REUSING LEARNED DEGRADATION MODELS VIA COUNTER-DEGRADATION GUIDANCE FOR DIFFUSION DEHAZING

Anonymous authors
Paper under double-blind review

ABSTRACT

Recent diffusion-based dehazing methods increasingly rely on learned haze generators to produce realistic training data for real-world restoration. However, once the restoration model has been trained, the degradation model is typically discarded, despite retaining explicit knowledge of how haze affects the diffusion prediction. We revisit this unused knowledge and propose Haze Counter-Degradation Guidance (HazeCDG), a training-free framework that reuses a pretrained haze generator to improve an already trained diffusion restorer at inference time. HazeCDG compares the pretrained haze generator with its frozen backbone at the same diffusion state, so that their prediction difference reveals how the learned haze process locally changes the diffusion prediction. It then constructs a counter-degradation reference by reversing this learned degradation change and guides the restoration prediction toward it while preserving the relative prediction offset between the restoration and degradation models. Across real-world haze benchmarks, HazeCDG substantially reduces residual haze while preserving the perceptual strengths of the original restorer, bringing DiffDehaze into the performance range of recent dedicated dehazing methods. Importantly, these gains are obtained while keeping both the degradation and restoration models frozen; HazeCDG requires no retraining, test-time parameter adaptation, or manually tuned guidance scale. Our results demonstrate that degradation models learned for data synthesis can remain valuable after training as complementary sources of inference-time knowledge for diffusion restoration. Our code is available at https://anonymous.4open.science/r/HAZECDG-E2C1/README.md.

1 INTRODUCTION

Diffusion models have emerged as powerful generative priors for image restoration, demonstrating strong performance across denoising, deblurring, deraining, dehazing, super-resolution, and general inverse problems (Kawar et al., 2022; Chung et al., 2023b; Wang et al., 2023b; Luo et al., 2023; Yue et al., 2023; Lan et al., 2025a). Recent studies have further extended diffusion priors toward blind and all-in-one restoration using conditional control, pretrained generative priors, multimodal prompting, and degradation-aware representations (Liu et al., 2024; Lin et al., 2024; Jiang et al., 2024; Ai et al., 2024; Luo et al., 2025; Cheng et al., 2026). However, real-world restoration remains sensitive to discrepancies between training and inference degradation distributions. This issue is particularly restrictive for real-world dehazing, where paired hazy-clean data are scarce and synthetic haze only approximates the heterogeneous degradations observed in real scenes, resulting in a substantial synthetic-to-real gap (Chen et al., 2021; Wu et al., 2023; Ma et al., 2025; Wang et al., 2025; Chen et al., 2025; Shin et al., 2025). Advancing real-world dehazing therefore requires not only powerful restoration models, but also an effective understanding of haze degradation itself.

To mitigate this gap, real-world dehazing methods have explored strategies that reduce dependence on idealized synthetic supervision. PSD (Chen et al., 2021) adapts a synthetically pretrained model to unlabeled real hazy images, while CoA (Ma et al., 2025) compresses the model in the synthetic domain and adapts the resulting parameter space to real scenes. RIDCP (Wu et al., 2023) introduces a more realistic degradation pipeline together with high-quality codebook priors. Recent studies have placed greater emphasis on degradation synthesis: HazeFlow (Shin et al., 2025) adopts non-

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.

homogeneous haze synthesis, and GenHaze (Chen et al., 2025) adapts a pretrained latent diffusion model into a controllable haze generator. Of particular relevance, LHD (Wang et al., 2025) explicitly couples learned haze generation and restoration: its diffusion-based HazeGen (G_D) generates realistic hazy images for training the restoration model DiffDehaze (G_R).

Despite the growing role of learned haze generation, the resulting degradation model is typically discarded at inference. This separation is explicit in LHD, where G_D generates training data for G_R, while test-time restoration relies on G_R alone. We therefore ask whether the degradation knowledge retained in G_D can still benefit restoration. To this end, we propose Haze Counter-Degradation Guidance (HazeCDG), which reuses the pretrained degradation model during reverse diffusion while keeping both networks fixed. HazeCDG probes how the learned haze-generation branch changes the diffusion prediction relative to its frozen backbone, revealing a local haze differential at the current restoration state. From this differential, it constructs a counter-degradation reference and derives a correction to the restoration prediction by preserving the relative prediction offset between the restoration and degradation models. The resulting correction is then introduced directly into the diffusion prediction without retraining, test-time parameter adaptation, or an additional guidance-strength hyperparameter.

We instantiate HazeCDG on the pretrained LHD framework (Wang et al., 2025) and evaluate it on RTTS, URHI, and Fattal. Across these real-world settings, HazeCDG consistently addresses a characteristic weakness of the pretrained G_R: residual haze. The proposed guidance strengthens haze suppression while preserving the perceptual quality provided by the restoration model. This complementary effect makes the unchanged LHD restorer competitive with recent real-world dehazing methods across haze-sensitive and perceptual-quality criteria, without modifying the restoration network or updating model parameters. These results support our central hypothesis that a degradation model learned for data synthesis can retain restoration-relevant information complementary to the restoration model after training. HazeCDG therefore views learned degradation models not as disposable training components, but as reusable sources of inference-time guidance.

Our contributions are summarized as follows:

- We introduce an inference-time perspective on learned degradation models for real-world dehazing, extending their role beyond training-data synthesis to restoration guidance. We show that the degradation knowledge retained by a pretrained haze generator can complement an already trained diffusion restorer and improve haze removal while both networks remain frozen.
- We propose Haze Counter-Degradation Guidance (HazeCDG), a training-free framework that uses a counter-degradation reference to correct the restoration prediction while preserving the restoration-degradation prediction offset. This reference is constructed through an equal-magnitude reversal of the local prediction change induced by the learned haze-generation branch relative to its frozen backbone. Both networks remain frozen, and no retraining, test-time parameter adaptation, or manually tuned guidance scale is required.
- We validate HazeCDG on RTTS, URHI, and Fattal and demonstrate that HazeCDG substantially strengthens the frozen DiffDehaze restorer. Our method improves haze removal over the original DiffDehaze baseline while maintaining competitive performance across multiple perceptual quality metrics.

2 Related Work

2.1 Real-World Dehazing and Learned Degradation Modeling

Single-image dehazing has progressed from hand-crafted physical priors to increasingly powerful CNN, attention, and Transformer-based restoration models (He et al., 2009; Li et al., 2017; Liu et al., 2019; Dong et al., 2020; Qin et al., 2020; Wu et al., 2021; Song et al., 2023; Qiu et al., 2023). Despite these advances, the scarcity of paired real hazy–clean data and the resulting synthetic-to-real gap remain major obstacles to real-world generalization (Shao et al., 2020; Chen et al., 2021; Wu et al., 2023). Existing approaches address this gap through domain adaptation, real-image priors, unpaired learning, and parameter-efficient real-domain adaptation (Shao et al., 2020; Chen et al., 2021; Liu et al., 2022; Wu et al., 2023; Lan et al., 2025a; Fu et al., 2025; Ma et al., 2025; Zhang et al., 2026b).

A complementary direction instead improves the degradation distribution itself, using diffusion-based multi-degradation synthesis (Rajagopalan et al., 2025), controllable haze generation (Chen et al., 2025), non-homogeneous haze synthesis (Shin et al., 2025), or a dedicated haze generator paired with a diffusion restorer (Wang et al., 2025). While these approaches primarily improve the restoration model or the degradation distribution used for training, HazeCDG instead asks whether knowledge retained in a pretrained degradation model can remain useful after training, directly reusing G_D to guide the frozen G_R at inference.

2.2 Inference-Time Adaptation and Diffusion Guidance

Another line of work improves restoration by exploiting degradation information during inference. Test-time dehazing adapts restoration parameters to individual test observations using self-supervision (Liu et al., 2022), while TAO (Gou et al., 2024) introduces a test-time degradation adapter for previously unseen corruptions. Among adaptation-based approaches, DCTTA (Tang et al., 2026b) employs a diffusion-based re-degradation generator to construct pseudo degraded–clean pairs and selectively updates degradation-sensitive restoration parameters. HazeCDG differs from these methods in that both G_R and G_D remain fixed; degradation knowledge affects only the reverse diffusion trajectory rather than model parameters.

