SEPARATE TO RESTORE: HAZE-GUIDED FEATURE EXCHANGE FOR IMAGE DEHAZING

Anonymous authors
Paper under double-blind review

ABSTRACT

Single-image dehazing requires preserving scene content while suppressing haze-induced degradation. Modeling these objectives separately can encourage specialized representations, but restoration also requires communication between them. We propose a dual-branch framework that models clean scene content and a signed haze residual through separate pathways coordinated by haze-guided feature exchange. Haar wavelet cues provide complementary frequency biases, guiding the content branch toward structural details and the haze branch toward low-frequency degradation cues without imposing a fixed decomposition. Multiscale gated exchanges enable selective cross-branch communication. At the bottleneck, a routing module combines image-derived haze priors with learned corrections, prior-consistency weighting, and a learnable scaling factor to regulate where and how strongly messages are exchanged. Paired additive and subtractive updates redistribute features between the branches, while clean-image supervision, residual supervision, and input reconstruction consistency jointly encourage task-specific specialization. Under a controlled reimplementation protocol, the framework achieves 27.13 dB PSNR on O-Haze, exceeding the strongest compared baseline by 0.75 dB, with a mean PSNR of 26.98 ± 0.21 dB across five training seeds. Ablations and branch interventions suggest partial functional specialization, while evaluations on three additional real-haze benchmarks reveal remaining limitations under dense and nonuniform haze.

Keywords: single-image dehazing, two-path network, functional roles, Haar cues, haze evidence, sum-preserving exchange.

1 INTRODUCTION

Single-image dehazing aims to recover scene content while removing haze from an observed image. Content restoration depends on edges, textures, and colors, while haze removal depends on brightness, contrast, and spatial patterns. These signals are related: a model should give them different roles while still allowing the information exchange needed for restoration.

Recent dehazing methods use convolutional networks, multiscale fusion, and Transformers to improve feature modeling (Liu et al., 2019; Dong et al., 2020; Guo et al., 2022; Song et al., 2023). Wavelet methods use frequency bands to preserve image structure and detail, and recent work also uses low-frequency cues to model haze (Fu et al., 2021; Hwang et al., 2024; Sun et al., 2025). We focus on how to exchange content and haze features across scales and use haze cues to control where and how strongly messages pass at the bottleneck.

We propose STR, which models scene content and a signed haze residual along two connected feature paths. The Haar Decomposition Block (HDB) gives the haze path low-frequency information from LL and the content path detail information from LH, HL, and HH. Learnable projections can adjust these signals; they do not force a fixed image split. The Feature Distribution Block (FDB) maps shared features to the content and haze paths. The model predicts a clean image J and a signed residual R, linked by the input reconstruction constraint $J + R \approx I$. The paths are learned features, not statistically independent factors or a physical haze layer.

The Haze Feature Exchange Block (HFEB) and HERB control information exchange in different places. HFEB uses gates at multiple scales and paired add-subtract updates that preserve the sum of

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.

the two feature paths. At the bottleneck, HERB forms relative haze evidence D from image priors and learned corrections. It combines D with prior consistency C to form a routing gate G. In our current tests, the full HERB performs better than Base and several simpler variants. This result supports the combined design, but does not show that adding D alone improves performance.

On O-Haze, STR reaches 27.13 dB PSNR, 0.75 dB above the strongest compared baseline. Its mean PSNR over five training seeds is 26.98 ± 0.21 dB. Branch swaps and feature probes show different effects on the clean-image and residual outputs. This supports partial division of roles, but does not establish statistical independence or a physical haze layer.

Our contributions are:

1. We propose STR for joint scene-content restoration and signed haze-residual modeling. HDB provides learnable Haar-frequency cues, while clean-image, residual, and input-reconstruction losses jointly supervise the outputs.
2. We propose a multiscale feature exchange strategy that coordinates scene and haze features. HFEB implements this strategy with gated paired updates, passing information between paths while preserving their combined feature sum.
3. We introduce the haze evidence routing block (HERB). It combines relative haze evidence, prior consistency, and a learnable strength to route messages at the bottleneck. Ablations show that the full HERB performs better than several tested variants under the current protocol.

2 Related Work

2.1 Single Image Dehazing

Single-image dehazing has moved from hand-crafted parameter estimation to end-to-end image restoration. DehazeNet and MSCNN learn transmission maps with CNNs and use the atmospheric scattering model to recover clean images (Cai et al., 2016; Ren et al., 2016). GridDehazeNet passes information across scales with a grid structure and attention; FFA-Net uses feature attention; and MSBDN uses dense fusion and progressive enhancement (Liu et al., 2019; Qin et al., 2020; Dong et al., 2020). DeHamer and DehazeFormer use Transformers to model nonuniform haze over a larger area (Liu et al., 2021b; Guo et al., 2022; Song et al., 2023).

Many methods restore scene content and remove haze in one feature space or one backbone. The same feature can then mix scene structure with haze, which may lead to smooth textures, color shifts, or residual haze. Our focus is to assign different roles to two feature paths and let them exchange useful information.

2.2 Content and Haze Representations

Prior work assigns different roles to scene content and haze through physical models or feature learning. DID-Net uses the atmospheric scattering model to estimate a clean image, transmission, and airlight, with physical constraints to make these outputs easier to interpret (Liu et al., 2021a). CDD-GAN uses adversarial and contrastive learning to distinguish clean content from haze-related features. AECR-Net uses clean and hazy images as positive and negative examples, pushing restored features toward the clean distribution and away from the haze distribution (Chen et al., 2022; Wu et al., 2021). These designs use different objectives; their names alone do not show that their features are statistically independent.

