



GenHaze: Pioneering Controllable One-Step Realistic Haze Generation for Real-World Dehazing

Sixiang Chen¹ Tian Ye¹ Yunlong Lin² Yeying Jin³ Yijun Yang¹
Haoyu Chen¹ Jianyu Lai¹ Song Fei¹ Zhaohu Xing¹ Fugee Tsung^{1,4} Lei Zhu^{1,4*}

¹The Hong Kong University of Science and Technology (Guangzhou)

²Xiamen University ³Tencent

⁴The Hong Kong University of Science and Technology

Generative Model Empowers Realistic Haze Generation, Driving Real-World Dehazing to Peak!



Figure 1. **We propose GenHaze, a controllable one-step haze generation model.** The core principle of GenHaze is to use the existing latent diffusion models (LDM) [67], and adapt it to the haze generation task via our clean-to-haze generation protocol. Based on GenHaze, we can significantly unleash the real-world dehazing potential of existing baselines in a simple way!

Abstract

Real-world image dehazing is crucial for enhancing visual quality in computer vision applications. However, existing physics-based haze generation paradigms struggle to model the complexities of real-world haze and lack controllability, limiting the performance of existing baselines on real-world images. In this paper, we introduce **GenHaze**, a pioneering haze generation framework that enables the one-step generation of high-quality, reference-controllable hazy images. GenHaze leverages the pre-trained latent diffusion model (LDM) with a carefully designed clean-to-haze generation protocol to produce realistic hazy images. Additionally, by leveraging its fast, controllable generation of paired high-quality hazy images, we illustrate that existing dehazing

baselines can be unleashed in a simple and efficient manner. Extensive experiments indicate that GenHaze achieves visually convincing and quantitatively superior hazy images. It also significantly improves multiple existing dehazing models across 7 non-reference metrics with minimal fine-tuning epochs. Our work demonstrates that LDM possesses the potential to generate realistic degradations, providing an effective alternative to prior generation pipelines.

1. Introduction

Image dehazing aims to restore a clean image from a haze image, ultimately enhancing real-world visual quality. However, acquiring paired real-world haze and clear images is virtually impossible. Consequently, existing methods

*corresponding authors.

rely heavily on synthetic datasets constructed using physics-based haze imaging models [38, 56]. Despite achieving state-of-the-art performance on these synthetic datasets, many advanced methods [19, 22, 63, 72] cannot generalize effectively to real-world hazy due to the inability to precisely control the parameters within these imaging models.

To address this issue, some works have attempted manual fine-tuning [79] or incorporated additional factors [21] to develop novel generation pipelines. Nevertheless, they remain constrained by the limitations of the physical imaging formulas. Existing pipelines struggle to encompass the diverse factors contributing to haze, and synthetic methods based on physical models are difficult to control due to the difficulty in precisely manipulating all the involved parameters. Meanwhile, complicated model designs and training strategies [14, 21, 79] have been proposed to enhance real-world dehazing performance, yet challenges remain.

To address these challenges, we propose a novel assumption: real hazy images can serve as references to generate high-quality, similar hazy images from clean backgrounds. Based on it, using these hazy images to fine-tune existing baseline models should significantly improve dehazing performance in a simple (i.e., without modifying architectures) and efficient (i.e., requiring only 1–2 epochs) manner. Therefore, this paper explores how to achieve high-quality, controllable haze generation and effectively enhance the performance of existing dehazing models.

We are inspired by the observation that current pre-trained diffusion models [62, 67] are trained on vast and diverse natural images, and are designed to generate high-quality visual images. Therefore, we raise a question: *If clean images can be derived from haze images, conversely, can the high-quality image generation capabilities of diffusion models be extended to generate haze degradation?* If so, we should be able to obtain a broadly applicable and controllable haze generator from pre-trained diffusion models, which can replace existing synthetic methods. In this work, we explore this possibility and develop **GenHaze**, an LDM [67] based method, which utilizes our clean-to-haze generation protocol to produce high-quality, controllable hazy images in a single step. Moreover, GenHaze can enhance the dehazing performance of existing baseline models in a straightforward and rapid manner (see Tab.1).

Our experiments demonstrate that the hazy images generated by GenHaze achieve highly realistic visual comparisons and quantitative results. Furthermore, a small number of fine-tuning epochs for existing baseline models with our generated hazy images leads to considerable improvements across **7 common non-reference metrics**. This also validates our above assumptions that existing dehazing baselines have sufficient potential to be unlocked. In summary, our contributions are as follows:

- By using the clean-to-haze generation protocol, we pro-

pose GenHaze, a pioneering haze generation pipeline based on pre-trained LDM, which is capable of producing high-quality, controllable hazy images in one step.

- We show that simply fine-tuning dehazing baseline models in 1-2 epochs with our generated hazy images significantly improves their performance, without modifying their architectures.
- GenHaze attains SOTA performance in generating realistic haze, while assisting existing baselines to achieve significant gains in 7 non-reference metrics, highlighting its practical value.

2. Related Works

Image Generation. Text-to-image diffusion models [1, 11, 23, 24, 39, 40, 80] became central to generating high-quality images from prompts. Among them, Stable Diffusion [67] excelled by performing diffusion in latent space (LDM). Recent works further extended its capabilities: DreamBooth [68] and Custom-Diffusion [36] fine-tuned models for specific content, while T2I-Adapter [57] and ControlNet [87] improved precise image control. For other tasks, InstructPix2Pix [4] adapted Stable Diffusion for image editing, and CycleGAN-Turbo [59] integrated GANs for diverse image translation. GenDeg [65] leveraged Instruct-Pix2Pix to construct large-scale degradation datasets. In this paper, we focus on the one-step generation of realistic haze images with controllable references. We further show that our carefully designed pipeline effectively supports a simple and efficient paradigm for baseline fine-tuning.

Diffusion Acceleration. Recent advances in diffusion model acceleration focus on two strategies: fast samplers and diffusion distillation. Fast samplers [35, 52, 53, 71] reduced sampling steps from thousands to 10-50. Diffusion distillation methods [2, 54, 69] trained student models to replicate teacher outputs within 1-4 steps. Early methods like [54] required pre-computed trajectories, while Progressive Distillation [69] and Consistency Distillation [55, 73] improved efficiency via step reduction and consistency constraints. Adversarial Distillation [70, 81] and Variational Distillation [86] further enhanced quality by reducing artifacts through discriminators and probabilistic frameworks.

Single Image Dehazing. Similar to other image restoration methods [7–9, 31, 32, 37, 41, 43–46, 84], single-image dehazing aims to restore clear images from hazy ones. Learning-based methods have dominated this field [5, 16, 19, 30, 47, 48, 51, 84]. For example, MSBDN [19] leveraged boosting and error feedback for iterative refinement. Recently, transformer and diffusion models further advanced this task [12, 22, 63, 72, 82], while adverse weather restoration methods also demonstrated dehazing capabilities [12, 13, 75, 76, 85, 89]. However, recent research has shifted toward real-world haze, typically requiring specialized designs and training yet still facing performance limitations.

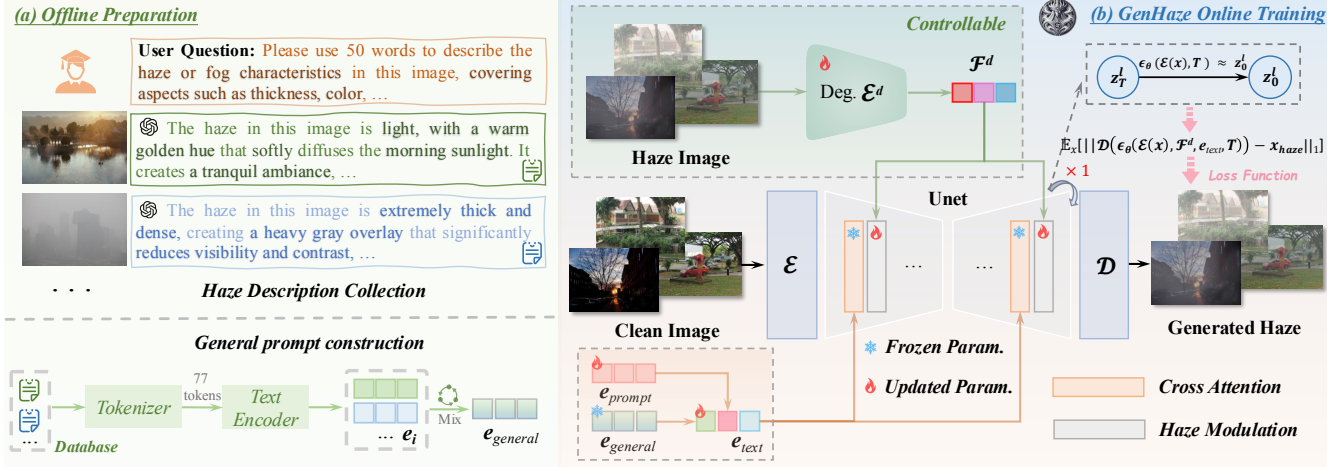


Figure 2. **Overview of the proposed GenHaze and dehazing baseline fine-tuning.** (a) and (b) illustrate our targeted fine-tuning strategy based on LDM [67], the clean-to-haze generation protocol (from text embedding and reference-based control perspectives). It converts SD to enable fast, controllable, and high-quality haze generation with a one-step strategy.

