GEOHAZE: GEOMETRY-CONDITIONED SPECTRAL PROPAGATION FOR STRUCTURE-PRESERVING IMAGE DEHAZING

Anonymous authors
Paper under double-blind review

ABSTRACT

Image dehazing in severe natural haze and remote-sensing atmospheric degradation suffers from severe information suppression, making structure-consistent restoration depend on information beyond the degraded neighborhood. Monocular depth provides such nonlocal geometry, yet depth inferred from hazy inputs may contain shifted or missing boundaries; directly fusing it with RGB features can therefore entangle geometric errors with the appearance being reconstructed. We introduce GeoHaze based on a different principle: uncertain geometry should condition how RGB evidence propagates, rather than serve as reconstruction content. GeoHaze realizes this principle through a projected damped-wave spectral operator whose RGB state is driven by a spatially gated depth response, enabling global interaction while retaining RGB as the appearance carrier. Within this operator, two-level RGB–depth wavelet statistics predict sample-wise wave speed and damping, while a residual gate regulates how much of the propagated response reenters the RGB stream; a frequency-decoupled decoder then uses geometric guidance to separately recover low-frequency layout and refine high-frequency detail, translating wave responses into structurally coherent reconstructions. Across 11 natural and remote-sensing benchmarks, GeoHaze achieves competitive restoration quality.Controlled wave variants and depth perturbations characterize the contribution and sensitivity of geometry conditioning, with an approximately 64% reduction in mean PSNR drop relative to SSDG across the three evaluated depth corruptions on RW^{2}AH. Code is available in the supplementary materials.

1 INTRODUCTION

Image dehazing recovers clear scene radiance from hazy observations and supports perception in both natural scenes and remote-sensing imagery (Gui et al., 2023; Li et al., 2019; Liu et al., 2021; Zhou et al., 2026). Under dense atmospheric haze or thick aerial contamination, local RGB evidence is often too weak for structure-consistent restoration, where textures vanish, and even coarse layout cues can be attenuated (Wu et al., 2023; Fan et al., 2025a; Zhu et al., 2025). Natural and remote-sensing settings both need long-range structural consistency, yet they differ in imaging geometry and haze formation, so restoration cannot rely on a single transferable depth–transmission coupling (Liu et al., 2021; Zhou et al., 2026; Lihe et al., 2024).

Monocular geometry from vision foundation models such as Depth Anything (Yang et al., 2024a;b; Lin et al., 2026) provides relative layout and boundary cues beyond corrupted RGB appearance (Zhang et al., 2025). Depth-guided dehazing already exploits this cue through joint depth–restoration optimization (Zhang et al., 2024b), depth prompting (Ye et al., 2025), self-supervised depth guidance (Liang et al., 2025), depth-centric architectures (Fan et al., 2025b), and recent depth-prior networks such as UDPNet (Zuo et al., 2026). Depth estimated from hazy inputs may nevertheless contain shifted or missing boundaries. When geometry is injected mainly as an appearance feature to be fused with RGB, those errors can entangle with the content being reconstructed, as illustrated in Fig. 1. The open question is therefore how to condition restoration on imperfect monocular geometry while limiting direct appearance transfer.

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.

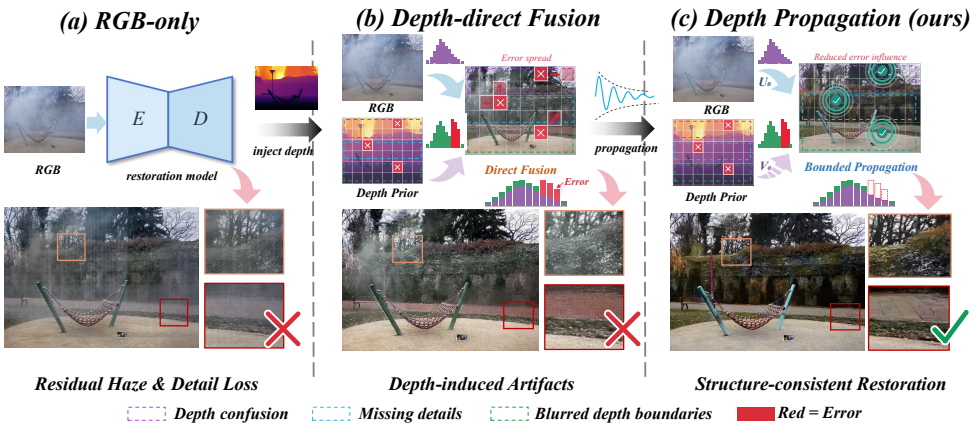

Figure 1: Depth use in image dehazing. (a) RGB-only restoration leaves residual haze and missing detail. (b) Direct depth fusion can transfer geometric errors into the restored image. (c) GeoHaze conditions bounded RGB propagation on depth to preserve scene structure. Framed insets compare corresponding regions.

A complementary design choice is the long-range operator itself. Existing restorers extend feature interaction beyond local convolutions using windowed attention (Song et al., 2023), linearized spatial attention (Qiu et al., 2023), and state-space modeling (Guo et al., 2024). WaveFormer (Shu et al., 2026) instead formulates feature evolution with a damped wave equation and derives an efficient projected spectral wave-propagation operator with $\mathcal{O}(N \log N)$ global interaction, complementing heat-like diffusion that favors stronger low-pass smoothing (Wang et al., 2025). For dehazing, this operator is attractive because velocity and damping provide controllable long-range interaction while RGB remains the appearance carrier. The remaining gap is not inventing a new wave PDE, but conditioning the driving response and sample-wise propagation parameters on external geometry estimated under haze.

To this end, we present *GeoHaze*, which follows a principle that uncertain geometry conditions how RGB evidence propagates, rather than serving as reconstruction content. Building on WaveFormer (Shu et al., 2026), a Heterogeneous Wave Propagation Block (HWPB) applies a projected damped-wave spectral step whose initial driving response is spatially gated by depth features; a Wavelet-Driven Kinematics Estimator (WDKE) predicts sample-wise wave speed and damping from image-level RGB–depth wavelet statistics, while a residual gate regulates how much of the propagated response re-enters the RGB stream; and a Geometry-Gated Frequency Decoupling Block (GFDB) uses geometric guidance to recover low-frequency layout and refine high-frequency detail separately. As shown in Fig. 1, this design seeks structure-preserving restoration without treating depth as an unconstrained fusion token.

The main contributions are as follows:

- ***Methodologically***, we study imperfect monocular geometry for structure-preserving dehazing and treat fusion-induced error transfer as a motivating risk rather than a proven paradox, asking how geometry should condition restoration while limiting direct appearance transfer.
- ***Technically***, we introduce GeoHaze, which conditions a projected damped-wave spectral operator on monocular geometry through depth-gated initial driving, sample-wise cross-modal kinematics with residual admission, and frequency-decoupled decoding, building on the wave-propagation operator.
- ***Empirically***, across 11 natural-image and remote-sensing benchmark settings, GeoHaze achieves competitive restoration quality. Controlled wave variants and three test-time depth corruptions characterize the contribution and sensitivity of geometry conditioning, with an approximately 64% reduction in mean PSNR drop relative to baselines across the evaluated corruptions on RW^{2}AH. Downstream detection on RTTS also shows a modest overall gain.

2 Related Work

Image Dehazing and Cross-Scenario Restoration. Learning-based dehazing has progressed from CNN and Transformer restorers with multi-scale context (Qin et al., 2020; Song et al., 2023) to real-world and nighttime haze (Wu et al., 2023; Cong et al., 2024; Fan et al., 2025a), as well as remote-sensing haze or thin-cloud removal under near-overhead imaging (Lihe et al., 2024; Zhu et al., 2025; Zhou et al., 2026). These settings share the need for structure-consistent recovery under weak local RGB evidence, yet they differ in imaging geometry and haze formation. GeoHaze is evaluated across natural-image and remote-sensing protocols without assuming a transferable depth–transmission coupling.

Geometry Priors for Dehazing. Depth has long served as a structural cue for dehazing. Prior work jointly estimates depth and clean radiance (Zhang et al., 2024b; Fan et al., 2025b), prompts restoration with foundation monocular geometry such as Depth Anything (Yang et al., 2024a; Ye et al., 2025), or fuses RGB and depth features with gates and cross-attention, including SSDG and UDPNet (Zhang et al., 2025; Liang et al., 2025; Zuo et al., 2026). Relative geometry can stabilize layout under haze, but depth estimated from hazy inputs may still contain shifted or missing boundaries. GeoHaze uses depth to condition a spectral propagation operator and a frequency-decoupled decoder, rather than treating depth mainly as an appearance feature fused into RGB representations.

Spectral Operators for Restoration. Haze corrupts layout and detail asymmetrically across frequency bands, motivating frequency-selective restorers such as AdaIR and WaveletFormerNet (Cui et al., 2025c; Zhang et al., 2024a). Dense haze further requires long-range interaction beyond local convolutions (Song et al., 2023; Guo et al., 2024). Physics-inspired operators provide another route: heat-conduction networks favor semantic diffusion with a strong low-pass bias (Wang et al., 2025), whereas WaveFormer formulates feature evolution with a damped wave equation and derives an efficient projected spectral wave-propagation operator with $\mathcal{O}(N \log N)$ global interaction (Shu et al., 2026). GeoHaze builds on this operator and conditions its driving response and sample-wise kinematics on monocular geometry. The frequency-decoupled decoder complements propagation rather than posing frequency decoupling as an independent paradigm.

3 Method

3.1 Problem Formulation

Let $I \in \mathbb{R}^{3\times H\times W}$ be a hazy observation. Classical atmospheric scattering (Narasimhan & Nayar, 2002) writes

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

where J is the clean radiance, t is transmission, and A is atmospheric light. When $t(x) \to 0$, local RGB evidence is dominated by A, so pixel-wise inversion is highly ill-posed. We therefore introduce a relative geometry prior from Depth Anything 3 (Lin et al., 2026),

$$D = \mathcal{F}_{\mathrm{DA3}}(I), \qquad \hat{J} = \mathcal{G}_\theta(I, D), \tag{2}$$

and seek a restoration map $\mathcal{G}_\theta$ that uses D as a conditioning signal rather than as reconstruction content. Here D is a monocular relative-depth estimate from the hazy input; it is neither transmission nor ground-truth depth. Direct RGB–depth fusion can entangle geometric errors with appearance (Liang et al., 2025; Zuo et al., 2026). Our design principle is accordingly: uncertain geometry should condition how RGB features propagate, while RGB remains the appearance carrier.

Building on the damped-wave spectral operator of WaveFormer (Shu et al., 2026), let $U(x, \tau)$ denote an RGB feature state evolving over an internal time τ:

$$\frac{\partial^2 U}{\partial \tau^2} + \alpha \frac{\partial U}{\partial \tau} = v^2 \Delta U, \tag{3}$$

with planar Laplacian Δ, initial state $U(0) = U_0$ derived from RGB features, and initial velocity $\dot{U}(0) = \widetilde{V}_0$. Relative to heat-like diffusion (Wang et al., 2025), the wave step retains an oscillatory component that better preserves high-frequency structure. Geometry enters through a spatial driving response

$$\widetilde{V}_0 = \Gamma_d \odot \phi_v(F_d), \qquad \Gamma_d = \sigma\big(\phi_g(F_d)\big), \tag{4}$$

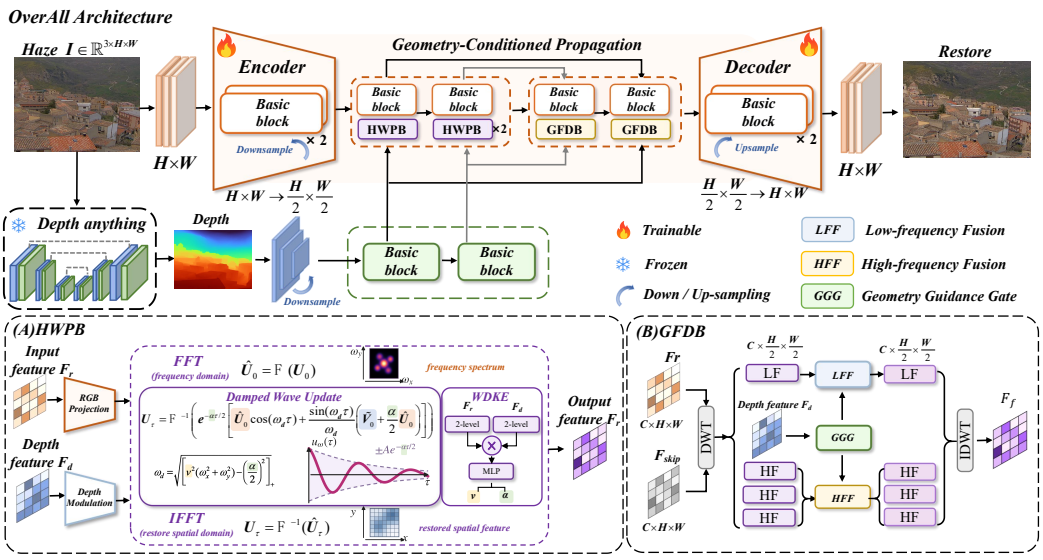

Figure 2: Pipeline of GeoHaze. A hazy RGB image is paired with a monocular depth prior from Depth Anything 3. HWPB applies a geometry-conditioned projected wave step, WDKE predicts sample-wise wave speed and damping from RGB–depth wavelet statistics, and GFDB performs frequency-decoupled skip fusion in the decoder.

where F_d is a depth feature and Γ_d is a spatial conditioning map (not a calibrated reliability estimate). Sample-wise kinematics and residual admission complete the control:

$$\begin{aligned}[a_v, a_\alpha] &= \phi_p\big([s_1(F_r), s_1(F_d), s_2(F_r), s_2(F_d)]\big),\\ v &= v_{\min} + (v_{\max} - v_{\min})\sigma(a_v), \qquad \alpha = \alpha_{\min} + (\alpha_{\max} - \alpha_{\min})\sigma(a_\alpha),\\ F_r' &= F_r + r_s \odot \phi_o\big(\mathrm{LN}(U_\tau)\big) \odot \mathrm{SiLU}(G_r^0),\end{aligned} \tag{5}$$

where F_r is an RGB feature, $s_k(\cdot)$ denotes a k-level Haar magnitude summary (Appendix A), r_s is a per-channel learned scale initialized at 0.1, and G_r^0 is predicted from RGB. The pair (v, α) has shape $B \times 1 \times 1 \times 1$ and is broadcast over space and frequency; it is not a spatial coefficient field. After propagation, decoding must still handle frequency-asymmetric haze; we therefore use geometry-gated Haar fusion on decoder skips, with a final RGB head producing $\hat{J}$ (Sec. 3.3). Equations (3)–(5) define the design targets instantiated by HWPB, WDKE, and the residual path in Fig. 2.