Physical decomposition depends on its chosen assumptions, while a learned two-path model does not automatically give each path a clear role. Scene content and haze are related, so useful information must still pass between paths. FDB maps shared features to the content and haze paths; HFEB exchanges features across scales; and HERB controls haze messages at the bottleneck. Clean-image and input-reconstruction losses guide the outputs and encourage different main roles for the two paths.

2.3 Multiresolution Analysis for Image Dehazing

Wavelets split an image into low-frequency and directional high-frequency bands, so they are useful for multiscale restoration. DW-GAN adds a wavelet branch to combine spatial and frequency features and uses high-frequency bands to retain texture (Fu et al., 2021). WHNet uses wavelet decomposition and reconstruction for local detail and global structure. WaveDH uses wavelet bands for feature downsampling, upsampling, and refinement, while WDMamba restores low-frequency content before adding high-frequency detail (Hwang et al., 2024; Sun et al., 2025).

Unlike methods that mainly use wavelets for resizing or reconstruction, we use Haar bands to set an initial frequency bias for each path. The local low-frequency band LL cues the haze path; the directional detail bands LH, HL, and HH cue the content path. Learned projections can adjust or ignore these cues. They are only initial biases: no band is fixed as scene content or a physical haze layer. HDB applies the fixed Haar transform, while FDB maps shared features to two learned paths; neither imposes statistical independence.

In short, prior dehazing models provide strong feature learning, role-based methods explore how to handle different information, and wavelets provide local frequency information. STR uses Haar bands to set initial path biases, FDB to map shared features into two paths, and HFEB, HERB, and paired supervision to preserve useful exchange.

3 Method

We describe the STR architecture, the roles of its feature paths, and the training loss. Haar bands first set an initial bias. Controlled exchange and haze evidence routing then keep the paths connected, while joint clean-image and haze-residual objectives guide the outputs.

3.1 Overall Architecture

Given a hazy image $I \in [0,1]^{3\times H\times W}$ and its paired clean image J^*, the goal is to predict $J = F_\theta(I)$. The atmospheric scattering model is

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

Here, $t(x)$ is transmission and A is airlight. STR does not directly estimate both. Instead, it learns a content branch B and a haze branch H in feature space. These are feature tensors, not RGB background and physical haze images.

Figure 1 shows the network. A shared feature extractor feeds HDB and FDB, which provide different signals to the content and haze branches. The channel widths are 68, 136, and 272 at full, half, and quarter resolution. The content branch uses 2, 2, and 4 refinement blocks (DRBs) across the three scales; the haze branch uses 1, 1, and 2. Both bottlenecks use transposed channel attention to model long-range relations at lower spatial cost. HFEB exchanges features at several scales, and HERB routes messages at the bottleneck. The decoder upsamples and joins features from matching encoder scales, then predicts the two output features.

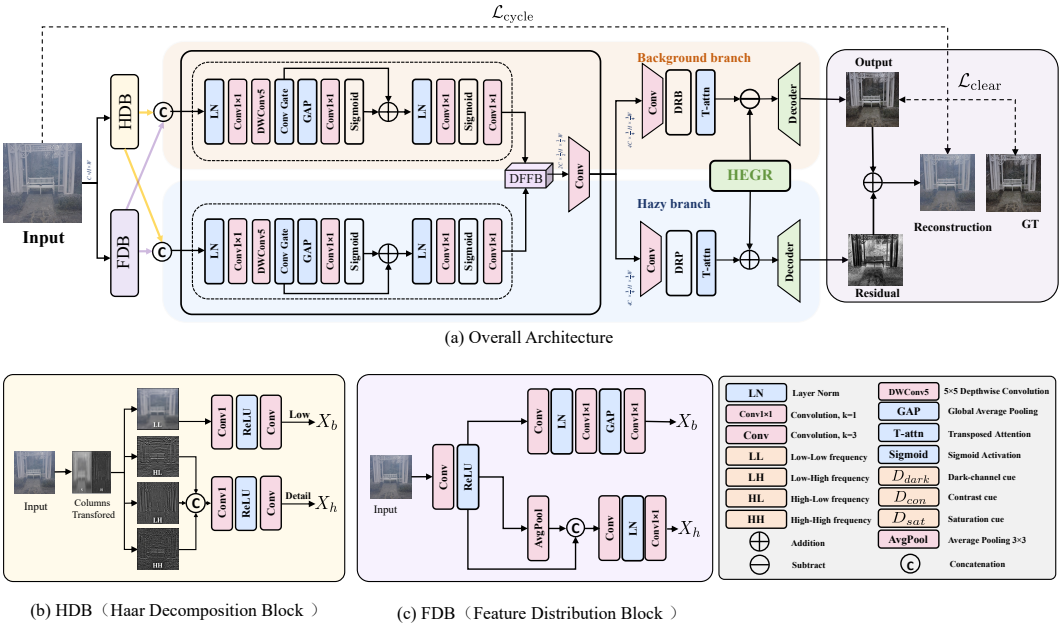
3.2 Feature Distribution Block (FDB)

FDB maps shared shallow features to the content and haze paths. It does not enforce orthogonality or statistical independence. Each encoder and decoder unit uses a refinement block (DRB). It uses LayerNorm, a 1×1 convolution, a 5×5 depthwise convolution, and SimpleGate to extract local features. Global average pooling and Sigmoid then recalibrate channels, followed by a light feed-forward path. Both transformations use learned residual scales:

$$Y = X + \beta \mathcal{F}_a(X), \tag{2}$$

$$X' = Y + \gamma \mathcal{F}_f(Y), \tag{3}$$

The parameters β and γ start small, which helps stabilize early training. The content branch uses more DRB units to restore structure and texture. The lighter haze branch mainly models degradation patterns and residual signals.



