PRISM: Rethinking Atmospheric Scattering Reconstruction as a Unified Understanding and Restoration Model for Real-world Dehazing

Chengyu Fang[1,2], **Chunming He**[2], **Yuelin Zhang**[3], **Chubin Chen**[1], **Chenyang Zhu**[1], **Hongqiu Wang**[5], **Longxiang Tang**[4], **Xiu Li**[1*], **Sina Farsiu**[2]
[1]Tsinghua University, [2]Duke University, [3]CUHK, [4]HKUST, [5]HKUST(GZ)
https://github.com/cnyvfang/PRISM

Abstract

Real-world image dehazing (RID) aims to remove haze-induced degradation from real scenes. This task remains challenging due to non-uniform haze distribution, spatially varying color shifts, and the scarcity of paired real hazy-clean data. In PRISM, we propose *Proximal Scattering Atmosphere Reconstruction (PSAR)*, a physically structured framework that jointly reconstructs the clear scene and scattering variables under the atmospheric scattering model, making the restoration process more interpretable in complex real-world conditions. To bridge the synthetic-to-real gap, we design an online non-uniform haze synthesis pipeline and a *Selective Self-Distillation Adaptation (SSDA)* scheme for unpaired real-world scenarios, which enables the model to selectively learn from high-quality perceptual targets while leveraging its intrinsic scattering understanding to audit residual haze and guide self-refinement. Experiments on real-world benchmarks demonstrate that PRISM achieves competitive performance on RID tasks.



Figure 1: Our proposed PRISM framework enables effective non-uniform haze removal, producing results with more realistic colors and fewer artifacts than prior methods.

1 Introduction

Image dehazing is a long-standing problem in intelligent transportation, remote sensing, and general photography. In real-world scenes, haze density is highly non-uniform, and multiple degradations

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

create spatially varying color shifts. These factors make the joint recovery of the clear scene and scattering variables from a single observation highly ill-posed. Many learning-based methods treat dehazing as a direct hazy-to-clear mapping, which often reduces physical interpretability and leads to artifacts in challenging regions. Alternatively, conventional methods based on the atmospheric scattering model (ASM) typically estimate transmission T and atmospheric light A, and then invert the model to restore scene radiance J. This sequential approach often treats scattering variables as auxiliary byproducts, failing to fully exploit their mutual constraints with the scene content.

To achieve more interpretable restoration, recent works have introduced unfolding network [1, 2] and proximal gradient descent into the ASM framework. For instance, PDN [3] optimizes the dark channel and transmission map within an unfolded proximal framework. However, its recombination strategy for the final output struggles with the complex composite degradations found in real scenes. Similarly, CORUN [4] unfolds gradient steps to jointly update transmission and the scene, improving their coupling. Yet, it assumes a globally constant atmospheric light, which limits its ability to model spatially varying haze and illumination effects in real images.

In real-world scenarios, data distributions pose substantial challenges. Real-world haze is inherently non-homogeneous, often coupled with complex color shifts caused by region-wise color-tinted haze, car lights, and building windows, as well as camera noise and compression artifacts. Existing synthetic haze distributions are typically over-simplified and overly consistent, failing to capture such spatial diversity and stochastic noise. This mismatch, combined with the scarcity of paired hazy-clean images in the wild, creates a synthetic-to-real domain gap that hinders robust adaptation.

To explicitly model haze formation while restoring the scene, we formulate image dehazing as a joint reconstruction problem, where the clear scene and scattering variables are optimized simultaneously. In PRISM, we propose **Proximal Scattering Atmosphere Reconstruction (PSAR)**, a physically structured framework that performs dehazing through a sequence of optimization-inspired stages. Each stage couples closed-form proximal updates for J, T, and A with lightweight refinement blocks, transforming the dehazing task into a compact, physics-guided reconstruction process. By representing both A and T as spatial variables, PSAR can better model non-uniform haze and color shifts, instead of forcing these effects to be implicitly absorbed by the restored radiance.

To further improve real-world adaptation, PRISM introduces a **Selective Self-Distillation Adaptation (SSDA)** scheme together with a lightweight online non-uniform haze synthesis framework. In this framework, an exponential moving average (EMA) teacher produces pseudo-labels, which are evaluated alongside the student's predictions by an ensemble of perceptual metrics. The student distills only from targets that pass a conservative quality gate, reducing its exposure to erroneous pseudo-labels. Beyond this external guidance, the ASM-based model naturally unifies haze modeling and clean-image restoration by estimating transmission and atmospheric light while reconstructing a clean scene. Leveraging this property, we use the teacher network to audit the student's output for residual haze and guide further refinement. This scattering-based audit complements quality-gated distillation, providing an additional physical constraint for more stable unsupervised adaptation.

Our core contributions are threefold:

(1) We propose PSAR, a physically structured dehazing framework that decouples atmospheric scattering variables from inputs and performs reconstruction through proximal optimization steps.

(2) We introduce an online non-uniform haze synthesis pipeline and SSDA scheme that leverages quality-gated distillation and a physics-guided scattering prior for real-world adaptation.

(3) We evaluate our PRISM framework on real-world image dehazing tasks. Experiments on widely used benchmarks demonstrate that our approach achieves competitive performance.

2 Related Works

2.1 Single Image Dehazing

Early single image dehazing methods rely on hand-crafted priors. DCP [5] requires at least one low-intensity channel in local patches, while other priors [6, 7] use regularities to estimate transmission, but they often fail in sky regions and non-uniform haze. With data-driven learning, deep models have become mainstream [8–15] and typically learn direct hazy-to-clear mappings [16–19], yet the lack of physical constraints reduces interpretability and robustness under complex real degradations.

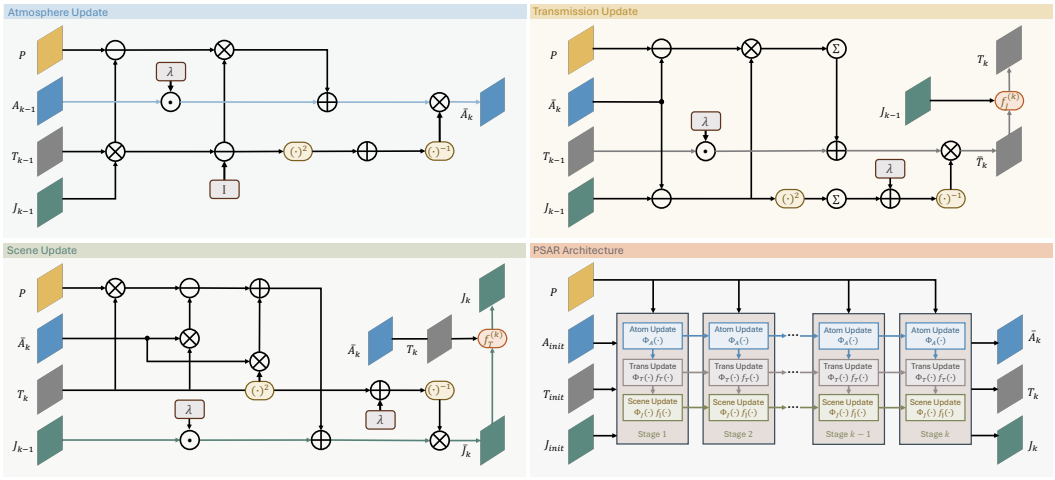

Figure 2: Architecture of PSAR at the k-th stage. Given the previous estimates $(J_{k-1}, T_{k-1}, A_{k-1})$, PSAR computes proximal updates for atmospheric light, transmission, and radiance under the same ASM objective, followed by lightweight refinements for T and J.

Diffusion models [20–23] have achieved high perceptual quality in dehazing owing to their strong generative capabilities. However, they may occasionally hallucinating unfaithful details, which limiting the credibility issues in downstream tasks. This highlights the necessity of developing physically grounded frameworks that explicitly model the degradation process for reliable reconstruction.

2.2 Real-world Image Dehazing

The domain gap between synthetic and real haze remains a key bottleneck for real-world dehazing [11, 12, 16–18, 24]. Real haze is entangled with spatially varying degradations, while paired real hazy-clean data are scarce, motivating increasing efforts in real-world settings [25–29]. To better model real-world haze [30], PDN [3] and CORUN [4] introduce explicit physical modeling, but incomplete atmospheric scattering modeling still limits their color fidelity and robustness. Other studies narrow the gap by improving haze synthesis realism, such as RIDCP [31] and Wang *et al.* [32]. HazeFlow [33] further improves realism but requires an additional non-uniform haze field synthesis step, leading to a trade-off between process complexity and haze authenticity. Domain adaptation methods improve generalization [34–39], but adversarial training may introduce artifacts and instability [40]. Some methods combine synthetic and real data [25, 41, 42], yet noisy pseudo-labels in self-distillation may accumulate and degrade restoration quality. Although Colabator [4] reduces this issue by up-weighting high-quality regions, it still relies on noisy pseudo-labels and a relatively complex training pipeline. Therefore, an unsupervised adaptation framework is still needed to reduce pseudo-label error propagation while using explicit physical constraints to bridge the synthetic-to-real domain gap.

3 Methodology

3.1 PSAR: Proximal Scattering Atmosphere Reconstruction

We build our model directly on the atmospheric scattering equation:

$$P(x) = T(x)\,J(x) + (1 - T(x))\,A(x), \tag{1}$$

where $P(x), J(x) \in \mathbb{R}^{3\times1}$ are the hazy and clear RGB column vectors at pixel x, $A(x) \in \mathbb{R}^{3\times1}$ is the atmospheric light field, and $T(x) \in \mathbb{R}$ is the scalar transmission. PSAR treats single image dehazing as the joint reconstruction of (J, T, A) under Eq. (1). At each stage k, the network starts from coupled proximal updates induced by the same scattering model and then applies lightweight learnable corrections, so that each stage remains explicitly tied to the same physical objective.

