Hihaze: One-Step Diffusion for High-Resolution Real-World Image Dehazing

Anonymous authors
Paper under double-blind review

Abstract

Diffusion models have shown promising performance in real-world image dehazing. However, restoring high-resolution images remains challenging due to spatially non-uniform haze and inconsistent predictions across local patches. To address these issues, we propose Hihaze, a one-step diffusion framework for high-resolution real-world dehazing. Hihaze adopts two haze-aware modulation modules: attentive skip compensation (ASC) and a spatial degradation adapter (SDA). ASC selectively transfers reliable encoder information to the variational autoencoder (VAE) decoder, preserving structural details while suppressing haze-degraded features. SDA extracts multi-scale spatial degradation representations and injects them into the U-Net to guide restoration according to local haze variations. Furthermore, we design a training-free global-guided patch fusion (GPF) method. It uses a global prediction to align patch appearance and downweight unreliable predictions, followed by structure-guided detail compensation. Notably, this training-free design also makes GPF a plug-and-play module for other dehazing models. Evaluations on four real-world benchmarks demonstrate that Hihaze outperforms state-of-the-art methods in both restoration quality and inference efficiency. Hihaze achieves PSNR gains of **3.6 dB** on Dense-Haze and **3.6 dB** on NH-HAZE over the best competing methods. Code will be made publicly.

1 Introduction

Single-image dehazing aims to recover a clear image from a hazy observation. Haze reduces contrast, alters colors, and obscures fine details. Its density also varies across the scene. The atmospheric scattering model (Narasimhan & Nayar, 2002; McCartney & Hall Jr, 1977) describes this process as

$$\mathbf{I}(x) = \mathbf{I}_{\mathrm{c}}(x)\, t_{\mathrm{h}}(x) + \mathbf{A}\big(1 - t_{\mathrm{h}}(x)\big), \tag{1}$$

where $\mathbf{I}$ and $\mathbf{I}_{\mathrm{c}}$ denote the hazy image and its clean counterpart. $\mathbf{A}$ is the atmospheric light, and $t_{\mathrm{h}}(x)$ is the transmission at pixel x. Recovering $\mathbf{I}_{\mathrm{c}}$ is ill-posed because the scattering parameters are unknown. Early methods use statistical priors, such as dark-channel sparsity (He et al., 2010), to constrain restoration. However, these assumptions may fail in complex scenes. Later CNN- and Transformer-based methods learn restoration mappings from data (Cai et al., 2016; Liu et al., 2019; Qin et al., 2020; Guo et al., 2022; Song et al., 2022). Despite this progress, recovering realistic textures remains difficult under dense, spatially varying haze.

Recently, large pretrained diffusion models have provided rich generative priors for image restoration (Wang et al., 2024a; Liu et al., 2024). They help recover realistic textures from degraded inputs. For dehazing, existing approaches exploit these priors through realistic haze synthesis, unpaired training, and physical or textual guidance (Wang et al., 2025; Lan et al., 2025a). To reduce computation, latent diffusion performs denoising in a compact latent space (Rombach et al., 2022). Further reducing the number of denoising steps accelerates inference and makes diffusion-based dehazing more practical (Song et al., 2020; Wang et al., 2025).

Nevertheless, two difficulties remain in high-resolution real-world dehazing: **(1) Spatially non-uniform haze.** VAE compression can discard fine details. Encoder–decoder skip connections help retain input information (Lan et al., 2025a). However, encoder features contain both scene structures and haze degradation. Directly transferring them may leave residual haze in the result. Moreover, high-level semantic guidance does not explicitly describe local degradation. It therefore provides

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.



Figure 1: Visual results on NH-HAZE 51 (left), I-HAZE 34 (upper right), and O-HAZE 01 (lower right). Each scene shows the hazy input on the left of the diagonal and the Hihaze result on the right. Bottom: matched details from the marked region of NH-HAZE 51. Hihaze removes haze while recovering scene structures and textures in the restored images.

limited guidance for spatially varying restoration. Effective dehazing requires reliable detail transmission together with spatial degradation guidance. **(2) Inconsistent patch predictions.** High-resolution inference often uses patches to limit memory use (Wang et al., 2024a). Each patch sees only part of the scene. Adjacent patches may therefore differ in color, illumination, and dehazing strength. Overlapping inference and local aggregation can reduce boundary artifacts (Özdenizci & Legenstein, 2023; Ding et al., 2024). However, they lack an explicit global reference. Averaging inconsistent predictions may also blur fine textures.

To address these issues, we propose **Hihaze** for high-resolution image dehazing. Hihaze adapts a pretrained latent diffusion model to restore hazy images in one step. We introduce two haze-aware modulation modules: attentive skip compensation (ASC) and a spatial degradation adapter (SDA). **(1) ASC** uses decoder features to retrieve reliable encoder information. It compensates for missing structures and fine details while suppressing haze-degraded features. **(2) SDA** extracts spatial degradation features from the hazy input. It injects these features into the U-Net at multiple scales. This guides restoration according to local haze variations.

Moreover, we design training-free global-guided patch fusion (GPF) for high-resolution inference. A prediction from the downsampled full image provides a shared appearance reference. GPF aligns local patch predictions with this reference and reduces the contribution of inconsistent patches. Structural agreement between the fused result and the hazy input then guides detail compensation. This improves cross-patch consistency while preserving reliable textures. GPF can also be integrated into other dehazing models as a plug-and-play module.

Extensive experiments on four real-world benchmarks demonstrate the effectiveness of Hihaze. As shown in Fig. 1, Hihaze removes more haze than the compared methods. It also recovers clearer edges and textures, even in severely degraded regions. Our main contributions are as follows:

- We propose Hihaze, a one-step diffusion dehazing framework with attentive skip compensation and a spatial degradation adapter for high-resolution images.
- We introduce global-guided patch fusion to improve consistency and preserve fine textures. It is a training-free, plug-and-play module for other dehazing models.
- We validate Hihaze on four real-world benchmarks. It achieves superior reconstruction fidelity and perceptual quality under diverse haze conditions.

2 Related Work

2.1 Image Dehazing

Early dehazing methods typically derive physical priors from natural-image statistics. These priors are then incorporated into the atmospheric scattering model to recover clear images from hazy observations (Fattal, 2008; He et al., 2010; Meng et al., 2013; Fattal, 2014; Berman et al., 2016). For example, He et al. (2010) introduced the Dark Channel Prior (DCP), based on the near-zero dark-channel values commonly observed in haze-free images. Similarly, Fattal (2014) proposed the Color-Line Prior, leveraging the one-dimensional distributions formed by pixels in local image patches in RGB color space to estimate scene transmission. However, handcrafted priors often struggle to model complex haze in real-world scenes, potentially causing color distortions.

With the development of deep learning, numerous learning-based methods have been proposed for single-image dehazing. By designing task-specific architectures, these methods directly learn the mapping from hazy images to their clear counterparts (Liu et al., 2019; Dong et al., 2020; Guo et al., 2022; Song et al., 2022). Although effective on synthetic benchmarks, these methods often generalize poorly to real-world scenes, where haze distributions are considerably more complex and diverse. To address this limitation, recent studies have increasingly focused on real-world image dehazing by exploiting prior knowledge or adaptively refining the restoration process (Wu et al., 2023b; Fang et al., 2024; Chen et al., 2024; Fu et al., 2025). For example, RIDCP (Wu et al., 2023b) leverages high-quality codebook priors to guide the restoration of real-world hazy images. More recently, IPC-Dehaze (Fu et al., 2025) employs an iterative predictor-critic framework that retains reliable codes and progressively refines code predictions. However, conventional CNN- and Transformer-based methods lack sufficient generative capability, resulting in limited restoration performance under dense and spatially non-uniform haze conditions.

2.2 Diffusion Model-Based Image Restoration

Recent years, advances in diffusion models have demonstrated their powerful generative capabilities (Ho et al., 2020; Song et al., 2020; Rombach et al., 2022; Ruiz et al., 2023). Owing to their strong ability to model natural image distributions, diffusion models have been increasingly applied to image restoration, achieving promising performance (Saharia et al., 2022b;a; Lugmayr et al., 2022; Zheng et al., 2024; Liu et al., 2024). In the context of real-world image dehazing, DiffDehaze (Wang et al., 2025) combines realistic haze synthesis with accelerated fidelity-preserving sampling to achieve efficient and faithful restoration. Diff-Dehazer (Lan et al., 2025a) combines unpaired bidirectional diffusion mapping with physical priors to improve real-world generalization. Despite these advances, existing methods remain limited in handling spatially non-uniform haze in real-world scenes, often resulting in uneven restoration across image regions and residual haze.

Moreover, the high memory cost of diffusion models often necessitates patch-wise inference for high-resolution images, resulting in inconsistent predictions across patches and visible boundary artifacts (Wang et al., 2025; Lan et al., 2025a). Existing approaches mitigate boundary artifacts through overlapping patch inference and local aggregation (Özdenizci & Legenstein, 2023; Ding et al., 2024; Wang et al., 2024a). Specifically, Patch-DM (Ding et al., 2024) reduces boundary artifacts by collaging features from neighboring patches to predict shifted patches. However, such aggregation mainly enforces consistency between neighboring patches. Averaging predictions in overlapping regions may also attenuate fine high-frequency textures in the restored images. In this paper, we introduce a one-step diffusion framework for high-resolution real-world image dehazing that enables memory-efficient patch-wise inference and effectively handles non-uniform haze.

3 Method

In this section, We present Hihaze, a one-step diffusion framework for high-resolution real-world dehazing. We first introduce the overall framework in Sec. 3.1. We then describe attentive skip compensation (ASC) in Sec. 3.2 and the spatial degradation adapter (SDA) in Sec. 3.3. ASC retrieves reliable encoder features for VAE decoding, while SDA guides denoising according to local haze variations. Finally, Sec. 3.4 presents training-free global-guided patch fusion (GPF) for appearance consistency and structure-guided detail compensation. Training details are provided in Appendix A.