3.2 Geometry-Conditioned Spectral Propagation

GeoHaze uses a dual-branch encoder: patch embeddings $F_r^0 = P_r(I)$ and $F_d^0 = P_d(D)$ keep appearance and geometry in separate streams. At interaction stages, a Heterogeneous Wave Propagation Block (HWPB) updates RGB features with depth-conditioned propagation, while the depth branch remains a compact conditioning pathway. Shallow stages encode each modality independently so that low-level RGB texture is formed before geometry begins to modulate the wave step.

HWPB realizes Eq. (3) as one projected spectral step. From normalized features $\bar{F}_r$ and $\bar{F}_d$, it predicts $[U_0, G_r^0] = \phi_u(\mathrm{DWConv}(\bar{F}_r))$ and the depth-driven velocity in Eq. (4). With Wavelet-Driven Kinematics Estimator (WDKE) parameters (v, α) from Eq. (5), the spectral update is

$$U_\tau = \mathcal{F}^{-1}\left(e^{-\alpha\tau/2}\left[\hat{U}_0\cos(\rho\tau) + \frac{\sin(\rho\tau)}{\rho}\left(\widehat{\tilde{V}}_0 + \tfrac{\alpha}{2}\hat{U}_0\right)\right]\right), \tag{6}$$

where $\rho = \sqrt{\max\big(v^2(\omega_x^2+\omega_y^2) - (\alpha/2)^2, \epsilon\big)}$ and the ratio $\sin(\rho\tau)/\rho$ uses its continuous limit as $\rho \to 0$. The closed-form modal branches and discrete FFT grid are given in Appendix A. The residual path in Eq. (5) then re-enters the RGB stream; because $\mathrm{SiLU}(G_r^0)$ is signed, it is not a $[0, 1]$ hard gate. Initial-drive trajectories are visualized in Fig. 3, and internal HWPB activations in Appendix Fig. 8; controlled wave variants in Sec. 4.4 test the contribution of geometry conditioning.

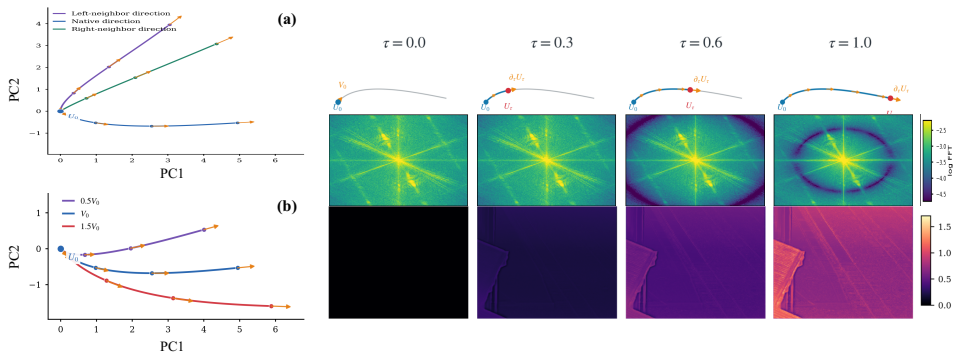

Figure 3: **Initial-drive sensitivity and HWPB state evolution on RW^{2}AH.** Left: PCA trajectories relative to U_0, with fixed U_0 and wave coefficients. (a) Driving direction varies at a fixed 64-dimensional norm. (b) Drive magnitude is scaled by 0.5, 1, and 1.5 with its initial direction fixed. Orange arrows denote $\partial_\tau U_\tau$. Right: four time slices ($\tau = 0, 0.3, 0.6, 1$) show state–velocity projections (top), log spectral-amplitude maps (middle), and channel RMS of $U_\tau - U_0$ (bottom); each map row uses a shared color scale across time.

3.3 Frequency-Decoupled Decoding

Haze veils low-frequency layout while mixing high-frequency edges with noise. After encoder-side propagation, a Geometry-Gated Frequency Decoupling Block (GFDB) applies a one-level Haar decomposition to each encoder skip S and upsampled decoder feature U, together with an aligned depth skip Z. The low-frequency band is aggregated and modulated by channel–spatial gates and a depth multiplier; the three high-frequency bands share one geometry-derived gate $G^H = \sigma(\phi_h(\downarrow Z))$ while refining residual detail descriptors. Inverse Haar reconstruction yields a fused skip feature

$$F_{\text{fuse}} = \phi_f\big([\mathcal{D}^{-1}(L^*, H^*_{LH}, H^*_{HL}, H^*_{HH}); U]\big) + U, \tag{7}$$

which complements geometry-conditioned propagation rather than defining an end-to-end wavelet inversion of the whole network. Full gate definitions appear in Appendix A. The decoder finally maps the full-resolution RGB stream to $\hat{J}$ through a convolutional head with an RGB residual.

3.4 Training Objective

Training supervises a three-scale pyramid. Let $r(i) = \min(1, i/T_w)$ be a warm-up weight on iteration i. The executed objective is

$$\begin{aligned}\mathcal{L} = \|\hat{J} - J\|_1 + 0.5\|\hat{J}_{1/2} - J_{1/2}\|_1 + 0.25\|\hat{J}_{1/4} - J_{1/4}\|_1 \\ + r(i)\Big[\|\nabla\hat{J} - \nabla J\|_1 + 0.5\big(1 - \text{SSIM}(\hat{J}, J)\big)\Big],\end{aligned} \tag{8}$$

So multi-scale ℓ_1 acts at full, half, and quarter resolution, while gradient and SSIM terms apply only at full resolution after warm-up, matching the implementation in Appendix B. At inference, $\hat{J}$ is taken from the full-resolution decoder output.

4 Experiments

4.1 Experimental Setup

We evaluate natural real haze, aerial haze-density splits, and remote-sensing cloud/haze under different imaging geometries. Beyond fidelity comparisons, we report mechanism baselines against depth-fusion and homogeneous-wave variants, test-time depth corruptions, and a modest RTTS detection check. Table 11 in Appendix C summarizes the 11 datasets: RW^{2}AH (Fang et al., 2025), I-HAZE (Ancuti et al., 2018b), O-HAZE (Ancuti et al., 2018a), and NH-HAZE (Ancuti et al., 2020) for real haze; RRSHID (Zhu et al., 2025) for thin/moderate/thick aerial splits; and

Table 1: Quantitative comparison on RW^{2}AH (Fang et al., 2025), I-HAZE (Ancuti et al., 2018b), O-HAZE (Ancuti et al., 2018a), and NH-HAZE (Ancuti et al., 2020). Higher is better for PSNR/SSIM; lower is better for MSE/LPIPS. Best and second-best entries are highlighted.

Method	Venue	RW^{2}AH				I-HAZE				O-HAZE				NH-HAZE			
		PSNR ↑	SSIM ↑	MSE ↓	LPIPS ↓	PSNR ↑	SSIM ↑	MSE ↓	LPIPS ↓	PSNR ↑	SSIM ↑	MSE ↓	LPIPS ↓	PSNR ↑	SSIM ↑	MSE ↓	LPIPS ↓
DCP (He et al., 2009)	CVPR'09	15.53	.4865	.0365	.3442	11.02	.5809	.0820	.5213	14.05	.5685	.0421	.5233	12.92	.4678	.0532	.5833
DehazeFormer (Song et al., 2023)	TIP'23	20.94	.6120	.0131	.3560	18.63	.5454	.0174	.3796	21.74	.7188	.0077	.4037	18.28	.5556	.0155	.4442
RIDCP (Wu et al., 2023)	CVPR'23	19.87	.6270	.0158	.3393	18.48	.7734	.0159	.3967	22.83	.7273	.0055	.3799	16.64	.6010	.0243	.4236
DEA-Net (Chen et al., 2024)	TIP'24	20.84	.5408	.0145	.3290	**20.85**	.7656	**.0096**	**.3072**	22.93	.7129	.0055	**.2674**	19.25	.6631	.0126	**.2792**
DCMPNet (Zhang et al., 2024b)	CVPR'24	20.13	.5870	.0231	.6726	17.55	.5107	.0217	.5393	20.39	.6799	.0103	.4940	17.43	.5330	.0189	.4248
PoolNet (Cui et al., 2025a)	TIP'25	22.15	.6438	**.0105**	.3189	18.17	.8002	.0192	.3979	22.52	.7394	.0078	.3853	16.24	.5807	.0246	.3879
SFMN (Shen et al., 2025)	TIP'25	20.43	.5765	.0139	.3875	17.39	.7619	.0236	.3804	22.66	.7304	.0069	.4187	16.52	.5473	.0229	.4539
SSDG (Liang et al., 2025)	PR'25	17.93	.6073	.0197	.4745	17.79	.5539	.0174	.5326	20.74	.6117	.0086	.3829	17.55	.6021	.0184	.3230
WaveFormer (Shu et al., 2026)	AAAI'26	20.57	.5632	.0146	.3257	15.78	.7367	.0292	.3654	21.70	.7122	.0089	.3936	17.18	.5584	.0198	.4982
PhDnet (Lihe et al., 2024)	InfoFusion'24	20.77	.5766	.0134	.3641	16.75	.6686	.0241	.4109	22.04	.7247	.0069	.3268	16.07	.4723	.0262	.5472
SFRDP-Net (Sun et al., 2025)	TGRS'25	19.69	.5640	.0163	.3691	16.81	.7369	.0231	.4355	19.86	.6866	.0114	.3388	16.57	.5721	.0228	.4409
ThiefCloud (Zhao et al., 2025)	TCSVT'25	19.63	.5610	.0165	.3265	15.51	.7402	.0298	.4126	20.31	.7079	.0110	.3709	17.59	.6259	.0182	.3375
GeoHaze (ours)	–	**22.41**	**.6610**	.0106	**.3011**	19.14	**.8209**	.0135	.3407	**24.97**	**.7888**	**.0034**	.3658	**19.64**	**.6896**	**.0122**	.3100

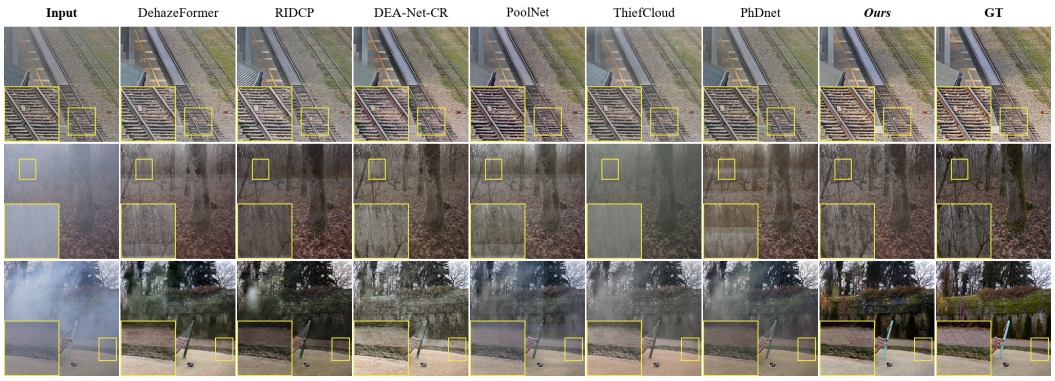

Figure 4: Qualitative comparison on real-haze benchmarks. Rows 1–3 correspond to RW^{2}AH (Fang et al., 2025), O-HAZE (Ancuti et al., 2018a), and NH-HAZE (Ancuti et al., 2020), respectively. Columns align DehazeFormer (Song et al., 2023), RIDCP (Wu et al., 2023), DEA-Net-CR (Chen et al., 2024), PoolNet (Cui et al., 2025a), ThiefCloud (Zhao et al., 2025), PhDnet (Lihe et al., 2024), and GeoHaze, together with the input and ground truth. The lower-left framed insets enlarge the regions marked by color-matched boxes. Additional RW^{2}AH scenes are provided in the supplementary material.

RICE1/RICE2 (Lin et al., 2019), RSID (Chi et al., 2023), and SateHaze1k-thick (Huang et al., 2020) for remote-sensing cloud/haze.