Stage-wise scattering objective. We measure the mismatch to Eq. (1) by the quadratic data term:

$$\mathcal{D}(J,T,A) = \frac{1}{2}\sum_{x} \|P(x) - T(x)\,J(x) - (1-T(x))A(x)\|_2^2, \tag{2}$$

where $\|\cdot\|_2$ is the Euclidean norm in RGB space. At stage k, PSAR first computes the atmospheric light and transmission proximal points, keeps the atmospheric light branch purely proximal with $A_k = \bar{A}_k$, refines the transmission as $T_k = \bar{T}_k + f_T^{(k)}(\bar{T}_k, J_{k-1})$, and then updates the clear-scene proximal point under (T_k, A_k):

$$\begin{aligned}
\bar{A}_k &= \arg\min_{A}\ \mathcal{D}(J_{k-1}, T_{k-1}, A) + \frac{\lambda_A}{2}\|A - A_{k-1}\|_2^2, \\
\bar{T}_k &= \arg\min_{T}\ \mathcal{D}(J_{k-1}, T, \bar{A}_k) + \frac{\lambda_T}{2}\|T - T_{k-1}\|_2^2, \\
\bar{J}_k &= \arg\min_{J}\ \mathcal{D}(J, T_k, A_k) + \frac{\lambda_J}{2}\|J - J_{k-1}\|_2^2,
\end{aligned} \tag{3}$$

where $\lambda_A, \lambda_T, \lambda_J > 0$ are learnable proximal weights. The quadratic terms serve as stage-to-stage trust regions around $(J_{k-1}, T_{k-1}, A_{k-1})$. We use bars to denote the analytic proximal solutions $(\bar{A}_k, \bar{T}_k, \bar{J}_k)$ of these subproblems. The propagated variables are then refined as $T_k = \bar{T}_k + f_T^{(k)}(\bar{T}_k, J_{k-1})$ and $J_k = \bar{J}_k + f_J^{(k)}(\bar{J}_k, T_k, A_k)$, and (T_k, J_k) denote the refined variables passed to the next stage.

We keep $A_k = \bar{A}_k$ without learnable refinement. Atmospheric light mainly captures low-frequency atmospheric veil and color bias, and its update is a closed-form proximal step anchored to the previous-stage estimate. Adding a learnable refinement to A would increase parameters in a weakly constrained branch and may amplify errors inherited from the current J_{k-1} and T_{k-1} estimates. We therefore reserve learnable refinements for $T(x)$ and $J(x)$, where residual structural and radiance errors are more pronounced, and keep the A branch as a bounded proximal update.

Atmospheric light proximal update. The first step updates the atmospheric light while keeping J_{k-1} and T_{k-1} fixed. Most classical formulations use a global atmospheric light A, which works under near-uniform color shifts and haze. In real scenes, spatially varying color shifts and non-uniform scattering make A spatially varying, so a single RGB vector can force illumination mismatch into $T(x)$ or $J(x)$. We therefore model atmospheric light as a field $A(x) \in \mathbb{R}^3$ at all stages, enabling locally adaptive updates under the same scattering constraint. For each pixel x, the atmospheric light proximal problem in Eq. (3) becomes:

$$\begin{aligned}
\bar{A}_k(x) = \arg\min_{A(x)\in\mathbb{R}^3} \frac{1}{2}\|J_{k-1}(x)\,T_{k-1}(x) + (1-T_{k-1}(x))A(x) - P(x)\|_2^2 \\
+ \frac{\lambda_A}{2}\|A(x) - A_{k-1}(x)\|_2^2,
\end{aligned} \tag{4}$$

which is a quadratic least-squares problem in $\mathbb{R}^3$. Solving the normal equation yields the update:

$$\bar{A}_k(x) = \frac{(1-T_{k-1}(x))(P(x) - J_{k-1}(x)\,T_{k-1}(x)) + \lambda_A\,A_{k-1}(x)}{(1-T_{k-1}(x))^2 + \lambda_A}. \tag{5}$$

Here $(1-T_{k-1}(x))$ modulates the contribution of the observation term, while the denominator is at least $\lambda_A > 0$, making the update bounded and anchored to $A_{k-1}(x)$.

Transmission proximal update. With J_{k-1} and A_k fixed, the scattering model becomes quadratic in $T(x)$, so the update focuses on estimating how much scene radiance is preserved at each pixel under the current atmospheric light. For each pixel x, the transmission update from Eq. (3) reads:

$$\bar{T}_k(x) = \arg\min_{T(x)\in\mathbb{R}} \frac{1}{2}\|T(x)\,J_{k-1}(x) + (1-T(x))A_k(x) - P(x)\|_2^2 + \frac{\lambda_T}{2}\|T(x) - T_{k-1}(x)\|_2^2. \tag{6}$$

This is a scalar quadratic in $T(x)$. Writing explicitly over the three RGB channels $c \in \{1, 2, 3\}$, the analytic solution is:

$$\bar{T}_k(x) = \frac{\lambda_T \, T_{k-1}(x) + \sum_{c=1}^{3} (A_k^{(c)}(x) - J_{k-1}^{(c)}(x))(A_k^{(c)}(x) - P^{(c)}(x))}{\lambda_T + \sum_{c=1}^{3} (A_k^{(c)}(x) - J_{k-1}^{(c)}(x))^2}. \tag{7}$$

Here $P^{(c)}(x)$, $J_{k-1}^{(c)}(x)$ and $A_k^{(c)}(x)$ denote the c-th components of the corresponding RGB vectors. The numerator combines the current scattering residual with the previous transmission $T_{k-1}(x)$, and the denominator normalizes by the local contrast between $A_k(x)$ and $J_{k-1}(x)$, yielding a closed-form, data-dependent transmission update that remains anchored to the previous stage.

Clear-scene proximal update. After refining the transmission, the third step updates the clear image while keeping T_k and A_k fixed. With transmission and atmospheric light fixed, the scattering model becomes linear in $J(x)$, so we can solve for the radiance by a closed-form least-squares update under the same physical constraint. For each pixel x, the clear-scene update from Eq. (3) becomes:

$$\bar{J}_k(x) = \arg\min_{J(x) \in \mathbb{R}^3} \frac{1}{2} \| T_k(x) \, J(x) + (1 - T_k(x)) A_k(x) - P(x) \|_2^2 + \frac{\lambda_J}{2} \| J(x) - J_{k-1}(x) \|_2^2. \tag{8}$$

This is a quadratic least-squares problem in $\mathbb{R}^3$. Since $T_k(x)$ is scalar, all three channels share the same scalar coefficients. Solving the normal equation yields:

$$\bar{J}_k(x) = \frac{T_k(x) \, P(x) + T_k(x)^2 \, A_k(x) - T_k(x) \, A_k(x) + \lambda_J \, J_{k-1}(x)}{T_k(x)^2 + \lambda_J}, \tag{9}$$

where the scalar denominator $T_k(x)^2 + \lambda_J$ divides each RGB component of the numerator. This update balances adherence to the scattering model with temporal consistency of the clear estimate across stages.

From proximal updates to PSAR stages. The three proximal operators in Eqs. (5), (7) and (9) define an optimization-inspired backbone. One PSAR stage is summarized as:

$$\begin{cases} \bar{A}_k = \Phi_A(P, J_{k-1}, T_{k-1}, A_{k-1}), & A_k = \bar{A}_k \\ \bar{T}_k = \Phi_T(P, J_{k-1}, T_{k-1}, A_k), & T_k = \bar{T}_k + f_T^{(k)}(\bar{T}_k, J_{k-1}), \\ \bar{J}_k = \Phi_J(P, J_{k-1}, T_k, A_k), & J_k = \bar{J}_k + f_J^{(k)}(\bar{J}_k, T_k, A_k), \end{cases} \tag{10}$$

where Φ_A, Φ_T, Φ_J denote the closed-form proximal maps induced by Eq. (2), and $f_T^{(k)}, f_J^{(k)}$ are shallow learnable refinements. Stacking several such stages yields a deep architecture in which each stage is built from explicit proximal updates for the same scattering objective, while the lightweight corrections allow the model to adapt to the statistics of real images. The detailed algebraic derivations from Eqs. (4), (6), (8) to Eqs. (5), (7), (9) are deferred to the appendix.

3.2 Stage 1: Pretraining with Online Non-Uniform Haze Synthesis

We pretrain the network on synthetic hazy images generated by a depth-guided online scattering model (Alg. 1), inspired by RIDCP [31] and extended to non-uniform density fields and controllable near-camera haze. The design aims to approximate depth-dependent and spatially varying haze observed in real scenes, rather than restricting the model to simple uniform synthetic haze.

Online non-uniform haze synthesis. We synthesize diverse haze patterns online without precomputation, as summarized in Alg. 1. Given a clean image J_{gt} and normalized depth D, the algorithm samples a base density β_{init} and, with probability p, adds a non-negative low-frequency perturbation:

$$\Delta\beta = \mathcal{S}(G_{\sigma_1} * \mathcal{U}(G_{\sigma_0} * z_0)), \tag{11}$$

where $\mathcal{U}(\cdot)$ denotes upsampling and $\mathcal{S}(\cdot)$ is a monotone affine rescaling. This gives $\beta(x) = \beta_{\text{init}} + \Delta\beta(x)$ for non-uniform haze and $\beta(x) = \beta_{\text{init}}$ for uniform haze, exposing the model to both heterogeneous and homogeneous degradations.

Algorithm 1: Online Non-Uniform Haze Synthesis

Input: Clear image J_{gt}, normalized depth D, probability p, near-haze range $[h_{\min}, h_{\max}]$
Output: Hazy image P, transmission T_{gt}, atmospheric light offset A_{rend}

$J_{gt} \leftarrow \text{RandAddGaussianNoise}(\text{RandAdjustLuminance}(J_{gt}))$, $\beta_{\text{init}} \sim \mathcal{U}(\beta_{\min}, \beta_{\max})$;
if *with probability* p **then**
  Generate low-resolution noise $z_0 \sim \mathcal{N}(0, 1)$,
  $\beta(x) \leftarrow \beta_{\text{init}} + \mathcal{S}(G_{\sigma_1} * \mathcal{U}(G_{\sigma_0} * z_0))$; `// Non-uniform haze`
else
  $\beta(x) \leftarrow \beta_{\text{init}}$; `// Uniform haze`
$h_{\text{near}} \sim \mathcal{U}(h_{\min}, h_{\max})$, $d_0 \leftarrow -\log(1 - h_{\text{near}})/(\beta_{\text{init}} + \varepsilon)$; `// Foreground haze`
$T_{gt}(x) \leftarrow \exp(-\beta(x)\,[(1 - D(x)) + d_0])$;
$A_{\text{rend}} \leftarrow \text{Clip}\big(\mathcal{U}(A_{\min}, A_{\max}) \in \mathbb{R}^1 + \mathcal{U}(-\delta, \delta) \in \mathbb{R}^3,\, 0,\, 1\big)$;
$P \leftarrow \text{RandCompress}(T_{gt} \odot J_{gt} + (1 - T_{gt}) \odot A_{\text{rend}})$; `// Haze rendering`
return $P, T_{gt}, A_{\text{rend}}$;

To control foreground haze, we sample h_{near} and convert it into a depth offset $d_0 = -\log(1 - h_{\text{near}})/(\beta_{\text{init}} + \varepsilon)$, which is then used with $\beta(x)$ and the normalized depth D to compute T_{gt}. This allows the density field to vary spatially while keeping the near-camera haze strength controllable. After obtaining T_{gt}, we sample the atmospheric light A_{rend} and combine it with J_{gt} to synthesize the hazy image P by Equation (1), followed by random compression to mimic real image degradation.