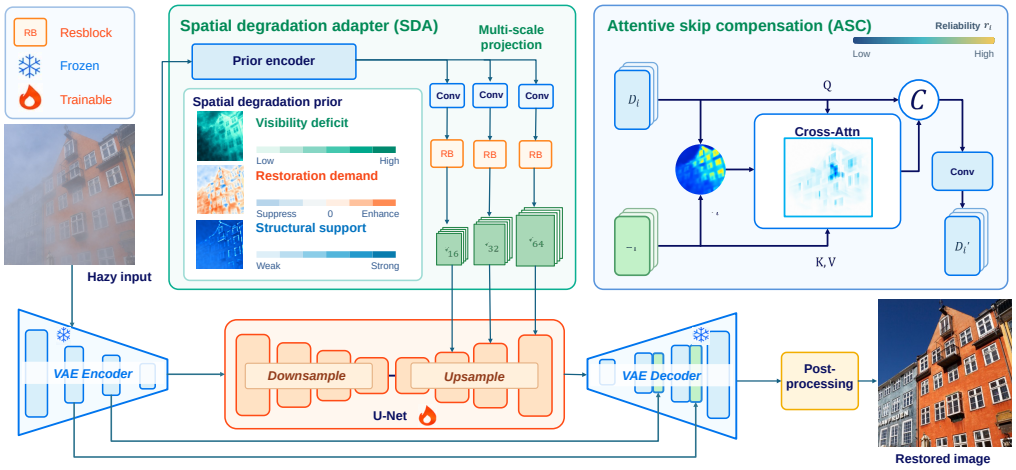


Figure 2: Overview of Hihaze. SDA guides spatially adaptive denoising, while ASC transfers reliable encoder features to the VAE decoder. GPF combines global and local predictions for high-resolution restoration with consistent appearance and preserved fine details.

3.1 Overall Framework

As illustrated in Fig. 2, we build our dehazing network on the pretrained Stable Diffusion 2.1-base model. Given a hazy image $\mathbf{I}$, the VAE encoder $\mathcal{E}$ produces a latent representation $\mathbf{z}_I$ and intermediate features for subsequent reconstruction. At the fixed timestep t=399, we use this encoded representation as the diffusion state, i.e., $\mathbf{z}_t$=$\mathbf{z}_I$. The standard forward formulation is

$$\mathbf{z}_t = \sqrt{\bar{\alpha}_t}\mathbf{z} + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}, \qquad \bar{\alpha}_t = \prod_{s=1}^{t}(1-\beta_s), \tag{2}$$

where $\mathbf{z}$ is a clean latent, $\boldsymbol{\epsilon}$ is Gaussian noise, and β_s denotes the noise variance.

To account for spatially varying haze, SDA extracts multi-scale degradation features directly from the hazy input. These features guide multiple U-Net decoder stages and refine the output noise prediction. The resulting prediction $\boldsymbol{\epsilon}'$ is used to recover the clean latent in one denoising step:

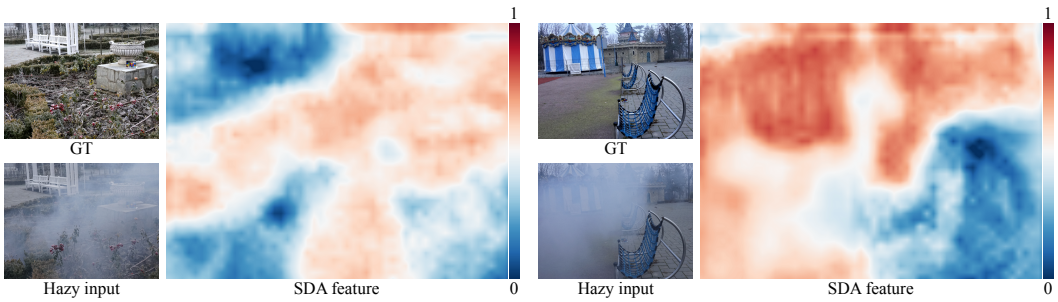
$$\widehat{\mathbf{z}}_0 = \frac{\mathbf{z}_t - \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}'}{\sqrt{\bar{\alpha}_t}}. \tag{3}$$

The VAE decoder $\mathcal{D}$ then reconstructs the restored image from $\widehat{\mathbf{z}}_0$. During decoding, ASC uses decoder features to query encoder information. This supports detail reconstruction while reducing the influence of unreliable haze-degraded encoder features during image reconstruction.

For high-resolution inputs, we apply the same network to a downsampled full image and overlapping patches at the original resolution. The resulting global reference $\mathbf{G}$ guides the alignment and fusion of local predictions $\{\mathbf{P}_i\}$. GPF further uses structural agreement with the hazy input to compensate fine details, producing the final result $\widehat{\mathbf{I}}$. Together, these designs enable spatially adaptive dehazing while maintaining global appearance consistency and preserving local structural details.

3.2 Attentive Skip Compensation

Latent diffusion models reduce computational cost through VAE compression, but this may attenuate fine spatial details. Encoder–decoder skip connections help recover these details by reusing input features (Wang et al., 2024a; Lan et al., 2025a; Ai et al., 2024; Zhou et al., 2022; Chen et al., 2025). However, spatially non-uniform haze causes the reliability of encoder features to vary across regions. Some regions retain useful scene details, whereas others are more severely affected by haze. Direct feature transfer can therefore introduce degraded information along with useful details. To address this variation, we propose attentive skip compensation (ASC). It selectively retrieves reliable encoder information and adaptively refines less reliable features during reconstruction.



(a) NH-HAZE 55 (b) NH-HAZE 14

Figure 3: Visualization of the first SDA feature channel on NH-HAZE 55 (left) and 14 (right). Both examples use the same color scale; warmer colors indicate stronger feature responses.

At stage ℓ, ASC takes encoder and decoder features $\mathbf{E}_\ell$ and $\mathbf{D}_\ell$. A lightweight convolutional predictor with a sigmoid output estimates a spatial reliability map $\mathbf{R}_\ell$ from both $\mathbf{E}_\ell$ and $\mathbf{D}_\ell$. Decoder features then query encoder features through local cross-attention, with reliability guiding the selection of source information. For each window and attention head, the retrieved feature is

$$\mathbf{U}_\ell = \mathrm{softmax}\left(\frac{\mathbf{Q}_\ell \mathbf{K}_\ell^\top}{\sqrt{d}} + \mathbf{B}_\ell + \log\big(\max\{\mathbf{R}_\ell^\top, \varepsilon\}\big)\right)\mathbf{V}_\ell, \tag{4}$$

where $\mathbf{Q}_\ell$ and $(\mathbf{K}_\ell, \mathbf{V}_\ell)$ are projected from normalized decoder and encoder features, respectively. $\mathbf{B}_\ell$ is the relative positional bias, d is the head dimension, and $\varepsilon{>}0$ ensures numerical stability.

After merging the attention heads and restoring the spatial layout, we refine the encoder features using the difference between retrieved and source features, $\mathbf{U}_\ell{-}\mathbf{V}_\ell$:

$$\widehat{\mathbf{E}}_\ell = \mathbf{E}_\ell + (1 - \mathbf{R}_\ell) \odot \mathrm{Conv}_{1\times1}(\mathbf{U}_\ell - \mathbf{V}_\ell)\,, \tag{5}$$

where $\mathrm{Conv}_{1\times1}$ is a zero-initialized convolution and $\odot$ denotes element-wise multiplication. The factor $1{-}\mathbf{R}_\ell$ increases correction weights at less reliable locations. The refined encoder features are concatenated with $\mathbf{D}_\ell$ and processed by residual convolutional blocks, whose output is added to $\mathbf{D}_\ell$.

This design retains reliable encoder information and focuses refinement on less reliable regions. It supports detail reconstruction while reducing the influence of haze-degraded features.

3.3 Spatial Degradation Adapter

Spatially non-uniform haze attenuates contrast and structural information to varying degrees across an image, leading to different restoration needs. To effectively handle spatially varying haze degradation, we introduce the spatial degradation adapter (SDA). It extracts multi-scale haze-aware features from the hazy input and injects them into the U-Net to guide spatially adaptive restoration.

Haze representation. A convolutional encoder extracts multi-scale haze features $\{\mathbf{F}_\ell\}_{\ell=1}^{3}$ from the hazy input $\mathbf{I}$, supervised by local degradation and restoration needs. The supervision targets and training details are provided in Appendix A.2.1 of the supplementary material.

To illustrate the haze information captured by SDA, we visualize the first channel of its finest-scale features in Fig. 3. On NH-HAZE 55, the visible haze across the central region exhibits stronger responses, whereas the clearer upper-left region generally shows weaker responses. This spatial correspondence suggests that the learned features are sensitive to variations in haze distribution. Additional SDA feature visualizations on NH-HAZE and O-HAZE are provided in Appendix D.2.

U-Net adaptation. We align the haze features with three U-Net decoder stages, where convolutional adapters produce spatially gated residual corrections. At each scale, we inject features as:

$$\mathbf{F}'_{\mathrm{d}} = \mathbf{F}_{\mathrm{d}} + \mathrm{Adapter}(\mathbf{F}_{\mathrm{d}}, \mathbf{F}), \tag{6}$$

where $\mathbf{F}$ denotes haze features, $\mathbf{F}_{\mathrm{d}}$ and $\mathbf{F}'_{\mathrm{d}}$ are decoder features before and after injection, and $\mathrm{Adapter}$ denotes the haze-guided residual mapping.