Training follows the two-stage schedule and the objective in Eq. (8): 256×256 crops (batch 10/GPU, LR 2.5×10^{-4}, 200k iterations), then 384×384 crops (batch 6, LR 5×10^{-5}, 40k iterations), with AdamW $(\beta_1, \beta_2) = (0.9, 0.99)$, weight decay 10^{-4}, gradient clipping 1.0, EMA, and cosine annealing. Unless noted, each dataset uses an independently trained checkpoint under its official split; we evaluate the predefined final Stage 2 EMA checkpoint on fixed test manifests with a shared metric implementation. Seed protocol, split manifests, and the depth-assisted checkpoint ledger are in Appendix B–C. Metrics are PSNR/MSE, SSIM (Wang et al., 2004), and LPIPS (Zhang et al., 2018).

4.2 Real Haze Benchmarks

Table 1 compares GeoHaze with physical-prior, transformer, depth-assisted, spectral, and remote-sensing restorers on RW^{2}AH, I-HAZE, O-HAZE, and NH-HAZE.

On RW^{2}AH, GeoHaze reaches 22.41 dB PSNR / 0.6610 SSIM / 0.3011 LPIPS. It also leads on most metrics on O-HAZE and NH-HAZE, and ranks first in SSIM on I-HAZE, where DEA-Net remains strongest in PSNR/MSE/LPIPS. Fig. 4 shows representative real-haze restorations with cleaner boundaries and more stable local contrast; additional scenes are in the supplementary ma-

Table 2: Quantitative comparison on RRSHID-thin/thick (Zhu et al., 2025), RICE2 (Lin et al., 2019), and RSID (Chi et al., 2023). Arrows indicate the preferred metric direction; brown-gold, blue, and green mark the top three. Complete results are in the supplementary material.

Method	Venue	Complexity		RRSHID-thin				RRSHID-thick				RICE2				RSID			
		Param. (M)	FLOPs (G)	PSNR ↑	SSIM ↑	LPIPS ↓	MSE ↓	PSNR ↑	SSIM ↑	LPIPS ↓	MSE ↓	PSNR ↑	SSIM ↑	LPIPS ↓	MSE ↓	PSNR ↑	SSIM ↑	LPIPS ↓	MSE ↓
DCP (He et al., 2009)	CVPR'09	–	**.60**	18.46	.4564	.4851	.0192	18.39	.4843	.4996	.0208	14.10	.3898	.6234	.0656	13.87	.6892	.2615	.0538
FFA-Net (Qin et al., 2020)	AAAI'20	4.46	287.80	17.08	.4452	.5761	.0326	16.71	.4792	.5573	.0377	17.77	.6261	.5378	.0323	18.31	.8582	.1449	.0204
4KDehazing (Zheng et al., 2021)	CVPR'21	34.55	105.80	22.83	.6177	.4352	.0063	22.55	.6912	.4754	.0099	25.21	.8604	.3704	.0087	23.61	.9415	.0693	.0053
DehazeFormer (Song et al., 2023)	TIP'23	**1.21**	39.76	22.74	.6005	.4609	.0071	24.69	.7143	.4377	.0051	34.54	.8875	.2035	.0010	25.04	.9393	.0557	.0040
OKNet (Cui et al., 2024b)	AAAI'24	4.72	39.67	22.20	.5916	.4797	.0081	23.66	.6968	.4782	.0064	29.78	.8714	.2779	.0018	25.72	.9342	.0369	.0035
DCMPNet (Zhang et al., 2024b)	CVPR'24	7.16	62.89	21.75	.4100	.7106	.0080	23.27	.5270	.5427	.0062	31.67	.8101	.2750	.0015	25.28	.9402	.0471	.0038
EENet (Cui et al., 2025b)	PR'25	5.44	49.83	22.34	.6006	.4791	.0078	23.98	.7068	.4752	.0060	31.48	.8852	.2562	.0014	23.89	.9376	.0695	.0052
PoolNet (Cui et al., 2025a)	TIP'25	6.92	66.38	23.80	.6584	.3346	.0054	25.07	.7291	.3504	.0049	36.57	.9169	.1754	**.0007**	25.88	.9522	.0363	.0035
SFMN (Shen et al., 2025)	TIP'25	3.43	36.57	23.64	.6587	.4146	.0062	25.09	.7331	.4601	.0052	**36.61**	.9142	.1815	**.0007**	26.82	.9583	.0331	.0030
SSDG (Liang et al., 2025)	PR'25	45.02	60.60	21.79	.5731	**.3019**	.0083	22.94	.6490	**.2972**	.0065	33.71	.8600	.1931	.0013	23.94	.9247	.0504	.0066
WaveFormer (Shu et al., 2026)	AAAI'26	43.54	118.40	21.31	.5255	.3794	.0103	22.56	.6268	.4357	.0075	32.12	.8331	.3130	.0011	22.50	.8434	.2092	.0070
Trinity-Net (Chi et al., 2023)	TGRS'23	20.14	30.78	20.51	.5728	.4578	.0120	24.11	.7234	.4206	.0058	18.81	.7781	.2287	.0200	23.60	.9322	.0692	.0060
PhDnet (Lihe et al., 2024)	InfoFusion'24	10.03	33.24	22.50	.5949	.4558	.0074	24.64	.7203	.4563	.0054	34.52	.8836	.2604	.0010	23.77	.9179	.0865	.0054
SFRDP-Net (Sun et al., 2025)	TGRS'25	3.51	51.37	23.02	.6225	.4708	.0064	25.01	.7312	.4318	.0047	34.33	.9011	.2041	.0011	26.60	.9562	.0349	.0031
ThiefCloud (Zhao et al., 2025)	TCSVT'25	2.06	40.83	23.01	.6144	.3592	.0066	25.05	.7187	.3610	**.0041**	34.91	.9000	.1786	.0009	25.13	.9507	.0390	.0043
GeoHaze (ours)	–	11.49	57.64	**24.43**	**.6926**	.3539	**.0050**	**25.52**	**.7519**	.4120	.0046	36.48	**.9254**	**.1709**	.0009	**27.09**	**.9625**	**.0287**	**.0027**

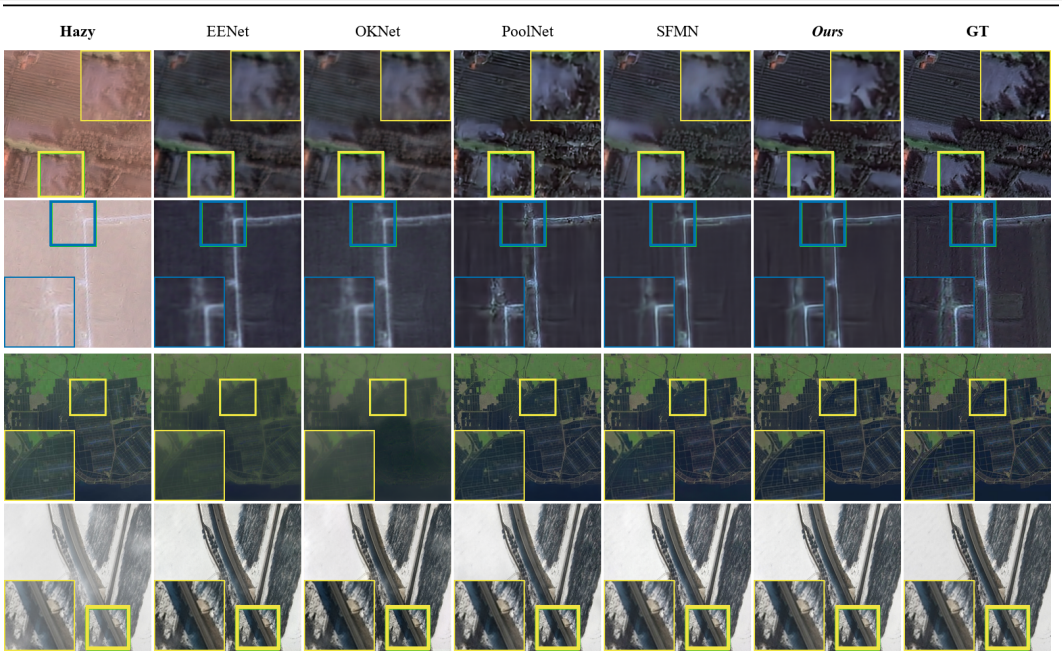

Figure 5: Qualitative comparison on four remote-sensing settings. Rows 1–4 correspond to RRSHID-thin, RRSHID-thick, RICE2, and RSID, respectively. The framed insets enlarge the regions marked by color-matched boxes. Additional scenes are provided in the supplementary material.

terial. These results support competitive restoration across indoor, outdoor, and non-homogeneous real haze when geometry conditions a spectral operator rather than serving only as a fused appearance feature.

4.3 Remote-Sensing Benchmarks

Table 2 reports four representative remote-sensing settings: RRSHID-thin/thick, RICE2, and RSID. RRSHID exposes the same overhead geometry to a haze-density gradient, while RICE2 and RSID test thick cloud obstruction and synthetic satellite haze. Complete RRSHID-moderate, RICE1, and SateHaze1k-thick results are in the supplementary material.

GeoHaze obtains the best PSNR/SSIM on RRSHID-thin (24.43 dB / 0.6926) and RRSHID-thick (25.52 dB / 0.7519). On RICE2, it ranks first in SSIM/LPIPS (0.9254 / 0.1709) and third in PSNR (36.48 dB). On RSID, it reaches 27.09 dB PSNR and 0.0287 LPIPS, improving over SFMN by

Table 3: Mechanism comparison on RW^{2}AH and RSID. DCMPNet (Zhang et al., 2024b) and SSDG (Liang et al., 2025) use Depth Anything 3 (Lin et al., 2026) priors at test time; fixed-wave variants retain the spectral step.

Method	Mechanism	RW^{2}AH (Fang et al., 2025)		RSID (Chi et al., 2023)	
		PSNR↑	SSIM↑	PSNR↑	SSIM↑
DCMPNet	depth fusion	20.13	0.5870	25.28	0.9402
SSDG	attention fusion	17.93	0.6073	23.94	0.9247
Fixed wave	fixed kinematics	22.18	0.6573	26.82	0.9598
Zero depth velocity	no depth drive	20.22	0.5861	26.43	0.9585
GeoHaze (ours)	geometry-guided wave	**22.41**	**0.6610**	**27.09**	**0.9625**

Controls. Fixed wave uses $v = 1.0$ and $\alpha = 0.1$; zero depth velocity removes the drive in Eq. (4). Both retain the spectral step and use the same test manifests and metrics as the full model.

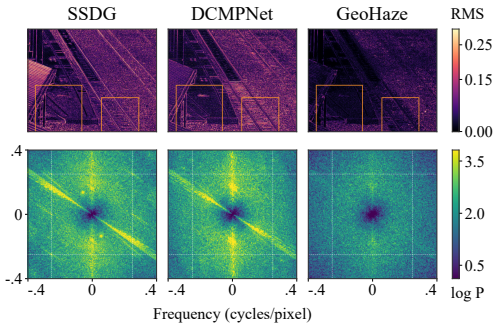

Figure 6: High-frequency residuals (top, Gaussian high-pass with $\sigma = 1.5$) and residual-power spectra (bottom). Color scales are shared across methods; numerical diagnostics are reported in Sec. 4.4.

Table 4: Test-time depth corruption on RW^{2}AH (Fang et al., 2025). SSDG (Liang et al., 2025) and GeoHaze share the clean-depth protocol; $|D|$ is the absolute PSNR drop.

Depth setting	Method	PSNR↑	SSIM↑	LPIPS↓	$\lvert D\rvert$↓
Clean depth	SSDG	17.93	0.6073	0.4745	0.0000
	GeoHaze	22.41	0.6610	0.3011	0.0000
Gaussian noise	SSDG	17.61	0.5674	0.5407	0.3233
	GeoHaze	22.38	0.6600	0.3016	**0.0260**
Gaussian blur	SSDG	17.83	0.6070	0.4641	0.1015
	GeoHaze	22.33	0.6591	0.3009	**0.0783**
Edge shift	SSDG	17.82	0.6043	0.4673	0.1137
	GeoHaze	22.32	0.6601	0.3008	**0.0899**

Table 5: Reported YOLOv8s (Varghese & M., 2024) detection mAP$_{50}$ on RTTS (Li et al., 2019) for hazy inputs and restorations by PoolNet (Cui et al., 2025a) and GeoHaze (Ours). Larger mAP$_{50}$ scores indicate better detection.

Class	mAP$_{50}$		
	Hazy	PoolNet	Ours
person	**0.754**	0.6980	0.745
bicycle	**0.518**	0.4203	0.513
car	0.700	0.6219	**0.707**
motorcycle	0.489	0.4258	**0.512**
bus	0.425	0.2870	**0.427**
all	0.577	0.4906	**0.581**

0.27 dB PSNR and 13.3% relative LPIPS. Fig. 5 shows recovered land-cover boundaries under thin/thick aerial haze, clearer structure under cloud contamination on RICE2, and finer contour continuity on RSID.

4.4 Mechanism Comparison: Depth Fusion vs. Geometry-Conditioned Waves

Table 3 compares retrained DCMPNet and SSDG with controlled homogeneous-wave variants on RW^{2}AH and RSID. The baselines retain their respective checkpoints and protocols; all depth-using methods share test manifests and metrics.

On RW^{2}AH, the full model reaches 22.41 dB, 0.23 and 2.19 dB above the fixed-wave and zero-depth-velocity variants. On RSID, it reaches 27.09 dB, versus 25.28 / 23.94 dB for DCMPNet / SSDG and 26.82 / 26.43 dB for the wave variants. These results support geometry-conditioned driving and sample-wise kinematics under the reported protocols, but do not establish that every attention-fusion method is inferior on a matched backbone. Figure 6 shows representative high-frequency residuals: after Gaussian high-pass filtering ($\sigma = 1.5$), HF MSE for SSDG / DCMPNet / GeoHaze is 0.00976 / 0.00978 / 0.00249; the top 0.1% HF spectral-energy shares are 5.66% / 14.22% / 1.57%.