A related family of diffusion methods incorporates degradation or observation information directly into sampling (Kawar et al., 2022; Chung et al., 2023b; Wang et al., 2023b; Zhu et al., 2023). BlindDPS (Chung et al., 2023a) extends this paradigm to blind inverse problems by introducing a separate diffusion prior for the forward operator and jointly estimating the unknown operator and image. More recent approaches improve posterior sampling through frequency-aware measurement guidance (Thaker et al., 2025) or decoupled noise-annealing trajectories (Zhang et al., 2025). For real-world dehazing, FreeDehaze (Wu et al., 2026) exploits haze representations internalized by a pretrained diffusion model and constrains denoising using a pseudo-clean reference within a PCA-based haze subspace. In contrast, HazeCDG neither estimates an unknown degradation operator nor derives guidance solely from generic diffusion representations. It reuses the task-specific G_D learned for haze generation, isolates the local haze-induced prediction change relative to its frozen backbone, and constructs a counter-degradation reference while preserving the relative prediction offset established by G_R. Thus, HazeCDG connects learned degradation modeling with training-free diffusion guidance while leaving both pretrained networks unchanged.

3 Proposed Method

3.1 Haze Differential Sensing

We instantiate HazeCDG on the pretrained LHD framework (Wang et al., 2025), using its degradation model G_D and restoration model G_R. Both models are built on the same frozen latent-diffusion backbone and operate in a common noise-prediction space, which allows G_D to be queried directly at intermediate states of the restoration trajectory. Our key observation is that the learned haze-generation branch of G_D provides a direct way to measure how the degradation model locally alters the underlying diffusion prediction.

In G_D, haze generation is modeled by a trainable IRControlNet branch attached to the frozen latent-diffusion backbone (Wang et al., 2025; Lin et al., 2024; Rombach et al., 2022). LHD trains this branch using both synthetic hazy-clean pairs and real hazy images; for the latter, the clean-image condition is removed ($c_{\text{img}} = 0$) while the haze text condition is retained (Wang et al., 2025). This image-unconditional training allows the learned degradation response to be queried at test time without access to the unknown clean image.

At reverse step t, let $\mathbf{x}_t$ denote the current noisy latent along the restoration trajectory. To isolate the effect learned by the degradation branch, we evaluate G_D twice at the same $\mathbf{x}_t$, timestep, and haze text condition, changing only whether the learned IRControlNet contribution is included. Without the IRControlNet residuals, the frozen latent-diffusion backbone produces a reference prediction, denoted by $\epsilon_{\text{base}}(\mathbf{x}_t, t)$. Activating the trained degradation branch yields $\epsilon_D(\mathbf{x}_t, t)$. Because the two evaluations differ only by the contribution of the learned haze-generation branch, we define

$$\mathbf{g}_t = \epsilon_D(\mathbf{x}_t, t) - \epsilon_{\text{base}}(\mathbf{x}_t, t). \tag{1}$$

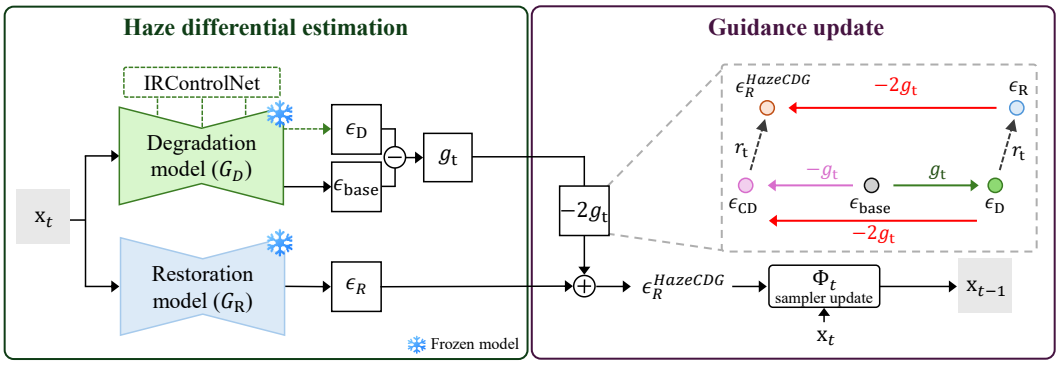


Figure 1: In its dense formulation, HazeCDG reuses a pretrained haze generator to isolate the haze differential g_t relative to its frozen backbone. Reversing this differential defines a counter-degradation reference; preserving the restoration–degradation prediction offset then yields $\epsilon_R^{\text{HazeCDG}} = \epsilon_R - 2g_t$. The guided prediction is passed to the unchanged sampler, with all model parameters kept frozen.

We call $\mathbf{g}_t$ the *haze differential*. Rather than treating it as a physical haze vector, we interpret $\mathbf{g}_t$ as the local change in diffusion prediction induced by the branch learned for haze generation. Under the standard diffusion score parameterization, this differential directly corresponds to the degradation-induced change in the local score field; a formal derivation is provided in Appendix A.

3.2 Counter-Degradation Guidance

The haze differential $\mathbf{g}_t$ describes how the learned degradation branch locally changes the diffusion prediction at the current restoration state. This provides not only a degradation direction, but also its observed prediction-space displacement. We therefore obtain a counter-degradation reference by reversing this displacement while preserving its magnitude:

$$\epsilon_{\text{CD}}(\mathbf{x}_t, t) = \epsilon_{\text{base}}(\mathbf{x}_t, t) - \mathbf{g}_t. \tag{2}$$

Here, ϵ_{CD} serves as a parameter-free reference for guiding restoration, obtained by reversing the local prediction change induced by the learned haze-generation branch. It is not assumed to equal the unknown clean prediction.

Directly replacing ϵ_R with ϵ_{CD} is undesirable because this reference is constructed from the degradation model and does not account for the prediction already formed by G_R. To incorporate the reference without discarding that prediction, we express ϵ_R relative to ϵ_D as $\epsilon_R = \epsilon_D + \mathbf{r}_t$, where $\mathbf{r}_t$ denotes the relative prediction offset between the restoration and degradation models evaluated at the same $\mathbf{x}_t$ and timestep. The haze differential characterizes the change induced by the degradation branch relative to its frozen backbone, rather than the difference between the restoration and degradation predictions. Therefore, it provides the basis for changing the degradation reference, but does not prescribe a corresponding change to $\mathbf{r}_t$. We accordingly adopt a correction that replaces ϵ_D with ϵ_{CD} while retaining $\mathbf{r}_t$. In this way, the reference displacement is transferred to the restoration prediction instead of replacing it with the counter-degradation reference:

$$\mathbf{r}_t = \epsilon_R(\mathbf{x}_t, t) - \epsilon_D(\mathbf{x}_t, t) = \epsilon_R^{\text{HazeCDG}}(\mathbf{x}_t, t) - \epsilon_{\text{CD}}(\mathbf{x}_t, t). \tag{3}$$

Using $\epsilon_D = \epsilon_{\text{base}} + \mathbf{g}_t$ together with Eq. 2, the reference displacement $\epsilon_{\text{CD}} - \epsilon_D$ is $-2\mathbf{g}_t$. Applying this change while preserving Eq. 3 gives

$$\epsilon_R^{\text{HazeCDG}}(\mathbf{x}_t, t) = \epsilon_R(\mathbf{x}_t, t) - 2\mathbf{g}_t. \tag{4}$$

The factor of two is therefore not introduced as a tunable guidance strength. It follows from replacing the degradation reference $\epsilon_{\text{base}} + \mathbf{g}_t$ by its equal-magnitude counter-degradation counterpart $\epsilon_{\text{base}} - \mathbf{g}_t$, while keeping the relative prediction offset of the restoration model unchanged.