The finest-scale haze features, input latent, and U-Net noise prediction jointly produce a residual correction to obtain $\boldsymbol{\epsilon}'$. Both feature and noise correction branches use zero-initialized output projections to preserve the original network output at initialization.

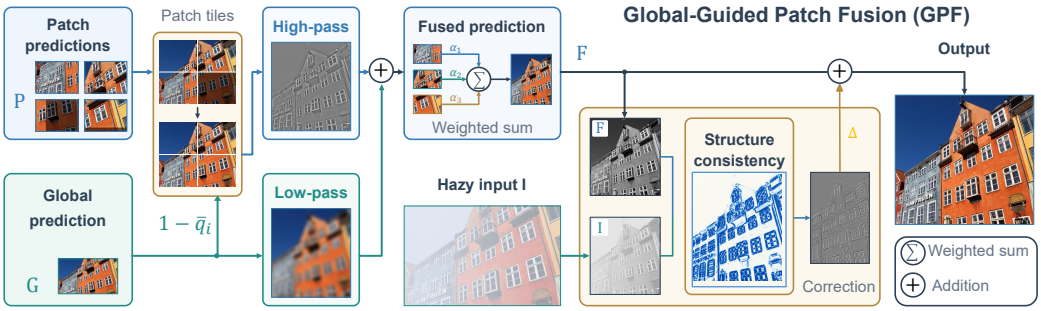


Figure 4: Overview of global-guided patch fusion (GPF). A global prediction guides patch alignment and reliability-aware fusion. Structural cues from the hazy input guide detail compensation.

Through multi-scale feature injection and noise prediction refinement, SDA incorporates spatially varying haze information into one-step denoising, adapting restoration to local degradation.

3.4 Global-Guided Patch Fusion

High-resolution diffusion-based dehazing often relies on patch-wise inference to accommodate limited GPU memory. However, each patch observes only partial scene context, which can lead to inconsistent color and illumination across regions. Averaging overlapping predictions can alleviate boundary artifacts, but cannot ensure global appearance consistency and may smooth fine details.

To address these limitations, we propose training-free global-guided patch fusion (GPF), which combines global appearance guidance with local detail preservation. As illustrated in Fig. 4, a prediction from the downsampled full image provides a common appearance reference, while original-resolution patch predictions supply complementary fine details. GPF consists of three components: appearance alignment, reliability-aware fusion, and detail compensation.

Appearance alignment. For each local prediction $\mathbf{P}_i$, we combine its multiscale wavelet high-frequency component with the low-frequency component of the corresponding global region $\mathbf{G}_i$. This preserves local structures while establishing a consistent global appearance.

Reliability-aware fusion. Patch-wise inference restricts each patch to local context, potentially producing abnormal predictions, such as excessively dark regions. Directly aggregating these predictions propagates such artifacts into the final image. To address this issue, we use the global prediction, which captures the full scene, as a reference for assessing patch reliability.

We downsample each raw patch prediction to the native resolution of the global branch. We then compute its PSNR-Y against the corresponding global region. A sigmoid mapping converts this score into a reliability weight q_i. Consistent patches receive higher weights to preserve local details. Inconsistent patches contribute less and rely more on the global reference.

To avoid abrupt transitions between patches, we aggregate their reliability weights using cosine windows $\mathbf{w}_i$ and apply Gaussian smoothing $\mathcal{S}$. The resulting spatial reliability map $\bar{\mathbf{q}}$ controls the blending of aligned local predictions with the global reference:

$$\bar{\mathbf{q}} = \mathcal{S}\left(\frac{\sum_i \mathbf{w}_i q_i}{\sum_i \mathbf{w}_i}\right), \quad \widehat{\mathbf{P}}_i = \bar{\mathbf{q}}_i \odot [\mathcal{W}_{\mathrm{H}}(\mathbf{P}_i) + \mathcal{W}_{\mathrm{L}}(\mathbf{G}_i)] + (1 - \bar{\mathbf{q}}_i) \odot \mathbf{G}_i. \tag{7}$$

$\mathbf{G}_i$ and $\bar{\mathbf{q}}_i$ denote the corresponding crops of the global prediction and reliability map, respectively. The operators $\mathcal{W}_{\mathrm{H}}$ and $\mathcal{W}_{\mathrm{L}}$ extract the wavelet high- and low-frequency components, respectively. Combining the two wavelet components yields the prediction produced by appearance alignment. Reliable regions retain more local information, while unreliable regions rely more on the global prediction. Finally, we aggregate the corrected patches using reliability weights and cosine windows:

$$\mathbf{I}^{\mathrm{f}} = \frac{\sum_i q_i \mathbf{w}_i \odot \widehat{\mathbf{P}}_i}{\sum_i q_i \mathbf{w}_i}. \tag{8}$$

Detail compensation. Aggregating patch predictions may attenuate high-frequency details. We therefore use structural information retained in the hazy input to compensate for detail loss in the

Table 1: Quantitative comparison with state-of-the-art methods on four real-world dehazing benchmarks. The best and second-best results are colored in red and blue, respectively.

Dataset	Metric	Non-diffusion			Diffusion / Flow					
		PTTD	IPC-Dehaze	BiLaLoRA	DiffUIR	FCDM	Diff-Dehazer	DiffDehaze	HazeFlow	Ours
I-HAZE	PSNR ↑	15.84	15.68	15.67	15.89	14.09	15.83	13.89	16.12	19.02
	SSIM ↑	0.6508	0.7985	0.7075	0.7705	0.4737	0.6709	0.7138	0.7675	0.7888
	LPIPS ↓	0.5072	0.4595	0.4638	0.3336	0.6258	0.4421	0.3725	0.4559	0.3217
	DISTS ↓	0.2172	0.2033	0.1566	0.2448	0.2370	0.1849	0.1861	0.1974	0.0946
	FID ↓	197.12	118.08	217.57	120.02	254.86	157.31	153.27	168.12	78.99
	Q-Align ↑	3.3117	3.4398	3.6040	2.8698	2.9932	3.1771	3.3588	3.4755	4.2306
O-HAZE	PSNR ↑	18.78	13.98	12.75	15.12	15.35	16.62	15.77	16.63	21.28
	SSIM ↑	0.6241	0.6601	0.6185	0.6413	0.5442	0.6332	0.5399	0.6484	0.6992
	LPIPS ↓	0.3922	0.5383	0.4617	0.4558	0.4820	0.4085	0.3987	0.4291	0.3236
	DISTS ↓	0.1914	0.2406	0.1998	0.2975	0.2059	0.1935	0.1763	0.2054	0.1087
	FID ↓	130.09	139.47	170.16	156.33	187.18	144.15	124.64	152.20	66.86
	Q-Align ↑	3.4973	3.3362	3.2017	2.9872	3.0500	3.2273	3.2613	3.3200	3.9219
Dense-Haze	PSNR ↑	13.67	8.46	11.54	9.90	12.69	10.94	13.18	12.10	17.29
	SSIM ↑	0.4536	0.4548	0.5055	0.4432	0.4088	0.4656	0.4636	0.5320	0.5676
	LPIPS ↓	0.5865	0.7131	0.6492	0.7676	0.7813	0.7149	0.6786	0.6374	0.4774
	DISTS ↓	0.3174	0.4617	0.4178	0.6137	0.4036	0.5034	0.4580	0.3834	0.2124
	FID ↓	256.00	271.59	262.50	272.00	275.20	256.24	210.41	236.11	172.18
	Q-Align ↑	2.4576	2.2399	2.9652	1.9165	2.0283	2.1314	2.5013	2.4468	3.3325
NH-HAZE	PSNR ↑	14.53	11.18	12.07	11.79	14.04	12.99	14.26	14.38	18.22
	SSIM ↑	0.6119	0.5850	0.5989	0.4788	0.5955	0.6079	0.3067	0.6343	0.6664
	LPIPS ↓	0.3934	0.4445	0.4398	0.5424	0.4264	0.3936	0.6763	0.4180	0.2863
	DISTS ↓	0.1938	0.2400	0.2106	0.3462	0.2159	0.2094	0.3374	0.2021	0.1272
	FID ↓	187.93	200.18	199.34	215.83	198.70	151.51	170.95	162.45	90.10
	Q-Align ↑	3.4309	3.4721	3.6929	3.1777	3.3738	3.4304	3.2871	3.1193	3.8066

fused result. First, we estimate a local structural agreement map $\mathbf{s}$ between the hazy input $\mathbf{I}$ and the fused result $\mathbf{I}^{\mathrm{f}}$. Its values lie in $[0, 1]$ and guide spatially adaptive detail compensation.

We then extract high-frequency information from both images using self-guided filtering, obtaining detail layers $\mathbf{H}$ and $\mathbf{H}^{\mathrm{f}}$, respectively. Guided by structural agreement, we adaptively retain restored details and incorporate input details as follows:

$$\widehat{\mathbf{H}} = \big[\alpha + \mathbf{s}(1-\alpha)\big] \odot \mathbf{H}^{\mathrm{f}} + \beta \mathbf{s} \odot \mathbf{H}. \tag{9}$$

The fixed coefficients $\alpha \in (0, 1)$ and $\beta > 0$ control the retention of restored details and compensation from input details, respectively. Regions with stronger structural agreement retain more restored details and receive more detail from the input. Regions with weaker agreement have attenuated detail responses, reducing the influence of inconsistent structures. Finally, we combine compensated details with low-frequency information from the fused result to reconstruct the output $\widehat{\mathbf{I}}$.

Through appearance alignment, reliability-aware fusion, and structure-guided detail compensation, GPF improves cross-patch consistency while preserving fine structures in high-resolution dehazing.