Optimization. We train PSAR with an objective that follows its physical decomposition of haze formation. The restored radiance J and transmission T are supervised by their targets, while atmospheric light is regularized by a smoothness prior rather than direct pixel-wise supervision:

$$\mathcal{L}_{\text{total}} = \|J - J_{gt}\|_1 + \|T - T_{gt}\|_1 + \lambda_A^{\text{tv}} \sum_x \|\nabla A(x)\|_1 + \lambda_c \frac{d(J, J_{gt})}{d(J, P) + \epsilon} + \lambda_{\text{adv}} \mathcal{L}_{\text{adv}}(J), \quad (12)$$

where $d(\cdot, \cdot)$ denotes the LPIPS distance [43], and ϵ is a small constant. We do not impose direct ℓ_1 supervision on A_{rend}. During synthesis, A_{rend} only serves as an image-level anchor for haze rendering, while J_{gt} already contains spatial illumination variations. Matching $A(x)$ to this anchor would bias the atmospheric light branch toward flat fields and limit its adaptation to spatial illumination. Instead, supervision on J and T anchors the scattering reconstruction, while the closed-form A update and total variation prior encourage a smooth yet spatially varying atmospheric light field and prevent A from absorbing high-frequency textures. Beyond pixel-level supervision, the contrastive LPIPS term increases the perceptual separation between the restored radiance J and hazy input P while aligning J with J_{gt}. Since LPIPS gradients may also strengthen structured high-frequency artifacts, the adversarial objective encourages natural-image statistics and helps suppress these artifacts.

3.3 Stage 2: Selective Self-Distillation Adaptation

Models pretrained on synthetic data often degrade in real-world scenarios due to domain shifts. Since blind unsupervised adaptation can collapse into degenerate solutions or amplify artifacts, we propose the Selective Self-Distillation Adaptation (SSDA) framework to bridge this domain gap (Figure 3). Built upon a mean-teacher architecture, SSDA processes a weakly augmented real hazy image P_w through the EMA teacher $\mathcal{F}_{\text{ema}}$ to generate a stable pseudo-label J_{ema}, which guides the student $\mathcal{F}_\theta$ operating on a strongly augmented view P_s. To ensure robust domain adaptation, SSDA uses two complementary constraints: Quality-Gated Self-Distillation (QGSD) for external perceptual guidance, and a Physically-Grounded Scattering Regularization (PGSR) for internal physical consistency.

Quality-Gated Self-Distillation (QGSD). Directly distilling all teacher predictions may propagate unreliable pseudo-labels on out-of-distribution data. To reduce dependence on any single filtering criterion, we introduce a strict filtering mechanism guided by an ensemble of No-Reference Image Quality Assessment (NR-IQA) metrics, denoted as $\Omega = \{\mathcal{Q}_i\}_{i=1}^N$. This module authorizes distillation if and only if the teacher's prediction strictly outperforms the student's across all selected criteria.

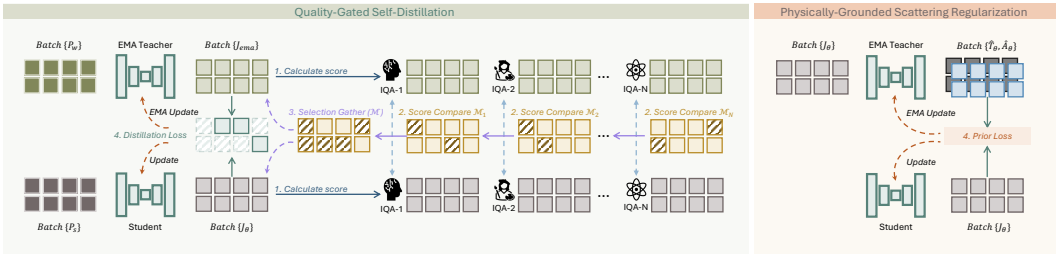

Figure 3: Pipeline of Selective Self-Distillation Adaptation (SSDA). QGSD selects teacher pseudo-labels only when they outperform the student under all selected perceptual metrics, while PGSR uses the teacher to audit residual haze in the student's output through estimated scattering variables.

Specifically, we define a binary update mask $\mathcal{M}$ for each sample:

$$\mathcal{M} = \prod_{i=1}^{N} \mathbb{I}\left(\mathcal{Q}_i(J_{\text{ema}}) \succ \mathcal{Q}_i(J_\theta)\right), \tag{13}$$

where $\mathbb{I}(\cdot)$ is the indicator function. In our implementation, $\Omega = \{\text{MUSIQ}, \text{NIMA}\}$ [44, 45] serves as a comprehensive perceptual safeguard, evaluating both multi-scale visual quality and high-level aesthetics. Gated by $\mathcal{M}$, our selective distillation objective integrates a pixel-wise loss and a contrastive perceptual loss:

$$\mathcal{L}_{\text{distill}} = \mathcal{M} \cdot (\|J_\theta - \text{sg}(J_{\text{ema}})\|_1 + \lambda_c \frac{d(J_\theta, \text{sg}(J_{\text{ema}}))}{d(J_\theta, P_s) + \epsilon}), \tag{14}$$

where $\text{sg}(\cdot)$ denotes the stop-gradient operation, and $d(\cdot,\cdot)$ computes the LPIPS distance. By restricting updates to pseudo-labels that pass all quality criteria, this intersection logic reduces the risk of unreliable supervision during adaptation. Because QGSD relies entirely on final dehazed outputs, it functions as a model-agnostic module for robust real-world adaptation.

Physically-Grounded Scattering Regularization (PGSR). While quality-gated distillation provides reliable external perceptual guidance, it does not inherently enforce internal physical consistency. Since transmission and atmospheric light estimation encode explicit scene-dependent scattering cues, they provide physically meaningful constraints for haze-free image reconstruction. To exploit this, we introduce a self-reinforced scattering prior. We utilize the stable teacher network $\mathcal{F}_{\text{ema}}$ to physically audit the student's dehazed prediction J_θ for any residual haze:

$$(\hat{J}_{\text{ema}}, \hat{T}_{\text{ema}}, \hat{A}_{\text{ema}}) = \mathcal{F}_{\text{ema}}(J_\theta). \tag{15}$$

If J_θ is completely haze-free, the estimated transmission $\hat{T}_{\text{ema}}$ should ideally approach $\mathbf{1}$. However, uniformly penalizing $\hat{T}_{\text{ema}}$ often induces artifacts in smooth or overexposed regions, where the scattering prior is less reliable. To address this, we construct a spatially aware brightness-and-atmospheric light weight map, W_{BAW}, for reliable physical auditing:

$$W_{\text{BAW}} = \max(W_{\text{dist}}, W_{\text{tex}}) \odot W_{\text{dist}} \odot W_{\text{high}}, \tag{16}$$

Here, W_{dist} serves as the primary reliability gate by favoring pixels sufficiently separated from the estimated atmospheric light. W_{tex} enhances textured regions, while W_{high} suppresses bright or near-saturated pixels. These three terms are defined as:

$$W_{\text{dist}} = \mathcal{S}\left(\frac{1}{3}\sum_c \left|J_\theta^c - \hat{A}_{\text{ema}}^c\right|\right), \quad W_{\text{tex}} = \mathcal{S}(\|\nabla \text{Gray}(J_\theta)\|), \quad W_{\text{high}} = 1 - \mathcal{S}(\max_c J_\theta^c). \tag{17}$$

In this way, W_{BAW} uses atmospheric light deviation as the main condition, texture as an auxiliary cue, and highlight masking as a final exclusion term.

Guided by W_{BAW}, we further apply a robust directional hinge loss to $\hat{T}_{\text{ema}}$. Instead of rigidly forcing the transmission toward $\mathbf{1}$, which may over-suppress natural atmospheric perspective and introduce visual artifacts, we adopt a relaxed target $T_{\text{target}} = 0.9$:

$$\mathcal{L}_{\text{prior}} = \frac{1}{\|W_{\text{BAW}}\|_1 + \epsilon} \sum (W_{\text{BAW}} \odot \max(0, T_{\text{target}} - \hat{T}_{\text{ema}})). \tag{18}$$

Table 1: Quantitative results on RTTS [46] and Fattal's [7] datasets. The best result of each metric is marked in **red**, and the second-best result is marked in **blue**.

Method	Venue	RTTS [46]			Fattal's [7]		
		FADE↓	PAQ2PIQ↑	CLIPIQA↑	FADE↓	PAQ2PIQ↑	CLIPIQA↑
Hazy Inputs	—	2.484	66.05	0.388	1.061	71.54	0.506
PDN [3]	ECCV 2018	0.876	67.11	0.319	0.316	73.73	0.557
MSBDN [12]	CVPR 2020	1.363	66.82	0.356	0.557	72.79	0.520
Dehamer [18]	CVPR 2022	1.895	66.70	0.365	0.667	72.73	0.496
DAD [36]	CVPR 2020	1.130	66.93	0.254	0.484	71.56	0.413
PSD [25]	CVPR 2021	0.920	72.06	0.275	0.416	76.02	0.547
D4 [35]	CVPR 2022	1.358	66.84	0.340	0.411	73.13	0.561
RIDCP [31]	CVPR 2023	0.944	70.82	0.337	0.408	74.64	0.470
CORUN [4]	NeurIPS 2024	0.824	72.95	**0.450**	0.338	76.28	**0.564**
CoA [47]	CVPR 2025	0.859	70.38	0.349	0.314	74.27	0.536
HazeFlow [33]	ICCV 2025	**0.583**	**72.96**	0.437	**0.264**	**76.44**	0.549
PRISM	—	**0.470**	**74.05**	**0.456**	**0.225**	**76.86**	**0.590**

By penalizing underestimated transmission only in reliable, non-saturated, atmospheric-light-separated regions, this prior encourages physically consistent scenes with fewer artifacts.

Overall Adaptation Objective. The complete objective for the SSDA stage seamlessly integrates the QGSD, the PGSR, and the atmospheric total variation regularization defined in Stage 1:

$$\mathcal{L}_{\text{adapt}} = \mathcal{L}_{\text{distill}} + \mathcal{L}_{\text{prior}} + \lambda_A^{\text{tv}} \sum_x \|\nabla A(x)\|_1. \tag{19}$$

During adaptation, the student updates the EMA teacher, and the teacher provides quality-gated pseudo-labels and scattering-based residual-haze audits for subsequent iterations. This controlled loop is designed to improve real-world adaptation while limiting updates from low-quality pseudo-labels.

4 Experiments

4.1 Experimental Setup

We evaluate PRISM on real-world dehazing benchmarks, using unlabeled real hazy images for adaptation and standard test sets for evaluation. Experiments are mainly conducted on RTTS [46], with cross-dataset results on Fattal's dataset [7].