(a) Overall Architecture

(b) HDB（Haar Decomposition Block ）

(c) FDB（Feature Distribution Block ）

Figure 1: STR overview and input blocks. (a) Network architecture. (b) Haar Decomposition Block (HDB) maps the input to frequency bands. (c) Feature Distribution Block (FDB) maps shared features to the content and haze branches. HFEB exchanges features across scales, and HERB routes messages at the bottleneck.

3.3 Haar Decomposition Block (HDB)

The Haar Decomposition Block (HDB) applies fixed Haar filters to produce low-frequency and directional detail bands. Learned projections then decide how to send these cues to the two paths. Here, decomposition means only the transform into Haar bands; it does not separate latent scene factors or recover a physical haze layer. A shared feature is first computed with a 3×3 convolution and PReLU:

$$S = \text{PReLU}(\text{Conv}_{3\times3}(I)). \tag{4}$$

To give each path an initial frequency bias, the Haar router splits every 2×2 image patch into a low-frequency band LL and three detail bands LH, HL, and HH. Each band passes through its own 1×1 convolution, PReLU, and 3×3 convolution, then is upsampled to the input size:

$$L = P_L(LL), \tag{5}$$
$$E = P_E([LH, HL, HH]), \tag{6}$$
$$B_0 = f_B(S) + \sigma(s_E)E, \tag{7}$$
$$H_0 = f_H([S, \text{AvgPool}(S)]) + \sigma(s_L)L. \tag{8}$$

Here, E cues the content branch to use edges and textures, while L cues the haze branch to use low-frequency haze patterns. The projections and cue strengths are learned, so the network can adjust or ignore cues that do not help an image.

3.4 Haze Feature Exchange Block (HFEB)

HFEB exchanges features between the content and haze paths at multiple scales. It uses gates and paired updates. The branches need to share information. Scene structure can help in dense haze, while content restoration also needs to know where haze is strong. For features B and H at the same scale, each branch first uses the other branch to produce a gate:

$$g_B = \sigma(f_B(\text{LN}(H))), \tag{9}$$
$$g_H = \sigma(f_H(\text{LN}(B))), \tag{10}$$

Each f_B and f_H uses a 3×3 depthwise convolution and a 1×1 convolution. The gated messages are $T_B = g_B \odot B$ and $T_H = g_H \odot H$. The update is:

$$B' = B - \lambda_B T_B + \lambda_H T_H, \tag{11}$$

$$H' = H - \lambda_H T_H + \lambda_B T_B. \tag{12}$$

Each exchange term is subtracted from one branch and added to the other, so $B\prime + H\prime = B + H$. This sum-preserving form does not mean the branches are statistically independent.

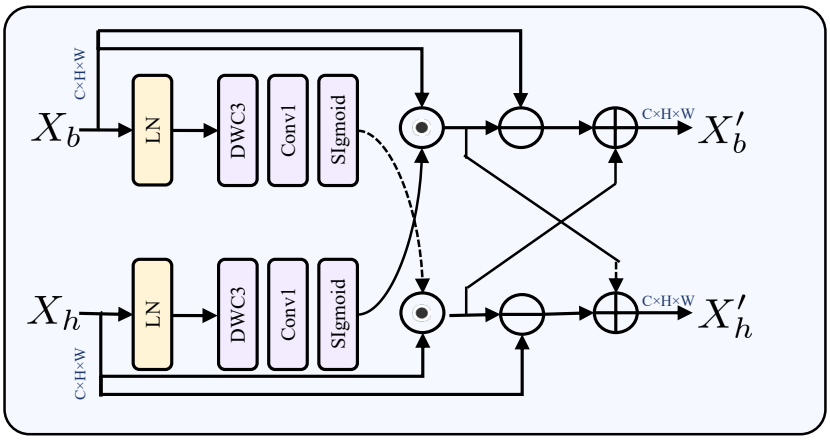


Figure 2: Haze Feature Exchange Block (HFEB). The content and haze features produce gates after normalization. These messages update the two paths through paired add-subtract operations.

3.5 Haze Evidence Routing Block (HERB)

3.5.1 Relative Haze Evidence

At the bottleneck, the network decides which messages from the haze branch should reach the content branch. From the input, we compute three simple cues: dark channel D_{dark}, inverse local contrast D_{contrast}, and inverse saturation D_{sat}. The initial prior is

$$P = \text{clip}(0.45D_{\text{dark}} + 0.30D_{\text{contrast}} + 0.25D_{\text{sat}}, 0, 1). \tag{13}$$

A fixed prior can fail on white objects, sky, and bright lights, so the network learns a bounded correction from bottleneck features:

$$\Delta D = f_D([\text{LN}(B_2^0), \text{LN}(H_2^0), P]), \tag{14}$$

$$D = \text{clip}(P + 0.25 \tanh(\Delta D), 0, 1). \tag{15}$$

Thus, D is a relative cue that a location may need stronger dehazing. It is not a physical haze concentration.

3.5.2 Prior Consistency and Message Routing

If the corrected evidence differs greatly from the image prior, the model should be less confident. We define a consistency score

$$C = \exp\left(-\frac{|D - P|}{0.20}\right), \tag{16}$$

The final routing gate is

$$G = D \odot C \odot \sigma(\eta), \tag{17}$$

where η is a learned strength. The haze branch proposes a correction message

$$M = 0.25 \tanh\big(f_{H \to B}(\text{LN}(H_2^0))\big). \tag{18}$$

HERB then updates the branches as follows:

$$B_2^+ = B_2^0 + G \odot M, \tag{19}$$

$$H_2^+ = H_2^0 - G \odot M. \tag{20}$$

The gate suppresses messages when D is small or C is low. When haze evidence is clear and consistent, the content branch receives a stronger correction. Figure 3 shows the routing process.