4.5 Ablation Study

Component ablations. Table 6 ablates depth encoding, HWPB, WDKE, and GFDB on RSID. Removing the depth encoder lowers PSNR from 27.09 to 26.37 dB. Removing HWPB or replacing it with local fusion yields 26.63 / 26.25 dB. Disabling depth-induced velocity or fixing (v, α) yields

Table 6: Architecture ablation on RSID (Chi et al., 2023). Variants remove or constrain the geometry encoder, HWPB, WDKE, and GFDB. Arrows indicate the preferred metric direction.

Variant	Depth Enc.	HWPB				GFDB		RSID (Chi et al., 2023)			
		Block	Wave Prop.	Depth Vel.	WDKE (v, α)	Block	Depth Gate	PSNR↑	SSIM↑	MSE↓	LPIPS↓
w/o depth encoder	×	✓	✓	×	✓	✓	×	26.37	0.9581	0.0031	0.0348
w/o HWPB	✓	×	×	×	×	✓	✓	26.63	0.9591	0.0029	0.0348
local HWPB	✓	✓	×	×	×	✓	✓	26.25	0.9573	0.0033	0.0358
zero depth-induced velocity	✓	✓	✓	×	✓	✓	✓	26.43	0.9585	0.0032	0.0347
fixed (v, α)	✓	✓	✓	✓	×	✓	✓	26.82	0.9598	0.0028	0.0336
w/o GFDB	✓	✓	✓	✓	✓	×	×	26.12	0.9575	0.0033	0.0371
GFDB w/o geometry gate	✓	✓	✓	✓	✓	✓	×	26.38	0.9581	0.0032	0.0355
Full model	✓	✓	✓	✓	✓	✓	✓	**27.09**	**0.9625**	**0.0027**	**0.0287**

Table 7: Loss ablation on RSID (Chi et al., 2023). The full objective combines multi-scale reconstruction, gradient consistency, and SSIM regularization; arrows indicate the preferred metric direction.

Variant	Loss setting				RSID (Chi et al., 2023)			
	L1	Multi-scale	Grad	SSIM	PSNR↑	SSIM↑	MSE↓	LPIPS↓
full-resolution L1	✓	×	×	×	26.20	0.9547	0.0032	0.0383
multi-scale L1	✓	✓	×	×	26.48	0.9560	0.0031	0.0373
multi-scale L1+SSIM	✓	✓	×	✓	26.53	0.9584	0.0031	0.0351
multi-scale L1+grad	✓	✓	✓	×	26.51	0.9574	0.0031	0.0363
Full objective	✓	✓	✓	✓	**27.09**	**0.9625**	**0.0027**	**0.0287**

26.43 / 26.82 dB. Appendix Fig. 7 visualizes single-HWPB damping and wave-speed controls with the initial state and drive fixed. Removing GFDB gives the largest architecture-side drop (26.12 dB); removing its geometry gate gives 26.38 dB. Table 7 shows 27.09 dB with multi-scale ℓ_1, gradient, and SSIM terms, versus 26.20 dB with full-resolution ℓ_1 alone. These ablations characterize module contributions without implying physical wave travel along object geometry.

Robustness under Depth Corruption. Table 4 corrupts only the test-time depth prior on RW^{2}AH while keeping trained checkpoints fixed. SSDG and GeoHaze share the clean-depth evaluation protocol; operator definitions appear in Appendix E.1. Across Gaussian noise, blur, and edge shift, the mean absolute PSNR drop D of GeoHaze is 0.0647, which is 63.94% lower than SSDG under the same three corruptions (Appendix Table 16). We interpret this as reduced sensitivity under the tested corruptions, not as robustness to arbitrary depth failure or as a proof that gates actively detect prior faults.

Downstream Object Detection on RTTS. We apply a fixed YOLOv8s (Varghese & M., 2024) detector to hazy RTTS (Li et al., 2019) images and to restorations from the RW^{2}AH-trained GeoHaze checkpoint. Relative to hazy inputs, GeoHaze raises all-class mAP$_{50}$ from 0.577 to 0.581, with gains on car/motorcycle/bus and decreases on person/bicycle (Table 5). The effect is therefore mixed class-wise with only a modest overall gain. PoolNet numbers in the same table are reported reference values and are not claimed under an identical detector/preprocessing protocol. Visual examples are in Appendix Fig. 9.

4.6 Failure Cases and Limitations

GeoHaze can fail when both RGB and depth cues are uninformative: large-area depth hallucination, near-uniform dense haze, thin repetitive textures, or ambiguous monocular depth in near-orthographic scenes. In such cases, the model cannot invent missing scene content.

5 Conclusion

GeoHaze uses monocular geometry to condition RGB propagation while retaining RGB as the appearance carrier. HWPB performs a gated damped-wave spectral update, WDKE predicts wave speed and damping, and GFDB separates layout and detail recovery. Across 11 natural-image and remote-sensing settings, GeoHaze achieves competitive restoration, supported by ablations and depth-corruption tests.

AI USE STATEMENT

In this work, we have not used generative AI tools for any of the tasks that require disclosure under the ICLR 2027 AI Policy for Authors, including generating synthetic datasets, developing theoretical models or conceptual frameworks, formulating mathematical claims, providing critical ingredients for proofs, writing proofs, proposing or refining research hypotheses, designing or providing feedback on research methodology or experiments, implementing the proposed methods, translating the manuscript, cleaning or reformatting datasets, supporting qualitative or thematic data analysis, or interpreting experimental results. Additionally, we used generative AI tools for recommended-disclosure tasks: drafting and editing manuscript text to improve readability, suggesting phrasing and section structure, assisting with LaTeX formatting and reference formatting, summarizing publicly available related work to support literature organization, and helping prepare analysis or visualization scripts. We have reviewed all AI-assisted work: AI-suggested text and references were checked against primary sources; quantitative claims and tables were verified against experimental logs; and any AI-assisted scripts were inspected and tested by the authors. We take responsibility for the final content of this work, including text, claims, and artifacts produced with the aid of generative AI.

ETHICS STATEMENT

This work studies structure-preserving image dehazing on publicly available natural-image and remote-sensing benchmarks (including RW^{2}AH, I-HAZE, O-HAZE, NH-HAZE, RRSHID, RICE1/RICE2, RSID, and SateHaze1k) and does not collect new human-subject data or personally identifiable information beyond what these datasets already contain. Monocular geometry is obtained from a publicly released foundation depth estimator (Depth Anything 3) applied to the same hazy inputs used for restoration; we do not claim calibrated scene metrology or privacy-preserving sensing. Dehazing and remote-sensing restoration methods can be misused to enhance deceptive imagery or to support secondary analysis of overhead scenes beyond the intended restoration setting; the reported results concern restoration fidelity and a limited downstream detection check under the specified protocols, and do not establish safety guarantees for surveillance, forensic, or security-critical applications. The authors adhered to the ICLR Code of Ethics and are responsible for the scientific claims, methodology, experiments, and presentation of this work.

REPRODUCIBILITY STATEMENT

To facilitate reproducibility, we describe datasets, metrics, training schedules, and evaluation protocols in Section 4, and detail the geometry-conditioned spectral operator, kinematics controls, and training objective in Section 3. Appendix A derives the spectral wave step and records operator visualizations; Appendix B lists network widths, Depth Anything 3 usage, optimizer settings, augmentation, seed protocol, and checkpoint selection; Appendix C audits split manifests and pair counts for all 11 evaluation settings. Additional remote-sensing tables, depth-corruption and RTTS detection protocols, and expanded qualitative examples appear in Appendices D–G. Anonymized source code and configuration files are provided in the supplementary materials to support replication of the reported results.

REFERENCES

Codruta O. Ancuti, Cosmin Ancuti, Radu Timofte, and Christophe De Vleeschouwer. O-HAZE: A dehazing benchmark with real hazy and haze-free outdoor images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pp. 754–762, 2018a.

Codruta O. Ancuti, Cosmin Ancuti, and Radu Timofte. NH-HAZE: An image dehazing benchmark with non-homogeneous hazy and haze-free images. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp. 1798–1805, 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 *Advanced Concepts for Intelligent Vision Systems*, volume 11182 of *Lecture Notes in Computer Science*, pp. 620–631. Springer, 2018b.

Zixuan Chen, Zewei He, and Zhe-Ming Lu. DEA-Net: Single image dehazing based on detail-enhanced convolution and content-guided attention. *IEEE Transactions on Image Processing*, 33: 1002–1015, 2024.

Kaichen Chi, Yuan Yuan, and Qi Wang. Trinity-Net: Gradient-guided Swin transformer-based remote sensing image dehazing and beyond. *IEEE Transactions on Geoscience and Remote Sensing*, 61:1–14, 2023. Art. no. 4702914.

Xiaofeng Cong, Jie Gui, Jing Zhang, Junming Hou, and Hao Shen. A semi-supervised nighttime dehazing baseline with spatial-frequency aware and realistic brightness constraint. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2631–2640, 2024.

Yuning Cui, Wenqi Ren, Xiaochun Cao, and Alois Knoll. Revitalizing convolutional network for image restoration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(12): 9423–9438, December 2024a.

Yuning Cui, Wenqi Ren, and Alois Knoll. Omni-Kernel network for image restoration. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 1426–1434, 2024b.

Yuning Cui, Wenqi Ren, and Alois Knoll. Exploring the potential of pooling techniques for universal image restoration. *IEEE Transactions on Image Processing*, 34:3403–3416, 2025a.

Yuning Cui, Qiang Wang, Chaopeng Li, Wenqi Ren, and Alois Knoll. EENet: An effective and efficient network for single image dehazing. *Pattern Recognition*, 158:111074, 2025b.

Yuning Cui, Syed Waqas Zamir, Salman Khan, Alois Knoll, Mubarak Shah, and Fahad Shahbaz Khan. AdaIR: Adaptive all-in-one image restoration via frequency mining and modulation. In Y. Yue, A. Garg, N. Peng, F. Sha, and R. Yu (eds.), *International Conference on Learning Representations*, volume 2025, pp. 101306–101327, 2025c.

Junkai Fan, Xiang Li, Jianjun Qian, Jun Li, and Jian Yang. Non-aligned supervision for real image dehazing. *IEEE Transactions on Circuits and Systems for Video Technology*, 35(11):10705–10715, 2025a.

Junkai Fan, Kun Wang, Zhiqiang Yan, Xiang Chen, Shangbing Gao, Jun Li, and Jian Yang. Depth-centric dehazing and depth-estimation from real-world hazy driving video. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 2852–2860, 2025b.

Wenxuan Fang, Junkai Fan, Yu Zheng, Jiangwei Weng, Ying Tai, and Jun Li. Guided real image dehazing using YCbCr color space. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 2906–2914, 2025.

Jie Gui, Xiaofeng Cong, Yuan Cao, Wenqi Ren, Jun Zhang, Jing Zhang, Jiuxin Cao, and Dacheng Tao. A comprehensive survey and taxonomy on single image dehazing based on deep learning. *ACM Computing Surveys*, 55(13s):1–37, December 2023. Article 279, 37 pages.

Hang Guo, Jinmin Li, Tao Dai, Zhihao Ouyang, Xudong Ren, and Shu-Tao Xia. MambaIR: A simple baseline for image restoration with state-space model. In *Computer Vision – ECCV 2024*, volume 15076 of *Lecture Notes in Computer Science*, pp. 222–241. Springer, 2024.

Kaiming He, Jian Sun, and Xiaoou Tang. Single image haze removal using dark channel prior. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1956–1963, 2009.

Binghui Huang, Zhi Li, Chao Yang, Fuchun Sun, and Yixu Song. Single satellite optical imagery dehazing using SAR image prior based on conditional generative adversarial networks. In *2020 IEEE Winter Conference on Applications of Computer Vision*, pp. 1795–1802, 2020.

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, January 2019.

Yudong Liang, Shaoji Li, De Cheng, Wenjian Wang, Deyu Li, and Jiye Liang. Image dehazing via self-supervised depth guidance. *Pattern Recognition*, 158:111051, 2025.

Ziyang Lihe, Jiang He, Qiangqiang Yuan, Xianyu Jin, Yi Xiao, and Liangpei Zhang. PhDnet: A novel physic-aware dehazing network for remote sensing images. *Information Fusion*, 106: 102277, 2024.

Daoyu Lin, Guangluan Xu, Xiaoke Wang, Yang Wang, Xian Sun, and Kun Fu. A remote sensing image dataset for cloud removal. *arXiv preprint arXiv:1901.00600*, 2019.

Haotong Lin, Sili Chen, Jun Hao Liew, Donny Y. Chen, Zhenyu Li, Yang Zhao, Sida Peng, Hengkai Guo, Xiaowei Zhou, Guang Shi, Jiashi Feng, and Bingyi Kang. Depth Anything 3: Recovering the visual space from any views. In C. Vondrick, B. Hariharan, C. Raffel, L. Pinto, D. Yang, and A. Faust (eds.), *International Conference on Learning Representations*, volume 2026, pp. 141261–141285, 2026.

Juping Liu, Shiju Wang, Xin Wang, Mingye Ju, and Dengyin Zhang. A review of remote sensing image dehazing. *Sensors*, 21(11):3926, 2021.

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*, pp. 7313–7322, 2019.

Srinivasa G. Narasimhan and Shree K. Nayar. Vision and the atmosphere. *International Journal of Computer Vision*, 48(3):233–254, 2002.

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 *2023 IEEE/CVF International Conference on Computer Vision*, pp. 12756–12767, 2023.