4.2 Comparative Evaluation

Quantitative Comparison. On the RTTS benchmark (Table 1), PRISM consistently outperforms existing methods, including physics-based models such as CORUN [4] and generative methods such as HazeFlow [33]. PRISM achieves the lowest FADE score, indicating its strong ability to remove dense real-world haze, while also obtaining the best results across all adopted non-reference perceptual metrics. On the out-of-distribution Fattal's dataset, PRISM also maintains strong generalization, achieving the best FADE score and leading perceptual quality. These results demonstrate the effectiveness and robustness of PRISM under the adopted real-world dehazing protocol.

Qualitative Comparison. As shown in Figures 1, 4 and 8, PRISM visually outperforms state-of-the-art methods. Baselines like D4 [35] and RIDCP [31] struggle with dense, non-uniform haze, leaving a noticeable residual veil. While CORUN [4] and CoA [47] improve visibility, they often introduce unnatural color shifts and dark artifacts. Similarly, HazeFlow [33] exhibits severe color degradation, hallucinating unrealistic purple hues in background skies. In contrast, PRISM removes substantial haze while preserving more natural colors, sharp textures, and illumination, supporting the effectiveness of our physically constrained adaptation framework.

4.3 Ablation Study

Ablations on PSAR modules. As reported in Table 2, removing f_J^k degrades all metrics, confirming the importance of scene-radiance refinement. Removing f_T^k slightly improves PAQ2PIQ but worsens



Figure 4: Visual comparison on the RTTS dataset [46] and Fattal's dataset [7].

Table 2: Ablation studies on data pipeline and PSAR.

Datasets	Metrics	w/o non-uni	w/o f_T^k	w/o f_J^k	w/ f_A^k	w/o *spatial* A	PSAR	PRISM
RTTS	FADE↓	0.955	0.875	1.404	0.885	0.927	0.869	0.470
	PAQ2PIQ↑	69.84	71.01	60.92	70.56	70.31	70.43	74.05
	CLIPIQA↑	0.395	0.380	0.259	0.356	0.381	0.436	0.456

Table 3: Ablation studies on PSAR stage and SSDA.

Datasets	Metrics	1 *Stage*	2 *Stage*	6 *Stage*	PSAR	w/o QGSD	w/o PGSR	PRISM
RTTS	FADE↓	0.998	0.931	0.840	0.869	0.549	0.652	0.470
	PAQ2PIQ↑	70.11	70.52	70.68	70.43	73.19	73.04	74.05
	CLIPIQA↑	0.355	0.408	0.370	0.436	0.442	0.463	0.456

Table 4: Generalization studies of SSDA.

Datasets	Metrics	DGUNet [48]	with QGSD	CORUN$_{s1}$ [4]	with SSDA	PSAR	PRISM
RTTS	FADE↓	1.111	0.940	1.044	0.745	0.869	0.470
	PAQ2PIQ↑	69.92	72.48	70.62	73.56	70.43	74.05
	CLIPIQA↑	0.360	0.375	0.419	0.453	0.436	0.456

FADE and CLIPIQA, so we keep it for a better overall trade-off. Adding f_A^k degrades FADE and CLIPIQA, supporting our closed-form proximal A update. Replacing spatial A with a global constant consistently degrades all metrics, confirming the need for local atmospheric-light estimation.

Ablations on haze synthesis. As shown in Table 2, replacing our online non-uniform haze synthesis with a standard uniform synthesis pipeline (w/o non-uni) leads to clear performance degradation. This suggests that exposing the model to heterogeneous density and depth-aware variations during pretraining helps it capture complex real-world haze patterns.

Ablations on stage number. The stage number in a deep unfolding network controls the trade-off between computation and reconstruction quality. We evaluate PSAR with $k \in \{1, 2, 4, 6\}$. As reported in Table 3, 4 stages already provide strong dehazing quality. Further increasing k does not bring consistent gains in dehazing and perceptual quality, while it increases model depth and optimization difficulty, which may slow convergence and accumulate stage-wise errors.

Effect and Generalization of SSDA. As shown in Table 3, both QGSD and PGSR improve PRISM adaptation, while removing either weakens the balance between haze removal and perceptual quality. Without PGSR, CLIPIQA slightly increases, suggesting more natural textures, whereas FADE worsens notably, indicating weaker haze removal. This reflects the trade-off between perceptual texture quality and physical haze removal. We further evaluate transferability in Table 4. For DGUNet [48], which lacks explicit scattering variables, QGSD alone yields clear gains, while full SSDA improves all metrics for physically grounded CORUN [4]. These results show that SSDA generalizes across architectures, with greater gains when physical variables are available.



Figure 5: Visualization of different training phases.

Table 5: Additional ablations on atmospheric-light rendering and EMA teacher update.

Datasets	Metrics	S.R. A_{rend}	S.R. $A_{\text{rend}} + \ell_1$	$A_{\text{rend}} + \ell_1$	PSAR	Fixed Teacher	w/o EMA	PRISM
RTTS	FADE↓	0.918	0.920	0.894	0.869	0.583	0.526	0.470
	PAQ2PIQ↑	69.63	69.11	70.18	70.43	72.67	73.12	74.05
	CLIPIQA↑	0.388	0.376	0.396	0.436	0.438	0.447	0.456

Ablation on EMA Update. In self-distillation, low-quality teacher predictions may propagate when the teacher and student become too tightly coupled. Table 5 compares two teacher-update controls. *Fixed Teacher* freezes the teacher after synthetic pretraining, removing feedback from the adapted student while retaining QGSD and PGSR losses. *w/o EMA* removes temporal weight averaging and directly uses the online model for target generation and physical auditing. The results indicate that PRISM's EMA teacher stabilizes candidate targets under SSDA rather than merely reinforcing its own pseudo-labels through unchecked teacher-student coupling.

Ablation on Atmospheric Light Offset and Supervision. Direct ℓ_1 supervision on A_{rend} provides no gain and may slightly degrade performance by restricting the model's representation capacity and adaptability. More complex pixel-level spatial random transformations also yield no improvement and may introduce content-irrelevant counterfactual illumination patterns that hurt performance. We therefore omit ℓ_1 supervision on A_{rend} and retain its current random pattern as an atmospheric-light shift to improve adaptation to real scenes with spatially varying haze and illumination.

Visualization of Different Training Phases As shown in Figure 5, pretrained PSAR already removes much of the input haze but leaves residual haze and color distortion, especially in heavily degraded regions. After SSDA adaptation, PRISM produces fewer haze artifacts, more natural colors, and better detail preservation in dense haze or strong color-cast areas. These results demonstrate that selective self-distillation and residual haze auditing further improve PSAR on real-world hazy images.

5 Conclusion

In this paper, we propose PRISM, a unified understanding and restoration framework for real-world image dehazing. In PRISM, we introduce PSAR to jointly optimize the clear scene and scattering variables via an unfolded proximal process. To bridge the synthetic-to-real domain gap, we design an online non-uniform haze synthesis pipeline and a SSDA scheme. This enables quality-gated real-image adaptation while using estimated scattering variables to regularize residual haze removal. Experiments show that PRISM is competitive under the adopted real-world dehazing protocol, delivering dehazed results with fewer artifacts in complex scenes.

Acknowledgments and Disclosure of Funding

This work was supported by STI 2030-Major Projects under Grant 2021ZD0201404 and ZDSYS20210623092001004. The computing resources used in this work were provided by the ICLab, Tsinghua University. We sincerely thank reviewers for their valuable comments.

References

[1] Chunming He, Rihan Zhang, Fengyang Xiao, Chenyu Fang, Longxiang Tang, Yulun Zhang, Linghe Kong, Deng-Ping Fan, Kai Li, and Sina Farsiu. Run: Reversible unfolding network for concealed object segmentation. *ICML*, 2025. 2

[2] Chunming He, Fengyang Xiao, Rihan Zhang, Chengyu Fang, Deng-Ping Fan, and Sina Farsiu. Reversible unfolding network for concealed visual perception with generative refinement. *arXiv preprint arXiv:2508.15027*, 2025. 2

[3] Dong Yang and Jian Sun. Proximal dehaze-net: A prior learning-based deep network for single image dehazing. In *Proceedings of the european conference on computer vision (ECCV)*, pages 702–717, 2018. 2, 3, 8, 19

[4] Chengyu Fang, Chunming He, Fengyang Xiao, Yulun Zhang, Longxiang Tang, Yuelin Zhang, Kai Li, and Xiu Li. Real-world image dehazing with coherence-based pseudo labeling and cooperative unfolding network. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024. 2, 3, 8, 9, 19, 20

[5] Kaiming He, Jian Sun, and Xiaoou Tang. Single image haze removal using dark channel prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 33(12):2341–2353, 2010. 2

[6] Qingsong Zhu, Jiaming Mai, and Ling Shao. Single image dehazing using color attenuation prior. In *British Machine Vision Conference (BMVC)*. Citeseer, 2014. 2

[7] Raanan Fattal. Dehazing using color-lines. *ACM transactions on graphics (TOG)*, 34(1):1–14, 2014. 2, 8, 9, 18

[8] Wenqi Ren, Si Liu, Hua Zhang, Jinshan Pan, Xiaochun Cao, and Ming-Hsuan Yang. Single image dehazing via multi-scale convolutional neural networks. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 154–169. Springer, 2016. 2

[9] 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 (TIP)*, 25(11):5187–5198, 2016. 18

[10] Boyi Li, Xiulian Peng, Zhangyang Wang, Jizheng Xu, and Dan Feng. Aod-net: All-in-one dehazing network. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pages 4770–4778, 2017. 18

[11] Xiaohong Liu, Yongrui Ma, Zhihao Shi, and Jun Chen. Griddehazenet: Attention-based multi-scale network for image dehazing. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pages 7314–7323, 2019. 3, 18

[12] 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 (CVPR)*, pages 2157–2167, 2020. 3, 8, 18, 19

[13] Yuxin Feng, Zhuo Su, Long Ma, Xin Li, Risheng Liu, and Fan Zhou. Bridging the gap between haze scenarios: A unified image dehazing model. *IEEE Transactions on Circuits and Systems for Video Technology*, 34(11):11070–11085, 2024.