While many models excelled on synthetic datasets, their effectiveness often failed in real scenarios, also posing a huge challenge for the plug-and-play approach [15]. In this work, we aim to significantly enhance real-world performance of existing dehazing models without structural modifications. By leveraging the power of LDM, we bridge this gap and unlock their potential for real-world tasks.

3. GenHaze Pipeline

Generative Model Formulation: Diffusion models generate high-quality samples by gradually refining noise into structured data. DDPM [27] maps a simple noise distribution p_T to a complex data distribution p_0 . In the forward process, Gaussian noise is incrementally added over T time steps, gradually corrupting the initial data \mathbf{x}_0 to produce a noisy sample \mathbf{x}_t at each step t , defined as:

$$\mathbf{x}_t = \sqrt{\bar{\alpha}_t} \mathbf{x}_0 + \sqrt{1 - \bar{\alpha}_t} \boldsymbol{\epsilon}, \quad (1)$$

where $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \mathbf{I})$ is Gaussian noise, and $\bar{\alpha}_t = \prod_{\tau=1}^t \alpha_\tau$ with $\alpha_t = 1 - \beta_t$. The reverse process, executed by a denoising network, progressively refines \mathbf{x}_t to reconstruct the original data by predicting \mathbf{x}_{t-1} from noisy inputs.

The training objective minimizes the discrepancy between actual noise $\boldsymbol{\epsilon}$ and predicted noise $\hat{\boldsymbol{\epsilon}}_\theta(\mathbf{x}_t, t)$:

$$\mathcal{L} = \mathbb{E}_{\mathbf{x}_0, \boldsymbol{\epsilon} \sim \mathcal{N}(0, \mathbf{I}), t \sim \mathcal{U}(T)} [\|\boldsymbol{\epsilon} - \hat{\boldsymbol{\epsilon}}_\theta(\mathbf{x}_t, t)\|^2]. \quad (2)$$

Latent Diffusion Model (LDM) operate in the latent space of a Variational Autoencoder (VAE), which consists of an encoder \mathcal{E} and a decoder \mathcal{D} that map data to a lower-dimensional latent space and back, where $\mathcal{D}(\mathcal{E}(\mathbf{x})) \approx \mathbf{x}$. In LDM, noise is added to the latent representation rather than the data itself, reducing computational complexity:

$$\mathbf{z}_t^l = \sqrt{\bar{\alpha}_t} \mathbf{z}^l + \sqrt{1 - \bar{\alpha}_t} \boldsymbol{\epsilon}, \quad (3)$$

where $\mathbf{z}^l = \mathcal{E}(\mathbf{x})$ is the latent representation of \mathbf{x} .

Effective One-Step Strategy: This paper aims to adapt the SD model as an efficient haze generator, which can be utilized to enhance the generalization of existing dehazing models. By efficiently generating controllable realistic haze images, we can improve downstream dehazing performance without adding significant computational cost. However, traditional SD models [62, 67] employ multi-step sampling, which introduces significant computational overhead. Thus, we target a reduction to a single inference step to achieve optimal efficiency.

Firstly, we choose SD-Turbo as our backbone, which condenses the multi-step sampling process to 1 – 4 steps via adversarial distillation [70], forming a solid basis for single-step inference. For haze generation, it is essential to generate haze while preserving background consistency with the input clean image, whose core is similar to InstructPix2Pix [4]. However, this task is challenging due to the distributional discrepancy between pure Gaussian noise and clean images. Fortunately, in the sampling phase, we can apply the trailing setting [42, 52] by fixing the timestep $t = T$. Rather than decrementing t step-by-step, we set t directly to its maximum value, T , enabling one-step generation. To maintain consistency and simulate this one-step sampling condition during training, we replace traditional Gaussian noise with the clean image \mathbf{x} as input. Notably, both the clean image and Gaussian noise share a mean of zero, making this substitution logically sense to some extent. This setup is defined as follows:

$$\mathbf{z}_T^l = \mathcal{E}(\mathbf{x}), \quad (4)$$

in a standard LDM, the denoising target $\epsilon_\theta(\mathbf{z}_t^l, t)$ approximates the noise $\boldsymbol{\epsilon}$ added at timestep t . In our single-step setting, we simplify this by training the model to predict the latent representation of the hazy image, $\mathbf{z}_0^l = \mathcal{E}(\mathbf{x}_{\text{hazy}})$,

directly from \mathbf{z}_T^l . The predicted distribution is:

$$p(\mathbf{x}_0|\mathcal{E}(\mathbf{x}), \mathcal{F}^d, e^{\text{text}}) = \mathcal{N}(\mu_\theta(\mathcal{E}(\mathbf{x}), \mathcal{F}^d, e^{\text{text}}, T), \sigma_\theta^2 \mathbf{I}), \quad (5)$$

where $\mu_\theta(\mathcal{E}(\mathbf{x}), \mathcal{F}^d, e^{\text{text}}, T)$ is the model’s output, approximating \mathbf{z}_0^l , and \mathcal{F}^d and e^{text} are inputs for haze control and prompt embeddings, respectively (see Sec.3.1).

To supervise one-step forward pass, we measure the difference between the decoded image and the target image:

$$\mathcal{L}_{\text{single-step}} = \mathbb{E}_{\mathbf{x}} \left[\left\| \mathcal{D}(\mu_\theta(\mathcal{E}(\mathbf{x}_{\text{clean}}), \mathcal{F}^d, e^{\text{text}}, T)) - \mathbf{x}_{\text{hazy}} \right\|_1 \right], \quad (6)$$

this formulation enables to learn a direct mapping from the clean image to the hazy image in a single inference step. By bypassing multi-step denoising, we achieve computational efficiency while retaining background details and ensuring consistency with the clean image.

3.1. Clean-to-Haze Generation Protocol

In this section, we introduce generation protocol to adapt SD-Turbo to haze generation. Fig.2 outlines this procedure. **Text Embedding with Offline Haze Database:** For controllable haze generation, *i)* manually annotating captions for each image is impractical, and using identical fixed text reduces generation quality. *ii)* Real-world haze characteristics are complex and hard to explicitly describe through text, while generating embeddings from individual captions online also slows generation. Considering these limitations, we ask: *Could we initially provide a general embedding that encompasses most haze-related attributes, and allowing the network to adaptively adjust to obtain a suitable final embedding?*

To address this challenge, we observe that SOTA MLLMs, such as GPT-4o [28], can effectively provide comprehensive descriptions for real-world haze images of varying degrees, as illustrated in Fig.3. Therefore, we utilize real haze images $\{\mathcal{I}_i\}_{i=1}^{\mathcal{N}}$, where $\mathcal{N} = 4000$, and employ GPT-4o as a captioning function $f^{\text{cap}}(\cdot)$ to generate captions $\{\mathcal{C}_i\}$ offline, where $\mathcal{C}_i = f^{\text{cap}}(\mathcal{I}_i), i = 1, 2, \dots, \mathcal{N}$. This allows us to construct a text database that contains general descriptions of different attributes related to many real haze images. Furthermore, we use CLIP [64] to map these captions to embeddings, which can be expressed as follows: $e_i = f^{\text{emb}}(\mathcal{C}_i)$. Notably, leveraging the linear space properties of CLIP embeddings [3], we can average these multiple haze embeddings to obtain a single embedding enriched with prior knowledge of real-world haze:

$$e^{\text{general}} = \frac{1}{\mathcal{N}} \sum_{i=1}^{\mathcal{N}} e_i, \quad (7)$$

where e^{general} encapsulates the diverse and complicated features found under various haze conditions. For different



Figure 3. Example descriptions generated by GPT-4o. It is capable of providing targeted captions for different characteristics of real haze, including thickness, color, light diffusion, atmospheric layering, and other related attributes.

haze images that we need to generate, each haze image possesses numerous common characteristics, such as density, color, depth gradients, and light scattering, thus making it reasonable to have an embedding that reflects this diversity.