Hao Shen, Henghui Ding, Yulun Zhang, Zhong-Qiu Zhao, and Xudong Jiang. Spatial frequency modulation network for efficient image dehazing. *IEEE Transactions on Image Processing*, 34: 3982–3996, 2025.

Zishan Shu, Juntong Wu, Wei Yan, Xudong Liu, Hongyu Zhang, Chang Liu, Youdong Mao, and Jie Chen. WaveFormer: Frequency-time decoupled vision modeling with wave equation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 25428–25436, 2026.

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

Hang Sun, Zhiming Luo, Dong Ren, Wei Hu, Bo Du, Wen Yang, Jun Wan, and Lefei Zhang. Partial siamese with multiscale bi-codec networks for remote sensing image haze removal. *IEEE Transactions on Geoscience and Remote Sensing*, 61:1–16, 2023. Art. no. 4106516.

Hang Sun, Zhaoru Yao, Bo Du, Jun Wan, Dong Ren, and Lyuyang Tong. Spatial–frequency residual-guided dynamic perceptual network for remote sensing image haze removal. *IEEE Transactions on Geoscience and Remote Sensing*, 63:1–16, 2025. Art. no. 5612416.

Rejin Varghese and Sambath M. YOLOv8: A novel object detection algorithm with enhanced performance and robustness. In *2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems*, pp. 1–6, 2024.

Zhaozhi Wang, Yue Liu, Yunjie Tian, Yunfan Liu, Yaowei Wang, and Qixiang Ye. Building vision models upon heat conduction. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9707–9717, June 2025.

Zhou Wang, Alan C. Bovik, Hamid R. Sheikh, and Eero P. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4): 600–612, April 2004.

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.

Lihe Yang, Bingyi Kang, Zilong Huang, Xiaogang Xu, Jiashi Feng, and Hengshuang Zhao. Depth Anything: Unleashing the power of large-scale unlabeled data. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10371–10381, 2024a.

Lihe Yang, Bingyi Kang, Zilong Huang, Zhen Zhao, Xiaogang Xu, Jiashi Feng, and Hengshuang Zhao. Depth Anything V2. In A. Globerson, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomczak, and C. Zhang (eds.), *Advances in Neural Information Processing Systems*, volume 37, pp. 21875–21911. Curran Associates, Inc., 2024b.

Tian Ye, Sixiang Chen, Haoyu Chen, Wenhao Chai, Jingjing Ren, Zhaohu Xing, Wenxue Li, and Lei Zhu. PromptHaze: Prompting real-world dehazing via depth anything model. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 9454–9462, 2025.

Hongfei Zhang, Kun Zhou, Ruizheng Wu, and Jiangbo Lu. Can large pretrained depth estimation models help with image dehazing? *arXiv preprint arXiv:2508.00698*, 2025.

Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 586–595, June 2018.

Shengli Zhang, Zhiyong Tao, and Sen Lin. WaveletFormerNet: A transformer-based wavelet network for real-world non-homogeneous and dense fog removal. *Image and Vision Computing*, 146:105014, 2024a.

Yafei Zhang, Shen Zhou, and Huafeng Li. Depth information assisted collaborative mutual promotion network for single image dehazing. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2846–2855, 2024b.

Anqi Zhao, Ruitao Feng, and Xinghua Li. ThiefCloud: A thickness fused thin cloud removal network for optical remote sensing image with self-supervised learnable cloud prior. *IEEE Transactions on Circuits and Systems for Video Technology*, 35(12):11834–11848, December 2025.

Zhuoran Zheng, Wenqi Ren, Xiaochun Cao, Xiaobin Hu, Tao Wang, Fenglong Song, and Xiuyi Jia. Ultra-high-definition image dehazing via multi-guided bilateral learning. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 16180–16189, 2021.

Heng Zhou, Xiaoxiong Liu, Zhenxi Zhang, Jieheng Yun, Chengyang Li, Yunchu Yang, Dongyi Xia, Chunna Tian, and Xiao-Jun Wu. Remote sensing image dehazing: A systematic review of progress, challenges, and prospects. *ISPRS Journal of Photogrammetry and Remote Sensing*, 235: 318–348, 2026.

Zeng-Hui Zhu, Wei Lu, Si-Bao Chen, Chris H. Q. Ding, Jin Tang, and Bin Luo. Real-world remote sensing image dehazing: Benchmark and baseline. *IEEE Transactions on Geoscience and Remote Sensing*, 63:1–14, 2025.

Zengyuan Zuo, Junjun Jiang, Gang Wu, and Xianming Liu. UDPNet: Unleashing depth-based priors for robust image dehazing. *arXiv preprint arXiv:2601.06909*, 2026.

APPENDIX

This appendix provides the derivations, implementation details, evaluation protocols, and expanded visual evidence that support the main paper. Section A specifies the spectral operator, geometry-conditioned kinematics, and operator visualizations; Section B records the complete training and evaluation configuration; and Section C audits all dataset splits, beginning with the core evaluation inventory in Table 11. Section D restores the complete remote-sensing quantitative tables. Sections E and F describe the depth-corruption and downstream-detection protocols, respectively. Finally, Section G presents enlarged qualitative comparisons across the natural and remote-sensing benchmarks.

A SPECTRAL OPERATOR AND GEOMETRY GUIDANCE

A.1 MODAL DERIVATION

For one spatial Fourier mode with radial frequency $\kappa = (\omega_x^2 + \omega_y^2)^{1/2}$, the damped-wave model in the main paper reduces to

$$\ddot{\widehat{U}} + \alpha\dot{\widehat{U}} + v^2\kappa^2\widehat{U} = 0, \qquad \widehat{U}(0) = \widehat{U}_0, \quad \dot{\widehat{U}}(0) = \dot{\widehat{V}}_0. \tag{9}$$

Let

$$q = v^2\kappa^2 - (\alpha/2)^2, \qquad B = \dot{\widehat{V}}_0 + (\alpha/2)\widehat{U}_0. \tag{10}$$

The exact solution of Eq. (9) is

$$\widehat{U}(\tau) = e^{-\alpha\tau/2}\begin{cases} \widehat{U}_0\cos(\sqrt{q}\,\tau) + B\dfrac{\sin(\sqrt{q}\,\tau)}{\sqrt{q}}, & q > 0, \\ \widehat{U}_0 + B\tau, & q = 0, \\ \widehat{U}_0\cosh(\sqrt{-q}\,\tau) + B\dfrac{\sinh(\sqrt{-q}\,\tau)}{\sqrt{-q}}, & q < 0. \end{cases} \tag{11}$$

The three branches correspond to under-damped, critically damped, and over-damped modes, respectively.

A.2 PROJECTED SPECTRAL STEP USED BY HWPB

HWPB uses a single projected spectral filter. The continuous form is

$$\rho = \sqrt{[q]_+}, \qquad S_\tau(\rho) = \begin{cases} \sin(\rho\tau)/\rho, & \rho > 0, \\ \tau, & \rho = 0. \end{cases} \tag{12}$$

The implemented transfer functions are

$$H_0(\kappa) = e^{-\alpha\tau/2}\left[\cos(\rho\tau) + \frac{\alpha}{2}S_\tau(\rho)\right], \qquad H_1(\kappa) = e^{-\alpha\tau/2}S_\tau(\rho), \tag{13}$$

and the propagated feature is

$$U_\tau = \mathcal{F}^{-1}\left(H_0\widehat{U}_0 + H_1\dot{\widehat{V}}_0\right). \tag{14}$$

For $q > 0$, Eq. (14) equals the under-damped branch of Eq. (11). For $q \le 0$, the implementation uses an epsilon-regularized approximation to the critically damped continuation. The resulting HWPB operator is a PDE-inspired projected spectral filter rather than the over-damped branch of the exact solution.

A.3 DISCRETE REFERENCE REALIZATION

For a real feature map sampled at unit pixel spacing, the spectral step uses a two-dimensional discrete Fourier transform and its inverse. The horizontal axis retains the non-negative half spectrum, with frequencies

$$\omega_x(k) = \frac{2\pi k}{W}, \quad k = 0, \dots, \lfloor W/2 \rfloor, \qquad \omega_y(\ell) = \frac{2\pi\nu_H(\ell)}{H}, \quad \ell = 0, \dots, H-1, \tag{15}$$

Table 8: Tensor contract for one HWPB at stage s.

Tensor	**Shape / reduction**
F_r^s, U_0, G_r	$B \times C_s \times H_s \times W_s$
F_d^s	$B \times C_s^d \times H_s \times W_s$
$V_0, \widetilde{V}_0, G_r^0$	$B \times C_s \times H_s \times W_s$
Γ_d	$B \times 1 \times H_s \times W_s$, channel broadcast
$\widehat{U}_0, \widehat{\widetilde{V}}_0$	$B \times C_s \times H_s \times (\lfloor W_s/2 \rfloor + 1)$ complex
WDKE mean-absolute Haar magnitude	$B \times 1$, reduced over channel and space
v, α	$B \times 1 \times 1 \times 1$, broadcast
r_s	$1 \times C_s \times 1 \times 1$, initialized to 0.1
GFDB geometry gates	decoder-channel maps at sub-band resolution

where $\nu_H(\ell) = \ell$ for $\ell < \lceil H/2 \rceil$ and $\nu_H(\ell) = \ell - H$ otherwise. The vertical frequencies follow wrapped signed order, and the zero frequency remains at the origin. The parameters are $\tau = 1$ and $\epsilon = 10^{-6}$, with

$$\rho_\epsilon = \sqrt{\max(q, \epsilon)}, \qquad S(\rho_\epsilon) = \begin{cases} \tau, & \rho_\epsilon < 10^{-3}, \\ \sin(\rho_\epsilon \tau)/\rho_\epsilon, & \text{otherwise.} \end{cases} \tag{16}$$

The filter uses single precision for real values and for both components of complex values, even when the surrounding network uses automatic mixed precision. The forward transform is unscaled, and the inverse transform is scaled by $1/(HW)$. The inverse transform restores the original (H, W), so odd widths are reconstructed without ambiguity.

A.4 Geometry-Conditioned Kinematics

Let $F_r, F_d \in \mathbb{R}^{B \times C \times H \times W}$ be the aligned RGB and depth-stream features after layer normalization. Depth enters HWPB through the initial velocity and the sample-wise kinematic parameters:

$$\begin{aligned} U_0, G_r^0 = \mathrm{split}(\phi_u(\mathrm{DWConv}(F_r)))\,, \quad V_0 = \phi_v(F_d), \quad \Gamma_d = \sigma(\phi_g(F_d)), \quad \widetilde{V}_0 = \Gamma_d \odot V_0, \\ v = v_{\min} + (v_{\max} - v_{\min})\sigma(a_v), \quad \alpha = \alpha_{\min} + (\alpha_{\max} - \alpha_{\min})\sigma(a_\alpha). \end{aligned} \tag{17}$$

Here $\Gamma_d \in \mathbb{R}^{B \times 1 \times H \times W}$ controls the depth-derived driving term and broadcasts over channels, allowing depth discontinuities to spatially modulate the driving velocity. The executed bounds are $(v_{\min}, v_{\max}) = (0.5, 4.0)$ and $(\alpha_{\min}, \alpha_{\max}) = (0.01, 1.0)$. The RGB-derived residual admission map $G_r = \mathrm{SiLU}(G_r^0)$ acts after propagation:

$$F_r' = F_r + r_s \odot (\phi_o(\mathrm{LN}(U_\tau)) \odot G_r)\,, \tag{18}$$

where $r_s \in \mathbb{R}^{1 \times C \times 1 \times 1}$ is learned per channel and initialized to 0.1. This decomposition separates four roles: U_0 carries RGB appearance, $\widetilde{V}_0$ injects geometry-conditioned motion, (v, α) controls the spectral response, and G_r controls the returned residual.

Table 8 fixes the tensor contract at an interaction stage. The depth projections map the narrower depth stream to the RGB-stage channel count before modulation. The parameters v and α are predicted once per sample and HWPB instance, then broadcast over channel and frequency.

A.5 Wavelet Statistics and Frequency Decoding

For a Haar sub-band X_b, $b \in \{LL, LH, HL, HH\}$, WDKE defines $e_b = \mathrm{mean}_{C,H,W} |X_b|$ as its mean-absolute magnitude, not its squared energy. Each of the two decomposition levels records

$$s(X) = \big[e_{LL}, e_{LH}, e_{HL}, e_{HH}, (e_{LH} + e_{HL} + e_{HH})/(e_{LL} + \epsilon)\big]. \tag{19}$$

Each level contributes five RGB and five depth values. Concatenating two levels gives a 20-dimensional vector per sample, which an MLP $20 \to 64 \to 32 \to 2$ with SiLU hidden activations maps to (a_v, a_α). Consequently, WDKE predicts sample-wise wave speed and adaptively adjusts the overall damping coefficient from image-level RGB–depth statistics. The Haar transform uses

the orthonormal $1/2$ analysis and synthesis factors. Odd height or width is padded by one pixel on the bottom or right using reflection, with replication only for a degenerate dimension; the inverse transform crops to the recorded input size.

GFDB applies a one-level Haar transform in the decoder. Its low-frequency depth multiplier $1+\sigma(\cdot)$ lies in $(1, 2)$; its high-frequency geometry gate $\sigma(\cdot)$ lies in $(0, 1)$ and is shared across the three detail bands $\{LH, HL, HH\}$. The inverse Haar transform reconstructs the frequency-decoupled decoder feature. The high-frequency edge residual has a learned scalar coefficient β initialized to 0.5.