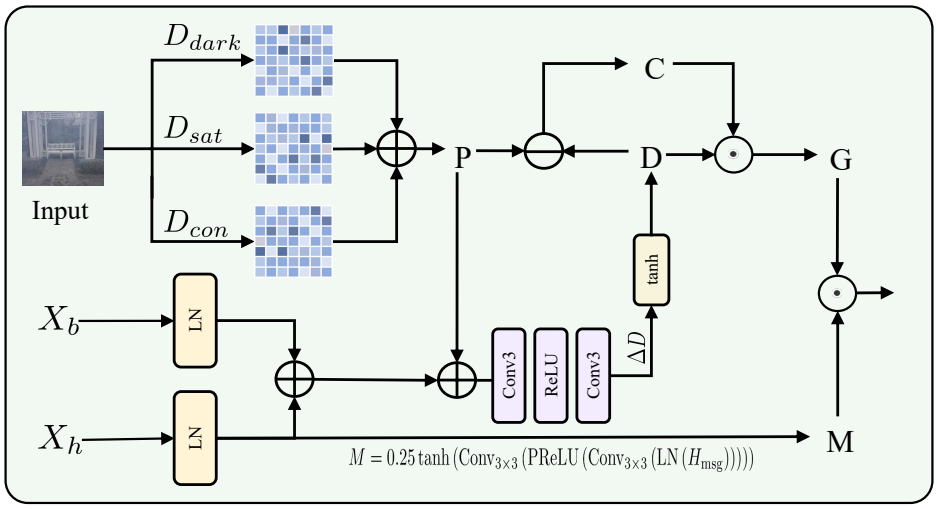


Figure 3: Haze Evidence Routing Block (HERB). The image prior P is corrected to form haze evidence D. The gate G combines D, prior consistency C, and a learned strength. Message M updates the content and haze branches with paired add-subtract operations.

3.6 Dual-Path Decoder and Image Outputs

At each scale, the decoder upsamples features and applies a 3×3 convolution before joining encoder features from the same scale. PixelSkipFusion uses a gate based on the decoder feature, the skip feature, and their difference instead of adding them directly:

$$q = \sigma(f_g([d, s, |d - s|])), \tag{21}$$

$$d' = d + f_p(q \odot s), \tag{22}$$

Here, d is the decoder feature and s is the skip feature from the same scale. The gate can reduce the use of shallow features when they contain strong haze.

The clean-image head and haze-residual head produce

$$J = I + \rho \tanh(h_B(B)), \tag{23}$$

$$R = 0.5 \tanh(h_H(H)), \tag{24}$$

$$I' = J + R, \tag{25}$$

Here, $\rho = 1$. The signed RGB residual R has the same size as the input, with target $R^* = I - J^*$. Thus, $I\prime = J + R$ can reconstruct the hazy input.

3.7 Joint Training Objective

The main experiments use the same `density_fusion` objective as the current code. We use Charbonnier and SSIM losses for the clean image:

$$\mathcal{L}_{\mathrm{char}} = \frac{1}{|\Omega|} \sum_{x \in \Omega} \sqrt{(J(x) - J^*(x))^2 + \epsilon^2}, \tag{26}$$

$$\mathcal{L}_{\mathrm{ssim}} = 1 - \mathrm{SSIM}(J, J^*), \tag{27}$$

where $\epsilon = 10^{-3}$. We define the haze-residual and input-consistency losses as

$$\mathcal{L}_{\mathrm{res}} = \mathrm{SmoothL1}(R, I - J^*), \tag{28}$$

$$\mathcal{L}_{\mathrm{cycle}} = \mathrm{SmoothL1}(J + R, I). \tag{29}$$

The haze evidence also uses a ranking loss $\mathcal{L}_{\mathrm{rank}}$ and an edge-aware smoothness loss $\mathcal{L}_{\mathrm{smooth}}$. The total loss is

$$\begin{aligned}\mathcal{L} = \mathcal{L}_{\mathrm{char}} + 0.2\mathcal{L}_{\mathrm{ssim}} + 0.04\mathcal{L}_{\mathrm{res}} + 0.02\mathcal{L}_{\mathrm{cycle}} \\ + 0.02\mathcal{L}_{\mathrm{rank}} + 0.001\mathcal{L}_{\mathrm{smooth}}.\end{aligned} \tag{30}$$

Equation (30) enforces three goals: J should match the clean target, R should match the hazy input minus the clean target, and $J + R$ should match the hazy input. Together, these losses limit how freely the two branches can assign information.

4 EXPERIMENTS

We report the data and setup, visual and quantitative results, ablations, branch analyses, cross-domain results, and runtime. This order first shows image quality, then tests what parts of the design contribute, and finally describes generalization and cost.

4.1 DATASETS AND IMPLEMENTATION DETAILS

We use Dense-Haze (Ancuti et al., 2019), I-Haze (Ancuti et al., 2018b), NH-Haze (Ancuti et al., 2020), O-Haze (Ancuti et al., 2018a), Haze4K (Liu et al., 2021a), and RESIDE-6K (Li et al., 2019). The first four contain real haze; Haze4K and RESIDE-6K contain synthetic paired images. RTTS is a real-haze subset of RESIDE. We also use hazy images from ACDC (Sakaridis et al., 2021) for cross-domain evaluation. Dataset sizes and splits are listed in Appendix Table 3.