4 Experiments

4.1 Experimental Settings

Datasets. We train our model on paired hazy and clear images from LMHaze (Zhang et al., 2024) and RW2AH (Fang et al., 2025). For evaluation, we use all image pairs from four real-world benchmarks: I-HAZE (Ancuti et al., 2018b) (30 pairs), O-HAZE (Ancuti et al., 2018a) (45 pairs), Dense-Haze (Ancuti et al., 2019) (55 pairs), and NH-HAZE (Ancuti et al., 2020) (55 pairs). I-HAZE and O-HAZE cover indoor and outdoor scenes, respectively. Dense-Haze contains densely hazed scenes, while NH-HAZE features spatially non-uniform haze.

Evaluation Metrics. We assess restoration quality using six metrics covering reconstruction fidelity, perceptual similarity, and no-reference image quality. PSNR and SSIM (Wang et al., 2004) measure fidelity to the ground truth, while LPIPS (Zhang et al., 2018) and DISTS (Ding et al., 2020) assess perceptual similarity between restored and clear images. We also report FID (Heusel et al., 2017) to measure the distance between the feature distributions of restored and clear images at the dataset level. For no-reference evaluation, we employ Q-Align (Wu et al., 2023a) to assess image quality.

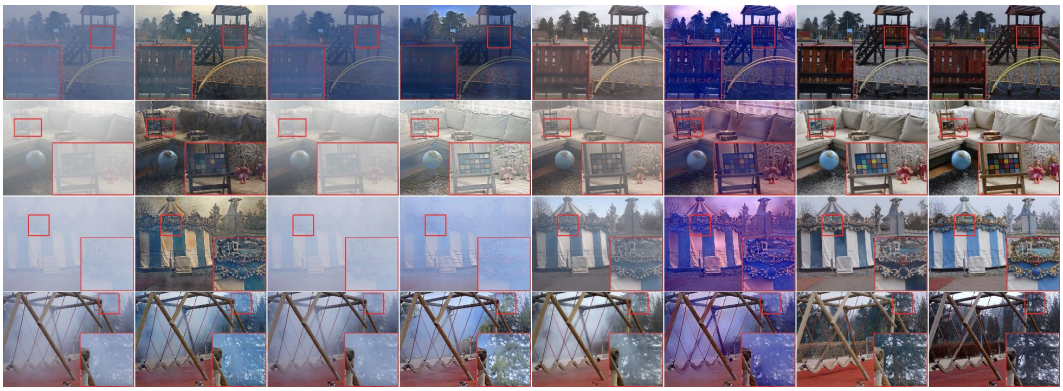


Figure 5: Visual comparisons on O-HAZE, I-HAZE, Dense-Haze, and NH-HAZE, from top to bottom. Red boxes mark the corresponding regions shown in the enlarged views.

Table 2: Inference time and model complexity for 3,328×4,096 inputs on an RTX A6000.

Metric	IPC-Dehaze	DiffUIR	FCDM	Diff-Dehazer	DiffDehaze	HazeFlow	Ours
Runtime (s)	185.4	43.5	146.6	32.5	1,046.8	15.2	12.9
FLOPs (T)	826.391	236.731	2,022.874	453.871	9,386.778	250.991	458.955
Params (M)	43.164	36.261	155.334	1,506.159	1,666.750	87.087	975.822

Implementation Details. We initialize the network from Stable Diffusion 2.1-base (Rombach et al., 2022) and use LoRA with rank 16. With AdamW (Loshchilov et al., 2017), we train the U-Net LoRA and ASC for 200,000 iterations on 512×512 crops. We then train the SDA encoder on 512×512 crops and the U-Net adaptation branches on $1{,}024\times1{,}024$ crops for 10,000 iterations each. Inference uses one denoising step at t=399; high-resolution images are processed with overlapping $1{,}024 \times 1{,}024$ patches and GPF. The complete training strategy, including loss functions, learning rates, batch sizes, and other settings, is provided in Appendix A of the supplementary material.

4.2 Comparison with State-of-the-Art Methods

Compared Methods. We compare our method with eight representative dehazing approaches. The non-diffusion methods include PTTD (Chen et al., 2024), IPC-Dehaze (Fu et al., 2025), and Bi-LaLoRA (Zhang et al., 2026). The diffusion-based methods include DiffUIR (Zheng et al., 2024), FCDM (Wang et al., 2024b), Diff-Dehazer (Lan et al., 2025a), DiffDehaze (Wang et al., 2025), and HazeFlow (Shin et al., 2025). Appendix B extends the comparison to five additional methods, reporting all six metrics on the same four benchmarks.

Quantitative Results. Table 1 presents quantitative comparisons on real-world dehazing benchmarks. Our method achieves the best PSNR across all benchmarks and the best SSIM on three of the four datasets. It also consistently obtains the lowest LPIPS, DISTS, and FID, along with the highest Q-Align scores. These results demonstrate that our framework achieves favorable reconstruction fidelity and perceptual quality across diverse real-world scenes, including challenging scenes affected by dense and spatially non-uniform haze distributions.

Qualitative Results. We provide visual comparisons on I-HAZE, O-HAZE, Dense-Haze, and NH-HAZE in Fig. 5. Our method produces clearer and more natural dehazed results across diverse real-world scenes. Compared with existing methods, it more effectively suppresses residual haze while preserving structural and textural details under spatially non-uniform haze. Dense-Haze remains particularly challenging: its extremely dense haze severely obscures scene appearance, and our restored result still exhibits some color discrepancies relative to the ground truth. The four examples show scene views and enlarged local details; more full-image comparisons appear in Appendix D.1.

Running Time Comparisons. We compare large-image algorithm latency and model complexity of different methods in Tab. 2. The reported runtimes use $3{,}328 \times 4{,}096$ inputs on an NVIDIA RTX A6000 GPU and include the complete deployed inference pipeline. Our method has the lowest runtime at 12.9 s, making it 1.18 times as fast as HazeFlow (15.2 s), the fastest competing method, and 2.52 times as fast as Diff-Dehazer (32.5 s). Although our method has more FLOPs and parameters than HazeFlow, it achieves lower end-to-end latency under this evaluation protocol.

Table 3: Ablation studies and GPF transferability on O-HAZE. (a) Base denotes the model without skip connections; Zero-conv uses zero-conv skip connections. (b) Conditioning strategies with ASC included in the base model. (c) GPF components: AA, appearance alignment; RF, reliability-aware fusion; DC, detail compensation. (d) GPF applied to other dehazing methods without retraining.

(a) Ablation of ASC.

Metric	base	+ Zero-conv	+ ASC
PSNR ↑	17.36	18.80	**19.79**
SSIM ↑	0.5636	0.6438	**0.6508**
LPIPS ↓	0.3837	0.3484	**0.3418**
DISTS ↓	0.1465	**0.1327**	0.1334

(b) Ablation of SDA.

Method	PSNR ↑	SSIM ↑	LPIPS ↓	DISTS ↓
base	19.79	0.6508	0.3418	0.1334
+ T2I-Adapter	19.7800	0.6500	0.3404	0.1235
+ Noise refinement	19.78	0.6483	0.3406	0.1291
+ SDA	**20.08**	**0.6598**	**0.3375**	**0.1216**

(c) Ablation of GPF.

AA	RF	DC	PSNR ↑	SSIM ↑	LPIPS ↓	DISTS ↓
–	–	–	20.08	0.6598	0.3375	0.1216
✓	–	–	21.08	0.6664	0.3282	0.1166
✓	✓	–	21.27	0.6979	0.3285	0.1101
✓	✓	✓	**21.28**	**0.6992**	**0.3236**	**0.1087**

(d) GPF transferability.

Method	PSNR ↑	SSIM ↑	LPIPS ↓	DISTS ↓
HazeFlow	16.63	0.6484	**0.4291**	0.2054
+ GPF	**17.95**	**0.6782**	0.4295	**0.1732**
Diff-Dehazer	16.62	0.6332	**0.4085**	0.1935
+ GPF	**16.78**	**0.6456**	0.4089	**0.1783**

4.3 Ablation Studies

We investigate the effectiveness of attentive skip compensation (ASC), spatial degradation adapter (SDA), and global-guided patch fusion (GPF) on O-HAZE. Specifically, we compare different VAE skip-connection designs and conditioning strategies, and progressively introduce the GPF components to examine their individual contributions. We further evaluate the transferability of GPF to two other dehazing methods without retraining. All variants are evaluated using PSNR, SSIM, LPIPS, and DISTS to assess fidelity and perceptual similarity, as shown in Tab. 3.

Attentive skip compensation. We compare three VAE variants in Tab. 3a: without skip connections, with zero-conv skip connections, and with ASC. Adding zero-conv improves all four metrics over the no-skip baseline, showing the benefit of introducing encoder features into reconstruction. ASC further improves PSNR, SSIM, and LPIPS by adaptively compensating encoder features, while zero-conv achieves the lowest DISTS. These results support the effectiveness of ASC for reconstruction.

Spatial degradation adapter. We compare different conditioning strategies using the model with ASC as the baseline. As shown in Tab. 3b, T2I-Adapter and noise refinement improve LPIPS and DISTS, but slightly reduce PSNR and SSIM. By incorporating spatially varying haze information, SDA improves all four metrics over the baseline and achieves the best results among the compared strategies. This suggests that haze-aware conditioning provides effective guidance for restoration.

Global-guided patch fusion. We progressively add the GPF components in Tab. 3c. First, appearance alignment improves all four metrics by aligning local predictions with the global reference. Then, reliability-aware fusion weights patches according to their consistency with the global prediction, further improving PSNR, SSIM, and DISTS. Finally, detail compensation uses structural cues from the hazy input and improves all four metrics, supporting the contribution of each component.

GPF transferability. We apply the same GPF procedure to HazeFlow (Shin et al., 2025) and Diff-Dehazer (Lan et al., 2025a) without retraining. As shown in Tab. 3d, GPF improves PSNR, SSIM, and DISTS for both methods on O-HAZE, although LPIPS increases slightly. These results indicate that the benefits of GPF extend to other dehazing methods. Appendix C provides results for all six metrics on all four benchmarks, together with qualitative comparisons before and after applying GPF to further examine its effectiveness across different methods and haze conditions.