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

[15] Risheng Liu, Shichao Cheng, Long Ma, Xin Fan, and Zhongxuan Luo. Deep proximal unrolling: Algorithmic framework, convergence analysis and applications. *IEEE Transactions on Image Processing*, 28(10):5013–5026, 2019. 2

[16] 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*, pages 10551–10560, 2021. 2, 3

[17] Xu Qin, Zhilin Wang, Yuanchao Bai, Xiaodong Xie, and Huizhu Jia. Ffa-net: Feature fusion attention network for single image dehazing. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(07), 2020. 18

[18] 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 (CVPR)*, pages 5812–5820, 2022. 3, 8, 19

[19] Yuda Song, Zhuqing He, Hui Qian, and Xin Du. Vision transformers for single image dehazing. *IEEE Transactions on Image Processing*, 32:1927–1941, 2023. 2, 18

[20] Junkai Fan, Hao Zheng, Kun Wang, Zhiqiang Yan, Jianjun Qian, Heyou Chang, Jun Li, and Jian Yang. Physics-guided posterior sampling for diffusion-based real-world dehazing and image enhancement. *IEEE Transactions on Circuits and Systems for Video Technology*, 2026. 3

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

[22] Ruiyi Wang, Yushuo Zheng, Zicheng Zhang, Chunyi Li, Shuaicheng Liu, Guangtao Zhai, and Xiaohong Liu. Learning hazing to dehazing: Towards realistic haze generation for real-world image dehazing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 23091–23100, 2025.

[23] Yunwei Lan, Zhigao Cui, Chang Liu, Jialun Peng, Nian Wang, Xin Luo, and Dong Liu. Exploiting diffusion prior for real-world image dehazing with unpaired training. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 4455–4463, 2025. 3

[24] Tian Ye, Yunchen Zhang, Mingchao Jiang, Liang Chen, Yun Liu, Sixiang Chen, and Erkang Chen. Perceiving and modeling density for image dehazing. In *European conference on computer vision*, pages 130–145. Springer, 2022. 3

[25] Zeyuan Chen, Yangchao Wang, Yang Yang, and Dong Liu. Psd: Principled synthetic-to-real dehazing guided by physical priors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 7180–7189, 2021. 3, 8, 19

[26] Yuwei Qiu, Kaihao Zhang, Chenxi Wang, Wenhan Luo, Hongdong Li, and Zhi Jin. Mb-taylorformer: Multi-branch efficient transformer expanded by taylor formula for image dehazing. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12802–12813, 2023.

[27] Yu Zheng, Jiahui Zhan, Shengfeng He, Junyu Dong, and Yong Du. Curricular contrastive regularization for physics-aware single image dehazing. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5785–5794, 2023.

[28] Yafei Zhang, Shen Zhou, and Huafeng Li. Depth information assisted collaborative mutual promotion network for single image dehazing. *arXiv preprint arXiv:2403.01105*, 2024.

[29] Yongzhen Wang, Xuefeng Yan, Fu Lee Wang, Haoran Xie, Wenhan Yang, Xiao-Ping Zhang, Jing Qin, and Mingqiang Wei. Ucl-dehaze: Toward real-world image dehazing via unsupervised contrastive learning. *IEEE Transactions on Image Processing*, 33:1361–1374, 2024. 3

[30] Yuxin Feng, Long Ma, Xiaozhe Meng, Fan Zhou, Risheng Liu, and Zhuo Su. Advancing real-world image dehazing: Perspective, modules, and training. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(12):9303–9320, 2024. 3, 20

[31] Ruiqi Wu, Zhengpeng Duan, Chunle 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*, 2023. 3, 5, 8, 18, 19

[32] Jing Wang, Songtao Wu, Zhiqiang Yuan, Qiang Tong, and Kuanhong Xu. Frequency compensated diffusion model for real-scene dehazing. *Neural Networks*, 175:106281, 2024. 3

[33] Junseong Shin, Seungwoo Chung, Yunjeong Yang, and Tae Hyun Kim. Hazeflow: Revisit haze physical model as ode and non-homogeneous haze generation for real-world dehazing. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6263–6272, 2025. 3, 8, 18, 19

[34] 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 Shai Avidan, Gabriel Brostow, Moustapha Cissé, Giovanni Maria Farinella, and Tal Hassner, editors, *Computer Vision – ECCV 2022*, pages 632–648, Cham, 2022. Springer Nature Switzerland. 3

[35] Yang Yang, Chaoyue Wang, Risheng Liu, Lin Zhang, Xiaojie Guo, and Dacheng Tao. Self-augmented unpaired image dehazing via density and depth decomposition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2037–2046, 2022. 8, 19

[36] Yuanjie Shao, Lerenhan Li, Wenqi Ren, Changxin Gao, and Nong Sang. Domain adaptation for image dehazing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2808–2817, 2020. 8, 19

[37] Yi Li, Yi Chang, Yan Gao, Changfeng Yu, and Luxin Yan. Physically disentangled intra- and inter-domain adaptation for varicolored haze removal. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5841–5850, 2022.

[38] Fu-Jen Tsai, Yan-Tsung Peng, Yen-Yu Lin, and Chia-Wen Lin. Phatnet: A physics-guided haze transfer network for domain-adaptive real-world image dehazing. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5591–5600, 2025.

[39] 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, pages 2906–2914, 2025. 3

[40] Ian J Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron C Courville, and Yoshua Bengio. Generative adversarial nets. In *Neural Information Processing Systems (NeruIPS)*, 2014. 3

[41] 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, 2024. 3

[42] Ming Tong, Yongzhen Wang, Peng Cui, Xuefeng Yan, and Mingqiang Wei. Semi-uformer: Semi-supervised uncertainty-aware transformer for image dehazing, 2022. 3

[43] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 586–595, 2018. 6

[44] Junjie Ke, Qifei Wang, Yilin Wang, Peyman Milanfar, and Feng Yang. Musiq: Multi-scale image quality transformer. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 5148–5157, 2021. 7, 19

[45] Hossein Talebi and Peyman Milanfar. Nima: Neural image assessment. *IEEE Transactions on Image Processing (TIP)*, 27(8):3998–4011, 2018. 7, 19

[46] 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. 8, 9, 18, 19

[47] Long Ma, Yuxin Feng, Yan Zhang, Jinyuan Liu, Weimin Wang, Guang-Yong Chen, Chengpei Xu, and Zhuo Su. Coa: Towards real image dehazing via compression-and-adaptation. In *Proceedings of the computer vision and pattern recognition conference*, pages 11197–11206, 2025. 8, 19

[48] Chong Mou, Qian Wang, and Jian Zhang. Deep generalized unfolding networks for image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17399–17410, 2022. 9

[49] 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 (CVPRW)*, pages 444–445, 2020. 18

[50] Lark Kwon Choi, Jaehee You, and Alan Conrad Bovik. Referenceless prediction of perceptual fog density and perceptual image defogging. *IEEE Transactions on Image Processing (TIP)*, 24(11):3888–3901, 2015. 18, 19

[51] Zhenqiang Ying, Haoran Niu, Praful Gupta, Dhruv Mahajan, Deepti Ghadiyaram, and Alan Bovik. From patches to pictures (paq-2-piq): Mapping the perceptual space of picture quality. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 3575–3585, 2020. 18, 19

[52] Jianyi Wang, Kelvin CK Chan, and Chen Change Loy. Exploring clip for assessing the look and feel of images. In *Proceedings of the AAAI conference on artificial intelligence*, volume 37, pages 2555–2563, 2023. 18, 19

[53] Dongdong Chen, Mingming He, Qingnan Fan, Jing Liao, Liheng Zhang, Dongdong Hou, Lu Yuan, and Gang Hua. Gated context aggregation network for image dehazing and deraining. In *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1375–1383, 2019. 18

[54] 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 (CVPR)*, pages 16180–16189. IEEE, 2021. 18

[55] Hu Yu, Jie Huang, Kaiwen Zheng, and Feng Zhao. High-quality image dehazing with diffusion model, 2024. 18

[56] Sidi Yang, Tianhe Wu, Shuwei Shi, Shanshan Lao, Yuan Gong, Mingdeng Cao, Jiahao Wang, and Yujiu Yang. Maniqa: Multi-dimension attention network for no-reference image quality assessment. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1190–1199. IEEE, 2022. 19

[57] Weixia Zhang, Guangtao Zhai, Ying Wei, Xiaokang Yang, and Kede Ma. Blind image quality assessment via vision-language correspondence: A multitask learning perspective. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 14071–14081, 2023. 19

[58] Chengyu Fang, Heng Guo, Zheng Jiang, Chunming He, Xiu Li, and Minfeng Xu. Photon: Speedup volume understanding with efficient multimodal large language models. In *The Fourteenth International Conference on Learning Representations*, 2026. 20

[59] Jeya Maria Jose Valanarasu, Rajeev Yasarla, and Vishal M Patel. Transweather: Transformer-based restoration of images degraded by adverse weather conditions. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2353–2363, 2022. 20

[60] John Y Chiang and Ying-Ching Chen. Underwater image enhancement by wavelength compensation and dehazing. *IEEE transactions on image processing*, 21(4):1756–1769, 2011. 20

[61] Long Ma, Tengyu Ma, Risheng Liu, Xin Fan, and Zhongxuan Luo. Toward fast, flexible, and robust low-light image enhancement. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5637–5646, 2022. 20

[62] Yeying Jin, Wenhan Yang, and Robby T Tan. Unsupervised night image enhancement: When layer decomposition meets light-effects suppression. In *European conference on computer vision*, pages 404–421. Springer, 2022.

[63] Yeying Jin, Beibei Lin, Wending Yan, Yuan Yuan, Wei Ye, and Robby T Tan. Enhancing visibility in nighttime haze images using guided apsf and gradient adaptive convolution. In *Proceedings of the 31st ACM international conference on multimedia*, pages 2446–2457, 2023.

[64] Chunming He, Chengyu Fang, Yulun Zhang, Kai Li, Longxiang Tang, Chenyu You, Fengyang Xiao, Zhenhua Guo, and Xiu Li. Reti-diff: Illumination degradation image restoration with retinex-based latent diffusion model. *arXiv preprint arXiv:2311.11638*, 2023. 20

[65] Xingyuan Li, Jinyuan Liu, Zhixin Chen, Yang Zou, Long Ma, Xin Fan, and Risheng Liu. Contourlet residual for prompt learning enhanced infrared image super-resolution. In *European Conference on Computer Vision*, pages 270–288. Springer, 2024. 20

[66] Xingyuan Li, Songcheng Du, Yang Zou, HaoYuan Xu, Zhiying Jiang, and Jinyuan Liu. Unifusion: A unified image fusion framework with robust representation and source-aware preservation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 33869–33880, 2026. 20

[67] Jinyuan Liu, Xingyuan Li, Zirui Wang, Zhiying Jiang, Wei Zhong, Wei Fan, and Bin Xu. Promptfusion: Harmonized semantic prompt learning for infrared and visible image fusion. *IEEE/CAA Journal of Automatica Sinica*, 2024.

