. Fig. 7 shows the distributions of $\mathcal{G}$ trained for 60 and 100 epochs with fixed ($\mathcal{F}_{fix}$) and dynamic ($\mathcal{F}_{dyn}$) encoders. After only 60 epochs of training, $\mathcal{G}$ is able to generate dehazed images that are close to the ground truths in the space of $\mathcal{F}_{fix}$. In contrast, in $\mathcal{F}_{dyn}$, the generated samples remain partially distinguishable from the ground truths throughout the learning process, continuously providing useful guidance for the learning of $\mathcal{G}$. The results also indicate that the gap between the ground

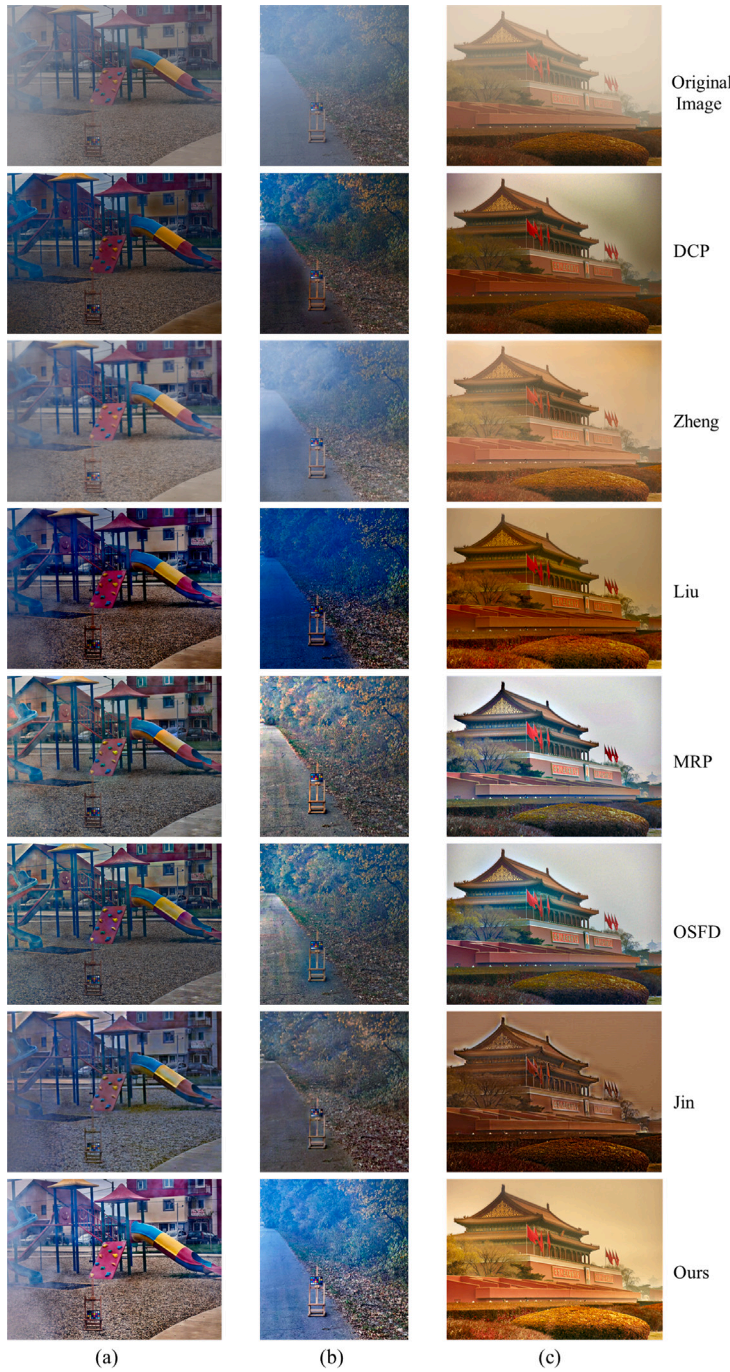
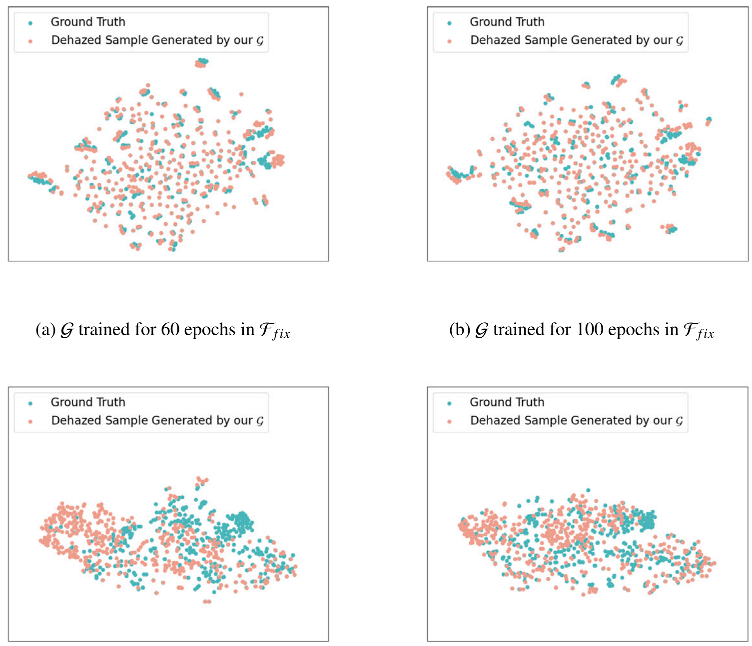