For the main O-Haze experiment, we train on random 256×256 crops with batch size 1 for 1,000 epochs using seed 2028. We validate every 10 epochs and use Adam with an initial learning rate of 6×10^{-5} and cosine decay. The model width is 68. We use gradient clipping and an exponential moving average. We scale test images to $[0, 1]$, clip outputs to that range, and report RGB PSNR and SSIM.

4.2 EVALUATION PROTOCOL AND BASELINES

For comparison, we keep each method backbone and use the same data, input size, training budget, and evaluation code. We call this a controlled reimplementation, not an exact reproduction of the official models. We mark results as – when original logs or final checkpoints are unavailable.

4.3 QUALITATIVE EXPERIMENTS

We first compare restored images, then inspect intermediate features. Figure 4 shows three O-Haze scenes with the hazy input, comparison methods, STR, and clean target. Insets show edges, color, and residual haze. STR preserves scene outlines and reduces haze in distant and dark areas, with colors closer to the target. These examples provide visual evidence only; they are not a statistical test.

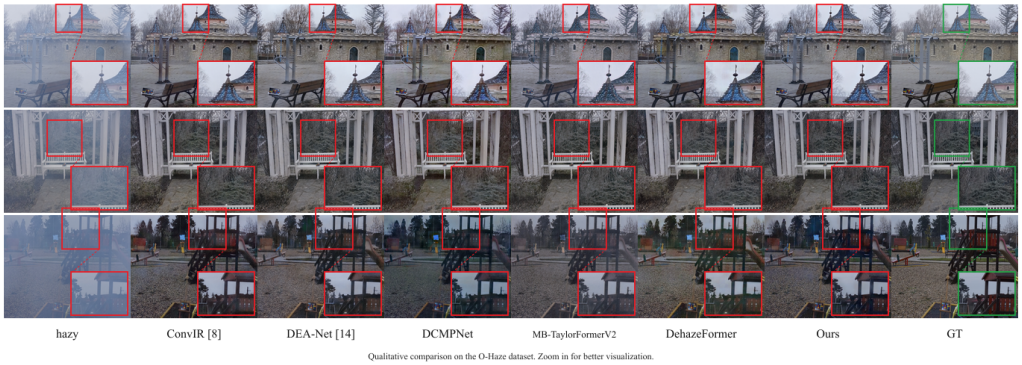


Figure 4: Visual comparison on O-Haze. Each row shows one test scene. Red boxes mark enlarged regions, and GT marks the clean target. Compared methods are ConvIR, DEA-Net, DCMPNet, MB-TaylorFormerV2, and DehazeFormer; STR is marked as Ours.

Table 1: PSNR/SSIM on four completed test sets. Best values are bold and second-best values are underlined. Ours denotes STR. Haze4K and RESIDE-6K are omitted until evaluation under a consistent protocol is complete.

Method	Venue/Year	O-Haze	Dense-Haze	I-Haze	NH-Haze
ConvIR	TPAMI 2024	26.1555/**0.9243**	16.5168/0.5962	22.5849/0.8583	**20.5669**/0.7708
DEA-Net	TIP 2024	25.5429/0.9198	16.4984/0.5766	22.3897/0.8665	20.5602/**0.7718**
DCMPNet	CVPR 2024	23.2200/0.8773	13.5576/0.4252	19.8352/0.8213	18.3886/0.7261
DehazeFormer	TIP 2023	26.3820/0.9185	**16.8821**/0.4893	**22.8655**/0.8253	20.3740/0.7698
MB-TaylorFormerV2	TPAMI 2025	25.7503/0.8987	16.2082/0.5204	21.5917/0.8548	18.6685/0.7358
UVM-Net	arXiv 2024	22.9512/0.8676	13.5568/0.4790	19.0269/0.7834	18.4268/0.7204
WDMamba	arXiv 2025	15.5677/0.5363	13.4954/0.4359	13.9966/0.5402	14.3292/0.5066
Ours	–	**27.1317**/0.9233	16.5499/**0.5979**	22.4013/**0.8754**	20.2124/0.7553

4.4 Quantitative Experiments

Table 1 reports RGB PSNR and SSIM on the four real-haze test sets with completed, consistent evaluations. We compare with ConvIR (Cui et al., 2024), DEA-Net (Chen et al., 2024), DCMP-Net (Zhang et al., 2024), DehazeFormer (Song et al., 2023), MB-TaylorFormerV2 (Jin et al., 2025), UVM-Net (Zheng & Wu, 2024), and WDMamba (Sun et al., 2025). Following the shared setup in Section 4.2, these are controlled reimplementations rather than results copied from the original papers. We omit Haze4K and RESIDE-6K until their training and evaluation follow the same protocol.

STR achieves the highest PSNR on O-Haze (27.1317 dB), 0.7497 dB above DehazeFormer, and the second-highest SSIM (0.9233). On Dense-Haze, it has the second-highest PSNR (16.5499) and the highest SSIM (0.5979). On I-Haze, it has the highest SSIM (0.8754), while DehazeFormer has the highest PSNR. On NH-Haze, STR remains below the strongest comparison results on both metrics. The results show gains on selected dataset and metric pairs, rather than consistent gains across all metrics and datasets.

The five-seed O-Haze evaluation has a mean PSNR of 26.9775 ± 0.2102 dB; detailed seed results are in Appendix Section A.3. This provides context for variation between training runs. All table values come from the current evaluation records. Haze4K and RESIDE-6K remain incomplete; we will add them after matching final checkpoints, logs, and per-image metrics are available.