5 Conclusion

In this paper, we propose a one-step diffusion framework for high-resolution real-world image dehazing. To handle spatially non-uniform haze, Hihaze employs ASC to preserve reliable encoder features and SDA to inject spatial degradation information into one-step denoising. Moreover, GPF uses a global prediction to guide local patch fusion, improving cross-patch consistency while preserving fine details. Extensive experiments on real-world benchmarks demonstrate favorable restoration quality and inference efficiency across diverse haze conditions.

AI Use Statement

Generative AI tools were used to aid and polish the writing of this paper, including translation, language editing, and improvements to clarity and presentation. All AI-assisted text was reviewed and revised by the authors, who take full responsibility for the final content of the paper. Generative AI tools were not used to generate data or to assist with the qualitative or thematic analysis of the data; these processes were conducted independently by the researchers.

Ethics Statement

This work studies single-image dehazing using existing research datasets and pretrained models. No human-subject studies were conducted. We do not identify any specific ethical concerns arising from this research on restoring hazy images using existing data.

Reproducibility Statement

Section 4.1 describes the training datasets, evaluation benchmarks, and implementation settings. Appendix A of the supplementary material provides the complete training strategy, loss functions, and detailed configurations. We plan to release code and trained weights upon acceptance.

References

Yuang Ai, Huaibo Huang, Xiaoqiang Zhou, Jiexiang Wang, and Ran He. Multimodal prompt perceiver: Empower adaptiveness, generalizability and fidelity for all-in-one image restoration. In *CVPR*, 2024.

Codruta Ancuti, Cosmin Ancuti, Radu Timofte, De Vleeschouwer, et al. O-haze: a dehazing benchmark with real hazy and haze-free outdoor images. In *CVPRW*, 2018a.

Codruta O Ancuti, Cosmin Ancuti, Mateu Sbert, and Radu Timofte. Dense-haze: A benchmark for image dehazing with dense-haze and haze-free images. In *ICIP*, 2019.

Codruta O Ancuti, Cosmin Ancuti, and Radu Timofte. Nh-haze: An image dehazing benchmark with non-homogeneous hazy and haze-free images. In *CVPRW*, 2020.

Cosmin Ancuti, Codruta O Ancuti, Radu Timofte, and Christophe De Vleeschouwer. I-haze: A dehazing benchmark with real hazy and haze-free indoor images. In *International conference on advanced concepts for intelligent vision systems*, 2018b.

Dana Berman, Shai Avidan, et al. Non-local image dehazing. In *CVPR*, 2016.

Bolun Cai, Xiangmin Xu, Kui Jia, Chunmei Qing, and Dacheng Tao. Dehazenet: An end-to-end system for single image haze removal. *IEEE TIP*, 2016.

I Chen, Wei-Ting Chen, Yu-Wei Liu, Yuan-Chun Chiang, Sy-Yen Kuo, Ming-Hsuan Yang, et al. Unirestore: Unified perceptual and task-oriented image restoration model using diffusion prior. In *CVPR*, 2025.

Zixuan Chen, Zewei He, Ziqian Lu, Xuecheng Sun, and Zhe-Ming Lu. Prompt-based test-time real image dehazing: a novel pipeline. In *ECCV*, 2024.

Keyan Ding, Kede Ma, Shiqi Wang, and Eero P Simoncelli. Image quality assessment: Unifying structure and texture similarity. *IEEE TPAMI*, 2020.

Zheng Ding, Mengqi Zhang, Jiajun Wu, and Zhuowen Tu. Patched denoising diffusion models for high-resolution image synthesis. In *ICLR*, 2024.

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 *CVPR*, 2020.

Chengyu Fang, Chunming He, Fengyang Xiao, Yulun Zhang, Longxiang Tang, Yuelin Zhang, Kai Li, and Xiu Li. Real-world image dehazing with coherence-based pseudo labeling and cooperative unfolding network. In *NeurIPS*, 2024.

Wenxuan Fang, Junkai Fan, Yu Zheng, Jiangwei Weng, Ying Tai, and Jun Li. Guided real image dehazing using ycbcr color space. In *AAAI*, 2025.

Raanan Fattal. Single image dehazing. *ACM TOG*, 2008.

Raanan Fattal. Dehazing using color-lines. *ACM TOG*, 2014.

Jiayi Fu, Siyu Liu, Zikun Liu, Chun-Le Guo, Hyunhee Park, Ruiqi Wu, Guoqing Wang, and Chongyi Li. Iterative predictor-critic code decoding for real-world image dehazing. In *CVPR*, 2025.

Chun-Le Guo, Qixin Yan, Saeed Anwar, Runmin Cong, Wenqi Ren, and Chongyi Li. Image dehazing transformer with transmission-aware 3d position embedding. In *CVPR*, 2022.

Kaiming He, Jian Sun, and Xiaoou Tang. Single image haze removal using dark channel prior. *IEEE TPAMI*, 2010.

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. In *NeurIPS*, 2017.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *NeurIPS*, 2020.

Yunwei Lan, Zhigao Cui, Chang Liu, Jialun Peng, Nian Wang, Xin Luo, and Dong Liu. Exploiting diffusion prior for real-world image dehazing with unpaired training. In *AAAI*, 2025a.

Yunwei Lan, Zhigao Cui, Xin Luo, Chang Liu, Nian Wang, Menglin Zhang, Yanzhao Su, and Dong Liu. When schrödinger bridge meets real-world image dehazing with unpaired training. In *ICCV*, 2025b.

Jianze Li, Jiezhang Cao, Zichen Zou, Xiongfei Su, Xin Yuan, Yulun Zhang, Yong Guo, and Xiaokang Yang. Unleashing the power of one-step diffusion based image super-resolution via a large-scale diffusion discriminator. In *NeurIPS*, 2026a.

Mingjia Li, Tianle Du, Hainuo Wang, Qiming Hu, and Xiaojie Guo. On the global photometric alignment for low-level vision. *arXiv preprint arXiv:2604.08172*, 2026b.

Xiaohong Liu, Yongrui Ma, Zhihao Shi, and Jun Chen. Griddehazenet: Attention-based multi-scale network for image dehazing. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 7313–7322. IEEE, 2019.

Yuhao Liu, Zhanghan Ke, Fang Liu, Nanxuan Zhao, and Rynson WH Lau. Diff-plugin: Revitalizing details for diffusion-based low-level tasks. In *CVPR*, 2024.

Ilya Loshchilov, Frank Hutter, et al. Fixing weight decay regularization in adam. *arXiv preprint arXiv:1711.05101*, 2017.

Andreas Lugmayr, Martin Danelljan, Andres Romero, Fisher Yu, Radu Timofte, and Luc Van Gool. Repaint: Inpainting using denoising diffusion probabilistic models. In *CVPR*, 2022.

Long Ma, Yuxin Feng, Yan Zhang, Jinyuan Liu, Weimin Wang, Guang-Yong Chen, Chengpei Xu, and Zhuo Su. Coa: Towards real image dehazing via compression-and-adaptation. In *CVPR*, 2025.

Earl J McCartney and Freeman F Hall Jr. Optics of the atmosphere: scattering by molecules and particles, 1977.

Gaofeng Meng, Ying Wang, Jiangyong Duan, Shiming Xiang, and Chunhong Pan. Efficient image dehazing with boundary constraint and contextual regularization. In *ICCV*, 2013.

Srinivasa G Narasimhan and Shree K Nayar. Vision and the atmosphere. *IJCV*, 2002.

Ozan Özdenizci and Robert Legenstein. Restoring vision in adverse weather conditions with patch-based denoising diffusion models. *IEEE TPAMI*, 2023.

Xu Qin, Zhilin Wang, Yuanchao Bai, Xiaodong Xie, and Huizhu Jia. Ffa-net: Feature fusion attention network for single image dehazing. In *AAAI*, 2020.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Bj¨orn Ommer. High-resolution image synthesis with latent diffusion models. In *CVPR*, 2022.

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 *CVPR*, 2023.

Chitwan Saharia, William Chan, Huiwen Chang, Chris Lee, Jonathan Ho, Tim Salimans, David Fleet, and Mohammad Norouzi. Palette: Image-to-image diffusion models. In *SIGGRAPH*, 2022a.

Chitwan Saharia, Jonathan Ho, William Chan, Tim Salimans, David J Fleet, and Mohammad Norouzi. Image super-resolution via iterative refinement. *IEEE TPAMI*, 2022b.

Junseong Shin, Seungwoo Chung, Yunjeong Yang, and Tae Hyun Kim. Hazeflow: Revisit haze physical model as ode and non-homogeneous haze generation for real-world dehazing. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 6263–6272, 2025.

Oriane Siméoni, Huy V Vo, Maximilian Seitzer, Federico Baldassarre, Maxime Oquab, Cijo Jose, Vasil Khalidov, Marc Szafraniec, Seungeun Yi, Michaël Ramamonjisoa, et al. Dinov3. *arXiv preprint arXiv:2508.10104*, 2025.

Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502*, 2020.

Yuda Song, Zhuqing He, Hui Qian, and Xin Du. Vision transformers for single image dehazing. *arXiv preprint arXiv:2204.03883*, 2022.

Jianyi Wang, Zongsheng Yue, Shangchen Zhou, Kelvin CK Chan, and Chen Change Loy. Exploiting diffusion prior for real-world image super-resolution. *IJCV*, 2024a.

Jing Wang, Songtao Wu, Zhiqiang Yuan, Qiang Tong, and Kuanhong Xu. Frequency compensated diffusion model for real-scene dehazing. *Neural Networks*, 2024b.