Fig. 6. A visual example of daytime hazy images from NH-HAZE, O-HAZE and 23 daytime hazy images dataset, handled by DCP [6], Zheng [34], Liu [43], MRP [7], OSFD [9], Jin [14] and our method.

truths and the samples generated by $\mathcal{G}$ after 100 epochs becomes more pronounced, as the performance of $\mathcal{G}$ improves with additional training epochs. This visualization may help explain why the quality of the dehazed images generated by our $\mathcal{G}$ is satisfactory.

5. Conclusion

In conclusion, our proposed deep learning framework for nighttime image dehazing successfully addresses the challenges posed by limited real hazy images and the disparity between synthetic and real data. By integrating contrastive and adversarial learning, our model fully exploits available information from the training data and generates dehazed images that closely resemble ground truths while remaining distinct from the original hazy images. Furthermore, the fine-tuning process using a small set of real nighttime hazy images, along with the introduced constraint, helps bridge the gap between synthetic and real data. Experimental evaluation against state-of-the-art methods on synthetic and real-world datasets demonstrates the superior performance of our model in nighttime and even daytime image dehazing. The framework effectively enhances visibility and reduces haze in nighttime images, offering a practical solution for various applications requiring image dehazing, regardless of the time of day. By leveraging



(a) $\mathcal{G}$ trained for 60 epochs in $\mathcal{F}_{fix}$ (b) $\mathcal{G}$ trained for 100 epochs in $\mathcal{F}_{fix}$

(c) $\mathcal{G}$ trained for 60 epochs in $\mathcal{F}_{dyn}$ (d) $\mathcal{G}$ trained for 100 epochs in $\mathcal{F}_{dyn}$

Fig. 7. t-SNE visualization of the feature space for dehazed images generated by $\mathcal{G}$, trained for 60 and 100 epochs, using the fixed encoder ($\mathcal{F}_{fix}$) and the encoder dynamically updated by our method ($\mathcal{F}_{dyn}$).

the advantages of both synthetic and real data, our framework showcases the potential of combining different data sources to improve the performance of image dehazing algorithms.