[68] Chengyu Fang, Chunming He, Longxiang Tang, Yuelin Zhang, Chenyang Zhu, Yuqi Shen, Chubin Chen, Guoxia Xu, and Xiu Li. Integrating extra modality helps segmentor find camouflaged objects well. *arXiv preprint arXiv:2502.14471*, 2025. 20

[69] Xingyuan Li, Haoyuan Xu, Xingyue Zhu, Jun Ma, Yang Zou, Zhiying Jiang, and Jinyuan Liu. Uncertainty-aware spatial-frequency registration and fusion for infrared and visible images. *arXiv preprint arXiv:2605.13049*, 2026. 20

[70] Chubin Chen, Jiashu Zhu, Xiaokun Feng, Nisha Huang, Chen Zhu, Meiqi Wu, Fangyuan Mao, Jiahong Wu, Xiangxiang Chu, and Xiu Li. Stochastic self-guidance for training-free enhancement of diffusion models. In *International Conference on Learning Representations*, volume 2026, pages 134462–134490, 2026. 20

[71] Chunming He, Kai Li, Yachao Zhang, Ziyun Yang, Longxiang Tang, Yulun Zhang, Linghe Kong, and Sina Farsiu. Segment concealed object with incomplete supervision. *TPAMI*, 2025. 20

[72] Chunming He, Dingming Zhang, Fengyang Xiao, Zhiwen Cao, and Sina Farsiu. Supervision-adaptive curriculum and guarded anti-curriculum learning for dense prediction. *arXiv preprint arXiv:2610.29519*, 2026. 20

[73] Chunming He, Dingming Zhang, Fengyang Xiao, Xinge Guo, Jingjia Feng, Yiqing Wang, and Sina Farsiu. Toward multimodal understanding in low-level vision: The rise of vision-language models. 2026. 20

[74] Sujie Hu, Huaisong Zhang, Pujing Yang, Qian Wang, Xiaomin Li, Xinhang Leng, Chen Li, Jing Lyu, Xiu Li, and Ling Pan. Blame errant words, not the canvas: Defect-aware spatial credit assignment for visual text generation, 2026. 20

[75] Sujie Hu, Chubin Chen, Jiashu Zhu, Jiahong Wu, Xiangxiang Chu, and Xiu Li. Embedding-perturbed exploration preference optimization for flow models. In *Forty-third International Conference on Machine Learning*, 2026. 20

[76] Chubin Chen, Sujie Hu, Jiashu Zhu, Meiqi Wu, Jintao Chen, Yanxun Li, Nisha Huang, Chengyu Fang, Jiahong Wu, Xiangxiang Chu, et al. Taming preference mode collapse via directional decoupling alignment in diffusion reinforcement learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12775–12786, 2026. 20

[77] Xingyuan Li, Zirui Wang, Yang Zou, Zhixin Chen, Jun Ma, Zhiying Jiang, Long Ma, and Jinyuan Liu. Difiisr: A diffusion model with gradient guidance for infrared image super-resolution. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 7534–7544, 2025. 20

A Detailed Derivations of the Proximal Updates

In this section, we provide the derivations for the closed-form proximal updates used in PSAR. At stage k, the network first computes the atmospheric light and transmission proximal points, keeps the atmospheric light branch purely proximal with $A_k = \bar{A}_k$, refines the transmission as $T_k = \bar{T}_k + f_T^{(k)}(\bar{T}_k, J_{k-1})$, and then computes the clear-scene proximal point under the refined transmission T_k and atmospheric light A_k.

Since the scattering data term is defined pixel-wise and the quadratic proximal penalties are also separable across pixels, all three subproblems can be solved independently at each pixel. Therefore, it is sufficient to derive the update at a generic pixel. For clarity, we omit the spatial index (x) in this appendix unless needed. Thus, throughout this section, $P, J, A \in \mathbb{R}^3$ denote RGB vectors at one pixel, and $T \in \mathbb{R}$ denotes the corresponding scalar transmission. Because $\lambda_A, \lambda_T, \lambda_J > 0$, each subproblem is a strictly convex quadratic problem with a unique minimizer, which is obtained by setting the first-order derivative or gradient to zero.

A.1 Closed-Form Solution for the Atmospheric Light Update $\bar{A}_k$

With J_{k-1} and T_{k-1} fixed, the atmospheric light proximal subproblem at one pixel is

$$E(A) = \frac{1}{2}\left\|J_{k-1}T_{k-1} + (1 - T_{k-1})A - P\right\|_2^2 + \frac{\lambda_A}{2}\|A - A_{k-1}\|_2^2. \tag{20}$$

Here the optimization variable is the RGB vector A, while J_{k-1}, T_{k-1}, P, and A_{k-1} are constants. Differentiating with respect to A gives

$$\frac{\partial E}{\partial A} = (1 - T_{k-1})\left(J_{k-1}T_{k-1} + (1 - T_{k-1})A - P\right) + \lambda_A(A - A_{k-1}). \tag{21}$$

Setting the gradient to $\mathbf{0}$ yields

$$(1 - T_{k-1})\left(J_{k-1}T_{k-1} + (1 - T_{k-1})A - P\right) + \lambda_A(A - A_{k-1}) = \mathbf{0}. \tag{22}$$

Collecting the terms involving A, we obtain

$$((1 - T_{k-1})^2 + \lambda_A)A = (1 - T_{k-1})(P - J_{k-1}T_{k-1}) + \lambda_A A_{k-1}. \tag{23}$$

Therefore, the unique minimizer is

$$\bar{A}_k = \frac{(1 - T_{k-1})(P - J_{k-1}T_{k-1}) + \lambda_A A_{k-1}}{(1 - T_{k-1})^2 + \lambda_A}. \tag{24}$$

A.2 Closed-Form Solution for the Transmission Update $\bar{T}_k$

After obtaining $A_k = \bar{A}_k$, the transmission proximal subproblem is

$$E(T) = \frac{1}{2}\left\|TJ_{k-1} + (1 - T)A_k - P\right\|_2^2 + \frac{\lambda_T}{2}(T - T_{k-1})^2. \tag{25}$$

Here the optimization variable is the scalar T, while J_{k-1}, A_k, P, and T_{k-1} are constants. To make the dependence on T explicit, we rewrite the residual as

$$TJ_{k-1} + (1 - T)A_k - P = T(J_{k-1} - A_k) + A_k - P. \tag{26}$$

Substituting this form into the objective and differentiating with respect to T gives

$$\frac{\partial E}{\partial T} = \langle J_{k-1} - A_k,\, T(J_{k-1} - A_k) + A_k - P\rangle + \lambda_T(T - T_{k-1}). \tag{27}$$

Setting the derivative to zero yields

$$\langle J_{k-1} - A_k,\, T(J_{k-1} - A_k) + A_k - P\rangle + \lambda_T(T - T_{k-1}) = 0. \tag{28}$$

Expanding the inner product gives

$$T\|J_{k-1} - A_k\|_2^2 + \langle J_{k-1} - A_k,\, A_k - P\rangle + \lambda_T T - \lambda_T T_{k-1} = 0. \tag{29}$$

To match the form used in the main text, we rewrite the expression using $A_k - J_{k-1}$. Since

$$J_{k-1} - A_k = -(A_k - J_{k-1}),$$

we have

$$\|J_{k-1} - A_k\|_2^2 = \|A_k - J_{k-1}\|_2^2$$

and

$$\langle J_{k-1} - A_k,\, A_k - P\rangle = -\langle A_k - J_{k-1},\, A_k - P\rangle .$$

Hence, the optimality condition becomes

$$T\|A_k - J_{k-1}\|_2^2 - \langle A_k - J_{k-1},\, A_k - P\rangle + \lambda_T T - \lambda_T T_{k-1} = 0. \tag{30}$$

Collecting the terms involving T, we obtain

$$T\left(\lambda_T + \|A_k - J_{k-1}\|_2^2\right) = \lambda_T T_{k-1} + \langle A_k - J_{k-1},\, A_k - P\rangle . \tag{31}$$

Therefore,

$$\bar{T}_k = \frac{\lambda_T T_{k-1} + \langle A_k - J_{k-1},\, A_k - P\rangle}{\lambda_T + \|A_k - J_{k-1}\|_2^2}. \tag{32}$$

Expanding the inner product and squared norm over the three RGB channels gives

$$\bar{T}_k = \frac{\lambda_T T_{k-1} + \sum_{c=1}^{3}(A_k^{(c)} - J_{k-1}^{(c)})(A_k^{(c)} - P^{(c)})}{\lambda_T + \sum_{c=1}^{3}(A_k^{(c)} - J_{k-1}^{(c)})^2}. \tag{33}$$

A.3 Closed-Form Solution for the Clear-Scene Update $\bar{J}_k$

After the transmission proximal update, the network refines the transmission as

$$T_k = \bar{T}_k + f_T^{(k)}(\bar{T}_k, J_{k-1}). \tag{34}$$

The clear-scene proximal step is then performed under the refined transmission T_k and the atmospheric light A_k. At one pixel, the corresponding subproblem is

$$E(J) = \frac{1}{2}\,\|T_k J + (1 - T_k)A_k - P\|_2^2 + \frac{\lambda_J}{2}\|J - J_{k-1}\|_2^2. \tag{35}$$

Here the optimization variable is the RGB vector J, while T_k, A_k, P, and J_{k-1} are constants. Differentiating with respect to J gives

$$\frac{\partial E}{\partial J} = T_k\,(T_k J + (1 - T_k)A_k - P) + \lambda_J(J - J_{k-1}). \tag{36}$$

Setting the gradient to $\mathbf{0}$ yields

$$T_k\,(T_k J + (1 - T_k)A_k - P) + \lambda_J(J - J_{k-1}) = \mathbf{0}. \tag{37}$$

Collecting the terms involving J, we obtain

$$T_k^2 J + \lambda_J J = T_k P - T_k(1 - T_k)A_k + \lambda_J J_{k-1}. \tag{38}$$

Using

$$-T_k(1 - T_k)A_k = -T_k A_k + T_k^2 A_k,$$

the above equation becomes

$$(T_k^2 + \lambda_J)J = T_k P + T_k^2 A_k - T_k A_k + \lambda_J J_{k-1}. \tag{39}$$

Therefore, the unique minimizer is

$$\bar{J}_k = \frac{T_k P + T_k^2 A_k - T_k A_k + \lambda_J J_{k-1}}{T_k^2 + \lambda_J}. \tag{40}$$

Since the denominator is scalar, it divides each RGB component of the numerator.