A.6 Operator Visualizations

Figures 3, 7, and 8 provide diagnostic views of the geometry-conditioned wave step used in the main paper. They illustrate feature-state trajectories under controlled initial drives, single-HWPB damping and wave-speed controls, and the internal activations of the first HWPB on an RW^{2}AH crop. Quantitative contribution tests appear in the main-paper mechanism baselines.

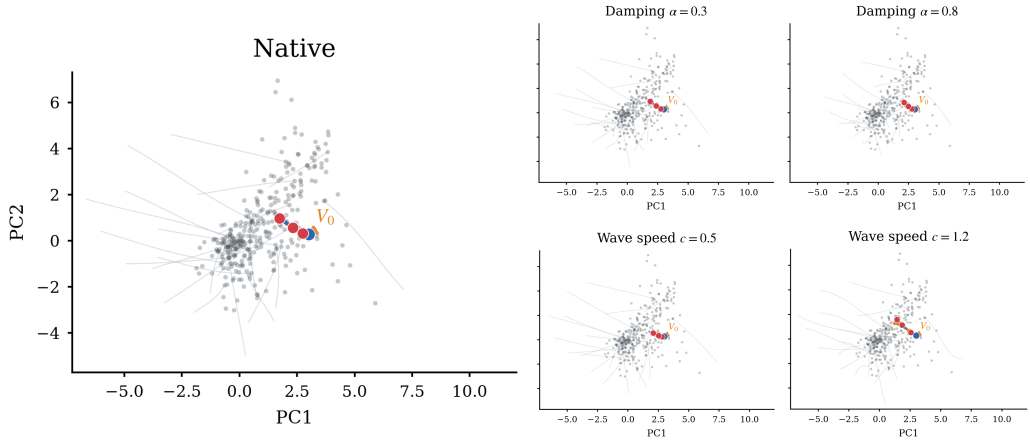

Figure 7: Single-HWPB controls in PCA space. The native response (left, enlarged) is compared with damping and wave-speed changes (right); U_0 and V_0 remain fixed.

B Reproducible Experimental Configuration

B.1 Network and Depth Prior

The hazy RGB image is the appearance input. Depth Anything 3 (Lin et al., 2026) estimates a relative-depth map from the same image. Each precomputed map has one channel and shares the corresponding RGB image's identifier and pixel grid. Depth and RGB dimensions must match, and samples follow image-identifier order. The cached depth values are used without per-image min–max normalization. All depth-assisted comparisons use the same precomputed maps. Table 9 records the executed restoration architecture.

B.2 Optimization and Objective

Training uses the two-stage schedule in Table 10. Each trainable configuration is run independently with seeds 42, 43, 44, 45, 46, and 47. Both stages use AdamW with $(\beta_1, \beta_2) = (0.9, 0.99)$ and weight decay 10^{-4}, AMP, gradient clipping at 1.0, and EMA decay 0.999. Stage 2 loads the final Stage 1 EMA parameters from the corresponding run. All quantitative tables report the arithmetic mean over these six runs.

Horizontal flips and 90° rotations are enabled in both stages. The Stage 1 brightness/contrast/saturation/hue amplitudes are $0.1/0.1/0.1/0.02$, and Stage 2 uses $0.05/0.05/0.05/0.01$. Depth augmentation applies scale, shift, additive noise, element dropout,

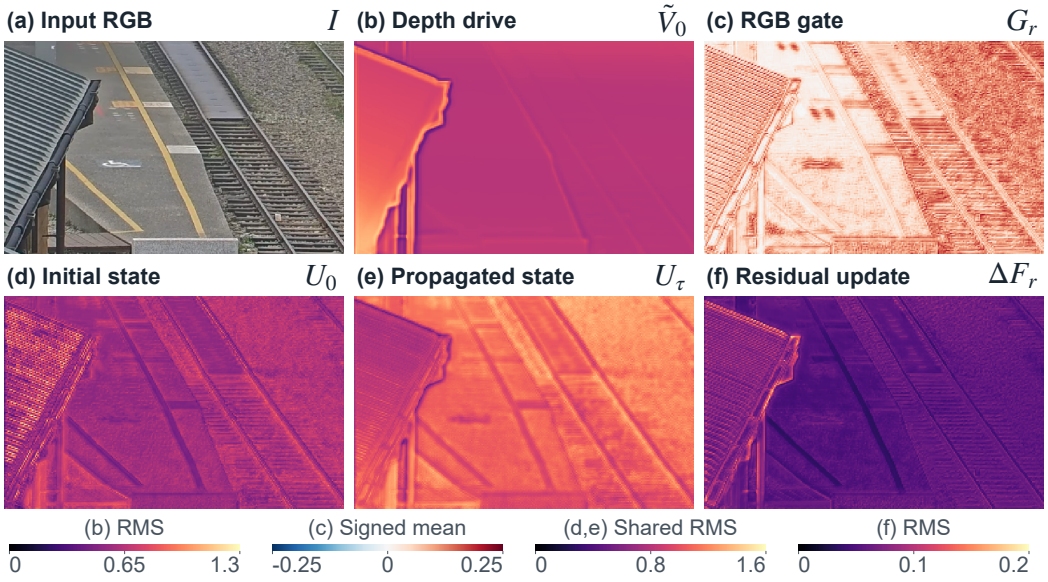

Figure 8: **Geometry-conditioned RGB feature updates.** Activations from the first HWPB on RW^{2}AH, with unmodified estimated depth. Top row: (a) RGB input, (b) gated depth-derived drive, and (c) RGB gate. Bottom row: (d) RGB initial state, (e) raw propagated state before normalization and projection, and (f) feature residual after normalization, projection, RGB modulation, and learned scaling. All panels use the same spatial crop. Drive, states, and residual are channel RMS; G_r is a signed channel mean. Panels (d,e) share a linear scale. These activations illustrate a normal forward pass, not propagation trajectories or causal ablations.

Table 9: Network configuration used in all GeoHaze experiments.

Component	**Configuration**
Base feature dimension	32
RGB encoder depths	$[4, 4, 6, 8]$
Depth encoder depths	$[2, 2, 3, 3]$
Decoder depths	$[8, 6, 4, 4]$
HWPB counts	$[1, 1, 2]$ at interaction stages
Depth estimator	Depth Anything 3
Depth input	Fixed aligned depth; no per-image normalization
Wave parameters	$\tau = 1$, $\epsilon = 10^{-6}$, ρ threshold 10^{-3}
Kinematic ranges	$v \in [0.5, 4.0]$, $\alpha \in [0.01, 1.0]$
WDKE head	$20 \to 64 \to 32 \to 2$, SiLU
Residual / GFDB initialization	$r_s = 0.1$ per channel; $\beta = 0.5$
Training outputs	Three restoration scales
Restoration model size	11.49M parameters
Restoration computation	57.64G FLOPs

and optional 3×3 blur before clipping to $[0, 1]$. The respective Stage 1 amplitudes or probabilities are 0.05, 0.02, 0.01, 0.05, and 0.10; Stage 2 uses 0.02, 0.01, 0.005, 0.02, and 0.05.

Let $r(i) = \min(1, i/T_w)$, where T_w is the loss warm-up in Table 10. The executed objective is

$$\mathcal{L} = \|\hat{J} - J\|_1 + 0.5\|\hat{J}_{1/2} - J_{1/2}\|_1 + 0.25\|\hat{J}_{1/4} - J_{1/4}\|_1 + r(i)\left[\|\nabla\hat{J} - \nabla J\|_1 + 0.5(1 - \mathrm{SSIM}(\hat{J}, J))\right]. \tag{20}$$

The half- and quarter-scale targets use bilinear interpolation with pixel-center alignment. The gradient term concatenates forward horizontal and vertical RGB differences and zero-pads the terminal column and row. Training computes SSIM using the pytorch-msssim implementation with unit data range, three channels, and batch-averaged output.

Table 10: Executed two-stage training schedule. Batch and effective batch are per GPU.

Configuration	Stage 1	Stage 2
Crop / batch	256^2 / 10	384^2 / 6
Accumulation / effective batch	2 / 20	2 / 12
Initial / minimum LR	2.5×10^{-4} / 10^{-6}	5×10^{-5} / 10^{-6}
Configured iterations	200k	40k
Scheduler warm-up	1k	1k
Loss warm-up	20k	5k
Evaluation / checkpoint interval	5k / 5k	2k / 2k
Color-jitter probability	0.5	0.2
Depth blur probability	0.10	0.05

All datasets follow the official training, validation, and test protocols of their source publications. In particular, RW^{2}AH uses its official validation and test splits without repartitioning or merging. Checkpoints are saved at the intervals in Table 10, but neither validation nor test metrics are used to select among them. Each run uses the predefined final checkpoint: the final Stage 1 EMA initializes Stage 2, and the final Stage 2 EMA is fixed before test evaluation. The official test split is evaluated only after training is complete.

B.3 Metric Definitions and Aggregation

Inference reflect-pads RGB and depth on the bottom and right to a multiple of 16, using replication only when reflection is undefined. The outputs are cropped back to the native size and clamped to $[0, 1]$; no tiling, resizing, or test-time augmentation is applied. All paired metrics operate on RGB images with a one-pixel border crop and no Y-channel conversion. For an image with N scalar RGB samples,

$$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N}(\hat{J}_i - J_i)^2, \qquad \mathrm{PSNR} = 10\log_{10}\frac{1}{\mathrm{MSE}}. \tag{21}$$

SSIM uses the registered BasicSR RGB evaluator. LPIPS uses the official implementation with AlexNet, version 0.1. RGB values are converted from $[0, 255]$ to $[0, 1]$ and then to $[-1, 1]$ before evaluation, with no additional input normalization. Each metric is computed per image and then averaged over the complete test manifest. Consequently, mean PSNR is not obtained by applying $-10\log_{10}(\cdot)$ to the dataset-level mean MSE. Tables report PSNR in dB, SSIM and LPIPS as unitless scores, and MSE on the $[0, 1]$ scale. Higher PSNR and SSIM are better; lower MSE and LPIPS are better.

C Dataset Split Audit

Table 11 lists the 11 real-haze and remote-sensing datasets used in the main evaluation. The remainder of this section records the official split protocols and pair counts that underlie those results.

C.1 Natural Real-Haze Benchmarks

RW^{2}AH contains 1,758 real hazy/clear pairs collected from stationary online webcams across twelve countries (Fang et al., 2025). We retain the official training, validation, and test splits defined by the source publication. I-HAZE, O-HAZE, and NH-HAZE are controlled real-haze benchmarks captured with fixed camera and illumination settings (Ancuti et al., 2018b;a; 2020). Table 12 records the split protocol for every natural-scene benchmark.

C.2 Remote-Sensing Benchmarks

RRSHID contains 3,053 geospatially aligned pairs divided by haze density (Zhu et al., 2025). RICE1 and RICE2 contain cloud/clear remote-sensing pairs (Lin et al., 2019); this work follows the processed RICE protocol used by RRSHID. RSID is the synthetic paired benchmark released

Table 11: Core real-haze and remote-sensing datasets used in the main evaluation.

Role	Dataset
paired real haze	RW^2AH (Fang et al., 2025)
indoor real haze	I-HAZE (Ancuti et al., 2018b)
outdoor real haze	O-HAZE (Ancuti et al., 2018a)
non-homogeneous real haze	NH-HAZE (Ancuti et al., 2020)
aerial thin haze	RRSHID-thin (Zhu et al., 2025)
aerial moderate haze	RRSHID-moderate (Zhu et al., 2025)
aerial thick haze	RRSHID-thick (Zhu et al., 2025)
thin cloud removal	RICE1 (Lin et al., 2019)
thick cloud removal	RICE2 (Lin et al., 2019)
synthetic RS haze	RSID (Chi et al., 2023)
thick RS haze	SateHaze1k-thick (Huang et al., 2020)

Table 12: Natural real-haze split protocols and counts. "Varied" denotes native webcam resolutions.

Dataset	Scene/degradation	Pairs	Resolution	Train	Val.	Test	Evaluation role
RW^2AH	Real webcam haze	1,758	Varied	Official train/validation/test splits			Primary paired real-haze benchmark
I-HAZE	Indoor homogeneous haze	35	5456×3632	25	5	5	Indoor real-haze evaluation
O-HAZE	Outdoor homogeneous haze	45	5456×3632	35	5	5	Outdoor real-haze evaluation
NH-HAZE	Outdoor non-homogeneous haze	55	5456×3632	45	5	5	Spatially varying real haze

Table 13: Remote-sensing split counts. RICE2 has 736 released entries; the published $588/147$ protocol defines 735 train/test pairs used by the retrained depth-assisted runs.

Dataset	Data type	Release	Resolution	Train	Val.	Test	Evaluated condition
RRSHID-thin	Real optical pair	763	256^2	610	77	76	Thin satellite haze
RRSHID-moderate	Real optical pair	1,526	256^2	1,220	154	152	Moderate satellite haze
RRSHID-thick	Real optical pair	764	256^2	611	77	76	Thick satellite haze
RICE1	Real cloud/clear pair	500	512^2	400	–	100	Cloud-layer removal
RICE2	Processed cloud/clear pair	736	512^2	588	–	147	Thick cloud and shadow
RSID	Synthetic haze/clear pair	1,000	256^2	900	–	100	Synthetic satellite haze
SateHaze1k-thick	Synthetic RGB/SAR/GT	400	512^2	320	35	45	Thick satellite haze

with Trinity-Net (Chi et al., 2023). SateHaze1k contains aligned hazy optical RGB, clear RGB, and SAR triplets (Huang et al., 2020); only its thick subset is evaluated.

The original RICE paper describes RICE2 as 450 image triplets (Lin et al., 2019). The later processed comparison protocol reports a 736-entry release and a $588/147$ train/test manifest (Zhu et al., 2025). The retrained SSDG, DCMPNet, and GeoHaze runs use the latter 147-pair test manifest.