Algorithm 1 Haze Counter-Degradation Guidance (HazeCDG)

Require: Frozen restoration model G_R; frozen degradation model G_D; initial noisy latent $\mathbf{x}_T$; reverse timesteps $\mathcal{T}$; sensing anchors $\mathcal{A} \subseteq \mathcal{T}$, $|\mathcal{A}| = K$; original sampler update operator Φ_t; restoration condition c_R; haze condition c_D

Ensure: Restored latent $\mathbf{x}_0$

Initialize $a_{\text{last}} \leftarrow$ and $\mathbf{g} \leftarrow \mathbf{0}$
for each reverse step $t \in \mathcal{T}$ in sampling order **do**
  $\epsilon_R \leftarrow G_R(\mathbf{x}_t, t, c_R)$
  if $t \in \mathcal{A}$ **then**
    $\epsilon_{\text{base}} \leftarrow G_D(\mathbf{x}_t, t, c_D; \text{branch off})$
    $\epsilon_D \leftarrow G_D(\mathbf{x}_t, t, c_D; \text{branch on})$
    $\mathbf{g}_t \leftarrow \epsilon_D - \epsilon_{\text{base}}$ ▷ update haze differential
    $a_{\text{last}} \leftarrow t$
  end if
  $a(t) \leftarrow a_{\text{last}}$ ▷ most recent sensing anchor
  $\epsilon_R^{\text{HazeCDG}} \leftarrow \epsilon_R - 2\mathbf{g}_{a(t)}$ ▷ counter-degradation guidance
  $\mathbf{x}_{t-1} \leftarrow \Phi_t\big(\mathbf{x}_t, \epsilon_R^{\text{HazeCDG}}\big)$ ▷ original sampler update
end for
return $\mathbf{x}_0$

3.3 Sparse Degradation Sensing with Causal Guidance Reuse

Applying HazeCDG at every reverse-diffusion step requires repeatedly evaluating G_D and its frozen diffusion backbone. To reduce this cost, we select K fixed sensing anchors along the reverse trajectory. For the 50-step sampler, we use a deterministic nested hierarchy of sensing schedules, $A_5 \subset A_{10} \subset A_{25} \subset A_{50}$, so that increasing K only adds new sensing anchors while preserving those used at smaller sensing budgets. At each anchor a, we evaluate the current restoration state and recompute the haze differential $\mathbf{g}_a = \epsilon_D - \epsilon_{\text{base}}$ as described in Section 3.1. Between consecutive anchors, G_D is not evaluated again; instead, the most recently sensed haze differential is reused.

Specifically, let $a(t)$ denote the most recent sensing anchor at reverse step t. Sparse HazeCDG uses

$$\epsilon_R^{\text{HazeCDG}}(\mathbf{x}_t, t) = \epsilon_R(\mathbf{x}_t, t) - 2\mathbf{g}_{a(t)}. \tag{5}$$

Thus, the degradation information is refreshed only at the K sensing anchors and causally reused between them in noise-prediction space. When $K = T$, every reverse step is an anchor, so $a(t) = t$ and Eq. 5 reduces exactly to dense HazeCDG. Algorithm 1 summarizes the complete inference procedure.

4 Experiments

4.1 Experimental Setup

Datasets. Following the evaluation protocols of LHD (Wang et al., 2025) and recent real-world dehazing studies such as BiLaLoRA (Zhang et al., 2026b), we evaluate HazeCDG on three widely used real-world haze benchmarks: RTTS, URHI, and Fattal. RTTS (Li et al., 2019) contains 4,322 real hazy images and serves as our primary evaluation set. For URHI, we use a fixed subset of 150 images, consistent with the evaluation scale adopted in prior work. We further evaluate on the 31-image Fattal real-haze dataset (Fattal, 2014). These benchmarks collectively cover diverse real-world haze conditions and allow us to evaluate HazeCDG without relying on synthetic hazy–clean pairs.

Evaluation metrics. Because these real-world datasets do not provide paired haze-free references, we evaluate restoration performance using no-reference criteria. We report FADE (Choi et al., 2015) as a haze-specific measure, together with Q-Align (Wu et al., 2024), CLIPIQA (Wang et al., 2023a), MUSIQ (Ke et al., 2021), and BRISQUE (Mittal et al., 2012) as complementary perceptual-quality measures. All methods are evaluated using the same metric implementations and the same set of

Table 1: Effect of HazeCDG on frozen DiffDehaze and DOD on RTTS. Panel (a) varies the sensing budget K for DiffDehaze, while Panel (b) evaluates the one-step application to DOD. **Bold** denotes the best result within each backbone.

(a) DiffDehaze (CVPR 2025)						
Dataset	Setting	FADE↓	Q-Align↑	CLIPIQA↑	MUSIQ↑	BRISQUE↓
RTTS	Baseline	1.135	3.592	0.416	**64.609**	24.290
	+ HazeCDG ($K = 5$)	**0.597**	**3.691**	0.505	64.589	**20.758**
	+ HazeCDG ($K = 10$)	0.610	3.683	**0.510**	64.342	21.479
	+ HazeCDG ($K = 25$)	0.626	3.661	0.509	63.936	22.021
	+ HazeCDG ($K = 50$)	0.626	3.651	0.509	63.812	22.125
(b) DOD (AAAI 2026)						
Dataset	Setting	FADE↓	Q-Align↑	CLIPIQA↑	MUSIQ↑	BRISQUE↓
RTTS	Baseline	1.508	**3.109**	0.393	**53.544**	**35.776**
	+ HazeCDG (K = 1)	**1.144**	3.074	**0.417**	51.858	36.963

input images. Reporting multiple no-reference criteria is particularly important in real-world dehazing, where improvement in a haze-specific score does not necessarily imply improved perceptual quality or naturalness.

Implementation details. We implement HazeCDG using the official LHD codebase and pretrained checkpoints (Wang et al., 2025), keeping both models frozen. For DiffDehaze, we use the standard 50-step spaced sampler. HazeCDG is applied immediately after the DiffDehaze noise prediction: we replace ϵ_R with $\epsilon_R - 2\mathbf{g}_{a(t)}$, following the sensing and reuse procedure in Section 3.3, before applying the unchanged sampler update. For degradation sensing, we use the haze prompt "hazy, foggy, misty, obscure, smoggy." and evaluate $K \in \{5, 10, 25, 50\}$ sensing anchors. DiffDehaze runs with and without HazeCDG use identical input images, restoration checkpoints, preprocessing, sampler settings, and output-resolution handling. All baseline and HazeCDG runs otherwise use identical inputs, preprocessing, checkpoints, sampler settings, and output-resolution handling. All inference experiments are conducted on a single NVIDIA RTX PRO 6000 GPU.

4.2 Effect of HazeCDG on Frozen Diffusion Restorers

We first evaluate HazeCDG on the pretrained DiffDehaze restorer from LHD (Wang et al., 2025), while keeping all model parameters fixed. As shown in Table 1, HazeCDG substantially reduces FADE on RTTS across all sensing budgets, indicating that the degradation guidance consistently suppresses residual haze in the restored images. Importantly, this gain is accompanied by improvements in Q-Align, CLIPIQA, and BRISQUE. Across $K \in \{5, 10, 25, 50\}$, these improvements are broadly comparable. Increasing the sensing budget requires additional degradation-model evaluations and higher inference time (Table 4), without providing consistent additional improvements in restoration quality. We therefore adopt $K = 5$, the smallest tested sensing budget, as the default configuration in subsequent DiffDehaze experiments to retain these restoration benefits while achieving the shortest inference time among the evaluated HazeCDG variants. Similar improvements in FADE, Q-Align, and CLIPIQA are observed on URHI and Fattal dataset, with complete results reported in Appendix C.

We additionally conduct a feasibility study with DOD (Tang et al., 2026a) to examine whether HazeCDG can be applied to another pretrained diffusion restorer. DOD performs restoration in a single diffusion step and is trained independently of LHD without using HazeGen-generated images. With both DOD and G_D frozen, HazeCDG with a single degradation-sensing operation ($K = 1$) reduces FADE on RTTS (Table 1). These results suggest that the degradation cue learned by G_D can improve haze suppression in this one-step restorer. Interestingly, CLIPIQA also improves, while the remaining perceptual-quality metrics show modest decreases, suggesting that further calibration may help preserve perceptual quality when transferring HazeCDG to one-step restorers. Future work

Table 2: Comparison with recent real-world dehazing methods on RTTS. DCTTA applies test-time adaptation to the same DiffDehaze backbone, whereas HazeCDG keeps all model parameters fixed. Methods are grouped into non-diffusion and diffusion-based approaches. Best and second-best results are shown in **bold** and underline, respectively.