A.4 Summary of the Stage-Wise Variable Flow

Putting the above derivations together, one PSAR stage follows the sequence

$$\begin{cases} \bar{A}_k = \Phi_A(P, J_{k-1}, T_{k-1}, A_{k-1}), \quad A_k = \bar{A}_k, \\ \bar{T}_k = \Phi_T(P, J_{k-1}, T_{k-1}, A_k), \quad T_k = \bar{T}_k + f_T^{(k)}(\bar{T}_k, J_{k-1}), \\ \bar{J}_k = \Phi_J(P, J_{k-1}, T_k, A_k), \quad J_k = \bar{J}_k + f_J^{(k)}(\bar{J}_k, T_k, A_k), \end{cases} \tag{41}$$

where Φ_A, Φ_T, and Φ_J denote the closed-form proximal maps derived above. Therefore, the appendix derivations are fully consistent with the stage-wise computation described in the main text.

B Experiments

B.1 Experiment Setup

Data Preparation. We perform initial pre-training of PSAR using the RIDCP dataset [31] with the same protocol as HazeFlow [33]. For the adaptation phase, the RESIDE-URHI [46] subset, containing 4,807 unpaired real-world hazy images, is used to generate pseudo-labels and refine the network parameters. We evaluate PRISM on the RTTS dataset [46], which contains over 4,000 images characterized by varying resolutions and complex degradation patterns. Additionally, we report results from Fattal's dataset [7] as cross-dataset evidence. To further assess PSAR under full-reference evaluation, we report results on the paired real-world NH-HAZE dataset [49], which contains hazy and corresponding clear image pairs captured under non-homogeneous haze conditions.

Metrics. To assess the performance of the proposed method, we report the results of FADE [50], PAQ2PIQ [51], and CLIPIQA [52]. FADE measures residual haze density, while PAQ2PIQ and CLIPIQA evaluate perceptual quality and naturalness. For paired real-world evaluation on NH-HAZE [49], we further report PSNR and SSIM to measure pixel-level fidelity and structural similarity against the haze-free reference images. To further support the comparison, we provide additional results in Section B.3 with more widely used metrics.

Implementation Details. Our framework is trained on four NVIDIA A100 GPUs. The PSAR network contains four unfolding stages. We pretrain the model for 35K iterations using AdamW. The generator and discriminator learning rates are initialized to 2×10^{-4} and 4×10^{-4}, respectively, and are halved after 20K iterations. During adaptation, the learning rate is fixed at 2×10^{-5} for 1.5K fine-tuning iterations. We set $\lambda_A^{\text{tv}} = \lambda_c = \lambda_{adv} = 0.5$. For efficiency, each $f_T^{(k)}$ and $f_J^{(k)}$ is implemented by 3 UNet-style layers with $[1, 1, 1]$ blocks and 30 input channels, where channels are doubled and spatial resolution is halved with depth.

B.2 Evaluate PSAR on the Paired Real-world Dehazing Dataset

To further validate the effectiveness of PSAR, we evaluate its performance on the paired real-world dehazing dataset NH-HAZE [49].

Table 6: Quantitative results on the paired real-world dataset NH-HAZE [49].

Dataset	Metric	AOD-Net [10]	MSBDN [12]	GCANet [53]	FFANet [17]	DehazeFormer [19]
NH-HAZE [49]	PSNR↑	15.31	17.34	16.64	18.48	18.15
	SSIM↑	0.458	0.556	0.558	0.618	0.607
Dataset	Metric	DehazeNet [9]	GridDehazeNet [11]	UHD [54]	DehazeDDPM [55]	PSAR
NH-HAZE [49]	PSNR↑	11.76	17.23	16.05	19.44	20.10
	SSIM↑	0.399	0.504	0.461	0.627	0.688

NH-HAZE contains paired hazy and corresponding clear images captured under real-world conditions, providing a direct evaluation of dehazing performance. We report the results in Table 6, where PSAR achieves competitive performance compared to state-of-the-art methods, demonstrating its ability to remove haze while preserving image details in real-world scenarios.

B.3 More Comparative Results

To reduce metric-specific bias and support a more diverse comparison, we further report RTTS [46] results with four additional no-reference image quality assessment (NR-IQA) metrics, complementing the main evaluation with FADE [50], PAQ2PIQ [51], and CLIPIQA [52]: NIMA [45], MUSIQ [44], ManIQA [56], and LIQE [57]. These metrics cover complementary quality aspects: NIMA focuses more on aesthetic quality, while MUSIQ, ManIQA, and LIQE focus more on general perceptual quality. Together, they provide an additional check on whether the observed improvement remains consistent beyond the main evaluation protocol.

Table 7: Supplementary quantitative results on RTTS [46] dataset using additional IQA metrics.

Metric	Hazy Inputs	PDN [3]	MSBDN [12]	Dehamer [18]	DAD [36]	PSD [25]
NIMA↑	4.857	4.880	4.896	4.898	4.778	4.783
MUSIQ↑	53.77	52.97	54.04	53.78	49.92	56.52
ManIQA↑	0.311	0.280	0.306	0.316	0.225	0.296
LIQE↑	1.982	1.959	2.108	2.037	1.724	1.693
Metric	D4 [35]	RIDCP [31]	CORUN [4]	CoA [47]	HazeFlow [33]	PRISM
NIMA↑	4.852	5.241	5.301	5.059	5.296	5.347
MUSIQ↑	53.57	59.38	67.60	53.43	63.92	70.02
ManIQA↑	0.300	0.275	0.357	0.243	0.346	0.362
LIQE↑	2.089	2.671	3.049	1.971	2.781	3.569

Since prior dehazing studies often do not specify the NIMA and LIQE evaluation implementations used for evaluation, for NIMA, we consistently adopt the VGG16-AVA version and retest all methods under the same implementation to ensure a fair and reproducible comparison. For LIQE, we only evaluate images whose shorter side is no smaller than 224 pixels, as required by the metric, instead of resizing the inputs. This avoids potential errors from different resizing kernels and the sensitivity of LIQE to resizing. As shown in Table 7, PRISM achieves the best results across all four additional NR-IQA metrics on RTTS. This trend is consistent with the main evaluation using FADE, PAQ2PIQ, and CLIPIQA, indicating that the advantage of PRISM is not tied to a specific metric choice. The results provide complementary evidence that PRISM produces visually higher-quality dehazed images under different quality assessment models.

B.4 Visualization of the Estimated A, T, and J

In Figure 7, we visualize the inputs and outputs of PSAR. For the inputs, we initialize J with the hazy image P, set the atmospheric light A to 0.9, and initialize the transmission map T to 0.5. The outputs A, T, and J correspond to the atmospheric light map, transmission map, and dehazed image estimated by PSAR from these initial states.

The visualization shows that PSAR produces transmission maps with clear spatial variation, indicating that it can capture non-uniform haze concentration across different regions of the scene. Regions with heavier haze usually correspond to lower transmission values, while clearer regions tend to have higher transmission. Meanwhile, the estimated atmospheric light maps reflect scene-dependent illumination and color bias rather than collapsing to a trivial constant pattern. This helps separate scattering-related illumination effects from the underlying scene radiance. As a result, PSAR produces clean and visually faithful dehazed images with more natural colors and fewer visible degradation artifacts. These results suggest that PSAR learns a meaningful decomposition of the hazy observation into atmospheric light, transmission, and scene radiance, instead of merely enhancing local contrast.

B.5 More Visualization Results

We provide more visual comparisons with RIDCP [31], CORUN [4], CoA [47], and HazeFlow [33] in Figure 8. Compared with these methods, PSAR recovers cleaner scene structures, more natural contrast, and more faithful colors under challenging real-world non-uniform haze. In particular, RIDCP and CORUN may still leave noticeable residual haze in heavily degraded regions, while CoA and HazeFlow can produce local color distortion or over-enhancement in some cases. By contrast, PSAR removes haze more thoroughly while better preserving the original scene appearance.

This advantage is especially clear in regions with dense haze, uneven illumination, or strong color cast, where competing methods often struggle to balance dehazing strength and visual fidelity. Overall, PSAR achieves a better trade-off between haze removal, structure preservation, and color realism, leading to higher-quality dehazed results.

C Failure Cases Analysis

Representative failure cases are shown in Figure 6. Although PRISM produces cleaner results than previous methods, extremely dense haze near object boundaries can still lead to mild residual artifacts. These regions are challenging because strong scattering, weak contrast, and depth discontinuities are tightly mixed. Nevertheless, compared with prior methods that often leave broader haze patches or introduce visible distortions, PRISM better suppresses haze while preserving scene structures.



Figure 6: Failure cases of our proposed PSAR.

D Limitations and Future Work

PSAR is formulated around atmospheric scattering [4, 30]. Its extension to medical imaging [58], snow and rain [59], underwater imaging [60], and low-light scenes [61–64] recommends adapting the image formation model. We will also investigate transferring the model-agnostic SSDA strategy to other image restoration tasks [65] and multimodal perception [66–68], preserving complementary information while accounting for cross-modal alignment uncertainty [69]. QGSD currently selects entire pseudo-labels using image-level quality scores. Future work will explore stochastic sub-network self-guidance from generative models [70] to refine pseudo-labels, combine reliability-aware selection [71, 72] with VLM-assisted assessment [73], and use defect-aware spatial feedback inspired by visual generation [74] to localize supervision. We will also investigate preference-guided adaptation inspired by generative model alignment [75], assessing reward bias [76], source fidelity, and physical consistency across domains. Real-world evaluation should also connect perceptual improvements with downstream utility. Inspired by task-oriented image restoration [77], we will develop haze-aware evaluation protocols that jointly assess color fidelity, halo suppression, structural faithfulness, and detection or segmentation performance. These protocols will help identify when restoration gains translate into more reliable scene understanding under spatially varying haze.

E Broader Impacts

Real-world image dehazing is an important image restoration task that removes haze degradation from images captured in real-world scenarios, with applications in autonomous driving, remote sensing, security monitoring, and general photography. This paper introduces PRISM, a physically structured dehazing framework that combines online non-uniform haze synthesis with selective self-distillation adaptation, achieving strong performance on real-world dehazing benchmarks. Image dehazing methods aim to restore degraded visual observations rather than generate new semantic content, and have not been associated with direct negative societal impacts.