Figure 9: Downstream object detection on RTTS (Li et al., 2019). The top and bottom rows show hazy inputs and restored outputs evaluated by the same YOLOv8s (Varghese & M., 2024) detector, showing that structure-preserving restoration reveals clearer vehicle boundaries under dense haze.

D Complete Remote-Sensing Quantitative Tables

The main paper reports a compact four-dataset remote-sensing comparison. Tables 14 and 15 restore the complete RRSHID density splits and the remaining RICE/SateHaze1k benchmarks, including complexity counts for RRSHID.

E Mechanism and Robustness Protocols

E.1 Depth-Corruption Operators

The trained checkpoints remain fixed. Only the aligned test-time depth map is changed. Let P_1 and P_{99} be its 1st and 99th percentiles, and let $s = \min(H, W)$. Gaussian noise adds zero-mean noise with $\sigma = 0.20(P_{99} - P_1)$. Gaussian blur uses $\sigma = 0.05s$. Edge shift uses grayscale dilation with an elliptical structuring element of radius $0.05s$. Clean depth uses the original aligned prior.

For the three corruptions, define mean absolute sensitivity $\bar{D} = \frac{1}{3}\sum_c |D_c|$, worst-case sensitivity $D_{\max} = \max_c |D_c|$, and normalized sensitivity $100\bar{D}/\mathrm{PSNR}_{\mathrm{clean}}$. The aggregate rows in Table 16 are derived from the retained unrounded values for the three corruption settings. The mean absolute PSNR sensitivity of GeoHaze is 63.94% lower than SSDG under this corruption suite.

F RTTS Downstream Detection Protocol

RTTS contains 4,322 real hazy traffic images and 41,203 boxes for person, bicycle, car, motorcycle, and bus (Li et al., 2019). The RW^{2}AH-trained GeoHaze checkpoint restores every RTTS image. A fixed YOLOv8s detector then evaluates the original and restored image sets with identical weights, image size, thresholds, and annotations. For precision or mAP_{50},

$$\Delta = 100(s_{\mathrm{restored}} - s_{\mathrm{hazy}}) \tag{22}$$

is reported in percentage points.

RTTS has no paired clear reference, so its evidence is detection-only; paired restoration metrics are not computed for this dataset.

Table 14: Quantitative comparison on RRSHID (Zhu et al., 2025). Higher is better for PSNR/SSIM; lower is better for LPIPS/MSE. The best, second-best, and third-best results for each metric are highlighted in bold brown-gold, blue, and green, respectively. These complete remote-sensing tables complement the compact four-dataset comparison in the main paper.

(a) RRSHID-thin (Zhu et al., 2025)

Method	Venue	Param. (M)	FLOPs (G)	PSNR↑	SSIM↑	LPIPS↓	MSE↓
DCP (He et al., 2009)	CVPR'09	–	**0.60**	18.46	0.4564	0.4851	0.0192
GridDehazeNet (Liu et al., 2019)	ICCV'19	0.96	85.72	22.77	0.6145	0.4123	0.0069
FFA-Net (Qin et al., 2020)	AAAI'20	4.46	287.80	17.08	0.4452	0.5761	0.0326
4KDehazing (Zheng et al., 2021)	CVPR'21	34.55	105.80	22.83	0.6177	0.4352	0.0063
DehazeFormer (Song et al., 2023)	TIP'23	1.21	39.76	22.74	0.6005	0.4609	0.0071
OKNet (Cui et al., 2024b)	AAAI'24	4.72	39.67	22.20	0.5916	0.4797	0.0081
DCMPNet (Zhang et al., 2024b)	CVPR'24	7.16	62.89	21.75	0.4100	0.7106	0.0080
EENet (Cui et al., 2025b)	PR'25	5.44	49.83	22.34	0.6006	0.4791	0.0078
PoolNet (Cui et al., 2025a)	TIP'25	6.92	66.38	23.80	0.6584	0.3346	0.0054
SFMN (Shen et al., 2025)	TIP'25	3.43	36.57	23.64	0.6587	0.4146	0.0062
SSDG (Liang et al., 2025)	PR'25	45.02	60.60	21.79	0.5731	**0.3019**	0.0083
WaveFormer (Shu et al., 2026)	AAAI'26	43.54	118.40	21.31	0.5255	0.3794	0.0103
Trinity-Net (Chi et al., 2023)	TGRS'23	20.14	30.78	20.51	0.5728	0.4578	0.0120
PhDnet (Lihe et al., 2024)	InfoFusion'24	10.03	33.24	22.50	0.5949	0.4558	0.0074
SFRDP-Net (Sun et al., 2025)	TGRS'25	3.51	51.37	23.02	0.6225	0.4708	0.0064
MCAF-Net (Zhu et al., 2025)	TGRS'25	**0.56**	19.82	23.32	0.6236	0.4023	0.0059
ThiefCloud (Zhao et al., 2025)	TCSVT'25	2.06	40.83	23.01	0.6144	0.3592	0.0066
GeoHaze (ours)	–	11.49	57.64	**24.43**	**0.6926**	0.3539	**0.0050**

(b) RRSHID-moderate (Zhu et al., 2025)

Method	Venue	Param. (M)	FLOPs (G)	PSNR↑	SSIM↑	LPIPS↓	MSE↓
DCP (He et al., 2009)	CVPR'09	–	**0.60**	17.80	0.4856	0.4700	0.0238
GridDehazeNet (Liu et al., 2019)	ICCV'19	0.96	85.72	22.62	0.6468	0.3833	0.0083
FFA-Net (Qin et al., 2020)	AAAI'20	4.46	287.80	17.40	0.5385	0.5450	0.0346
4KDehazing (Zheng et al., 2021)	CVPR'21	34.55	105.80	22.47	0.6505	0.4590	0.0083
DehazeFormer (Song et al., 2023)	TIP'23	1.21	39.76	23.06	0.6137	0.4495	0.0076
OKNet (Cui et al., 2024b)	AAAI'24	4.72	39.67	22.75	0.6519	0.4680	0.0080
DCMPNet (Zhang et al., 2024b)	CVPR'24	7.16	62.89	22.01	0.4707	0.5908	0.0086
EENet (Cui et al., 2025b)	PR'25	5.44	49.83	22.74	0.6542	0.4634	0.0081
PoolNet (Cui et al., 2025a)	TIP'25	6.92	66.38	24.09	0.6724	0.3632	0.0059
SFMN (Shen et al., 2025)	TIP'25	3.43	36.57	24.33	0.6926	0.4254	0.0057
SSDG (Liang et al., 2025)	PR'25	45.02	60.60	22.42	0.5949	**0.3107**	0.0082
WaveFormer (Shu et al., 2026)	AAAI'26	43.54	118.40	21.28	0.5846	0.4402	0.0110
Trinity-Net (Chi et al., 2023)	TGRS'23	20.14	30.78	22.46	0.5728	0.4314	0.0085
PhDnet (Lihe et al., 2024)	InfoFusion'24	10.03	33.24	23.12	0.6565	0.4296	0.0073
SFRDP-Net (Sun et al., 2025)	TGRS'25	3.51	51.37	23.74	0.6735	0.4383	0.0062
MCAF-Net (Zhu et al., 2025)	TGRS'25	**0.56**	19.82	23.60	0.6583	0.3799	0.0063
ThiefCloud (Zhao et al., 2025)	TCSVT'25	2.06	40.83	23.45	0.6562	0.3475	0.0063
GeoHaze (ours)	–	11.49	57.64	**24.67**	**0.7031**	0.3856	**0.0054**

(c) RRSHID-thick (Zhu et al., 2025)

Method	Venue	Param. (M)	FLOPs (G)	PSNR↑	SSIM↑	LPIPS↓	MSE↓
DCP (He et al., 2009)	CVPR'09	–	**0.60**	18.39	0.4843	0.4996	0.0208
GridDehazeNet (Liu et al., 2019)	ICCV'19	0.96	85.72	23.96	0.7112	0.3947	0.0061
FFA-Net (Qin et al., 2020)	AAAI'20	4.46	287.80	16.71	0.4792	0.5573	0.0377
4KDehazing (Zheng et al., 2021)	CVPR'21	34.55	105.80	22.55	0.6912	0.4754	0.0099
DehazeFormer (Song et al., 2023)	TIP'23	1.21	39.76	24.69	0.7143	0.4377	0.0051
OKNet (Cui et al., 2024b)	AAAI'24	4.72	39.67	23.66	0.6968	0.4782	0.0064
DCMPNet (Zhang et al., 2024b)	CVPR'24	7.16	62.89	23.27	0.5270	0.5427	0.0062
EENet (Cui et al., 2025b)	PR'25	5.44	49.83	23.98	0.7068	0.4752	0.0060
PoolNet (Cui et al., 2025a)	TIP'25	6.92	66.38	25.07	0.7291	0.3504	0.0049
SFMN (Shen et al., 2025)	TIP'25	3.43	36.57	25.09	0.7331	0.4601	0.0052
SSDG (Liang et al., 2025)	PR'25	45.02	60.60	22.94	0.6490	**0.2972**	0.0065
WaveFormer (Shu et al., 2026)	AAAI'26	43.54	118.40	22.56	0.6268	0.4357	0.0075
Trinity-Net (Chi et al., 2023)	TGRS'23	20.14	30.78	24.11	0.7234	0.4206	0.0058
PhDnet (Lihe et al., 2024)	InfoFusion'24	10.03	33.24	24.64	0.7203	0.4563	0.0054
SFRDP-Net (Sun et al., 2025)	TGRS'25	3.51	51.37	25.01	0.7312	0.4318	0.0047
MCAF-Net (Zhu et al., 2025)	TGRS'25	**0.56**	19.82	25.40	0.7221	0.3942	**0.0040**
ThiefCloud (Zhao et al., 2025)	TCSVT'25	2.06	40.83	25.05	0.7187	0.3610	0.0041
GeoHaze (ours)	–	11.49	57.64	**25.52**	**0.7519**	0.4120	0.0046

Table 15: Quantitative comparison on RICE1 (Lin et al., 2019), RICE2 (Lin et al., 2019), RSID (Chi et al., 2023), and SateHaze1k-thick (Huang et al., 2020). Higher is better for PSNR/SSIM; lower is better for LPIPS/MSE. The best, second-best, and third-best results for each metric are highlighted in bold brown-gold, blue, and green, respectively. These complete remote-sensing tables complement the compact four-dataset comparison in the main paper.

Method	Venue	RICE1 (Lin et al., 2019)				RICE2 (Lin et al., 2019)			
		PSNR↑	SSIM↑	LPIPS↓	MSE↓	PSNR↑	SSIM↑	LPIPS↓	MSE↓
DCP (He et al., 2009)	CVPR'09	15.68	0.6860	0.3766	0.0491	14.10	0.3898	0.6234	0.0656
FFA-Net (Qin et al., 2020)	AAAI'20	23.73	0.9068	0.1885	0.0096	17.77	0.6261	0.5378	0.0323
4KDehazing (Zheng et al., 2021)	CVPR'21	27.54	0.9425	0.0967	0.0024	25.21	0.8604	0.3704	0.0087
DehazeFormer (Song et al., 2023)	TIP'23	36.15	**0.9794**	0.0460	0.0005	34.54	0.8875	0.2035	0.0010
ConvIR (Cui et al., 2024a)	TPAMI'24	34.12	0.9555	0.0460	0.0008	34.12	0.9045	0.2022	0.0009
OKNet (Cui et al., 2024b)	AAAI'24	36.54	0.9629	0.0576	**0.0004**	29.78	0.8714	0.2779	0.0018
DCMPNet (Zhang et al., 2024b)	CVPR'24	31.38	0.9207	0.0797	0.0012	31.67	0.8101	0.2750	0.0015
EENet (Cui et al., 2025b)	PR'25	**36.93**	0.9636	0.0571	**0.0004**	31.48	0.8852	0.2562	0.0014
PoolNet (Cui et al., 2025a)	TIP'25	34.18	0.9575	0.0582	0.0010	36.57	0.9169	0.1754	**0.0007**
SFMN (Shen et al., 2025)	TIP'25	35.48	0.9607	0.0669	0.0006	**36.61**	0.9142	0.1815	**0.0007**
SSDG (Liang et al., 2025)	PR'25	33.78	0.9485	0.0503	0.0008	33.71	0.8600	0.1931	0.0013
WaveFormer (Shu et al., 2026)	AAAI'26	31.48	0.9107	0.1562	0.0015	32.12	0.8331	0.3130	0.0011
Trinity-Net (Chi et al., 2023)	TGRS'23	23.46	0.8796	0.0512	0.0100	18.81	0.7781	0.2287	0.0200
PSMB-Net (Sun et al., 2023)	TGRS'23	31.32	0.9444	0.0428	0.0006	30.21	0.8756	0.1983	0.0010
PhDnet (Lihe et al., 2024)	InfoFusion'24	35.81	0.9545	0.0511	**0.0004**	34.52	0.8836	0.2604	0.0010
SFRDP-Net (Sun et al., 2025)	TGRS'25	31.60	0.9420	0.0555	0.0021	34.33	0.9011	0.2041	0.0011
ThiefCloud (Zhao et al., 2025)	TCSVT'25	35.45	0.9551	**0.0392**	0.0006	34.91	0.9000	0.1786	0.0009
GeoHaze (ours)	–	36.67	0.9643	0.0591	0.0007	36.48	**0.9254**	**0.1709**	0.0009