	Method	FADE↓	Q-Align↑	CLIPIQA↑	MUSIQ↑	BRISQUE↓
Non-Diff	HazeFlow (ICCV'25)	**0.582**	3.369	<u>0.437</u>	63.923	<u>23.176</u>
	CoA (CVPR'25)	0.859	3.076	0.349	53.425	30.244
	PDDA (TIP'26)	0.735	3.338	0.370	59.721	28.479
	BiLaLoRA (CVPR'26)	0.752	3.470	0.396	61.773	25.067
	DehazeSB (ICCV'25)	0.824	3.030	0.345	59.552	23.553
Diff	HNDiff (ECCV'26)	0.699	2.933	0.377	56.717	35.750
	DiffDehaze (CVPR'25)	1.135	3.592	0.416	<u>64.609</u>	24.290
	DiffDehaze + DCTTA (CVPR'26)	1.170	<u>3.614</u>	0.427	**65.016**	23.541
	DiffDehaze + HazeCDG (Ours)	<u>0.597</u>	**3.691**	**0.505**	64.589	**20.758**

will explore this direction and further broaden HazeCDG's applicability to other diffusion restorers. Details of the schedule-normalized one-step implementation are provided in Appendix B.

4.3 Comparison with Recent Real-World Dehazing Methods

We compare HazeCDG with recent real-world dehazing methods, including HazeFlow (Shin et al., 2025), CoA (Ma et al., 2025), DehazeSB (Lan et al., 2025b), HNDiff (He et al., 2026), PDDA (Zhang et al., 2026a), and BiLaLoRA (Zhang et al., 2026b). We also include DCTTA (Tang et al., 2026b), a test-time adaptation method applied to the same DiffDehaze backbone. Although not specialized for dehazing, DCTTA provides a relevant post-training comparison because it also improves a pretrained restoration model at inference time. Table 2 reports results on RTTS using the default sparse HazeCDG configuration from Section 4.2. Results on URHI and Fattal are provided in Appendix C.

The results show that HazeCDG substantially improves the competitive position of the pretrained DiffDehaze restorer. DiffDehaze already provides strong perceptual quality, but its high FADE indicates residual haze in real scenes. The unchanged DiffDehaze backbone becomes competitive with recent real-world dehazing methods across both haze-sensitive and perceptual-quality criteria by reusing the degradation knowledge retained in G_D. These results suggest that HazeCDG addresses the residual-haze tendency while preserving the perceptual strengths of G_R. Similar improvements are observed on URHI and Fattal in Appendix C.

The comparison with DCTTA further distinguishes HazeCDG from inference-time adaptation. DCTTA improves the pretrained restorer through test-time optimization of degradation-sensitive parameters, but provides limited correction of the residual-haze tendency inherited from DiffDehaze. HazeCDG instead keeps both G_R and G_D fixed and introduces the learned counter-degradation cue directly into the diffusion prediction. Its gains therefore come from reusing complementary degradation knowledge rather than adapting the restoration model to each test instance, without test-time parameter updates or an additional optimization stage.

4.4 Ablation Study

Degradation neutralization vs. counter-degradation. Our formulation distinguishes between *canceling* the learned degradation effect and explicitly *reversing* it. The original DiffDehaze prediction is ϵ_R. Subtracting the haze differential once, $\epsilon_R - \mathbf{g}_t$, moves the degradation reference $\epsilon_D = \epsilon_{\text{base}} + \mathbf{g}_t$ back to ϵ_{base}; we refer to this as *degradation neutralization*. HazeCDG instead replaces the degradation reference with its counter-degradation counterpart while preserving the relative prediction offset, yielding $\epsilon_R - 2\mathbf{g}_t$. This comparison tests whether degradation neutralization is sufficient or whether counter-degradation provides additional benefit.

Table 3 shows progressive improvements in FADE, Q-Align, and CLIPIQA from DiffDehaze to degradation neutralization and counter-degradation guidance across all three datasets. These re-

Table 3: Ablation of degradation neutralization and counter-degradation guidance. Best results for each dataset and metric are shown in bold.

Dataset	Method	Prediction	FADE↓	Q-Align↑	CLIPIQA↑	MUSIQ↑	BRISQUE↓
RTTS	DiffDehaze	ϵ_R	1.135	3.592	0.416	64.609	24.290
	Neutralization	$\epsilon_R - \mathbf{g}_t$	0.855	3.672	0.460	**65.124**	22.377
	HazeCDG	$\epsilon_R - 2\mathbf{g}_t$	**0.597**	**3.691**	**0.505**	64.589	**20.758**
URHI	DiffDehaze	ϵ_R	1.293	3.743	0.470	65.938	24.193
	Neutralization	$\epsilon_R - \mathbf{g}_t$	1.012	3.833	0.514	**66.258**	21.910
	HazeCDG	$\epsilon_R - 2\mathbf{g}_t$	**0.736**	**3.844**	**0.552**	65.806	**20.141**
Fattal	DiffDehaze	ϵ_R	0.636	4.006	0.551	68.419	**21.090**
	Neutralization	$\epsilon_R - \mathbf{g}_t$	0.492	4.166	0.599	**69.066**	21.922
	HazeCDG	$\epsilon_R - 2\mathbf{g}_t$	**0.372**	**4.211**	**0.632**	68.770	21.858

Table 4: Runtime and peak GPU memory on RTTS. Relative runtime is normalized by DiffDehaze.

Method	Param. update	Time (s/img)↓	Relative	Peak mem. (GiB)↓
DiffDehaze	No	3.378	1.00×	8.370
HazeCDG ($K = 5$, default)	No	3.954	1.17×	9.732
HazeCDG ($K = 10$)	No	4.542	1.34×	9.732
HazeCDG ($K = 25$)	No	6.093	1.80×	9.732
HazeCDG ($K = 50$)	No	8.556	2.53×	9.732
DCTTA	Yes	16.033	4.75×	14.418

sults support the usefulness of $\mathbf{g}_t$ as a degradation-aware signal and show that counter-degradation provides additional haze-removal benefits over neutralization. The less uniform trends in the remaining no-reference metrics suggest that HazeCDG acts as degradation-specific guidance rather than generic perceptual enhancement.

Computational Efficiency of Sparse HazeCDG. Sparse HazeCDG reuses previously sensed degradation information instead of evaluating G_D at every reverse step. Thus, the sensing budget K controls the additional computation without changing the restoration network or the underlying 50-step sampler. Table 4 shows that the default $K = 5$ setting remains close to the DiffDehaze inference cost, while increasing K mainly increases runtime through additional degradation-model evaluations. Peak memory remains unchanged across HazeCDG configurations because the same frozen models are used regardless of sensing frequency. Compared with DCTTA (Tang et al., 2026b), which performs test-time parameter updates, HazeCDG requires only forward evaluations of frozen models and therefore incurs substantially lower runtime and memory overhead.

4.5 Qualitative Comparison

Figure 2 compares HazeCDG with HazeFlow (Shin et al., 2025) BiLaLoRA (Zhang et al., 2026b), PDDA (Zhang et al., 2026a), HNDiff (He et al., 2026), DCTTA (Tang et al., 2026b), and the original DiffDehaze (Wang et al., 2025) on a representative real-world hazy scene. Despite improving visibility, the compared methods still suffer from residual haze in distant regions, which obscures fine scene details. For example, DiffDehaze preserves the overall scene structure and appearance, but noticeable residual haze remains in the magnified distant region.

Compared with these methods, HazeCDG removes residual haze more effectively while retaining the overall scene structure and appearance. The resulting restoration reveals clearer details in distant regions that remain partially obscured in the other outputs. These observations are consistent with the quantitative results in Section 4.3 and support the effectiveness of reusing learned degradation knowledge to improve haze-removal performance in pretrained diffusion restorers. Additional qualitative comparisons on RTTS, URHI, and Fattal dataset are provided in Appendix D.