4.5 Ablation Study

Table 2 uses checkmarks to show which components each setting includes. Adding the Haar frequency prior (HFP) or D evidence alone to Base lowers PSNR to 26.0163 and 26.1942 dB, respectively, compared with 26.2299 dB for Base. The D-only gate, C-only gate, and HERB without sum-preserving updates also score below Ours, which reaches 27.1317 dB. These results show that the tested full combination performs best among these variants; they do not show that HFP or D alone improves performance. Some variants use three seeds while Ours uses five, so this comparison is descriptive.

The full model with prior weights A0 $(0.45, 0.30, 0.25)$ reaches 27.1317 dB/0.9233 on O-Haze and has a five-seed mean PSNR of 26.9775 ± 0.2102 dB. Detailed prior settings, seed results, and per-image tests are in Appendix Section A.3. Some variants use three seeds while the full model uses five, so we report their comparison as descriptive.

4.6 Further Analysis

Functional roles and interventions. Frozen-feature probes and bottleneck swaps on O-Haze and Dense-Haze show different main roles for the two paths. Swapping the content bottleneck B_2 has a larger effect on clean output J, while swapping the haze bottleneck H_2 has a larger effect on residual output R. Haze strength remains readable from both paths, so the evidence supports partial

Table 2: Ablations of STR on O-Haze. HFP is the Haar frequency prior. Checkmarks mark included components; blank cells mark omitted components. Rows show Base, the component variants described in the text, and the full model (Ours) in that order. Best values are bold and second-best values are underlined.

Model	HFP	D evidence	C consistency	Sum update	PSNR	SSIM	Epoch
Baseline					26.2299	0.9202	760
	✓				26.0163	0.9192	870
		✓			26.1942	0.9210	750
		✓		✓	25.7389	0.9185	590
			✓	✓	<u>26.4545</u>	<u>0.9216</u>	950
		✓	✓		26.0443	0.9185	740
	✓	✓	✓	✓	**27.1317**	**0.9233**	760

functional roles rather than independent representations. Full probe and swap results are in Appendix Tables 6 and 7; additional sensitivity results are in Figure 6.

Cross-domain generalization. We test a model trained on Haze4K, without retraining, on all 4,322 RTTS images and 1,000 paired ACDC images. On RTTS, NIQE, BRISQUE, and FADE are 4.8632, 26.5378, and 1.2016. On ACDC, PSNR is 11.9420 dB and SSIM is 0.4926. ACDC results are lower than those on the four evaluated test sets. RTTS uses no-reference metrics, while ACDC uses full-reference metrics; the full results are in Appendix Table 4. These metrics use different protocols and should not be compared directly.

Efficiency. Full complexity and speed results are in Tables 8 and 9. We use 512×512 RGB inputs, batch size 1, FP32, and an RTX 4090. STR focuses on structured feature roles and clear information exchange, not low runtime cost.

5 Discussion and Limitations

Our evidence supports partial functional roles, but not independent representations or a physical haze layer. Branch tests use one checkpoint and a small image set. Synthetic haze edits cannot capture all real-world depth, scattering, or color changes, and Haze4K and RESIDE-6K still lack matched evaluations. The model is computationally expensive, and the low ACDC scores need further study.

6 Conclusion

We presented STR, with Haar bands for the two paths, HFEB for feature exchange at multiple scales, and HERB for bottleneck routing. Paired updates preserve the feature sum, while joint losses link clean-image restoration, haze residuals, and input reconstruction. Current results support partial differences in path roles, but not independent representations or a physical haze layer. Future work will reduce computation and test more varied real haze.

AI Use Statement

A generative AI tool was used to translate this manuscript from Chinese into English and edit wording for readability. It also helped edit descriptions of experimental results based on records supplied by the authors. The authors are responsible for checking all AI-assisted text, verifying the claims and values, and approving the final manuscript.

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 Workshops, pp. 754–762, 2018a.

Codruta O. Ancuti, Cosmin Ancuti, Mateu Sbert, and Radu Timofte. Dense-haze: A benchmark for image dehazing with dense-haze and haze-free images. In *2019 IEEE International Conference on Image Processing (ICIP)*, pp. 1014–1018, 2019. doi: 10.1109/ICIP.2019.8803046.

Codruta O. Ancuti, Cosmin Ancuti, and Radu Timofte. Nh-haze: An image dehazing benchmark with non-homogeneous hazy and haze-free images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp. 444–445, 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*, pp. 620–631, 2018b.

Bolun Cai, Xiangmin Xu, Kui Jia, Chunmei Qing, and Dacheng Tao. Dehazenet: An end-to-end system for single image haze removal. *IEEE Transactions on Image Processing*, 25(11):5187–5198, 2016.

Xiang Chen, Zhentao Fan, Pengpeng Li, Longgang Dai, Caihua Kong, Zhuoran Zheng, Yufeng Huang, and Yufeng Li. Unpaired deep image dehazing using contrastive disentanglement learning. In *European Conference on Computer Vision*, pp. 632–648, 2022.

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. doi: 10.1109/TIP.2024.3354108.

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, 2024. doi: 10.1109/TPAMI.2024.3419007.

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.

Minghan Fu, Huan Liu, Yankun Yu, Jun Chen, and Keyan Wang. Dw-gan: A discrete wavelet transform gan for nonhomogeneous dehazing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp. 203–212, 2021.

Chun-Le Guo, Qixin Yan, Saeed Anwar, Runmin Cong, Wenqi Ren, and Chongyi Li. Image dehazing transformer with transmission-aware 3d position embedding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5812–5820, 2022.

Seongmin Hwang, Daeyoung Han, Cheolkon Jung, and Moongu Jeon. Wavedh: Wavelet sub-bands guided convnet for efficient image dehazing, 2024.