Ruiyi Wang, Yushuo Zheng, Zicheng Zhang, Chunyi Li, Shuaicheng Liu, Guangtao Zhai, and Xiaohong Liu. Learning hazing to dehazing: Towards realistic haze generation for real-world image dehazing. In *CVPR*, 2025.

Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE TIP*, 2004.

Haoning Wu, Zicheng Zhang, Weixia Zhang, Chaofeng Chen, Liang Liao, Chunyi Li, Yixuan Gao, Annan Wang, Erli Zhang, Wenxiu Sun, et al. Q-align: Teaching lmms for visual scoring via discrete text-defined levels. *arXiv preprint arXiv:2312.17090*, 2023a.

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 *CVPR*, 2023b.

Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *CVPR*, 2018.

Ruikun Zhang, Hao Yang, Yan Yang, Ying Fu, and Liyuan Pan. Lmhaze: intensity-aware image dehazing with a large-scale multi-intensity real haze dataset. In *ACM MM Asia*, 2024.

Yan Zhang, Long Ma, Yuxin Feng, Zhe Huang, Fan Zhou, and Zhuo Su. Bilevel layer-positioning lora for real image dehazing. *arXiv preprint arXiv:2603.10872*, 2026.

Dian Zheng, Xiao-Ming Wu, Shuzhou Yang, Jian Zhang, Jian-Fang Hu, and Wei-Shi Zheng. Selective hourglass mapping for universal image restoration based on diffusion model. In *CVPR*, 2024.

Shangchen Zhou, Kelvin Chan, Chongyi Li, and Chen Change Loy. Towards robust blind face restoration with codebook lookup transformer. *NeurIPS*, 2022.

A Training Strategy and Implementation Details

We train Hihaze in two stages. First, we train the restoration backbone and ASC. We then freeze this model to construct supervision for SDA, train the SDA encoder, and optimize its adaptation branches. The objectives and settings for both stages follow. GPF requires no training.

A.1 Stage 1: Backbone and ASC Training

We initialize the model from Stable Diffusion 2.1-base and train U-Net LoRA, ASC and its feature-fusion layers, and a shared learnable text embedding. The pretrained U-Net weights, VAE, and text encoder remain fixed. LoRA uses rank 16 and scaling parameter 16. ASC uses two intermediate VAE decoder stages with four attention heads, projected dimension 64, and 8×8 local windows.

Restoration objective. The restoration loss is defined as

$$\mathcal{L}_{\mathrm{b}} = \mathcal{L}_1 + \lambda_{\mathrm{pal}}\mathcal{L}_{\mathrm{pal}} + \lambda_{\mathrm{per}}\mathcal{L}_{\mathrm{per}} + \lambda_{\mathrm{adv}}\mathcal{L}_{\mathrm{adv}}, \tag{10}$$

where $\mathcal{L}_1$ is the L1 loss and $\mathcal{L}_{\mathrm{pal}}$ is the affinely aligned pixel loss (PAL) (Li et al., 2026b). The perceptual term $\mathcal{L}_{\mathrm{per}}$ sums DISTS on RGB images and their Scharr edge-magnitude maps (Li et al., 2026a). The adversarial term $\mathcal{L}_{\mathrm{adv}}$ is computed with a DINOv3-based discriminator (Siméoni et al., 2025). We set λ_{pal}=0.8, λ_{per}=1, and λ_{adv}=0.1. PAL fits an affine RGB transformation with a bias and ridge coefficient 10^{-3}. The transformation is detached before computing the L1 distance between the aligned prediction and clean reference.

ASC reliability supervision. In addition to the restoration objective, we supervise ASC's spatial reliability maps using paired hazy and clean images $\mathbf{I}, \mathbf{I}_{\mathrm{c}} \in [0,1]$. The target is defined as

$$\mathbf{R}^* = \exp\left[-\alpha_{\mathrm{rel}}\ \mathrm{Avg}_{11\times11}(\mathrm{Mean}_c\,|\mathbf{I}-\mathbf{I}_{\mathrm{c}}|)\right], \tag{11}$$

where Mean_c averages RGB channels, $\mathrm{Avg}_{11\times11}$ denotes local spatial averaging, and α_{rel}=4 controls the decay with image difference. Regions closer to the clean reference have higher target reliability. We average-pool $\mathbf{R}^*$ to each ASC scale and supervise the predicted map $\mathbf{R}_\ell$ with a mean L1 loss, averaged over layers and weighted by 0.05. Inputs to the reliability predictor and reliability values used for attention routing and gating are detached.

Training settings. Training runs for 200,000 iterations on 512×512 crops with a total batch size of four. The restoration learning rate is 5×10^{-5} and the weight decay is 0.01. The DINOv3 discriminator uses a frozen ConvNeXt-Large feature extractor; its head is trained at learning rate 10^{-5} and weight decay 0.01, alternating with restoration updates. Each optimizer uses 500 warmup updates followed by a constant learning rate for the remaining iterations.

A.2 Stage 2: SDA Training

We first train the SDA encoder to capture local degradation and restoration needs, and then train the adaptation branches to use these features. The Stage 1 model remains frozen in both phases.

A.2.1 SDA Encoder

Given a hazy input $\mathbf{I}$, its clean reference $\mathbf{I}_{\mathrm{c}}$, and the restored image $\mathbf{I}_{\mathrm{b}}$ produced by the frozen Stage 1 model, we construct three spatial supervision targets for training the SDA encoder:

$$\mathbf{T}_{\mathrm{d}} = 1 - \frac{\Phi(\mathbf{I})+\varepsilon_{\mathrm{str}}}{\Phi(\mathbf{I}_{\mathrm{c}})+\varepsilon_{\mathrm{str}}}, \quad \mathbf{T}_{\mathrm{r}} = \frac{\Phi(\mathbf{I}_{\mathrm{c}})-\Phi(\mathbf{I}_{\mathrm{b}})}{\Phi(\mathbf{I}_{\mathrm{c}})+\varepsilon_{\mathrm{str}}}, \quad \mathbf{T}_{\mathrm{s}} = \mathrm{sigmoid}\left(\frac{\Phi(\mathbf{I}_{\mathrm{c}})-\tau}{\kappa}\right). \tag{12}$$

Here, $\mathbf{T}_{\mathrm{d}}$, $\mathbf{T}_{\mathrm{r}}$, and $\mathbf{T}_{\mathrm{s}}$ represent structural attenuation, residual correction, and structural confidence, respectively. The function $\Phi(\cdot)$ combines local Sobel-gradient and contrast statistics over windows of 7, 15, and 31 pixels. We set $\varepsilon_{\mathrm{str}}$=$10^{-3}$ to stabilize the ratios. The parameters τ=0.03 and κ=0.01 control the threshold and smoothness of the sigmoid transition, respectively. Structural attenuation and residual correction are clipped to $[0,1]$ and $[-1,1]$, respectively.

The RGB input is resized to a long side of 512 while preserving aspect ratio and padded to multiples of eight. The three feature scales have 32 channels each, with spatial sizes 64×64, 32×32, and 16×16 for a square input. The three target maps are resized by area interpolation, concatenated, and repeated across channels to match each feature tensor. Encoder supervision uses only Smooth L1 with unit transition parameter, averaged over tensor elements and the three scales.

Training settings. We train only the SDA encoder for 10,000 updates on 512×512 crops with batch size one, learning rate 2×10^{-4}, and weight decay 10^{-4}. The learning rate uses 500 warmup updates followed by cosine decay to one tenth of its peak value.

A.2.2 SDA Adaptation Branches

After encoder training, we freeze the encoder and train only the three U-Net decoder adaptation branches and the noise-correction branch. Their output projections are zero-initialized. We use the restoration loss $\mathcal{L}_{\mathrm{b}}$ in Eq. (10) with the same weights as in Stage 1, and an additional 10^{-5}-weighted mean squared noise-correction penalty. The entire discriminator remains frozen.

Training settings. Training runs for 10,000 updates on $1{,}024 \times 1{,}024$ crops using four devices with one crop per device. The learning rate is 10^{-4}, with weight decay 0.01, 500 warmup updates, and cosine decay to zero. L1, PAL, and the discriminator objective are evaluated at $1{,}024 \times 1{,}024$; RGB and Scharr-edge DISTS are evaluated after resizing the images to 512×512.

A.3 Shared Data Processing and Numerical Settings

All phases use AdamW with $(\beta_1, \beta_2)=(0.9, 0.999)$ and seed 10. Hazy and clean images share the same crop coordinates and augmentation. Backbone training uses random flips and rotations; the two SDA phases use horizontal and vertical flips. Images smaller than the requested SDA crop are resized while preserving aspect ratio before cropping. During adaptation training, the smaller dataset partition is repeated to balance LMHaze and RW2AH sampling. The maximum gradient norm is 10 for backbone training and 1 for both SDA phases. SDA encoder training uses BF16 mixed precision; backbone and adaptation training use FP32 for restoration optimization.

Table 4: Additional quantitative comparisons on I-HAZE, O-HAZE, Dense-Haze, and NH-HAZE. The best and second-best values within this table are colored in red and blue, respectively.