Without relying on a large amount of real data for training, our method demonstrates strong generalization ability to complex haze conditions in real-world scenes. However, our method still has some limitations. When generating pseudo ground truth for real nighttime hazy images during the fine-tuning process, histogram equalization is employed to improve nighttime hazy images. This traditional method is simple and effective, but it encounters challenges in generating high quality image that has low noise when the image has extreme low-light region. One possible future work is to enhance the quality of the pseudo ground truth for real nighttime hazy images during the fine-tuning process by incorporating an existing dehazing method into the pipeline to prepare the pseudo ground truth. By leveraging the capabilities of established dehazing algorithms, the accuracy and quality of the ground truth can be significantly improved, leading to more effective fine-tuning of our model. In addition, exploring the application of models in various low visualization scenarios is also of great practical significance. For example, smoke is similar to haze and fog. But it presents a unique challenge due to its uneven localized nature and severe reduction in visibility, often accompanied by a dark or brownish tint. A feasible way is to integrate an attention mechanism that prioritizes areas affected by degradation factors (e.g., haze, smoke), which could enhance our model's performance and generalization under the inhomogeneous scenarios.

CRediT authorship contribution statement

Jingwen Deng: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Conceptualization. **Patrick P.K. Chan:** Writing – review & editing, Supervision, Methodology, Conceptualization. **Daniel S. Yeung:** Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

References

[1] J. Xu, L. Zhang, Hazy remote sensing images semantic segmentation for weakly annotation based on saliency-aware alignment strategy, in: ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP, 2024, pp. 4105–4109, http://dx.doi.org/10.1109/ICASSP48485.2024.10445947.

[2] C. Li, H. Zhou, Y. Liu, C. Yang, Y. Xie, Z. Li, L. Zhu, Detection-friendly dehazing: Object detection in real-world hazy scenes, IEEE Trans. Pattern Anal. Mach. Intell. 45 (7) (2023) 8284–8295, http://dx.doi.org/10.1109/TPAMI.2023.3234976.

[3] S. Banerjee, S. Sinha Chaudhuri, Nighttime image-dehazing: a review and quantitative benchmarking, Arch. Comput. Methods Eng. 28 (2021) 2943–2975.

[4] J. Zhang, Y. Cao, Z. Wang, Nighttime haze removal based on a new imaging model, in: 2014 IEEE International Conference on Image Processing, ICIP, IEEE, 2014, pp. 4557–4561.

[5] S. Kuanar, D. Mahapatra, M. Bilas, K. Rao, Multi-path dilated convolution network for haze and glow removal in nighttime images, Vis. Comput. (2022) 1–14.

[6] K. He, J. Sun, X. Tang, Single image haze removal using dark channel prior, IEEE Trans. Pattern Anal. Mach. Intell. 33 (12) (2010) 2341–2353.

[7] J. Zhang, Y. Cao, S. Fang, Y. Kang, C. Wen Chen, Fast haze removal for nighttime image using maximum reflectance prior, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 7418–7426.

[8] Visual attention dehazing network with multi-level features refinement and fusion, Pattern Recognit. 118 (2021) 108021.