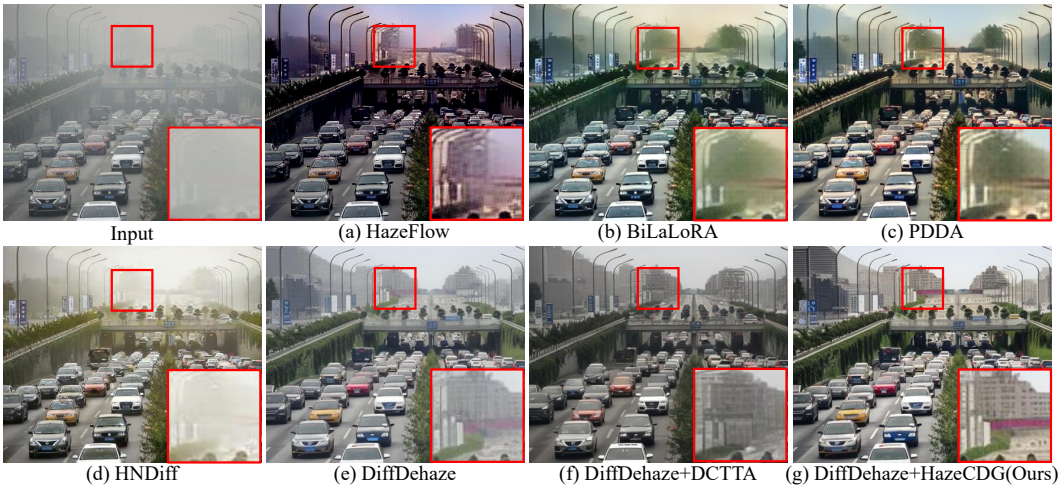


Figure 2: Visual comparisons on the RTTS dataset (Li et al., 2019). (a)-(g) present the outputs of HazeFlow, BiLaLoRA, PDDA, HNDiff, DiffDehaze, DiffDehaze with DCTTA, and DiffDehaze with HazeCDG, respectively. Red boxes indicate the regions magnified in the insets.

5 Limitations and Future Work

Our current study focuses on real-world dehazing. We choose dehazing as a particularly challenging testbed because the gap between synthetic and real haze is substantial, making learned degradation models especially important for constructing realistic training data and, consequently, a natural setting for studying whether their knowledge remains useful after training. While the successful transfer from the original LHD pipeline to an independently trained one-step diffusion restorer provides initial evidence that HazeCDG is not tied to a specific restoration architecture or sampling process, our experiments do not yet establish its effectiveness beyond haze degradation. An important direction for future work is therefore to extend the counter-degradation principle to other restoration problems, such as denoising, deblurring, deraining, and low-light enhancement. We are also interested in applying HazeCDG-style degradation reuse to recent all-in-one restoration models, where multiple degradation types are handled within a single model and the relevant degradation cue must be identified dynamically at inference time. Finally, since HazeCDG complements rather than replaces the restorer, we will investigate whether stronger restoration backbones can yield further gains under the same guidance principle.

6 Conclusion

We introduced Haze Counter-Degradation Guidance (HazeCDG), a training-free framework that reuses degradation knowledge retained by a pretrained haze generator to improve diffusion dehazing at inference time. HazeCDG identifies the local prediction change induced by the learned degradation model, constructs its counter-degradation counterpart, and guides the restoration prediction while preserving the relative prediction offset between the restoration and degradation models, without retraining or test-time parameter adaptation. Across real-world haze benchmarks, this simple reuse of degradation knowledge substantially alleviates residual haze while preserving the perceptual strengths of the pretrained restorer, bringing the frozen DiffDehaze model into the performance range of recent dedicated dehazing methods. Its effectiveness on the independently trained one-step DOD model provides further evidence that the learned degradation cue can remain useful beyond the restoration pipeline for which it was originally developed. These results show that degradation models learned for data synthesis need not be discarded after training, but can serve as complementary sources of inference-time guidance for pretrained diffusion restoration.

AI USE STATEMENT

We used generative AI tools to assist with writing and polishing the manuscript, including checking grammar, refining wording and sentence structure, and improving contextual consistency and coherence. We also used these tools to help check for missing punctuation and issues in mathematical expressions. All AI-assisted text was reviewed and revised by the authors, who take full responsibility for the final content of the manuscript.

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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 25432–25444, 2024.

Sixiang Chen, Tian Ye, Yunlong Lin, Yeying Jin, Yijun Yang, Haoyu Chen, Jianyu Lai, Song Fei, Zhaohu Xing, Fugee Tsung, and Lei Zhu. GenHaze: Pioneering controllable one-step realistic haze generation for real-world dehazing. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 9194–9205, 2025.

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*, pp. 7180–7189, 2021. doi: 10.1109/CVPR46437.2021.00710.

Zihan Cheng, Liangtai Zhou, Dian Chen, Ni Tang, Xiaotong Luo, Yuan Xie, and Yanyun Qu. Unildiff: Unlocking the power of diffusion priors for all-in-one image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 37465–37475, 2026.

Lark Kwon Choi, Jaehee You, and Alan Conrad Bovik. Referenceless prediction of perceptual fog density and perceptual image defogging. *IEEE Transactions on Image Processing*, 24(11): 3888–3901, 2015. doi: 10.1109/TIP.2015.2456502.

Hyungjin Chung, Jeongsol Kim, Sehui Kim, and Jong Chul Ye. Parallel diffusion models of operator and image for blind inverse problems. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6059–6069, 2023a.

Hyungjin Chung, Jeongsol Kim, Michael Thompson McCann, Marc Louis Klasky, and Jong Chul Ye. Diffusion posterior sampling for general noisy inverse problems. In *International Conference on Learning Representations*, 2023b.

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*, pp. 2157–2167, 2020.

Raanan Fattal. Dehazing using color-lines. *ACM Transactions on Graphics*, 34:13:1–13:14, 2014. doi: 10.1145/2651362.

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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 12700–12709, 2025.

Yuanbiao Gou, Haiyu Zhao, Boyun Li, Xinyan Xiao, and Xi Peng. Test-time degradation adaptation for open-set image restoration. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 16167–16177, 2024.

Jin-Ting He, Fu-Jen Tsai, Yan-Tsung Peng, Min-Hung Chen, Chia-Wen Lin, and Yen-Yu Lin. HNDiff: Haze-noise diffusion for image dehazing. In *European Conference on Computer Vision*, 2026.

Kaiming He, Jian Sun, and Xiaoou Tang. Single image haze removal using dark channel prior. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1956–1963, 2009. doi: 10.1109/CVPR.2009.5206515.

Yitong Jiang, Zhaoyang Zhang, Tianfan Xue, and Jinwei Gu. Autodir: Automatic all-in-one image restoration with latent diffusion. In *Proceedings of the European Conference on Computer Vision*, 2024. doi: 10.1007/978-3-031-73661-2_19.

Bahjat Kawar, Michael Elad, Stefano Ermon, and Jiaming Song. Denoising diffusion restoration models. In *Advances in Neural Information Processing Systems*, volume 35, 2022.

Junjie Ke, Qifei Wang, Yilin Wang, Peyman Milanfar, and Feng Yang. MUSIQ: Multi-scale image quality transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 5148–5157, 2021. doi: 10.1109/ICCV48922.2021.00510.

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. *Proceedings of the AAAI Conference on Artificial Intelligence*, 39(4):4455–4463, 2025a. doi: 10.1609/aaai.v39i4.32469.

Yunwei Lan, Zhigao Cui, Xin Luo, Chang Liu, Nian Wang, Menglin Zhang, Yanzhao Su, and Dong Liu. When schrodinger bridge meets real-world image dehazing with unpaired training. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 8756–8765, 2025b.

Boyi Li, Xiulian Peng, Zhangyang Wang, Jizheng Xu, and Dan Feng. AOD-Net: All-in-one dehazing network. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 4770–4778, 2017.

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, 2019.

Xinqi Lin, Jingwen He, Ziyan Chen, Zhaoyang Lyu, Bo Dai, Fanghua Yu, Yu Qiao, Wanli Ouyang, and Chao Dong. Diffbir: Toward blind image restoration with generative diffusion prior. In *Proceedings of the European Conference on Computer Vision*, 2024. doi: 10.1007/978-3-031-73202-7_25.

Huan Liu, Zijun Wu, Liangyan Li, Sadaf Salehkalaibar, Jun Chen, and Keyan Wang. Towards multi-domain single image dehazing via test-time training. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5831–5840, 2022.