In our pipeline, we remove the CLIP encoder [64] and directly initialize a learnable prompt embedding e^{prompt} . We then add this learnable embedding to the previously obtained general embedding to enable instance-specific adaptation. Also, we include positional encoding [74] $\mathcal{P}(\cdot)$ to consolidate semantic position:

$$e^{\text{text}} = e^{\text{general}} + \mathcal{P}(e^{\text{prompt}}), \quad (8)$$

this approach enables the network to freely adjust apposite embeddings for each image, while the rich prior knowledge from the database guides the learning direction of the learnable embedding. By integrating e^{text} into the model, the generated images effectively represent the right attributes of different haze scenarios without hand-craft writing.

Degradation Encoder and Haze Control: Controllability is key in haze image generation. Using a real haze image as a reference to generate similar haze effects for model fine-tuning can intuitively boost performance in downstream dehazing tasks. This outperforms uncontrolled haze generation through arbitrary parameter adjustments, which may introduce distribution mismatches and misguide model optimization. By referencing real haze images, we maintain consistency in haze characteristics, promoting more effective learning and enhancing model performance.

Leveraging real haze images as generation references is thus crucial. ControlNet [87] offers a straightforward way to guide generation, but it has notable limits: *i)* It replicates the UNet encoder, significantly reducing speed and increasing parameters; *ii)* It requires multi-step generation for high quality, which impairs performance in single-step strategies. In this work, we argue that haze generation does not demand such extensive parameters, given that

clean images serve as a foundation. Instead, we propose a lightweight degradation encoder with residual blocks to efficiently extract instance-level haze features, \mathcal{F}_d , from large training sets. The degradation encoder is formulated as:

$$\mathcal{F}^d = \mathcal{E}^d(\mathcal{I}^{\text{haze}}; \theta^d), \quad (9)$$

where $\mathcal{I}^{\text{haze}}$ is the haze image and θ^d represents the encoder parameters.

The next challenge is effectively integrating extracted haze features into SD-Turbo’s UNet, as these features often contain background information. Specifically, our unpaired generation hinders direct learning (e.g., via addition like ControlNet [87]) to isolate haze features from background interference. Inspired by style-transfer methods [17, 33, 34], we claim that using their technology to further transfer the characteristics of haze is a better choice. Thus, we apply the modulation-demodulation convolution from StyleGANv2 [34] as haze modulation, injecting haze-related features while reducing background interference. However, unlike most approaches [60, 61, 66] that tie modulation to the decoder, certain blocks at various SD stages influence generated style [77]. Therefore, we insert haze modulation after each stage, enabling real-time adjustment during generation. To be precise, given a degradation feature \mathcal{F}_i^d , we transform it into an embedding e_i^d and compute a style vector s to modulate each input feature map in the convolution layer. For each convolutional weight w_i , the modulated weight \hat{w}_i is defined as:

$$\hat{w}_i = s_i \cdot w_i, \quad (10)$$

where s_i is the modulation factor derived from the feature map after the i -th stage. To ensure stability and normalize the feature representations, we apply a demodulation step, rescaling the modulated weights for unit standard deviation:

$$\hat{w}'_i = \frac{\hat{w}_i}{\sqrt{\sum_i (\hat{w}_i)^2 + \epsilon}}, \quad (11)$$

where ϵ is a small constant for numerical stability. Finally, these demodulated weights \hat{w}'_i are used in the conv. layer, embedding haze-specific features into the network. This haze modulation enables dynamic adjustment at each stage, aligning the generated haze with target real-world style.

4. Advance Dehazing Baselines with GenHaze

Most existing dehazing models achieve state-of-the-art results on synthetic datasets [19, 63, 72], showing strong potential. However, the domain gap between synthetic and real-world haze limits their effectiveness. One approach to bridging this gap is synthesizing a large-scale dataset using more realistic physical models and training a new model from scratch [21, 79], but this is costly and cannot fully capture natural haze complexity. Another strategy is test-time adaptation (TTA) [15], adjusting network parameters

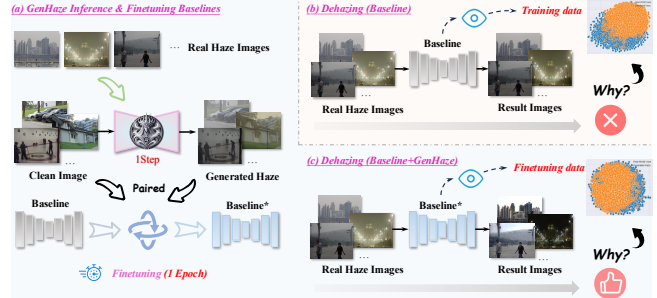


Figure 4. **The process of advancing dehazing baselines.** Leveraging the strengths of GenHaze, (a), (b) and (c) demonstrate a simple yet effective approach to unlock the potential of existing baselines in real-world scenarios.

per image, but this often proves ineffective when baselines entirely fail on real-world samples.

Leveraging GenHaze, we can generate realistic, controllable haze images from real references. We claim that dehazing baselines are capable of handling real haze with only minor adjustments. Fine-tuning baselines for 1-2 epochs on GenHaze-generated images resembling target real haze significantly could boost their performance (Tab. 2). Notably, our method allows processing of any number of images simultaneously, unlike TTA, which processes images individually. Moreover, generating multiple haze images from a single reference enhances model robustness and enables scalable dataset creation, highlighting GenHaze’s versatility. Mathematically, given j real haze images $\{r_j\}$, we generate i haze images $\{y^i\}$ from $\{x^i\}$ for each real image using GenHaze. For a baseline model f_θ , the fine-tuning process is defined as:

$$\theta' = \arg \min_{\theta} \sum_j \sum_i \mathcal{L}(f_\theta(y_j^i), x_j^i), \quad (12)$$

where \mathcal{L} represents the dehazing loss, and x_j^i is the ground truth clean image corresponding to the generated haze image y_j^i . This fine-tuning process adjusts model parameters θ over a limited number of epochs to improve handling of real-world haze. The dehazing results transition from the original model output $\hat{x}_{\text{base}} = f_\theta(\{r_j\})$ to the fine-tuned output $\hat{x}_{\text{base}'} = f_{\theta'}(\{r_j\})$, achieving significantly better adaptation to real-world haze samples. Please see the suppl. for more theoretical explanations on better fine-tuning.

5. Experiments

Implementation: In GenHaze, we construct a haze database using diverse real haze images (specific settings in suppl.). For the degradation encoder, we use a 4-residual block backbone to extract relevant features. To better adapt natural image generation to haze scenarios, we fine-tune UNet and VAE using LoRA ($r = 8$ and $r = 4$, respectively) on 512×512 images. Additionally, we incorporate skip connections [59] within the VAE’s encoder-decoder to

Table 1. **Quantitative comparison of GenHaze with existing degradation generation baselines on paired synthetic and unpaired real benchmarks.** Red triangle \triangle indicates the degradation generation pipeline adopted from the referenced work. Bold numbers denote the best performance, while underlined values indicate the second best.