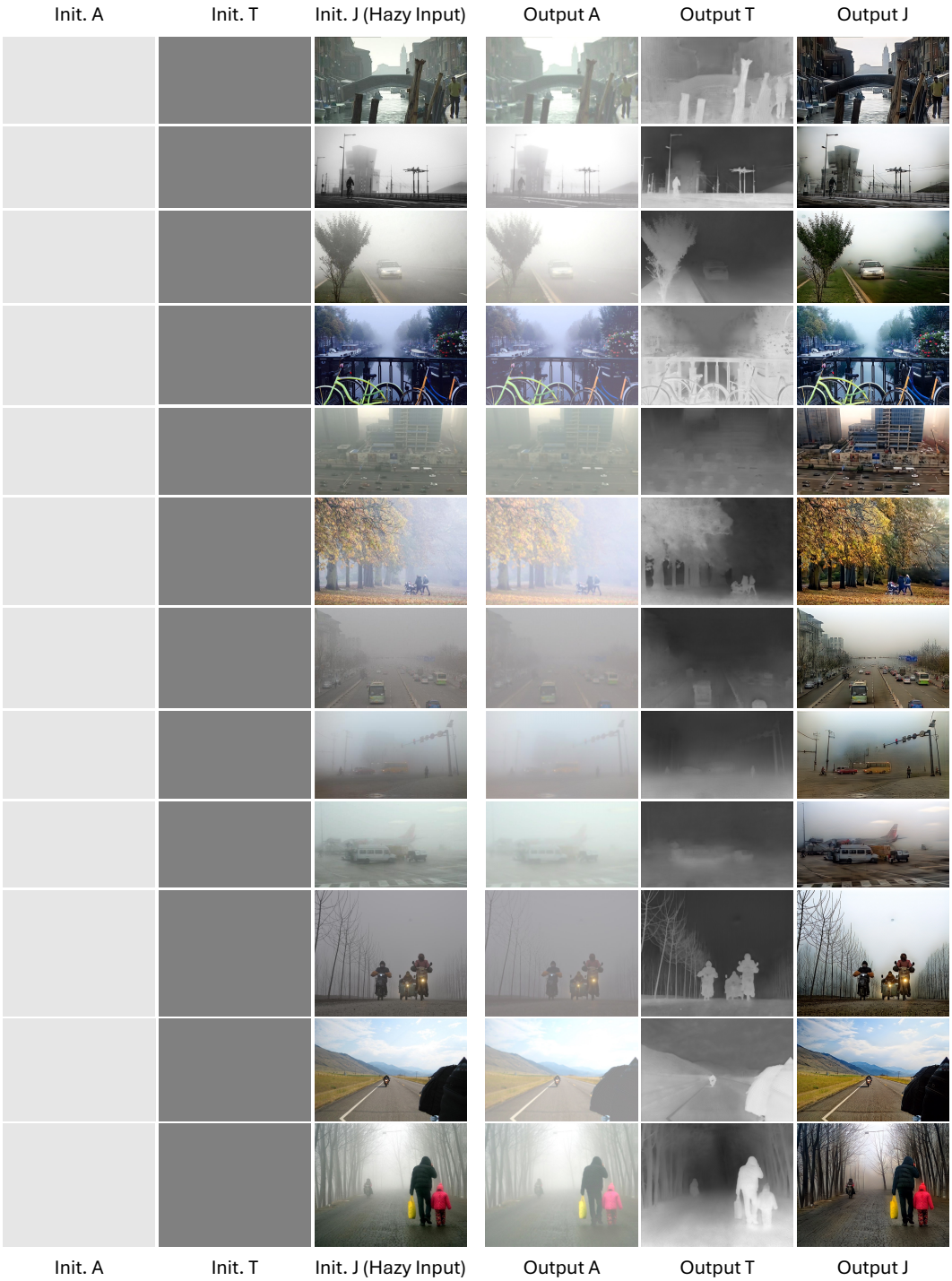

Figure 7: Visualization of the initialized inputs and the decomposed components estimated by the proposed PRISM, including atmospheric light A, transmission T, and restored scene radiance J.

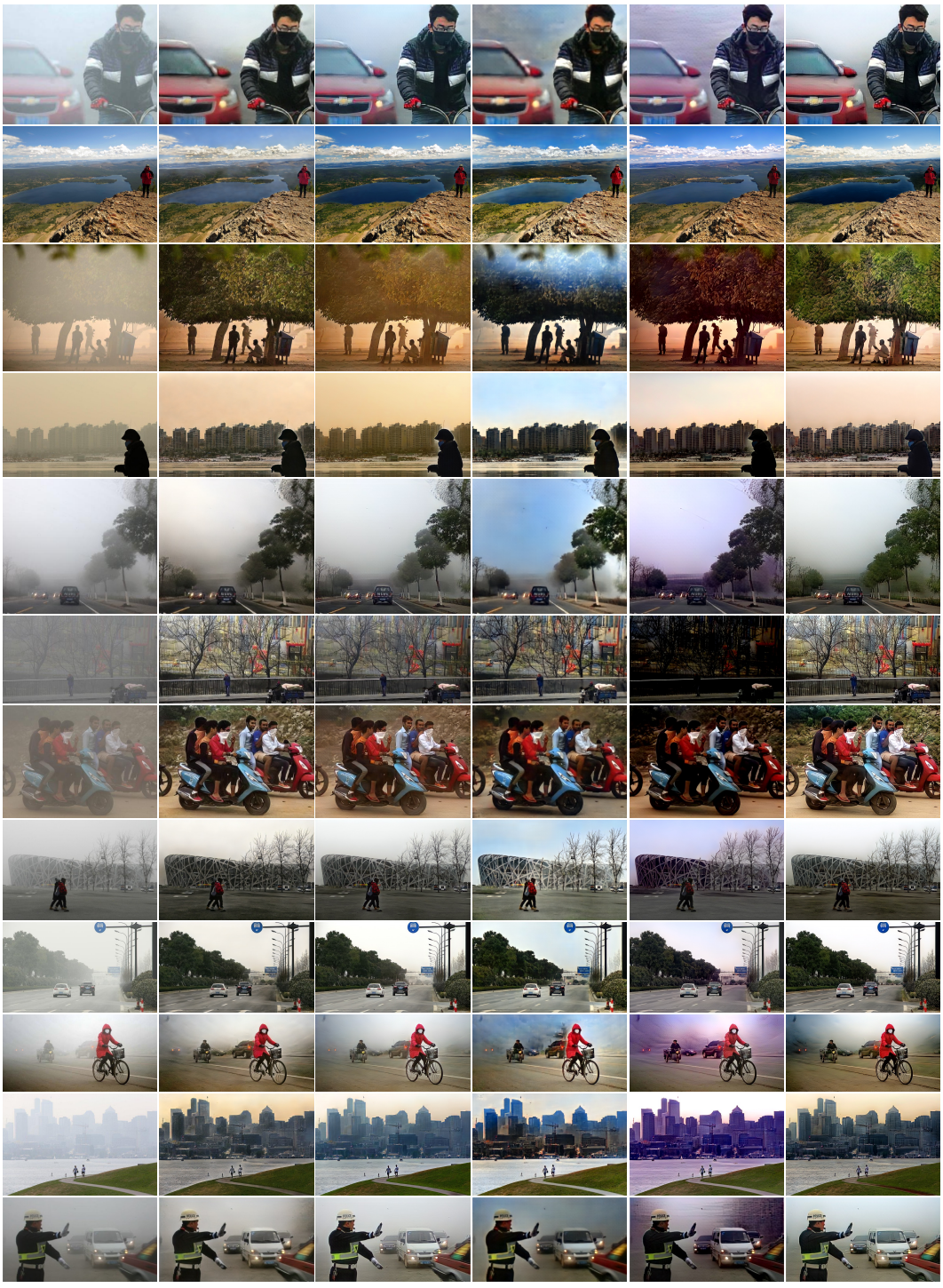

Figure 8: More visualization results of our PRISM and other state-of-the-art methods.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The abstract and section 1 state the paper's scope as real-world image dehazing and summarize PSAR, online haze synthesis, and SSDA. These claims are supported by the method in section 3, the comparative and ablation results in section 4 and tables 1 to 4, and the supplementary analyses in the appendix.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: section D discusses the current focus on haze-driven atmospheric scattering, the untested extension to other imaging and weather conditions, and the remaining limitations of real-world dehazing evaluation with existing IQA metrics.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [Yes]

Justification: The paper provides the full assumptions and derivations for the scattering objective and the closed-form proximal updates used by PSAR. The variables, positive regularization coefficients, and quadratic subproblems are defined in section 3.1, and the complete derivations for the atmospheric light, transmission, and clear-scene updates are given in section A.

Guidelines:

- The answer [N/A] means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [Yes]

Justification: The paper specifies the network structure and proximal updates in sections A and 3.1, the online haze synthesis in algorithm 1, the losses and adaptation objective in eqs. (12) and (19), and the datasets, metrics, and training hyperparameters in section B.1.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).

(d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: The current submission describes the method and experimental protocol in sections B.1 and 3 which can be followed to reproduce the results. The experiments use publicly cited datasets and standard evaluation metrics. Code and pretrained models will be released upon paper acceptance to preserve double-blind review integrity.

   Guidelines:

   - The answer [N/A] means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (https://neurips.cc/public/guides/CodeSubmissionPolicy) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (https://neurips.cc/public/guides/CodeSubmissionPolicy) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: section B.1 describe data preparation, training and adaptation datasets, evaluation benchmarks, metrics, optimizer, learning rates, training iterations, model stages, loss weights, and relevant implementation details, with additional derivation details in section A.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [No]

Justification: Following the convention in image restoration and image dehazing literature (e.g., RIDCP, CORUN,HazeFlow), the current submission reports single-run quantitative comparisons and ablations in tables 1 to 4 without error bars or statistical significance tests.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: section B.1 reports that the type and number of GPUs used for training and the iteration counts for training and adaptation.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics https://neurips.cc/public/EthicsGuidelines?

   Answer: [Yes]

   Justification: The work uses public image dehazing datasets described in section B.1 and standard model training/evaluation procedures in section B.1, does not involve private or sensitive data collection, and is intended for image restoration research consistent with the NeurIPS Code of Ethics.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.

- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

    Justification: section E discusses broader impacts of the proposed PRISM method.

    Guidelines:

    - The answer [N/A] means that there is no societal impact of the work performed.
    - If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
    - The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
    - The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
    - If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification: The paper introduces a dehazing method does not release a high-risk model, dataset, image generator, or scraped asset requiring special access controls or misuse safeguards.

    Guidelines:

    - The answer [N/A] means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [Yes]

Justification: The paper cites the existing datasets, metrics, and prior methods used for training and evaluation in sections B.1 and B.3.

Guidelines:

- The answer [N/A] means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [N/A]

Justification: The paper introduces the PRISM method and experimental pipeline in sections 3 and 4, but it does not release a new dataset, benchmark, code package, or model asset as part of the current submission. Code and pretrained checkpoints will be released upon acceptance with full documentation.

Guidelines:

- The answer [N/A] means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [N/A]

Justification: The research does not involve crowdsourcing experiments, user studies, or newly collected human-subject data.

Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

    Justification: The research does not involve crowdsourcing or human-subject experiments, so IRB approval or an equivalent human-subjects review is not applicable.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

    Answer: [N/A]

    Justification: LLMs were not used as a component of the core method development in this research. The paper's methodology does not involve LLMs at any important, original, or non-standard level.

    Guidelines:

    - The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.