Xiaohong Liu, Yongrui Ma, Zhihao Shi, and Jun Chen. Griddehazenet: Attention-based multi-scale network for image dehazing. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 7314–7323, 2019.

Yuhao Liu, Zhanghan Ke, Fang Liu, Nanxuan Zhao, and Rynson W. H. Lau. Diff-plugin: Revitalizing details for diffusion-based low-level tasks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4197–4208, 2024.

Wenyang Luo, Haina Qin, Zewen Chen, Libin Wang, Dandan Zheng, Yuming Li, Yufan Liu, Bing Li, and Weiming Hu. Visual-instructed degradation diffusion for all-in-one image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 12764–12777, 2025.

Ziwei Luo, Fredrik K. Gustafsson, Zheng Zhao, Jens Sjölund, and Thomas B. Schön. Image restoration with mean-reverting stochastic differential equations. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 23045–23066, 2023.

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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11197–11206, 2025.

Anish Mittal, Anush Krishna Moorthy, and Alan Conrad Bovik. No-reference image quality assessment in the spatial domain. *IEEE Transactions on Image Processing*, 21(12):4695–4708, 2012. doi: 10.1109/TIP.2012.2214050.

Xu Qin, Zhilin Wang, Yuanchao Bai, Xiaodong Xie, and Huizhu Jia. FFA-Net: Feature fusion attention network for single image dehazing. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pp. 11908–11915, 2020.

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*, pp. 12802–12813, 2023.

Sudarshan Rajagopalan, Nithin Gopalakrishnan Nair, Jay N. Paranjape, and Vishal M. Patel. GenDeg: Diffusion-based degradation synthesis for generalizable all-in-one image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 28144–28154, 2025. doi: 10.1109/CVPR52734.2025.02621.

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*, pp. 10684–10695, 2022.

Yuanjie Shao, Lerenhan Li, Wenqi Ren, Changxin Gao, and Nong Sang. Domain adaptation for image dehazing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2808–2817, 2020.

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.

Yuda Song, Zhuqing He, Hui Qian, and Xin Du. Vision transformers for single image dehazing. *IEEE Transactions on Image Processing*, 32:1927–1941, 2023. doi: 10.1109/TIP.2023.3256763.

Ni Tang, Xiaotong Luo, Zihan Cheng, Liangtai Zhou, Dongxiao Zhang, and Yanyun Qu. Diffusion once and done: Degradation-aware LoRA for all-in-one image restoration. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 9448–9456, 2026a. doi: 10.1609/aaai.v40i11.37905.

Ni Tang, Shenghao Nie, Xiaotong Luo, Yuan Xie, and Yanyun Qu. Degradation-consistent test-time adaptation for all-in-one image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 15476–15485, 2026b.

Darshan Thaker, Abhishek Goyal, and Rene Vidal. Frequency-guided posterior sampling for diffusion-based image restoration. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 12873–12882, 2025.

Jianyi Wang, Kelvin C. K. Chan, and Chen Change Loy. Exploring CLIP for assessing the look and feel of images. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pp. 2555–2563, 2023a. doi: 10.1609/aaai.v37i2.25353.

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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 23091–23100, 2025. doi: 10.1109/CVPR52734.2025.02150.

Yinhuai Wang, Jiwen Yu, and Jian Zhang. Zero-shot image restoration using denoising diffusion null-space model. In *International Conference on Learning Representations*, 2023b.

Haiyan Wu, Yanyun Qu, Shaohui Lin, Jian Zhou, Ruizhi Qiao, Zhizhong Zhang, Yuan Xie, and Lizhuang Ma. Contrastive learning for compact single image dehazing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10551–10560, 2021.

Haoning Wu, Zicheng Zhang, Weixia Zhang, Chaofeng Chen, Liang Liao, Chunyi Li, Yixuan Gao, Annan Wang, Erli Zhang, Wenxiu Sun, Qiong Yan, Xiongkuo Min, Guangtao Zhai, and Weisi Lin. Q-align: Teaching LMMs for visual scoring via discrete text-defined levels. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 54015–54029. PMLR, 2024.

Jiawei Wu, Yikun Ma, Wenqi Ren, Zhi Jin, and Xiaochun Cao. FreeDehaze: Towards training-free real-world image dehazing via diffusion degradation prior. *IEEE Transactions on Image Processing*, 35:7535–7549, 2026. doi: 10.1109/TIP.2026.3710485.

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*, pp. 22282–22291, 2023.

Zongsheng Yue, Jianyi Wang, and Chen Change Loy. Resshift: Efficient diffusion model for image super-resolution by residual shifting. In *Advances in Neural Information Processing Systems*, volume 36, 2023.

Bingliang Zhang, Wenda Chu, Julius Berner, Chenlin Meng, Anima Anandkumar, and Yang Song. Improving diffusion inverse problem solving with decoupled noise annealing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 20895–20905, 2025.

Yan Zhang, Yuxin Feng, Xin Li, Fan Zhou, and Zhuo Su. PDDA: Prompt-driven domain adaptation for real-world image dehazing. *IEEE Transactions on Image Processing*, 35:4280–4294, 2026a. doi: 10.1109/TIP.2026.3684402.

Yan Zhang, Long Ma, Yuxin Feng, Zhe Huang, Fan Zhou, and Zhuo Su. Bilevel layer-positioning LoRA for real image dehazing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 29969–29978, 2026b.

Yuanzhi Zhu, Kai Zhang, Jingyun Liang, Jiezhang Cao, Bihan Wen, Radu Timofte, and Luc Van Gool. Denoising diffusion models for plug-and-play image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp. 1219–1229, 2023.

APPENDIX

This appendix provides additional theoretical and experimental details for HazeCDG. We first present the score-space interpretation of the haze differential, followed by the evaluation protocol and implementation details for sparse HazeCDG and its one-step application to DOD. We then report additional quantitative results on URHI and Fattal, together with further qualitative comparisons on RTTS, URHI, and Fattal.

A SCORE-SPACE INTERPRETATION OF THE HAZE DIFFERENTIAL

Section 3.1 defines the haze differential $\mathbf{g}_t$ in Eq. 1 as the change in noise prediction produced by activating the learned degradation branch while keeping the noisy latent, timestep, and conditioning fixed. Here, we show that this differential directly corresponds to the local change in the model-estimated diffusion score induced by the degradation branch.

For an ϵ-parameterized diffusion model, the model-estimated score at reverse step t is related to the predicted noise by

$$\widehat{\mathbf{s}}(\mathbf{x}_t, t) = -\frac{\epsilon(\mathbf{x}_t, t)}{\sigma_t}, \tag{6}$$

where σ_t denotes the diffusion noise scale at timestep t. Because $\epsilon_D(\mathbf{x}_t, t)$ and $\epsilon_{\text{base}}(\mathbf{x}_t, t)$ in Eq. 1 are evaluated at the same $\mathbf{x}_t$ and t, the same σ_t applies to both predictions. Their difference in score space is therefore

$$\begin{aligned} \Delta\widehat{\mathbf{s}}_t^D &= \widehat{\mathbf{s}}_D(\mathbf{x}_t, t) - \widehat{\mathbf{s}}_{\text{base}}(\mathbf{x}_t, t) \\ &= -\frac{\epsilon_D(\mathbf{x}_t, t) - \epsilon_{\text{base}}(\mathbf{x}_t, t)}{\sigma_t} = -\frac{\mathbf{g}_t}{\sigma_t}. \end{aligned} \tag{7}$$

Equivalently, $\mathbf{g}_t = -\sigma_t \Delta\widehat{\mathbf{s}}_t^D$. Since the two predictions differ only by the contribution of the learned haze-generation branch, $\Delta\widehat{\mathbf{s}}_t^D$ isolates the local change in the model-estimated score induced by that branch. Thus, the haze differential in Eq. 1 is exactly the corresponding degradation-induced score change expressed in noise-prediction coordinates, up to the known diffusion scale σ_t and the sign imposed by the ϵ-to-score parameterization. This interpretation does not treat $\mathbf{g}_t$ as a physical haze vector; rather, it characterizes the local diffusion-prediction change learned by G_D in terms of its associated score displacement.