Method	Type	Controllable	Haze4K [50]					RTTS [38]	
			PSNR [29] \uparrow	SSIM [78] \uparrow	LPIPS [88] \downarrow	FID [25] \downarrow	sFID [58] \downarrow	FID [25] \downarrow	sFID [58] \downarrow
RIDCP \triangle [CVPR'23] [79]	Physics-based	N	-	-	-	-	-	<u>59.21</u>	108.81
UWNR [CVPRW'22] [83]	Regression	Y	20.04	0.6115	0.3082	94.79	180.83	75.07	114.59
Low-Res \triangle [CVPR'24] [6]	Regression	Y	<u>20.87</u>	<u>0.9338</u>	<u>0.0555</u>	22.94	57.33	60.65	111.76
InstructPix2Pix [CVPR'23] [4]	Diffusion (20 steps)	N	14.38	0.6283	0.4989	180.09	308.10	115.70	173.33
CycleGAN-Turbo [Arxiv'24] [59]	GAN	N	18.64	0.8536	0.1687	<u>22.80</u>	<u>56.34</u>	60.14	<u>108.73</u>
GenHaze (Ours)	Diffusion (1 step)	Y	27.06	0.9662	0.0309	15.92	47.76	57.12	105.43

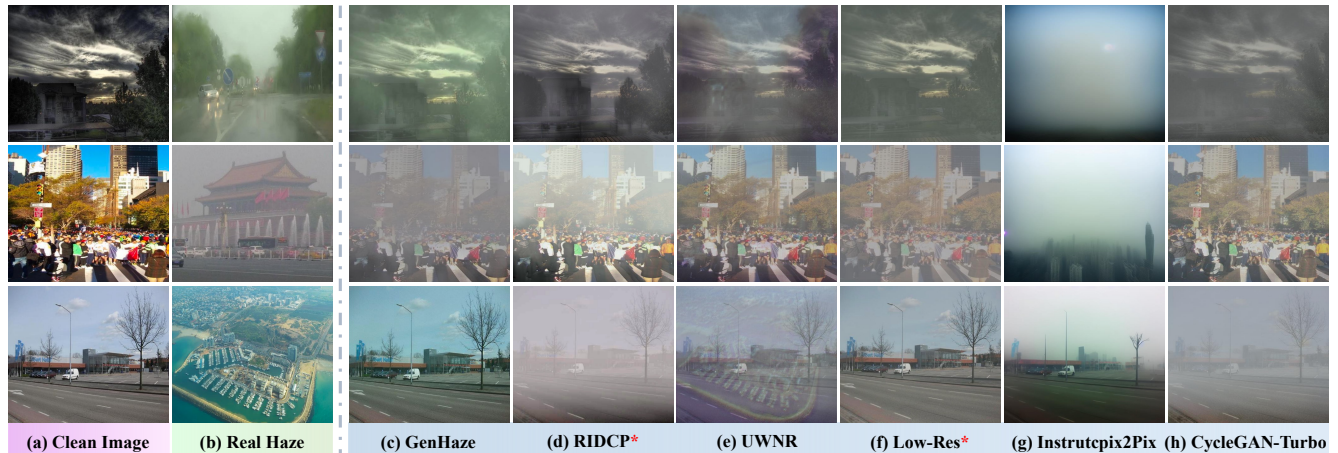


Figure 5. Visual comparisons of haze generation based on RTTS [38] and Fattal’s [20] datasets. Please zoom in for a better view. Our generated haze more closely matches the intensity and color of (b) real-world haze.

preserve finer details. Training runs on two A800 GPUs for 200K steps with Adam optimizer [18], using a batch size of 10 and a learning rate of 0.00008. For inference, we use the EulerDiscreteScheduler sampler [35] in trailing mode for single-step inference. For downstream fine-tuning, we only fine-tune dehazing baselines for 1-2 epochs. Due to our reference controllable design and speed pursuit, GenHaze generates 512×512 images without CFG [26], which are then resized to 256×256 for dehazing fine-tuning. No modifications are made to baselines, we only use generated paired images for direct fine-tuning to demonstrate the effectiveness of our approach. Further details are in the supp..

Datasets, Baselines and Metrics: For the training dataset, we use clean images from the Allweather dataset [76] to construct a paired haze dataset of 18,069 images via the RIDCP pipeline [79]. Clean images are fed into the SD-Turbo backbone to generate high-quality haze results, while haze images serve as references for controlling haze generation. For evaluation, we use the paired Haze4k [49] dataset and real-world haze dataset. For the latter, clean images are randomly selected to form an unpaired generation dataset. Haze4k enables us to assess pixel-level metrics, including PSNR [29], SSIM [78], and LPIPS [88], evaluating GenHaze’s generation capabilities. For haze generation, FID [25] and sFID [58] scores applied to unpaired real haze images measure the effectiveness and generalization. Baselines included available degradation generation methods [4, 6, 59, 79, 83], retrained on our dataset (ex-

cept Instructpix2pix [4], which requires text pairing) for fair comparison. In addition, we select 8 SOTA dehazing baselines, including CNN-based [16, 19], Transformer-based [22, 63, 72], Diffusion-based [12], and real-world dehazing methods [21, 79]. Evaluation is conducted using 7 common non-reference metrics on two well-known real haze datasets [20, 38], with before-and-after gains reported to thoroughly assess the improvements achieved by GenHaze. Additional details are provided in the supp..

5.1. Comparison with State-of-the-Arts

Haze Generation: We compare GenHaze against SOTA methods. As shown in Tab. 1, GenHaze outperforms other methods on the Haze4K dataset [49] regarding all metrics, highlighting its robust haze generation capability. On the RTTS dataset [38], GenHaze surpasses other baselines in FID (57.12) and sFID (105.43) scores, confirming its effectiveness in generating realistic haze that matches the distribution of real haze while demonstrating superior controllability. Notably, GenHaze is a diffusion-based, reference-controllable pipeline, contrasting with RIDCP [79] which relies on parameter adjustments, and non-controllable methods like CycleGAN-Turbo [59]. Visually, as shown in Fig. 7, GenHaze produces haze effects that resemble real haze in color, density, and other attributes. Compared to other pipelines (e.g., RIDCP [79] and Low-Res [6]), GenHaze avoids unnatural color shifts and artifacts. While approaches like UWNR [83] enable

Table 2. Performance gains of various baselines after fine-tuning for 1-2 epochs on controllable haze images generated by GenHaze.

Dataset	Metrics	CNN-based						Transformer-based					
		MSBDN [CVPR'2020] [19]	+ GenHaze	Gain	FocalNet [ICCV'2023] [16]	+ GenHaze	Gain	Dehazer [CVPR'2022] [22]	+ GenHaze	Gain	Dehazeformer [TIP'2023] [72]	+ GenHaze	Gain
RTTS [38]	FADE ↓	1.5818	0.6881	-0.8937	2.0583	1.4770	-0.5813	1.8926	0.8263	-1.0663	1.8817	0.9099	-0.9717
	BRISQUE ↓	28.51	20.23	-8.28	35.92	24.31	-11.61	33.07	25.12	-4.95	32.25	24.40	-7.85
	NIQE ↓	4.66	3.65	-1.01	5.11	4.25	-0.86	4.82	3.97	-0.85	4.81	4.02	-0.79
	PIQE ↓	44.96	28.26	-16.70	49.78	38.75	-11.03	45.34	37.02	-8.32	48.05	35.75	-12.30
	PaQ-2-PIQ ↑	66.83	69.74	+2.89	66.56	68.00	+1.44	66.75	70.68	+3.93	66.68	69.42	+2.74
	Metrics	TaylorFormer [ICCV'2023] [63]	+ GenHaze	Gain	T ³ -DiffWeather [ECCV'2024] [10]	+ GenHaze	Gain	PSD [CVPR'2021] [14]	+ GenHaze	Gain	KANet [TPAMI'2024] [21]	+ GenHaze	Gain
FADE ↓	1.9827	1.0156	-0.9671	2.3771	0.9771	-1.4000	1.0174	0.6533	-0.3641	0.8705	0.8505	-0.0200	
BRISQUE ↓	33.21	21.37	-11.84	30.14	22.01	-8.13	22.54	20.37	-2.17	18.82	17.87	-0.95	
NIQE ↓	4.89	3.84	-1.05	5.20	3.88	-1.32	3.81	3.51	-0.30	4.33	3.96	-0.37	
PIQE ↓	47.36	36.68	-10.68	41.36	28.56	-12.80	30.15	25.19	-4.96	20.71	19.12	-1.59	
PaQ-2-PIQ ↑	66.66	68.76	+2.10	65.82	66.94	+1.12	71.06	70.45	-0.61	68.79	69.14	+0.35	