[9] J. Zhang, Y. Cao, Z.-J. Zha, D. Tao, Nighttime dehazing with a synthetic benchmark, in: Proceedings of the 28th ACM International Conference on Multimedia, 2020, pp. 2355–2363.
[10] J. Gui, X. Cong, Y. Cao, W. Ren, J. Zhang, J. Zhang, J. Cao, D. Tao, A comprehensive survey and taxonomy on single image dehazing based on deep learning, ACM Comput. Surv. 55 (13s) (2023) 1–37.
[11] Y. Wang, X. Yan, F.L. Wang, H. Xie, W. Yang, X.-P. Zhang, J. Qin, M. Wei, Ucl-dehaze: Towards real-world image dehazing via unsupervised contrastive learning, IEEE Trans. Image Process. (2024).
[12] P. Shyam, H. Yoo, Data efficient single image dehazing via adversarial auto-augmentation and extended atmospheric scattering model, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023, pp. 227–237.
[13] H. Israël, F. Kasten, Koschmieders theorie der horizontalen sichtweite, Die Sichtweite Nebel Die Möglichkeiten Ihrer Künstlichen Beeinflussung (1959) 7–10.
[14] Y. Jin, B. Lin, W. Yan, Y. Yuan, W. Ye, R.T. Tan, Enhancing visibility in nighttime haze images using guided apsf and gradient adaptive convolution, in: Proceedings of the 31st ACM International Conference on Multimedia, 2023, pp. 2446–2457.
[15] Z. Lyu, Y. Chen, Y. Hou, MCPNet: Multi-space color correction and features prior fusion for single-image dehazing in non-homogeneous haze scenarios, Pattern Recognit. 150 (2024) 110290.
[16] Y. Liu, Z. Yan, J. Tan, Y. Li, Multi-purpose oriented single nighttime image haze removal based on unified variational retinex model, IEEE Trans. Circuits Syst. Video Technol. 33 (4) (2022) 1643–1657.
[17] E.H. Land, J.J. McCann, Lightness and retinex theory, Josa 61 (1) (1971) 1–11.
[18] B. Li, Y. Gou, J.Z. Liu, H. Zhu, J.T. Zhou, X. Peng, Zero-shot image dehazing, IEEE Trans. Image Process. 29 (2020) 8457–8466.
[19] Y.Z. Su, C. He, Z.G. Cui, A.H. Li, N. Wang, Physical model and image translation fused network for single-image dehazing, Pattern Recognit. 142 (2023) 109700.
[20] Y. Liu, A. Wang, H. Zhou, P. Jia, Single nighttime image dehazing based on image decomposition, Signal Process. 183 (2021) 107986.
[21] W. Yan, R.T. Tan, D. Dai, Nighttime defogging using high-low frequency decomposition and grayscale-color networks, in: European Conference on Computer Vision, Springer, 2020, pp. 473–488.
[22] N. Jiang, K. Hu, T. Zhang, W. Chen, Y. Xu, T. Zhao, Deep hybrid model for single image dehazing and detail refinement, Pattern Recognit. 136 (2023) 109227.
[23] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial networks, Commun. ACM 63 (11) (2020) 139–144.
[24] J. Ho, A. Jain, P. Abbeel, Denoising diffusion probabilistic models, Adv. Neural Inf. Process. Syst. 33 (2020) 6840–6851.
[25] M. Fu, H. Liu, Y. Yu, J. Chen, K. Wang, Dw-gan: A discrete wavelet transform gan for nonhomogeneous dehazing, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 203–212.
[26] J. Wang, S. Wu, Z. Yuan, Q. Tong, K. Xu, Frequency compensated diffusion model for real-scene dehazing, Neural Netw. 175 (2024) 106281.
[27] J. Xiong, X. Yan, Y. Wang, W. Zhao, X.-P. Zhang, M. Wei, RSHazeDiff: A unified Fourier-aware diffusion model for remote sensing image dehazing, 2024, arXiv preprint arXiv:2405.09083.
[28] X. Chen, Z. Fan, P. Li, L. Dai, C. Kong, Z. Zheng, Y. Huang, Y. Li, Unpaired deep image dehazing using contrastive disentanglement learning, in: European Conference on Computer Vision, Springer, 2022, pp. 632–648.
[29] Y. Shao, L. Li, W. Ren, C. Gao, N. Sang, Domain adaptation for image dehazing, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 2808–2817.
[30] H. Wu, Y. Qu, S. Lin, J. Zhou, R. Qiao, Z. Zhang, Y. Xie, L. Ma, Contrastive learning for compact single image dehazing, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 10551–10560.
[31] X. Qin, Z. Wang, Y. Bai, X. Xie, H. Jia, FFA-Net: Feature fusion attention network for single image dehazing, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 34, 2020, pp. 11908–11915.
[32] R. Wang, R. Shivanna, D. Cheng, S. Jain, D. Lin, L. Hong, E. Chi, Dcn v2: Improved deep & cross network and practical lessons for web-scale learning to rank systems, in: Proceedings of the Web Conference 2021, 2021, pp. 1785–1797.
[33] A.M. Reza, Realization of the contrast limited adaptive histogram equalization (CLAHE) for real-time image enhancement, J. VLSI Signal Process. Syst. Signal Image Video Technol. 38 (2004) 35–44.
[34] Z. Zheng, W. Ren, X. Cao, X. Hu, T. Wang, F. Song, X. Jia, Ultra-high-definition image dehazing via multi-guided bilateral learning, in: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR, IEEE, 2021, pp. 16180–16189.
[35] L.K. Choi, J. You, A.C. Bovik, Referenceless prediction of perceptual fog density and perceptual image defogging, IEEE Trans. Image Process. 24 (11) (2015) 3888–3901, http://dx.doi.org/10.1109/TIP.2015.2456502.
[36] A. Mittal, R. Soundararajan, A.C. Bovik, Making a "completely blind" image quality analyzer, IEEE Signal Process. Lett. 20 (3) (2013) 209–212, http://dx.doi.org/10.1109/LSP.2012.2227726.
[37] A. Mittal, A.K. Moorthy, A.C. Bovik, No-reference image quality assessment in the spatial domain, IEEE Trans. Image Process. 21 (12) (2012) 4695–4708, http://dx.doi.org/10.1109/TIP.2012.2214050.
[38] Z. Ying, H. Niu, P. Gupta, D. Mahajan, D. Ghadiyaram, A. Bovik, From patches to pictures (paq-2-piq): Mapping the perceptual space of picture quality, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 3575–3585.
[39] W. Zhang, K. Ma, G. Zhai, X. Yang, Uncertainty-aware blind image quality assessment in the laboratory and wild, IEEE Trans. Image Process. 30 (2021) 3474–3486.
[40] C.O. Ancuti, C. Ancuti, R. Timofte, NH-HAZE: an image dehazing benchmark with non-homogeneous hazy and haze-free images, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, in: IEEE CVPR 2020, 2020.
[41] C.O. Ancuti, C. Ancuti, R. Timofte, C. De Vleeschouwer, O-haze: a dehazing benchmark with real hazy and haze-free outdoor images, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2018, pp. 754–762.
[42] R. Fattal, Dehazing using color-lines, 34, (13) ACM, New York, NY, USA, 2014,
[43] X. Liu, H. Li, C. Zhu, Joint contrast enhancement and exposure fusion for real-world image dehazing, IEEE Trans. Multimed. 24 (2022) 3934–3946, http://dx.doi.org/10.1109/TMM.2021.3110483.