B EXPERIMENTAL DETAILS

Evaluation protocol. All evaluations use the official pretrained checkpoints and inference pipelines of the corresponding restoration models. For DiffDehaze (Wang et al., 2025), we preserve the original 50-step spaced sampler and restoration configuration and insert HazeCDG only between the restoration-model noise prediction and the subsequent sampler update. The restoration and degradation networks remain frozen throughout inference. The original DiffDehaze text condition, "remove dense fog," is retained, while degradation sensing uses the haze prompt "hazy, foggy, misty, obscure, smoggy." For RTTS and Fattal, we evaluate all available images, yielding 4,322 and 31 test images, respectively. For URHI, we construct a fixed 150-image subset from the available 4,809 images by sorting the image paths and uniformly sampling without replacement with random seed 123. The same subset is used for all compared methods and HazeCDG configurations.

We also preserve the baseline preprocessing and output-resolution handling for all DiffDehaze experiments. Images whose spatial dimensions fall below the model input scale are first resized to satisfy the minimum resolution requirement, after which the spatial size is adjusted to the nearest multiple of 64 using bicubic interpolation. No zero padding is introduced, and restored outputs are resized back to the original image resolution before evaluation. All reported methods are evaluated using identical input images and the same metric implementations.

Sparse HazeCDG implementation. For DiffDehaze, we evaluate $K \in \{5, 10, 25, 50\}$ degradation-sensing anchors along the 50-step reverse trajectory. For the 50-step sampler, these anchor sets are fixed and nested, $A_5 \subset A_{10} \subset A_{25} \subset A_{50}$, ensuring that a larger sensing budget preserves all anchors used by the corresponding smaller-budget setting. At each sensing anchor,

the haze differential $\mathbf{g}_t = \epsilon_D - \epsilon_{\text{base}}$ is recomputed from the current restoration state. At non-anchor steps, the degradation model and its frozen backbone are not evaluated; instead, the most recently sensed haze differential is reused directly in noise-prediction space according to Eq. 5. The guided restoration prediction is therefore obtained as $\epsilon_R^{\text{HazeCDG}} = \epsilon_R - 2\mathbf{g}$, where $\mathbf{g}$ denotes the latest sensed differential. Thus, the additional degradation-model evaluations scale with K, while the original DiffDehaze sampler remains unchanged. The $K = 50$ setting recomputes $\mathbf{g}_t$ at every reverse step and corresponds to the dense formulation.

One-step application to DOD. To examine whether HazeCDG depends on the multi-step LHD sampling process, we additionally apply it to the independently trained one-step diffusion restorer DOD (Tang et al., 2026a). DOD performs restoration using a single prediction at its native timestep $t = 999$, and therefore requires only one degradation-sensing operation ($K = 1$). We directly query the frozen HazeGen at the DOD latent state without re-encoding it and compute the haze differential $\mathbf{g} = \epsilon_D - \epsilon_{\text{base}}$. Following the same counter-degradation formulation as Sections 3.1 and 3.2, the corresponding HazeCDG update in noise-prediction space is $-2\mathbf{g}$.

Directly applying the multi-step correction magnitude to a one-step sampler would induce a larger scheduler-level perturbation because the correction is propagated through a single reverse transition. We therefore match the cumulative correction energy of the 50-step reference sampler using only the known diffusion schedules. Let h_k denote the linearized gain from an ϵ-space correction at reverse step k to the final latent. We define

$$H_{50} = \left(\sum_{k=1}^{50} h_k^2 \right)^{1/2}, \qquad \eta_{\text{DOD}} = \frac{H_{50}}{H_{\text{DOD}}}, \tag{8}$$

where H_{DOD} is the corresponding gain of the one-step DOD transition. For the SD2.1 schedule used by both models, $H_{50} = 6.48$ and $H_{\text{DOD}} = 14.61$, yielding $\eta_{\text{DOD}} = 0.443$. The DOD prediction is therefore corrected as

$$\epsilon_R^{\text{HazeCDG}} = \epsilon_R - 2\eta_{\text{DOD}}\mathbf{g}. \tag{9}$$

This coefficient is determined entirely by the diffusion schedules and is not tuned on validation data. The corrected prediction is passed to the original DOD scheduler update, after which the released Stage-2 decoder is used unchanged. Neither DOD nor HazeGen is retrained or adapted at test time.

C Additional Results on URHI and Fattal

The main paper reports RTTS results to keep the primary experimental presentation compact. Here, we provide the corresponding evaluations on URHI and Fattal for the frozen-restorer analysis in Section 4.2 and the comparison with recent real-world dehazing methods in Section 4.3.

Effect on frozen diffusion restorers. Table 5 extends the analysis of Table 1 to URHI and Fattal. For DiffDehaze, HazeCDG consistently improves haze suppression and perceptual alignment across different sensing budgets. As on RTTS, sparse sensing already captures most of the benefit, while denser sensing does not yield consistent gains across all quality criteria. This behavior supports the use of $K = 5$ as the default configuration and indicates that frequent degradation sensing is not necessary to exploit the learned haze differential.

The DOD results provide a complementary test of transferability. Despite its one-step sampling process and independently trained restoration model, HazeCDG again improves haze removal and CLIPIQA without modifying the DOD parameters. Together with the multi-step DiffDehaze results, this suggests that the counter-degradation cue is not tied to a particular reverse trajectory, but can also benefit a distinct one-step diffusion restorer.

Comparison with recent real-world dehazing methods. Table 6 compares HazeCDG with recent real-world dehazing methods on URHI and Fattal. On both datasets, HazeCDG substantially improves the competitive position of the frozen DiffDehaze backbone. Its gains are particularly consistent in perceptual alignment while also correcting much of the residual haze inherited from the original restorer. The relative advantage varies with the dataset, reflecting the different balance between aggressive haze removal and generic perceptual quality, but the overall pattern remains

Table 5: Additional HazeCDG results on frozen DiffDehaze and DOD for URHI and Fattal. **Bold** denotes the best result within each dataset and backbone.

(a) DiffDehaze (CVPR 2025)

Dataset	Setting	FADE↓	Q-Align↑	CLIPIQA↑	MUSIQ↑	BRISQUE↓
URHI	Baseline	1.293	3.743	0.470	**65.938**	24.193
	+ HazeCDG ($K = 5$)	**0.736**	3.844	0.552	65.806	**20.141**
	+ HazeCDG ($K = 10$)	0.737	**3.854**	**0.554**	65.651	21.296
	+ HazeCDG ($K = 25$)	0.748	3.848	0.553	65.468	21.691
	+ HazeCDG ($K = 50$)	0.747	3.845	**0.554**	65.420	21.769
Fattal	Baseline	0.636	4.006	0.551	68.419	**21.090**
	+ HazeCDG ($K = 5$)	**0.372**	**4.211**	**0.632**	**68.770**	21.858
	+ HazeCDG ($K = 10$)	0.382	4.190	0.627	68.717	21.712
	+ HazeCDG ($K = 25$)	0.390	4.170	0.629	68.659	22.645
	+ HazeCDG ($K = 50$)	0.387	4.161	**0.632**	68.588	23.074

(b) DOD (AAAI 2026)

Dataset	Setting	FADE↓	Q-Align↑	CLIPIQA↑	MUSIQ↑	BRISQUE↓
URHI	Baseline	1.583	**3.304**	0.468	**58.525**	**33.925**
	+ HazeCDG ($K = 1$)	**1.142**	3.263	**0.500**	56.997	35.272
Fattal	Baseline	0.488	**4.028**	0.528	**63.507**	**22.571**
	+ HazeCDG ($K = 1$)	**0.426**	3.987	**0.532**	62.456	24.183

Table 6: Comparison with recent real-world dehazing methods on URHI and Fattal. Best and second-best results are shown in **bold** and underline, respectively.