Dataset	Metric	Non-diffusion					Diffusion / Flow
		DehazeFormer	RIDCP	CORUN	CoA	DehazeSB	Ours
I-HAZE	PSNR ↑	16.38	15.63	17.10	13.82	16.04	19.02
	SSIM ↑	0.7835	0.7116	0.7658	0.4888	0.6135	0.7888
	LPIPS ↓	0.3360	0.4158	0.4093	0.6303	0.5019	0.3217
	DISTS ↓	0.2188	0.1645	0.2142	0.2512	0.1783	0.0946
	FID ↓	135.46	204.88	136.26	214.98	202.65	78.99
	Q-Align ↑	2.9887	3.6969	3.5317	3.7021	3.3250	4.2306
O-HAZE	PSNR ↑	16.36	14.14	15.02	9.99	15.20	21.28
	SSIM ↑	0.6657	0.6244	0.6179	0.5851	0.6346	0.6992
	LPIPS ↓	0.4431	0.4166	0.5104	0.5375	0.4569	0.3236
	DISTS ↓	0.2734	0.2008	0.2729	0.2492	0.1910	0.1087
	FID ↓	153.79	175.76	148.36	206.91	161.83	66.86
	Q-Align ↑	3.0723	3.5222	3.3406	3.3174	3.4124	3.9219
Dense-Haze	PSNR ↑	10.85	9.95	10.44	9.59	11.11	17.29
	SSIM ↑	0.4621	0.4686	0.4911	0.4874	0.4210	0.5676
	LPIPS ↓	0.7456	0.6827	0.6858	0.7166	0.6686	0.4774
	DISTS ↓	0.6006	0.4515	0.4090	0.4161	0.3499	0.2124
	FID ↓	270.10	294.78	251.70	265.38	277.73	172.18
	Q-Align ↑	1.9896	2.6232	2.5245	2.6543	2.9389	3.3325
NH-HAZE	PSNR ↑	12.29	12.50	12.43	11.04	12.69	18.22
	SSIM ↑	0.4959	0.5416	0.5738	0.5731	0.5907	0.6664
	LPIPS ↓	0.5300	0.4479	0.5664	0.4986	0.4421	0.2863
	DISTS ↓	0.3258	0.2306	0.2735	0.2262	0.1983	0.1272
	FID ↓	215.00	224.35	186.06	216.72	199.67	90.10
	Q-Align ↑	3.2526	3.5843	3.5282	3.5751	3.6618	3.8066

B Additional Quantitative Comparisons

Table 4 reports results for the five methods omitted from the compact comparison in the main paper: DehazeFormer (Song et al., 2022), RIDCP (Wu et al., 2023b), CORUN (Fang et al., 2024), CoA (Ma et al., 2025), and DehazeSB (Lan et al., 2025b). The table uses the same four benchmarks and six metrics as Sec. 4.2, and repeats our results to allow direct comparison. Red and blue identify the best and second-best values within this table, respectively.

Our method achieves the best values for all six metrics on each benchmark among the methods in Table 4. Together with Table 1, these results extend the comparison beyond the methods presented in the main paper, covering both non-diffusion and diffusion-based approaches.

Table 5: Results of applying GPF to HazeFlow on all four benchmarks. The best results for each metric within each dataset are highlighted in bold.

Dataset	Setting	PSNR ↑	SSIM ↑	LPIPS ↓	DISTS ↓	FID ↓	Q-Align ↑
O-HAZE	Raw patch	16.62	0.6484	**0.4291**	0.2054	152.20	3.3200
	+ GPF	**17.95**	**0.6782**	0.4295	**0.1732**	**133.54**	**3.4142**
NH-HAZE	Raw patch	14.38	0.6343	0.4180	0.2021	162.45	**3.1193**
	+ GPF	**14.72**	**0.6396**	**0.4105**	**0.1901**	**157.43**	3.0556
Dense-Haze	Raw patch	12.10	0.5320	**0.6374**	0.3834	236.11	**2.4468**
	+ GPF	**12.57**	**0.5416**	0.6430	**0.3828**	**225.49**	2.4267
I-HAZE	Raw patch	16.11	0.7675	0.4559	0.1974	168.12	3.4755
	+ GPF	**17.88**	**0.8115**	**0.4130**	**0.1453**	**131.19**	**3.7266**

Table 6: Results of applying GPF to Diff-Dehazer on all four benchmarks. The best results for each metric within each dataset are highlighted in bold.

Dataset	Setting	PSNR ↑	SSIM ↑	LPIPS ↓	DISTS ↓	FID ↓	Q-Align ↑
O-HAZE	Raw patch	16.62	0.6332	**0.4085**	0.1935	144.15	3.2273
	+ GPF	**16.78**	**0.6456**	0.4089	**0.1783**	**109.77**	**3.5442**
NH-HAZE	Raw patch	12.98	0.6079	0.3936	0.2094	151.50	**3.4304**
	+ GPF	**13.11**	**0.6156**	**0.3841**	**0.1981**	**145.86**	3.4129
Dense-Haze	Raw patch	10.94	0.4656	0.7149	0.5034	256.23	2.1314
	+ GPF	**11.25**	**0.4704**	**0.6946**	**0.4834**	**244.70**	**2.1626**
I-HAZE	Raw patch	15.83	0.6709	0.4421	0.1849	157.31	3.1771
	+ GPF	**16.50**	**0.6752**	**0.4293**	**0.1804**	**123.66**	**3.6387**

C GPF Transferability to Other Dehazing Methods

To evaluate its transferability, we apply GPF to HazeFlow (Shin et al., 2025) and Diff-Dehazer (Lan et al., 2025a) without retraining their restoration networks. Table 3d gives the O-HAZE results using four metrics; Tables 5–6 report all six metrics on all four benchmarks, including FID and Q-Align for O-HAZE. Raw patch denotes the original patch-based output, and + GPF denotes the final output after global-guided fusion and detail compensation.

GPF improves PSNR, SSIM, DISTS, and FID in all eight method–dataset pairs. LPIPS and Q-Align exhibit some method- and dataset-dependent trade-offs, as discussed below. Figures 6 and 7 complement these results with five selected O-HAZE examples per method. These examples illustrate reduced patch-dependent appearance variations; the tables report dataset-wide results.

HazeFlow.

The largest PSNR improvements occur on I-HAZE and O-HAZE, at 1.78 dB and 1.32 dB, respectively. LPIPS increases slightly on O-HAZE and Dense-Haze, while Q-Align decreases on NH-HAZE and Dense-Haze. Thus, improved global consistency does not uniformly benefit all perceptual metrics.

Diff-Dehazer.

Diff-Dehazer also benefits from GPF across all four benchmarks. LPIPS improves on three datasets but increases slightly on O-HAZE; Q-Align improves except on NH-HAZE. The selected visual examples highlight reductions in block-shaped brightness differences across adjacent regions.



Figure 6: Visual comparison of HazeFlow before and after applying GPF on O-HAZE. Rows show samples 03, 15, 17, 21, and 28, from top to bottom. Insets show matched crops.



Figure 7: Visual comparison of Diff-Dehazer before and after applying GPF on O-HAZE. Rows show samples 03, 07, 21, 24, and 33, from top to bottom. Insets show matched crops.

D Additional Qualitative Results

This appendix extends the comparisons in Sec. 4.2 and the SDA visualization in Sec. 3.3. Full-image views across all four benchmarks show scene-wide color, illumination, and residual haze.

D.1 Full-Image Qualitative Comparisons

Figures 8–11 present 12 full-image comparisons, three per benchmark. Each example contains two rows of four images: hazy input, PTTD, DiffUIR, and Diff-Dehazer in the first row; DiffDehaze, HazeFlow, our method, and ground truth (GT) in the second. Sample IDs appear in the GT labels.



Figure 8: Full-image qualitative comparisons on O-HAZE.

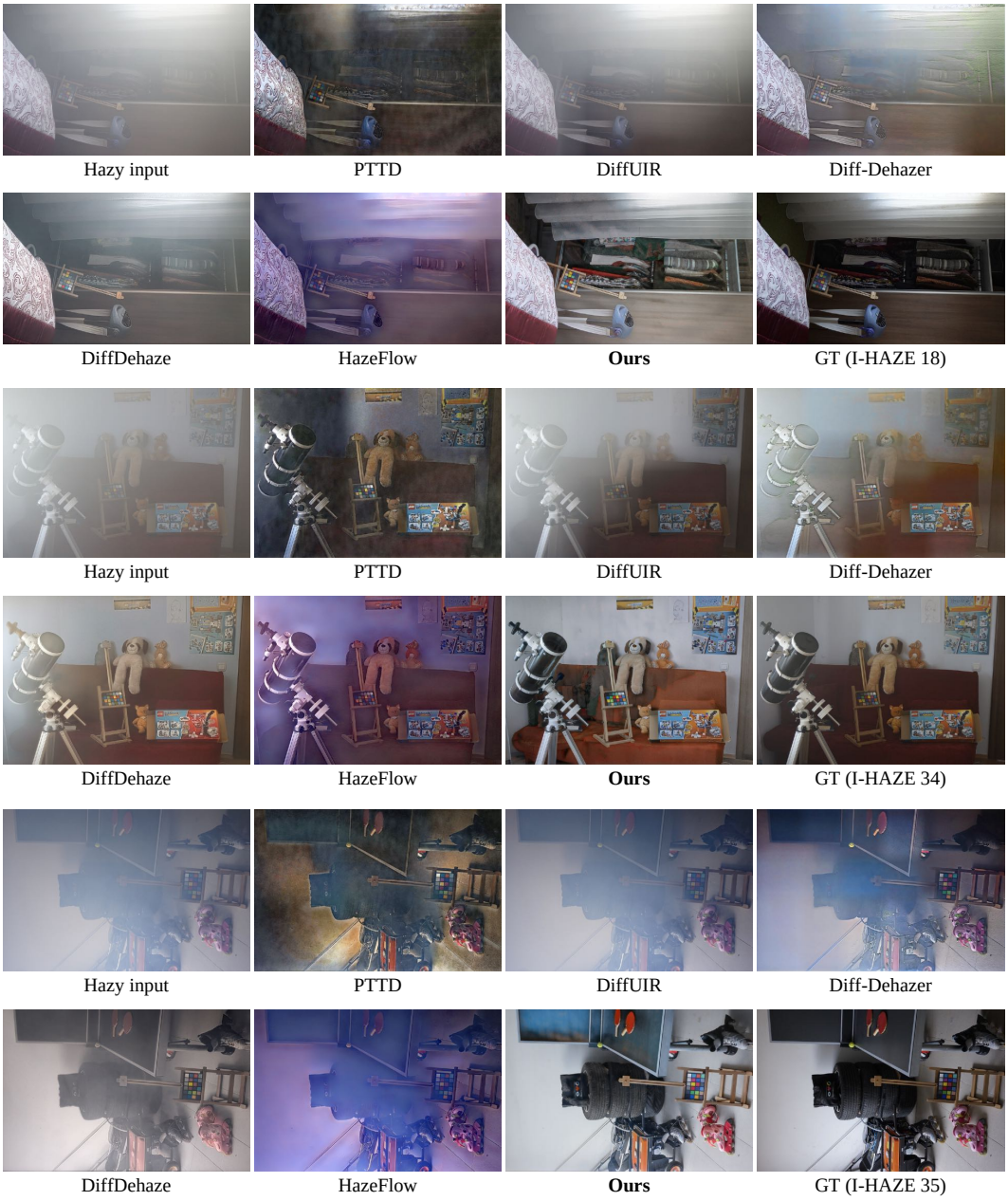


Figure 9: Full-image qualitative comparisons on I-HAZE.