Jingwen Deng received the bachelor's degree from South China University of Technology in 2023. She is currently studying for the M.S. degree in the School of Future Technology, South China University of Technology. Her research interests include deep learning and image editing.

Patrick P.K. Chan received the Ph.D. degree from Hong Kong Polytechnic University, Hong Kong, in 2009. He is currently serving as Deputy Dean and Associate Professor at the Shien-Ming Wu School of Intelligent Engineering, South China University of Technology, Guangzhou, China. His research interests include machine learning security, adversarial learning, deep learning, and large language models (LLMs). Dr. Chan serves as Editor-in-Chief Assistant and Associate Editor of Int'l Journal of Machine Learning and Cybernetics, and Associate Editors of IEEE Trans. on Cybernetics, Information Sciences and Advances in Computational Intelligence. He has actively involved in international activities. He was the Board of Director member of IEEE Systems, Man, and Cybernetics Society (SMCS) and the Chair of IEEE SMCS Hong Kong Chapter. He also serves as Organizing Committee Chairs of several international conferences.

Daniel S. Yeung received the Ph.D. degree from Case Western Reserve University, Cleveland, OH, USA. He was a Faculty Member with the Rochester Institute of Technology, Rochester, NY, USA, from 1974 to 1980. In the next ten years, he held industrial and business positions in the USA. In 1989, he joined the Department of Computer Science, City Polytechnic of Hong Kong, Hong Kong, as an Associate Head/Principal Lecturer. He served as the Founding Head and Chair Professor with the Department of Computing, Hong Kong Polytechnic University, Hong Kong, until his retirement in 2006. His research interests include neural network sensitivity analysis, large scale data retrieval problem, and cyber security.