Dataset	Metrics	CNN-based						Transformer-based					
		MSBDN [CVPR'2020] [19]	+ GenHaze	Gain	FocalNet [ICCV'2023] [16]	+ GenHaze	Gain	Dehazer [CVPR'2022] [22]	+ GenHaze	Gain	Dehazeformer [TIP'2023] [72]	+ GenHaze	Gain
Fattal's [20]	FADE ↓	0.5428	0.3362	-0.2066	0.6137	0.5211	-0.0926	0.6372	0.5098	-0.1274	0.6296	0.5070	-0.1226
	BRISQUE ↓	16.94	15.31	-1.63	17.03	16.89	-0.14	17.68	14.61	-3.07	16.12	16.69	+0.57
	NIMA ↑	5.28	5.33	+0.05	5.26	5.28	+0.02	5.25	5.35	+0.10	5.26	5.28	+0.02
	MUSIQ ↑	63.01	63.47	+0.46	63.51	63.32	-0.19	63.60	63.40	-0.20	62.45	62.93	+0.48
	PIQE ↓	28.85	26.55	-2.30	32.25	30.81	-1.44	31.80	30.44	-1.36	30.52	29.07	-1.45
	Metrics	TaylorFormer [ICCV'2023] [63]	+ GenHaze	Gain	T ³ -DiffWeather [ECCV'2024] [10]	+ GenHaze	Gain	PSD [CVPR'2021] [14]	+ GenHaze	Gain	KANet [TPAMI'2024] [21]	+ GenHaze	Gain
FADE ↓	0.6242	0.5307	-0.0935	1.0555	0.4284	-0.6271	0.3627	0.3363	-0.0263	0.3759	0.3821	-0.0062	
BRISQUE ↓	16.36	16.28	-0.08	17.65	16.32	-1.33	18.98	17.83	-1.15	15.17	15.04	-0.13	
NIMA ↑	4.96	5.00	+0.04	5.20	5.36	+0.16	4.91	5.20	+0.29	5.04	5.11	+0.07	
MUSIQ ↑	62.59	61.83	-0.76	62.46	63.75	+1.29	64.57	65.91	+1.34	63.85	63.59	-0.26	
PIQE ↓	30.16	28.76	-1.40	31.70	25.06	-6.64	28.77	28.20	-0.56	23.94	23.72	-0.22	

Table 3. Com. of the gains of different baselines using physical generation [79] and using GenHaze.

Method	MSBDN [CVPR'2020] [19]			Dehazeformer [TIP'2023] [72]			T ³ -Diff [ECCV'2024] [10]			KANet [TPAMI'2024] [21]		
	FA ↓	BRI ↓	PI ↓	FA ↓	BRI ↓	PI ↓	FA ↓	BRI ↓	PI ↓	FA ↓	BRI ↓	PI ↓
baseline	1.5818	28.51	44.96	1.8817	32.25	48.05	2.3771	30.14	41.36	0.8705	18.82	20.71
+Physical [79]	0.8532	24.14	35.42	1.0132	27.91	38.84	1.1975	26.52	36.94	0.8746	19.21	20.51
+GenHaze	0.6881	20.23	28.26	0.9099	24.40	35.75	0.9771	22.01	28.56	0.8505	17.87	19.12

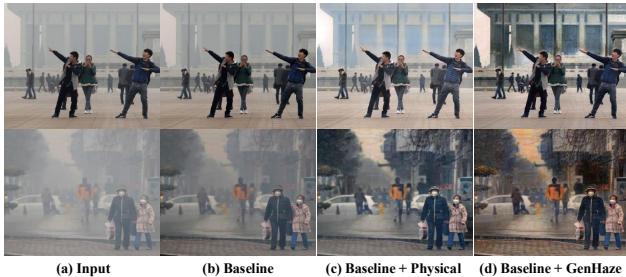


Figure 6. Visual Com. of fine-tuning using haze generated with physical method versus haze generated by GenHaze. The latter demonstrates significantly superior performance in dehazing.

controllable haze generation, they may introduce unwanted oversaturation or background interference.

Additionally, Tab.3 demonstrates that using a more advanced physics-based haze synthesis pipeline to generate training data improves baseline performance compared to training on data produced by inferior pipelines, though it still negatively impacts certain real-world baselines (e.g., KANet). In contrast, Tab.3 and Fig.6 show that GenHaze’s targeted and controllable high-quality haze generation substantially boosts real-world dehazing. For comparisons with training-from-scratch, please see the supp..

Dehazing Performance: For quantitative comparison, as

Table 4. Comparison of generation speed to other baselines at 512×512 resolution on an A800 graphics card.

Method	InstructP2P [4]	CycleGAN-Turbo [59]	TaylorFormer [63]	GenHaze
# Inference Speed	0.812s	0.211s	0.167s	0.169s

shown in Tab. 2, GenHaze consistently boosts the performance of various dehazing baselines across datasets (RTTS [38] and Fattal’s Dataset [20]) and metrics. In particular, GenHaze yields significant gains across CNN-based, Transformer-based, Diffusion-based, and real-world dehazing models. For example, integrating GenHaze with MSBDN [19] on the RTTS dataset achieves a substantial improvement, reducing the FADE metric by 0.89 (approximately 55%) and lowering BRISQUE by 8.28. Additionally, several baselines enhanced with GenHaze surpass the SOTA real-world dehazing model KANet [21], showcasing GenHaze’s ability to unlock the potential of baselines. For qualitative comparison, Fig. 5 shows the visual enhancements achieved with GenHaze. In each example, the ”+GenHaze” images exhibit more natural and realistic dehazing. In the top row, GenHaze-enhanced versions significantly outperform baseline versions, including KANet [21]. The bottom row further illustrates that GenHaze consistently produces cleaner dehazing results, enhancing overall performance across different scenes.

5.2. Generation Speed

To demonstrate GenHaze’s efficiency, we compare its performance in Tab.4. Our one-step method is faster than InstructPix2Pix [4] and CycleGAN-Turbo [59], thanks to its well-motivated design (e.g., removing the CLIP en-

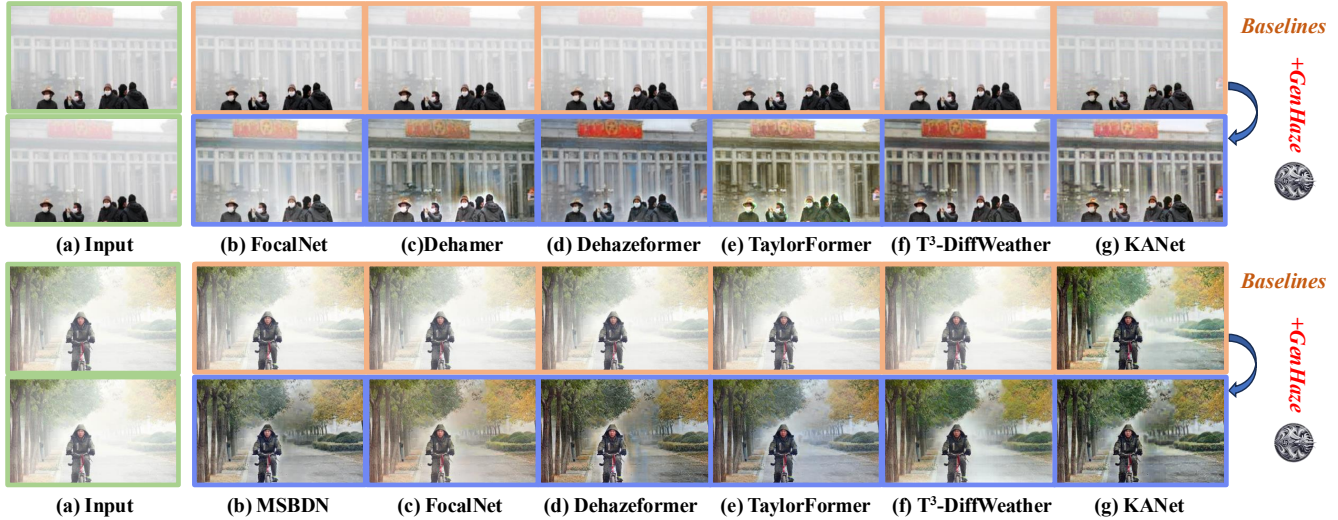


Figure 7. Visual com. of multiple baselines on real haze samples, before and after integrating GenHaze. Please zoom in to view details.