Figure 10: Full-image qualitative comparisons on Dense-Haze.

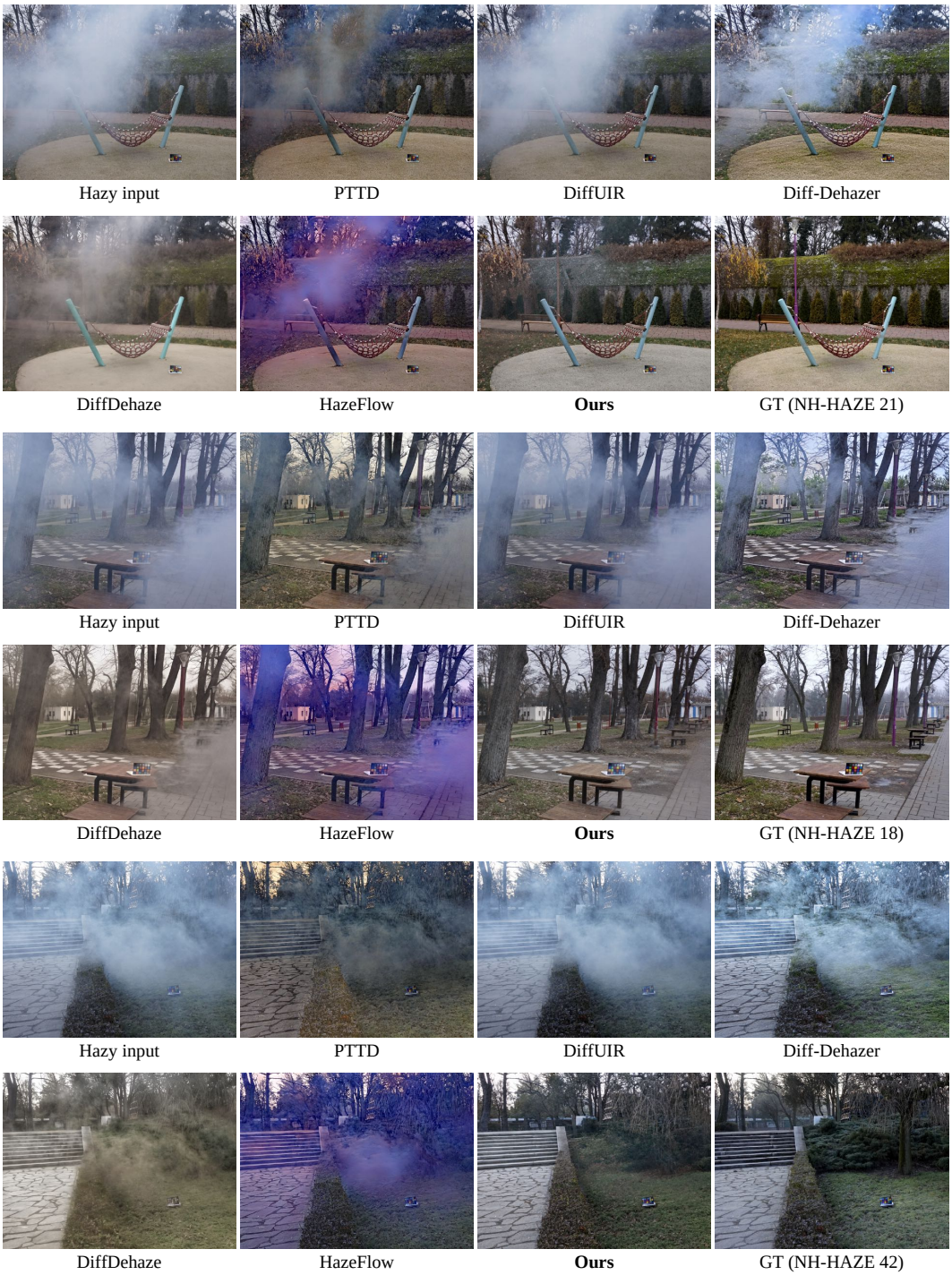


Figure 11: Full-image qualitative comparisons on NH-HAZE.

D.2 SDA Feature Visualizations

Figure 12 extends the visualization in Sec. 3.3 with eight NH-HAZE examples and two O-HAZE examples. Each pair shows the hazy RGB input on the left and the corresponding SDA feature visualization on the right, allowing the spatial structure of the learned representation to be inspected alongside the observed haze. The examples cover different scene structures and haze distributions, providing a broader view of the representation than the two examples in the main paper. Feature responses depend on both image content and degradation, so the maps are interpreted as learned restoration cues rather than calibrated measurements of physical haze density.

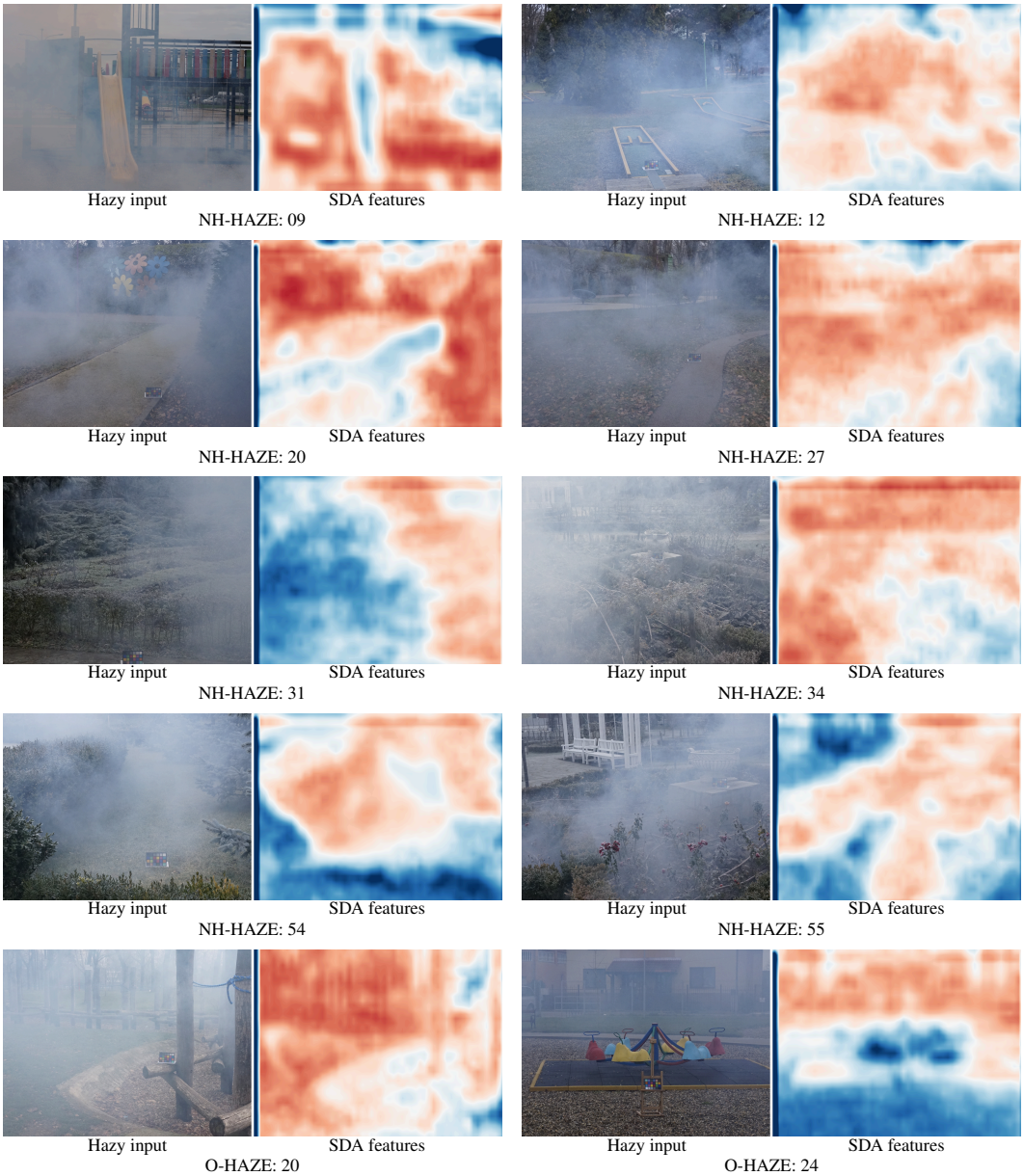


Figure 12: SDA feature visualizations on NH-HAZE and O-HAZE. Each pair shows the hazy input (left) and the corresponding learned feature map (right).