Dataset	Method	FADE↓	Q-Align↑	CLIPIQA↑	MUSIQ↑	BRISQUE↓
URHI	HazeFlow (ICCV'25)	**0.625**	3.620	0.505	65.818	24.201
	CoA (CVPR'25)	0.904	3.405	0.446	58.481	28.859
	DehazeSB (ICCV'25)	0.811	3.214	0.430	64.157	21.979
	HNDiff (ECCV'26)	0.669	3.041	0.452	61.501	35.104
	PDDA (TIP'26)	0.785	3.767	0.458	64.499	26.776
	BiLaLoRA (CVPR'26)	0.812	3.796	0.467	**65.971**	26.666
	DiffDehaze (CVPR'25)	1.293	3.743	0.470	65.938	24.193
	DiffDehaze + DCTTA (CVPR'26)	1.315	3.748	0.484	65.933	23.877
	DiffDehaze + HazeCDG (Ours)	0.736	**3.844**	**0.552**	65.806	**20.141**
Fattal	HazeFlow (ICCV'25)	**0.262**	4.188	0.606	68.552	24.420
	CoA (CVPR'25)	0.335	3.870	0.538	62.799	28.045
	DehazeSB (ICCV'25)	0.433	4.037	0.564	68.611	23.529
	HNDiff (ECCV'26)	0.330	3.865	0.543	66.748	25.817
	PDDA (TIP'26)	0.338	4.095	0.561	**69.441**	26.674
	BiLaLoRA (CVPR'26)	0.301	4.194	0.582	68.591	23.416
	DiffDehaze (CVPR'25)	0.636	4.006	0.551	68.419	21.090
	DiffDehaze + DCTTA (CVPR'26)	0.632	3.892	0.569	67.343	**19.689**
	DiffDehaze + HazeCDG (Ours)	0.372	**4.211**	**0.632**	68.770	21.858

consistent with the RTTS results: reusing G_D provides information that is complementary to the pretrained restoration model.

The comparison with DCTTA (Tang et al., 2026b) further highlights the difference between test-time adaptation and degradation reuse. DCTTA improves individual perceptual-quality measures through test-time optimization, but its gains in haze suppression are limited or inconsistent in these comparisons. HazeCDG keeps both pretrained networks fixed and directly modifies the diffusion prediction using the counter-degradation cue. The observed improvements support further evaluation of HazeCDG on other pretrained diffusion restorers to examine whether similar improvements can be obtained.

D ADDITIONAL QUALITATIVE RESULTS

We provide additional qualitative comparisons on RTTS, URHI, and Fattal dataset. These examples complement the quantitative results by illustrating how HazeCDG suppresses residual haze while preserving scene structure and perceptual quality across diverse real-world conditions.

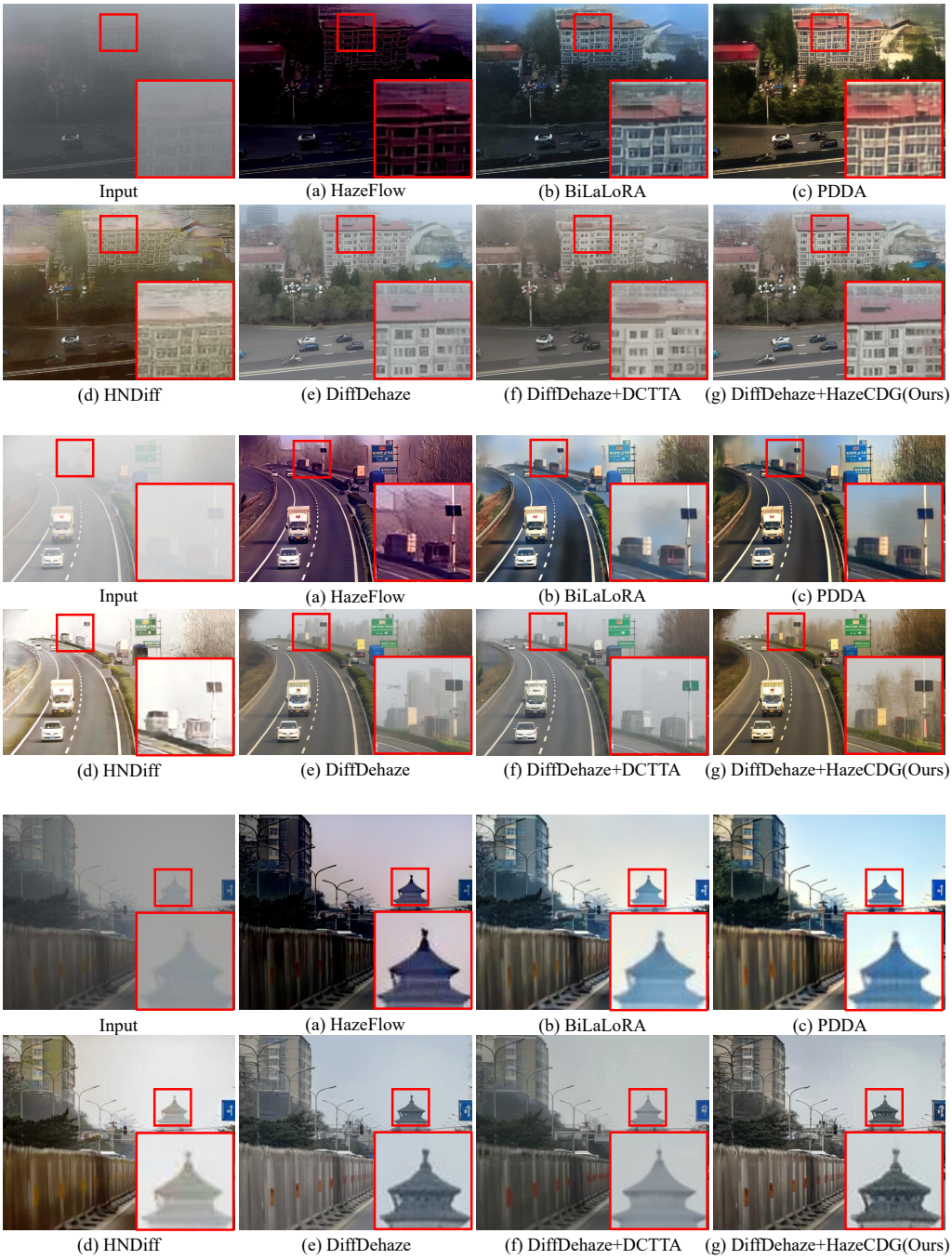


Figure 3: Visual comparisons on the RTTS dataset (Li et al., 2019). Red boxes indicate the regions magnified in the insets.



Figure 4: Visual comparisons on the URHI dataset. Red boxes indicate the regions magnified in the insets.

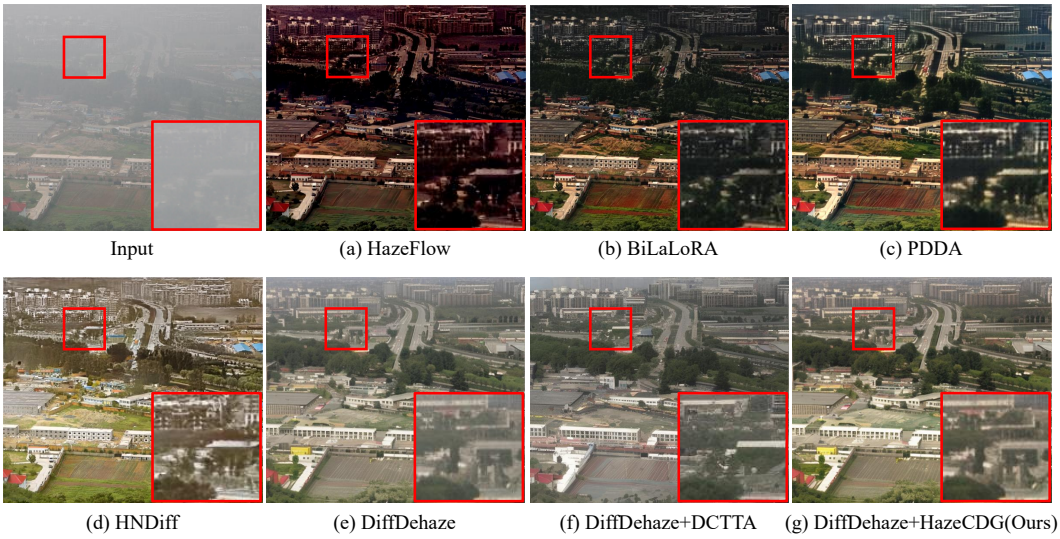


Figure 5: Visual comparisons on the Fattal dataset. Red boxes indicate the regions magnified in the insets.