Method	FID ↓	sFID ↓
Hand-craft Prompt	58.58	107.24
GenHaze w/o. Haze Database	58.05	106.81
GenHaze w/o. Larnable Prompt	57.99	106.53
GenHaze w/o. PE.	57.76	106.21
GenHaze (ours)	57.12	105.43

Table 5. Abl. of offline haze database and prompt embedding.

Method	FID ↓	sFID ↓
ControlNet [87]	58.29	107.06
HM as Decoder [34]	58.01	106.75
GenHaze w/o. HM	58.79	107.62
GenHaze (ours)	57.12	105.43

Table 6. Abl. on deg. encoder and haze modulation.

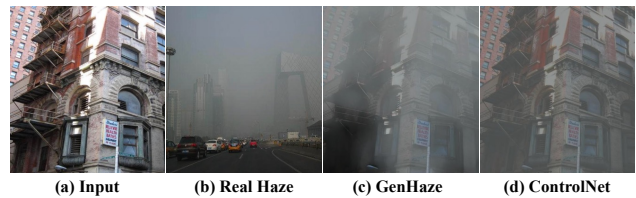


Figure 8. Com. of the quality and controllability of generated haze. GenHaze generate high-quality images that are more consistent with the properties of real haze, compared to ControlNet [87].

coder and using a lightweight degradation encoder). It also matches the speed of the latest dehazing baseline, TaylorFormer [63], highlighting its practicality for downstream tasks. Additional analysis is in the supp..

5.3. Ablation Studies

For ablation experiments related to GenHaze, we select real haze RTTS [38] as the test set, as our ultimate goal is to generate results for unpaired images in the real world.

One-step generation strategy. We validate our one-step strategy through ablation experiments. Unlike the concatenation approach in methods like InstructPix2Pix [4], our derivation shows that clean images can replace noise as input, eliminating the need for added noise.

Table 7. Abl. of one-step strategy.

Method	FID ↓	sFID ↓
Noise <i>cat</i> Clean	59.04	108.52
GenHaze (ours)	57.12	105.43

As shown in Tab. 7, the one-step generation strategy significantly enhances final output quality and

avoids the computational overhead of multi-step inference. In contrast, traditional denoising paradigms suffer considerable performance drops in single-step inference, likely because the excessive input noise deviates from the gradual denoising assumption expected by the network.

Offline Haze Database and Prompt embedding. We compare different text embedding strategies of protocol in Tab. 5. A fixed prompt (e.g., “a natural image with haze”) restricts SD’s extraction of relevant haze features, limiting performance. In contrast, guidance from the database helps

the model capture general haze characteristics effectively. Additionally, using a learnable prompt embedding increases optimization flexibility, enabling the model to better adapt to specific haze conditions.

Degradation Encoder and Haze Modulation. In this section, we explore the effectiveness of the degradation encoder and haze modulation. As shown in Tab. 6 and Fig. 8, combining degradation encoder and modulation approach achieves superior haze controllability and quality compared to ControlNet [87], yielding images that more closely match real-world distributions. Unlike previous methods [33, 34] that apply modulation only at the decoder stage, we incorporate it after each stage to dynamically control the generated style, enhancing generative performance overall. More ablation experiments can be found in the supp..

6. Conclusion

In this paper, we introduce GenHaze, a novel haze generation pipeline. It transforms the LDM into a framework capable of efficiently generating high-quality, controllable haze through a one-step strategy and clean-to-haze generation protocol. GenHaze also effectively unlocks the potential of existing dehazing baselines, leading to substantial gains in real-world dehazing tasks. Our work further demonstrates the potential of current LDM to be adapted as robust frameworks for controllabel degradation generation.

Acknowledgement: This work was supported by the Guangzhou Municipal Science and Technology Project (Grant No. 2024A04J4230) and the National Natural Science Foundation of China (Project No. 61902275).

References

- [1] Yogesh Balaji, Seungjun Nah, Xun Huang, Arash Vahdat, Jiaming Song, Qinsheng Zhang, Karsten Kreis, Miika Aittala, Timo Aila, Samuli Laine, et al. ediff-i: Text-to-image diffusion models with an ensemble of expert denoisers. *arXiv preprint arXiv:2211.01324*, 2022. 2
- [2] David Berthelot, Arnaud Autef, Jierui Lin, Dian Ang Yap, Shuangfei Zhai, Siyuan Hu, Daniel Zheng, Walter Talbott, and Eric Gu. Tract: Denoising diffusion models with transitive closure time-distillation. *arXiv preprint arXiv:2303.04248*, 2023. 2
- [3] Usha Bhalla, Alex Oesterling, Suraj Srinivas, Flavio P Calmon, and Himabindu Lakkaraju. Interpreting clip with sparse linear concept embeddings (splice). *arXiv preprint arXiv:2402.10376*, 2024. 4
- [4] Tim Brooks, Aleksander Holynski, and Alexei A Efros. Instructpix2pix: Learning to follow image editing instructions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18392–18402, 2023. 2, 3, 6, 7, 8
- [5] Bolun Cai, Xiangmin Xu, Kui Jia, Chunmei Qing, and Dacheng Tao. Dehazenet: An end-to-end system for single image haze removal. *IEEE Transactions on Image Processing*, 25(11):5187–5198, 2016. 2
- [6] Haoyu Chen, Wenbo Li, Jinjin Gu, Jingjing Ren, Haoze Sun, Xueyi Zou, Zhensong Zhang, Youliang Yan, and Lei Zhu. Low-res leads the way: Improving generalization for super-resolution by self-supervised learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 25857–25867, 2024. 6
- [7] Sixiang Chen, Tian Ye, Yun Liu, Erkang Chen, Jun Shi, and Jingchun Zhou. Snowformer: Scale-aware transformer via context interaction for single image desnowing. *arXiv preprint arXiv:2208.09703*, 2022. 2
- [8] Sixiang Chen, Tian Ye, Yun Liu, Taodong Liao, Yi Ye, and Erkang Chen. Msp-former: Multi-scale projection transformer for single image desnowing. *arXiv preprint arXiv:2207.05621*, 2022.
- [9] Sixiang Chen, Tian Ye, Jinbin Bai, Erkang Chen, Jun Shi, and Lei Zhu. Sparse sampling transformer with uncertainty-driven ranking for unified removal of raindrops and rain streaks. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 13106–13117, 2023. 2
- [10] Sixiang Chen, Tian Ye, Kai Zhang, Zhaohu Xing, Yunlong Lin, and Lei Zhu. Teaching tailored to talent: Adverse weather restoration via prompt pool and depth-anything constraint. *arXiv preprint arXiv:2409.15739*, 2024. 7
- [11] SiXiang Chen, Jianyu Lai, Jialin Gao, Tian Ye, Haoyu Chen, Hengyu Shi, Shitong Shao, Yunlong Lin, Song Fei, Zhaohu Xing, et al. Postercraft: Rethinking high-quality aesthetic poster generation in a unified framework. *arXiv preprint arXiv:2506.10741*, 2025. 2
- [12] Sixiang Chen, Tian Ye, Kai Zhang, Zhaohu Xing, Yunlong Lin, and Lei Zhu. Teaching tailored to talent: Adverse weather restoration via prompt pool and depth-anything constraint. In *European Conference on Computer Vision*, pages 95–115. Springer, 2025. 2, 6
- [13] Wei-Ting Chen, Zhi-Kai Huang, Cheng-Che Tsai, Hao-Hsiang Yang, Jian-Jiun Ding, and Sy-Yen Kuo. Learning multiple adverse weather removal via two-stage knowledge learning and multi-contrastive regularization: Toward a unified model. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17653–17662, 2022. 2
- [14] Zeyuan Chen, Yangchao Wang, Yang Yang, and Dong Liu. Psd: Principled synthetic-to-real dehazing guided by physical priors. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 7180–7189, 2021. 2, 7
- [15] Zixuan Chen, Zewei He, Ziqian Lu, Xuecheng Sun, and Zhe-Ming Lu. Prompt-based test-time real image dehazing: a novel pipeline. *arXiv preprint arXiv:2309.17389*, 2023. 3, 5
- [16] Yuning Cui, Wenqi Ren, Xiaochun Cao, and Alois Knoll. Focal network for image restoration. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 13001–13011, 2023. 2, 6, 7
- [17] Yingying Deng, Fan Tang, Weiming Dong, Chongyang Ma, Xingjia Pan, Lei Wang, and Changsheng Xu. Stytr2: Image style transfer with transformers. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11326–11336, 2022. 5
- [18] P Kingma Diederik. Adam: A method for stochastic optimization. (*No Title*), 2014. 6
- [19] Hang Dong, Jinshan Pan, Lei Xiang, Zhe Hu, Xinyi Zhang, Fei Wang, and Ming-Hsuan Yang. Multi-scale boosted dehazing network with dense feature fusion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2157–2167, 2020. 2, 5, 6, 7
- [20] Raanan Fattal. Dehazing using color-lines. *ACM transactions on graphics (TOG)*, 34(1):1–14, 2014. 6, 7
- [21] Yuxin Feng, Long Ma, Xiaozhe Meng, Fan Zhou, Risheng Liu, and Zhuo Su. Advancing real-world image dehazing: perspective, modules, and training. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024. 2, 5, 6, 7
- [22] Chun-Le Guo, Qixin Yan, Saeed Anwar, Runmin Cong, Wenqi Ren, and Chongyi Li. Image dehazing transformer with transmission-aware 3d position embedding. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5812–5820, 2022. 2, 6, 7
- [23] Jianshu Guo, Wenhao Chai, Jie Deng, Hsiang-Wei Huang, Tian Ye, Yichen Xu, Jiawei Zhang, Jenq-Neng Hwang, and Gaoang Wang. Versat2i: Improving text-to-image models with versatile reward. *arXiv preprint arXiv:2403.18493*, 2024. 2
- [24] Yaru Hao, Zewen Chi, Li Dong, and Furu Wei. Optimizing prompts for text-to-image generation. *Advances in Neural Information Processing Systems*, 36, 2024. 2