Zhi Jin, Yuwei Qiu, Kaihao Zhang, Hongdong Li, and Wenhan Luo. Mb-taylorformer v2: Improved multi-branch linear transformer expanded by taylor formula for image restoration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(7):5990–6005, 2025. doi: 10.1109/TPAMI.2025.3559891.

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. doi: 10.1109/TIP.2018.2867951.

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.

Ye Liu, Lei Zhu, Shunda Pei, Huazhu Fu, Jing Qin, Qing Zhang, Liang Wan, and Wei Feng. From synthetic to real: Image dehazing collaborating with unlabeled real data. In *Proceedings of the 29th ACM International Conference on Multimedia*, pp. 50–58, 2021a. doi: 10.1145/3474085.3475331.

Table 3: Datasets and evaluation protocol.

Dataset	Type	Train/test	Test size
Dense-Haze	Real, dense haze	$50/5$	512×512
I-Haze	Real, indoor haze	$25/5$	512×512
NH-Haze	Real, nonuniform haze	$50/5$	512×512
O-Haze	Real, outdoor haze	$40/5$	512×512
Haze4K	Synthetic pairs	$3000/1000$	512×512
RESIDE-6K	Synthetic pairs	$6000/1000$	512×512

Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 10012–10022, 2021b.

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.

Wenqi Ren, Si Liu, Hua Zhang, Jinshan Pan, Xiaochun Cao, and Ming-Hsuan Yang. Single image dehazing via multi-scale convolutional neural networks. In *European Conference on Computer Vision*, pp. 171–186, 2016.

Christos Sakaridis, Dengxin Dai, and Luc Van Gool. Acdc: The adverse conditions dataset with correspondences for semantic driving scene understanding. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 10765–10775, 2021. doi: 10.1109/ICCV48922.2021.01059.

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.

Jie Sun, Heng Liu, Yongzhen Wang, Xiao-Ping Zhang, and Mingqiang Wei. Wdmamba: When wavelet degradation prior meets vision mamba for image dehazing, 2025.

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.

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

Zhuoran Zheng and Chen Wu. U-shaped vision mamba for single image dehazing, 2024.

A Supplementary Experimental Details

This appendix gives the seed results, feature probes, bottleneck swaps, and runtime measurements omitted from the main text. It follows the same evaluation protocol and statistics as the main text.

A.1 Datasets and Cross-Domain Results

The dataset sizes, splits, and full cross-domain evaluation are listed below. The RTTS metrics are no-reference measures over all 4,322 images; the ACDC metrics use 1,000 paired images and clean targets.

Table 4: Full unseen-domain evaluation for a model trained on Haze4K. RTTS is unpaired and uses no-reference metrics; ACDC is paired and uses full-reference metrics.

Test set	Train → test	Images	Metric	Mean ± std.
RTTS	Haze4K→RTTS	4322	NIQE	4.8632±1.3511
RTTS	Haze4K→RTTS	4322	BRISQUE	26.5378±11.1237
RTTS	Haze4K→RTTS	4322	FADE	1.2016±0.6176
ACDC	Haze4K→ACDC	1000	PSNR (dB)	11.9420±1.0924
ACDC	Haze4K→ACDC	1000	SSIM	0.4926±0.0871

A.2 Feature Visualization

Figure 5 shows intermediate responses for one image. The content and haze features respond differently in the same region. Haze evidence D marks locations for stronger processing, consistency C lowers the gate when the evidence conflicts with the prior, and G controls where the message acts. The image illustrates module behavior; one feature map cannot show statistical independence.

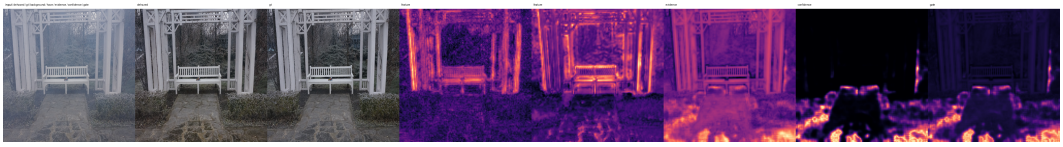


Figure 5: Input, restored image, clean GT, content and haze features, and visualizations of D, C, and G for O-Haze image 000043. Feature maps show relative response strength.

A.3 Prior Sensitivity and Per-Image Tests

The main prior uses weights A0 $(0.45, 0.30, 0.25)$. With seed 2028, uniform weights reach 26.5642/0.9232, a higher dark-channel weight reaches 26.1468/0.9194, and a higher contrast weight reaches 26.2551/0.9211. A0 works best in this data and training setup.

Across five full-HERB seeds, O-Haze PSNR is 26.9542, 27.1317, 26.6406, 26.9865, and 27.1744 dB, for a mean and sample standard deviation of 26.9775 ± 0.2102 dB. SSIM is 0.92316 ± 0.00105. Four key variants were each retrained with seeds 2028, 2029, and 2030; their results are in Table 5. Their three-seed means are below the five-seed full-model mean. Since seed counts differ, this is a descriptive comparison.

Per-image PSNR tests use the five O-Haze test images. The two-sided p-values for removing Cross0, Cross1, Cross2, or all Cross blocks, keeping only the density gate or consistency gate, and using non-sum-preserving HERB are 0.5332, 0.7181, 0.9480, 0.0438, 0.0898, 0.1183, and 0.4967. Only removing all Cross blocks reaches $p < 0.05$; the small test set limits this result.

A.4 Additional Branch Analysis

Feature probes use 15 generated examples per set. Bottleneck swaps use 12 paired observations per set. On O-Haze, mean B/H cosine similarity at cross2 is 0.0958 and channel correlation is 0.6041. Haze-probe accuracy for B/H is 1.000/1.000; content-probe accuracy ranges from 0.267 to 0.333. On Dense-Haze, mean cross2 cosine similarity is 0.1112 and channel correlation is 0.5568. Haze-probe accuracy for B/H is 0.933/1.000; content-probe accuracy is 0.333 for both.