Method	Venue	RSID (Chi et al., 2023)				SateHaze1k-thick (Huang et al., 2020)			
		PSNR↑	SSIM↑	LPIPS↓	MSE↓	PSNR↑	SSIM↑	LPIPS↓	MSE↓
DCP (He et al., 2009)	CVPR'09	13.87	0.6892	0.2615	0.0538	17.87	0.8481	0.1879	0.0169
FFA-Net (Qin et al., 2020)	AAAI'20	18.31	0.8582	0.1449	0.0204	19.45	0.9023	0.2357	0.0117
4KDehazing (Zheng et al., 2021)	CVPR'21	23.61	0.9415	0.0693	0.0053	20.75	0.7696	0.1674	0.0085
DehazeFormer (Song et al., 2023)	TIP'23	25.04	0.9393	0.0557	0.0040	22.02	**0.9306**	0.1707	0.0063
ConvIR (Cui et al., 2024a)	TPAMI'24	25.33	0.9512	0.0398	0.0038	19.92	0.8151	0.1766	0.0103
OKNet (Cui et al., 2024b)	AAAI'24	25.72	0.9342	0.0369	0.0035	21.15	0.8291	0.1812	0.0078
DCMPNet (Zhang et al., 2024b)	CVPR'24	25.28	0.9402	0.0471	0.0038	20.45	0.8377	0.1608	0.0091
EENet (Cui et al., 2025b)	PR'25	23.89	0.9376	0.0695	0.0052	**22.42**	0.8422	0.1630	**0.0058**
PoolNet (Cui et al., 2025a)	TIP'25	25.88	0.9522	0.0363	0.0035	21.94	0.8339	0.1702	0.0065
SFMN (Shen et al., 2025)	TIP'25	26.82	0.9583	0.0331	0.0030	21.24	0.8281	0.1786	0.0076
SSDG (Liang et al., 2025)	PR'25	23.94	0.9247	0.0504	0.0066	20.16	0.8297	0.1612	0.0097
WaveFormer (Shu et al., 2026)	AAAI'26	22.50	0.8434	0.2092	0.0070	16.77	0.5901	0.4493	0.0213
Trinity-Net (Chi et al., 2023)	TGRS'23	23.60	0.9322	0.0692	0.0060	20.43	0.8056	0.1764	0.0086
PSMB-Net (Sun et al., 2023)	TGRS'23	25.64	0.9447	0.0537	0.0030	21.55	0.8489	0.1642	0.0084
PhDnet (Lihe et al., 2024)	InfoFusion'24	23.77	0.9179	0.0865	0.0054	21.69	0.8255	0.1788	0.0069
SFRDP-Net (Sun et al., 2025)	TGRS'25	26.60	0.9562	0.0349	0.0031	21.64	0.8289	0.1696	0.0070
ThiefCloud (Zhao et al., 2025)	TCSVT'25	25.13	0.9507	0.0390	0.0043	20.63	0.8100	0.1751	0.0087
GeoHaze (ours)	–	**27.09**	**0.9625**	**0.0287**	**0.0027**	22.07	0.8517	**0.1420**	0.0065

Table 16: RW^{2}AH test-time depth-corruption results and aggregate sensitivity. $|D|$ is the absolute PSNR drop from each method's own clean-depth result. Displayed PSNR is rounded to two decimals; $|D|$ and aggregate values are computed from the evaluator's retained, unrounded PSNR.

Depth setting	Method	PSNR↑	SSIM↑	LPIPS↓	$\lvert D\rvert$↓
Clean	SSDG	17.93	0.6073	0.4745	0.0000
	GeoHaze	22.41	0.6610	0.3011	0.0000
Gaussian noise	SSDG	17.61	0.5674	0.5407	0.3233
	GeoHaze	22.38	0.6600	0.3016	**0.0260**
Gaussian blur	SSDG	17.83	0.6070	0.4641	0.1015
	GeoHaze	22.33	0.6591	0.3009	**0.0783**
Edge shift	SSDG	17.82	0.6043	0.4673	0.1137
	GeoHaze	22.32	0.6601	0.3008	**0.0899**

Aggregate sensitivity	$\bar{D}$ / $D_{\max}$	**Normalized** $\bar{D}$
SSDG	0.1795 / 0.3233	1.0009%
GeoHaze	**0.0647 / 0.0899**	**0.2889%**

Table 17: Class-wise RTTS detection with the same YOLOv8s detector before and after restoration.

Class	Precision			mAP_{50}		
	Hazy	Restored	Δ	Hazy	Restored	Δ
person	0.8000	0.7990	−0.10	0.754	0.745	−0.90
bicycle	0.5540	0.5770	+2.30	0.518	0.513	−0.50
car	0.8350	0.8240	−1.10	0.700	0.707	+0.70
motorcycle	0.6510	0.6820	+3.10	0.489	0.512	+2.30
bus	0.6410	0.6680	+2.70	0.425	0.427	+0.20
all	0.6960	0.7100	+1.40	0.577	0.581	+0.40

Evidence provenance. The GeoHaze ledger is computed per image from the frozen final checkpoint of each of the six runs, and its table entries are averaged across runs. SSDG/DCMPNet and other trainable baseline entries are likewise means from six locally retrained runs with the same seeds; mechanism variants use the same dataset and fixed-final-checkpoint protocol. Depth sensitivities retain evaluator precision and are averaged across the six corresponding checkpoints; RTTS deltas use the displayed source values. The expanded qualitative comparisons retain the fixed scenes and method outputs.

G Expanded Qualitative Model Comparisons

Figures 10–12 extend the qualitative comparisons in the main paper to RW^{2}AH, RRSHID, RICE2, and RSID. Each row shows one scene, and columns align the input, the compared restoration methods, and the ground truth. The RW^{2}AH panels also include the depth prior. As in the main paper, the framed insets enlarge the regions marked by color-matched boxes.

The RW^{2}AH representative in the main paper is scene s2, selected by the largest per-sample margin over the strongest baseline and by visual inspection of the marked structure. Figure 10 retains all four scenes from that fixed set. The main remote-sensing comparison retains RRSHID-thin s1, RRSHID-thick s3, RICE2 s1, and RSID s1; the other scenes are shown below. RICE2 and RSID use the seven-column comparison, while RW^{2}AH and RRSHID retain their larger method sets.

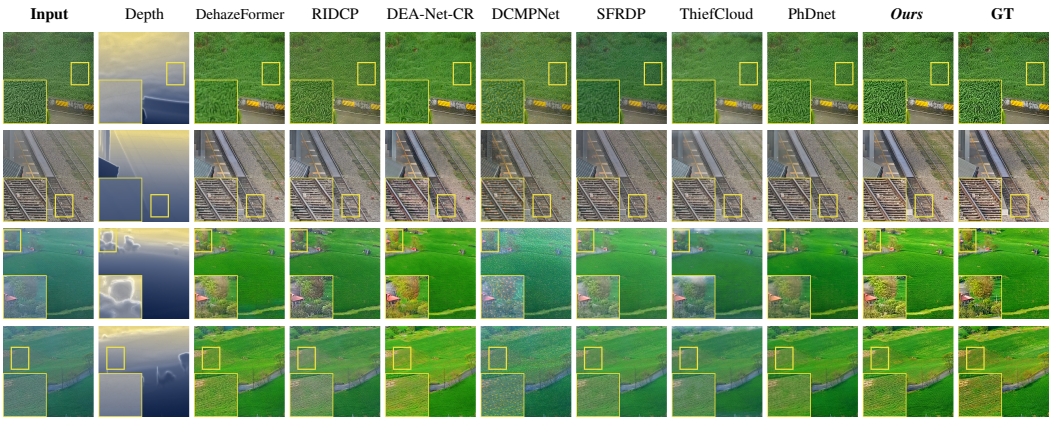

Figure 10: Expanded qualitative comparison on RW^{2}AH. Rows 1–4 show scenes s1–s4, respectively; scene s2 also appears in the main paper. The depth column shows the geometry prior used by GeoHaze. The framed insets enlarge the regions marked by color-matched boxes.

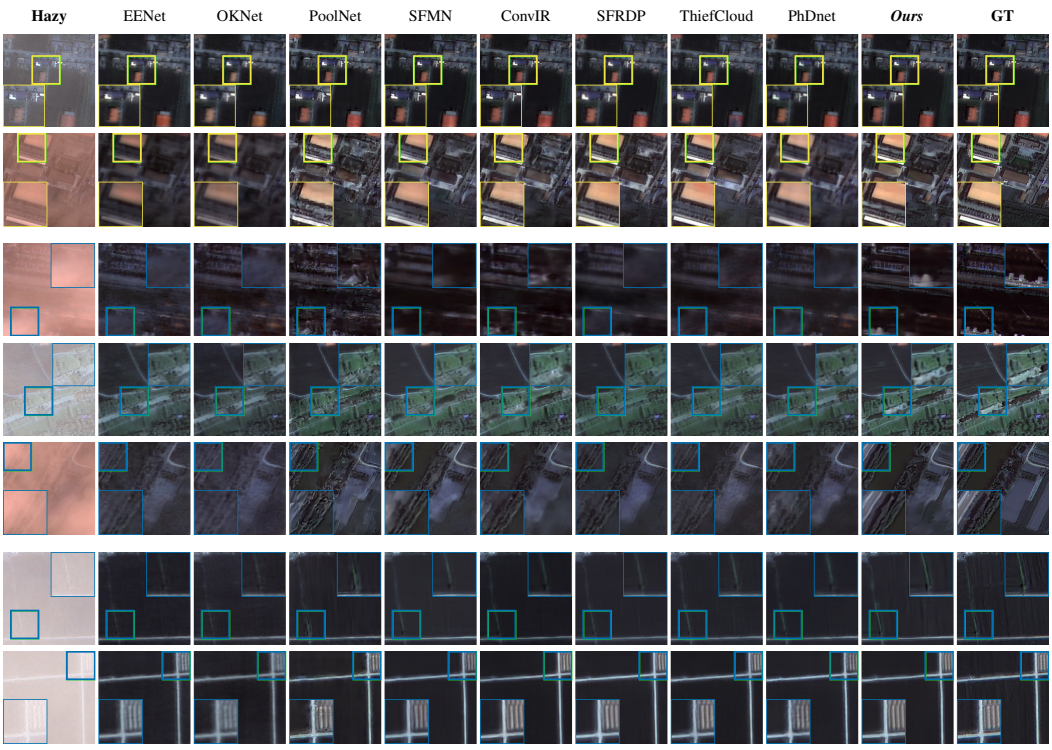

Figure 11: Expanded qualitative comparison on RRSHID. Rows 1–2 show thin-haze scenes s2/s3, rows 3–5 show moderate-haze scenes s1/s2/s3, and rows 6–7 show thick-haze scenes s1/s2. The framed insets enlarge the regions marked by color-matched boxes.

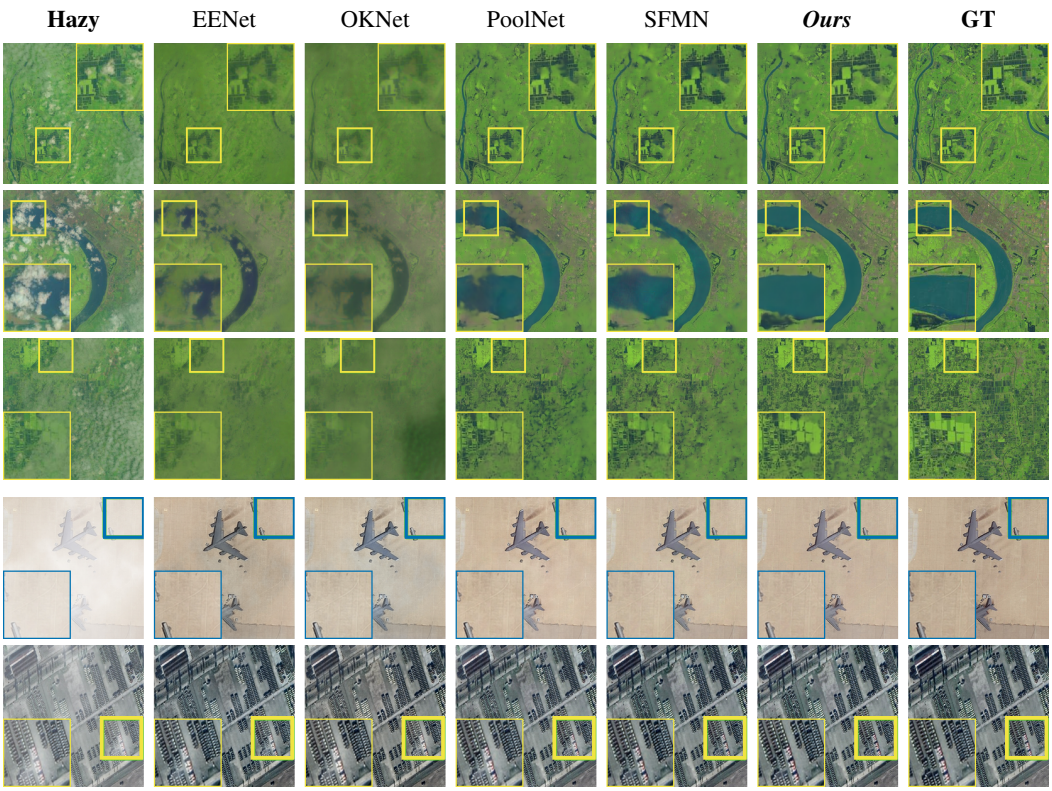

Figure 12: Expanded qualitative comparison on RICE2 and RSID. Rows 1–3 show RICE2 scenes s2/s3/s4, and rows 4–5 show RSID scenes s2/s3. The framed insets enlarge the regions marked by color-matched boxes.