- [25] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in neural information processing systems*, 30, 2017. 6
- [26] Jonathan Ho and Tim Salimans. Classifier-free diffusion guidance. *arXiv preprint arXiv:2207.12598*, 2022. 6
- [27] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020. 3
- [28] Aaron Hurst, Adam Lerer, Adam P Goucher, Adam Perelman, Aditya Ramesh, Aidan Clark, AJ Ostrow, Akila Welihinda, Alan Hayes, Alec Radford, et al. Gpt-4o system card. *arXiv preprint arXiv:2410.21276*, 2024. 4
- [29] Quan Huynh-Thu and Mohammed Ghanbari. Scope of validity of psnr in image/video quality assessment. *Electronics letters*, 44(13):800–801, 2008. 6
- [30] Yeying Jin, Wending Yan, Wenhan Yang, and Robby T Tan. Structure representation network and uncertainty feedback learning for dense non-uniform fog removal. In *Asian Conference on Computer Vision*, pages 155–172. Springer, 2022. 2
- [31] Yeying Jin, Wenhan Yang, and Robby T Tan. Unsupervised night image enhancement: When layer decomposition meets light-effects suppression. In *European Conference on Computer Vision*, pages 404–421. Springer, 2022. 2
- [32] Yeying Jin, Wenhan Yang, Wei Ye, Yuan Yuan, and Robby T Tan. Shadowdiffusion: Diffusion-based shadow removal using classifier-driven attention and structure preservation. *arXiv preprint arXiv:2211.08089*, 2022. 2
- [33] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4401–4410, 2019. 5, 8
- [34] Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Analyzing and improving the image quality of stylegan. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8110–8119, 2020. 5, 8
- [35] Tero Karras, Miika Aittala, Timo Aila, and Samuli Laine. Elucidating the design space of diffusion-based generative models. *Advances in neural information processing systems*, 35:26565–26577, 2022. 2, 6
- [36] Nupur Kumari, Bingliang Zhang, Richard Zhang, Eli Shechtman, and Jun-Yan Zhu. Multi-concept customization of text-to-image diffusion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1931–1941, 2023. 2
- [37] Jianyu Lai, Sixiang Chen, Yunlong Lin, Tian Ye, Yun Liu, Song Fei, Zhaohu Xing, Hongtao Wu, Weiming Wang, and Lei Zhu. Snowmaster: Comprehensive real-world image desnowing via mllm with multi-model feedback optimization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4302–4312, 2025. 2
- [38] Boyi Li, Wenqi Ren, Dengpan Fu, Dacheng Tao, Dan Feng, Wenjun Zeng, and Zhangyang Wang. Benchmarking single-image dehazing and beyond. *IEEE Transactions on Image Processing*, 28(1):492–505, 2018. 2, 6, 7, 8
- [39] Dongxu Li, Junnan Li, and Steven Hoi. Blip-diffusion: Pre-trained subject representation for controllable text-to-image generation and editing. *Advances in Neural Information Processing Systems*, 36, 2024. 2
- [40] Yuheng Li, Haotian Liu, Qingyang Wu, Fangzhou Mu, Jianwei Yang, Jianfeng Gao, Chunyuan Li, and Yong Jae Lee. Gligen: Open-set grounded text-to-image generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 22511–22521, 2023. 2
- [41] Huangxing Lin, Yunlong Lin, Jingyuan Xia, Linyu Fan, Feifei Li, Yingying Wang, and Xinghao Ding. Fusion2void: Unsupervised multi-focus image fusion based on image inpainting. *IEEE Transactions on Circuits and Systems for Video Technology*, 2024. 2
- [42] Shanchuan Lin, Bingchen Liu, Jiashi Li, and Xiao Yang. Common diffusion noise schedules and sample steps are flawed. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pages 5404–5411, 2024. 3
- [43] Yunlong Lin, Zhenqi Fu, Kairun Wen, Tian Ye, Sixiang Chen, Ge Meng, Yingying Wang, Yue Huang, Xiaotong Tu, and Xinghao Ding. Unsupervised low-light image enhancement with lookup tables and diffusion priors. *arXiv preprint arXiv:2409.18899*, 2024. 2
- [44] Yunlong Lin, Tian Ye, Sixiang Chen, Zhenqi Fu, Yingying Wang, Wenhao Chai, Zhaohu Xing, Lei Zhu, and Xinghao Ding. Aglldiff: Guiding diffusion models towards unsupervised training-free real-world low-light image enhancement. *arXiv preprint arXiv:2407.14900*, 2024.
- [45] Yunlong Lin, Zixu Lin, Haoyu Chen, Panwang Pan, Chenxin Li, Sixiang Chen, Kairun Wen, Yeying Jin, Wenbo Li, and Xinghao Ding. Jarvisir: Elevating autonomous driving perception with intelligent image restoration. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 22369–22380, 2025.
- [46] Yunlong Lin, Zixu Lin, Kunjie Lin, Jinbin Bai, Panwang Pan, Chenxin Li, Haoyu Chen, Zhongdao Wang, Xinghao Ding, Wenbo Li, et al. Jarvisart: Liberating human artistic creativity via an intelligent photo retouching agent. *arXiv preprint arXiv:2506.17612*, 2025. 2
- [47] Xiaohong Liu, Yongrui Ma, Zhihao Shi, and Jun Chen. Grid-dehazenet: Attention-based multi-scale network for image dehazing. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 7314–7323, 2019. 2
- [48] Xiaohong Liu, Yongrui Ma, Zhihao Shi, and Jun Chen. Grid-dehazenet: Attention-based multi-scale network for image dehazing. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 7314–7323, 2019. 2
- [49] Ye Liu, Lei Zhu, Shunda Pei, Huazhu Fu, Jing Qin, Qing Zhang, Liang Wan, and Wei Feng. From synthetic to real: Image dehazing collaborating with unlabeled real data. In *Proceedings of the 29th ACM International Conference on Multimedia*, pages 50–58, 2021. 6
- [50] Ye Liu, Lei Zhu, Shunda Pei, Huazhu Fu, Jing Qin, Qing Zhang, Liang Wan, and Wei Feng. From synthetic to real:

- Image dehazing collaborating with unlabeled real data. In *Proceedings of the 29th ACM International Conference on Multimedia*, pages 50–58, 2021. 6
- [51] Yun Liu, Zhongsheng Yan, Sixiang Chen, Tian Ye, Wenqi Ren, and Erkang Chen. Nighthazeforner: Single nighttime haze removal using prior query transformer. *arXiv preprint arXiv:2305.09533*, 2023. 2
- [52] Cheng Lu, Yuhao Zhou, Fan Bao, Jianfei Chen, Chongxuan Li, and Jun Zhu. Dpm-solver: A fast ode solver for diffusion probabilistic model sampling in around 10 steps. *Advances in Neural Information Processing Systems*, 35:5775–5787, 2022. 2, 3
- [53] Cheng Lu, Yuhao Zhou, Fan Bao, Jianfei Chen, Chongxuan Li, and Jun Zhu. Dpm-solver++: Fast solver for guided sampling of diffusion probabilistic models. *arXiv preprint arXiv:2211.01095*, 2022. 2
- [54] Eric Luhman and Troy Luhman. Knowledge distillation in iterative generative models for improved sampling speed. *arXiv preprint arXiv:2101.02388*, 2021. 2
- [55] Simian Luo, Yiqin Tan, Suraj Patil, Daniel Gu, Patrick von Platen, Apolinário Passos, Longbo Huang, Jian Li, and Hang Zhao. Lcm-lora: A universal stable-diffusion acceleration module. *arXiv preprint arXiv:2311.05556*, 2023. 2
- [56] EJ McCartney. Optics of the atmosphere: scattering by molecules and particles, 1976. 2
- [57] Chong Mou, Xintao Wang, Liangbin Xie, Yanze Wu, Jian Zhang, Zhongang Qi, and Ying Shan. T2i-adapter: Learning adapters to dig out more controllable ability for text-to-image diffusion models. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 4296–4304, 2024. 2
- [58] Charlie Nash, Jacob Menick, Sander Dieleman, and Peter W Battaglia. Generating images with sparse representations. *arXiv preprint arXiv:2103.03841*, 2021. 6
- [59] Gaurav Parmar, Taesung Park, Srinivasa Narasimhan, and Jun-Yan Zhu. One-step image translation with text-to-image models. *arXiv preprint arXiv:2403.12036*, 2024. 2, 5, 6, 7
- [60] Or Patashnik, Zongze Wu, Eli Shechtman, Daniel Cohen-Or, and Dani Lischinski. Styleclip: Text-driven manipulation of stylegan imagery. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 2085–2094, 2021. 5
- [61] Hamza Pehlivan, Yusuf Dalva, and Aysegul Dundar. Styleres: Transforming the residuals for real image editing with stylegan. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1828–1837, 2023. 5
- [62] Dustin Podell, Zion English, Kyle Lacey, Andreas Blattmann, Tim Dockhorn, Jonas Müller, Joe Penna, and Robin Rombach. Sdxl: Improving latent diffusion models for high-resolution image synthesis. *arXiv preprint arXiv:2307.01952*, 2023. 2, 3
- [63] Yuwei Qiu, Kaihao Zhang, Chenxi Wang, Wenhan Luo, Hongdong Li, and Zhi Jin. Mb-taylorformer: Multi-branch efficient transformer expanded by taylor formula for image dehazing. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12802–12813, 2023. 2, 5, 6, 7, 8
- [64] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pages 8748–8763. PMLR, 2021. 4
- [65] Sudarshan Rajagopalan, Nithin Gopalakrishnan Nair, Jay N Paranjape, and Vishal M Patel. Gendeg: Diffusion-based degradation synthesis for generalizable all-in-one image restoration. *arXiv preprint arXiv:2411.17687*, 2024. 2
- [66] Elad Richardson, Yuval Alaluf, Or Patashnik, Yotam Nitzan, Yaniv Azar, Stav Shapiro, and Daniel Cohen-Or. Encoding in style: a stylegan encoder for image-to-image translation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2287–2296, 2021. 5
- [67] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10684–10695, 2022. 1, 2, 3
- [68] Nataniel Ruiz, Yuanzhen Li, Varun Jampani, Yael Pritch, Michael Rubinstein, and Kfir Aberman. Dreambooth: Fine tuning text-to-image diffusion models for subject-driven generation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 22500–22510, 2023. 2
- [69] Tim Salimans and Jonathan Ho. Progressive distillation for fast sampling of diffusion models. *arXiv preprint arXiv:2202.00512*, 2022. 2
- [70] Axel Sauer, Dominik Lorenz, Andreas Blattmann, and Robin Rombach. Adversarial diffusion distillation. In *European Conference on Computer Vision*, pages 87–103. Springer, 2025. 2, 3
- [71] Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502*, 2020. 2
- [72] Yuda Song, Zhuqing He, Hui Qian, and Xin Du. Vision transformers for single image dehazing. *arXiv preprint arXiv:2204.03883*, 2022. 2, 5, 6, 7
- [73] Yang Song, Prafulla Dhariwal, Mark Chen, and Ilya Sutskever. Consistency models. *arXiv preprint arXiv:2303.01469*, 2023. 2
- [74] Jianlin Su, Murtadha Ahmed, Yu Lu, Shengfeng Pan, Wen Bo, and Yunfeng Liu. Roformer: Enhanced transformer with rotary position embedding. *Neurocomputing*, 568:127063, 2024. 4
- [75] Shangquan Sun, Wenqi Ren, Xinwei Gao, Rui Wang, and Xiaochun Cao. Restoring images in adverse weather conditions via histogram transformer. In *European Conference on Computer Vision*, pages 111–129. Springer, 2024. 2
- [76] Jeya Maria Jose Valanarasu, Rajeev Yasarla, and Vishal M Patel. Transweather: Transformer-based restoration of images degraded by adverse weather conditions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2353–2363, 2022. 2, 6
- [77] Haofan Wang, Matteo Spinelli, Qixun Wang, Xu Bai, Zekui Qin, and Anthony Chen. Instantstyle: Free lunch towards

- style-preserving in text-to-image generation. *arXiv preprint arXiv:2404.02733*, 2024. 5
- [78] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004. 6
- [79] Rui-Qi Wu, Zheng-Peng Duan, Chun-Le Guo, Zhi Chai, and Chongyi Li. Ridcp: Revitalizing real image dehazing via high-quality codebook priors. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 22282–22291, 2023. 2, 5, 6, 7
- [80] Jiazheng Xu, Xiao Liu, Yuchen Wu, Yuxuan Tong, Qinkai Li, Ming Ding, Jie Tang, and Yuxiao Dong. Imagereward: Learning and evaluating human preferences for text-to-image generation. *Advances in Neural Information Processing Systems*, 36, 2024. 2
- [81] Yanwu Xu, Yang Zhao, Zhisheng Xiao, and Tingbo Hou. Ufogen: You forward once large scale text-to-image generation via diffusion gans. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8196–8206, 2024. 2
- [82] Zizheng Yang, Hu Yu, Bing Li, Jinghao Zhang, Jie Huang, and Feng Zhao. Unleashing the potential of the semantic latent space in diffusion models for image dehazing. In *European Conference on Computer Vision*, pages 371–389. Springer, 2025. 2
- [83] Tian Ye, Sixiang Chen, Yun Liu, Yi Ye, Erkang Chen, and Yuche Li. Underwater light field retention: Neural rendering for underwater imaging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 488–497, 2022. 6
- [84] Tian Ye, Yunchen Zhang, Mingchao Jiang, Liang Chen, Yun Liu, Sixiang Chen, and Erkang Chen. Perceiving and modeling density for image dehazing. In *European Conference on Computer Vision*, pages 130–145. Springer, 2022. 2
- [85] Tian Ye, Sixiang Chen, Jinbin Bai, Jun Shi, Chenghao Xue, Jingxia Jiang, Junjie Yin, Erkang Chen, and Yun Liu. Adverse weather removal with codebook priors. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12653–12664, 2023. 2
- [86] Tianwei Yin, Michaël Gharbi, Richard Zhang, Eli Shechtman, Fredo Durand, William T Freeman, and Taesung Park. One-step diffusion with distribution matching distillation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6613–6623, 2024. 2
- [87] Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. Adding conditional control to text-to-image diffusion models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3836–3847, 2023. 2, 4, 5, 8
- [88] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 586–595, 2018. 6
- [89] Yurui Zhu, Tianyu Wang, Xueyang Fu, Xuanyu Yang, Xin Guo, Jifeng Dai, Yu Qiao, and Xiaowei Hu. Learning weather-general and weather-specific features for image restoration under multiple adverse weather conditions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 21747–21758, 2023. 2