On O-Haze, swapping B_2 changes J by 0.0178–0.0648, while swapping H_2 changes J by 0.0051–0.0135 and R by 0.0132–0.0251. Across both datasets, the H_2 swap affects R more, while the B_2 swap has a larger effect on J. Some cross-effects are not significant, so these tests support different main roles, not independence. Turning off all Cross blocks lowers PSNR from 25.6691 to 22.5050 dB. Keeping only the D or C gate gives 24.8163 and 24.9944 dB, respectively.

Table 5: Multi-seed O-Haze ablations. Each cell shows PSNR/SSIM at the best checkpoint. The final column gives mean PSNR ± sample standard deviation over three runs.

Setting	Seed 2028	Seed 2029	Seed 2030	Mean PSNR ± std.
No Haar	26.5042/0.9213	26.7410/0.9233	25.5785/0.9135	26.2746±0.6143
Non-sum-preserving HERB	26.8537/0.9227	26.2638/0.9212	26.2246/0.9223	26.4474±0.3524
No HERB	25.4982/0.9154	26.1287/0.9218	25.9355/0.9202	25.8541±0.3230
No Cross blocks	26.4532/0.9220	25.6262/0.9191	26.7424/0.9244	26.2739±0.5793

Table 6: Feature probes for the frozen model.

Stage	n	Mean cosine	$\lvert\rho_{\mathrm{ch}}\rvert$	Haze probe (B/H)	Content probe range
cross0	15	0.2367	0.7130	1.000/1.000	0.200–0.200
cross1	15	0.4158	0.7324	1.000/0.933	0.200–0.333
cross2	15	0.0958	0.6041	1.000/1.000	0.267–0.333
decoder1	15	0.1969	0.5703	1.000/1.000	0.133–0.200
decoder0	15	0.0978	0.5241	1.000/1.000	0.200–0.200

Table 7: Per-image paired statistics for bottleneck swaps in the frozen model. Changes are mean absolute output changes from the baseline.

Dataset	Comparison	n	Mean difference	t	Two-sided p
O-Haze	H_2: $\Delta R - \Delta J$	12	0.00955	6.500	4.43×10^{-5}
O-Haze	B_2: $\Delta J - \Delta R$	12	−0.00096	−0.705	0.49530
O-Haze	J: B_2 swap − H_2 swap	12	0.02633	2.993	0.01224
O-Haze	R: H_2 swap − B_2 swap	12	−0.01774	−2.168	0.05292
Dense-Haze	H_2: $\Delta R - \Delta J$	12	0.01147	6.017	8.71×10^{-5}
Dense-Haze	J: B_2 swap − H_2 swap	12	0.02320	3.233	0.00797
Dense-Haze	R: H_2 swap − B_2 swap	12	−0.01508	−2.288	0.04294

Table 8: Parameters, computation, and GPU speed for 512×512 inputs.

Model	Params (M)	FLOPs (G)	GPU (ms)	FPS
ConvIR	5.5305	335.702	16.286	61.403
DEA-Net	7.7893	156.376	20.017	49.956
DCMPNet	17.3583	550.170	44.596	22.424
UVM-Net	1003.9402	–	33.553	29.803
MB-TaylorFormerV2	2.6295	250.939	443.673	2.254
WDMamba	11.2544	308.484	61.047	16.381
DehazeFormer	1.2849	95.908	23.879	41.878
STR	**12.4212**	**1037.386**	**147.700**	**6.770**

Table 9: Peak memory and CPU speed for 512×512 inputs.

Model	Peak memory (MiB)	CPU (ms)
ConvIR	436.8	3934.69
DEA-Net	316.0	3721.46
DCMPNet	803.6	9088.37
UVM-Net	8273.7	–
MB-TaylorFormerV2	1894.2	45965.36
WDMamba	287.0	–
DehazeFormer	672.9	4437.95
STR	**1665.3**	**22600.29**

We do not estimate UVM-Net FLOPs or UVM-Net and WDMamba CPU times because their custom CUDA/Mamba operators cannot be measured reliably.

B Supplementary Sensitivity Analysis

Figure 6 shows response sensitivity to haze strength and scene content at different stages. It illustrates branch response trends only. The five-image sample does not replace multi-seed training ablations or per-image statistical tests.

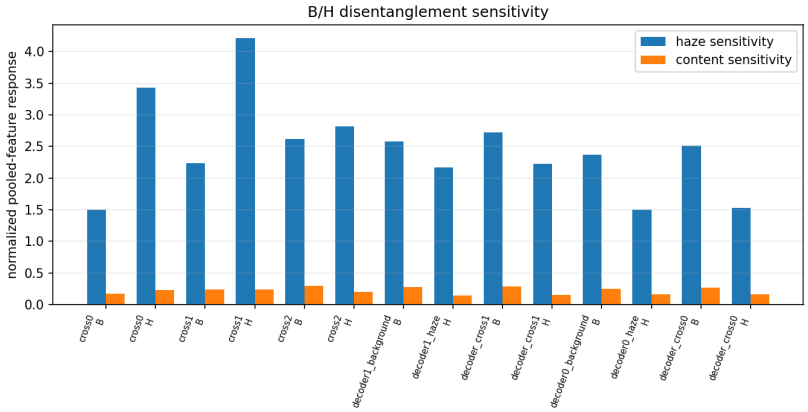


Figure 6: Sensitivity to haze strength and scene content at different stages.