UG-PRIOR: UNCERTAINTY-GUIDED PRIORS WITH ONLINE REFINEMENT FOR ZERO-SHOT LOW-LIGHT ENHANCEMENT

Anonymous authors
Paper under double-blind review

ABSTRACT

Pixel-level and distribution-level supervision for low-light image enhancement (LLIE) remain sensitive to scene and degradation in the training data, limiting generalization. Physical illumination-invariant priors bridge low/normal-light domains, posing a promising direction for improved generalization. However, their unreliable estimation from degraded observations hampers cross-domain robustness and faithful enhancement. To alleviate this issue, we propose **UG-Prior**, an Uncertainty-Guided zero-shot LLIE framework. Specifically, we propose uncertainty-guided multi-scale fusion to construct robust and informative priors with reliable structural and chromatic cues. An auxiliary decoder is proposed to regularize generative features by reconstructing clean intrinsic representations. During inference, its predictions enable online prior-image co-optimization, which refines the conditioning prior as the image evolves. Extensive experiments demonstrate that our method achieves state-of-the-art zero-shot LLIE. Relative to the recent SOTA, ZeroIDIR on LOL, our UG-Prior improves PSNR by 0.42 dB and LPIPS by 0.02, with strong generalization across diverse scenes and illumination conditions.

1 INTRODUCTION

Low-light image enhancement (LLIE) aims to improve visibility and perceptual quality while preserving scene content. Despite advances of deep-learning methods with Retinex theory (Wei et al., 2018; Wu et al., 2022; Cai et al., 2023), improved model architecture (Zamir et al., 2020; Wang et al., 2023b; Yan et al., 2025; Li et al., 2023), and generative modeling (Wang et al., 2022; Yi et al., 2023; Hou et al., 2023), supervised methods remain limited by the overfitting to scene and degradation of paired training data. Unsupervised methods relax paired supervision through distribution alignment and self-supervised constraints (Jiang et al., 2021; Yang et al., 2023; Liang et al., 2023; Jiang et al., 2024; Liu et al., 2021; Ma et al., 2022; Fu et al., 2023), but still depend on the low-light training domain. Generalizing to unseen illumination conditions remains difficult (Wang et al., 2024; Lin et al., 2025; Jiang et al., 2026).

Prior-guided enhancement introduces cross-illumination transferable knowledge to constrain the restoration process, thereby essentially alleviating the dependence on specific scenes or degradation. Exposure, spatial smoothness, and color preservation fundamentally enables enhancement without normal-light references (Guo et al., 2020; Li et al., 2021; Jung et al., 2025), however leaving enhanced exposure sensitive to manually specified parameters and training conditions (Wang et al., 2024). Based on this, some works incorporate pretrained generative priors (*e.g.*, DDPM (Ho et al., 2020), Stable Diffusion (Rombach et al., 2022)) to improve the enhancement process with more complicated constraints such as frequency decoupling (Lv et al., 2024; Lin et al., 2025), Retinex decomposition (Zhao & Shi, 2026; Shang et al., 2025; Huang et al., 2025b) and diffusion consistency (Fei et al., 2023; Jiang et al., 2026; Cho et al., 2026). Although large-scale generative models capture richer natural-image statistics, they still face the challenging generation-fidelity dilemma (Blau & Michaeli, 2018), particularly when the low-light conditions are severely degraded.

Recently, studies introduce **physics-derived illumination-invariant priors** that constrain generative enhancement (Wang et al., 2024; Huang et al., 2025a) by decoupling enhancement into two

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

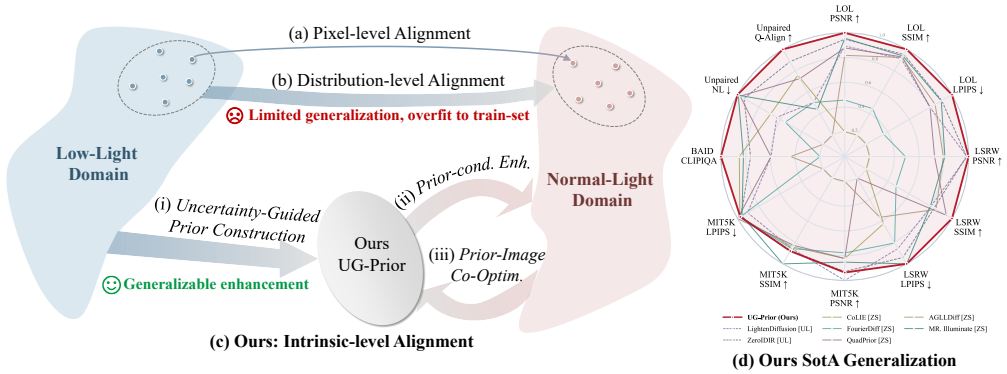


Figure 1: **Generalizable low-light image enhancement with UG-Prior.** **(a,b)** Pixel- and distribution-level alignment prone to overfitting, while **(c)** our UG-Prior bridges illumination domains through a **robust and informative intrinsic prior**. We first construct uncertainty-guided multi-scale prior, then propose decoding-regularization and online prior-image co-optimization for effective utilization. **(d)** Comparison across multiple benchmarks demonstrates that our method achieves state-of-the-art performance with superior generalization.

parts: first constructing a prior that preserves illumination-invariant scene content, then conditioning the generation of high-quality lighting with the prior. However, real-world degraded low-light observations may hamper the ideally-derived illumination invariance and leave recoverable scene cues heavily entangled with noise and color distortion, which further leads to a robustness-informativeness trade-off: suppressing unstable responses may discard useful details, whereas preserving them may amplify degradation and enlarge the cross-illumination gap.

Based on the above analysis, we identify the core challenge of prior-guided enhancement as two-fold. On one hand, we should first construct **a proper cross-illumination information bottleneck that balances robustness and informativeness**. On the other hand, we should utilize such bottleneck as a domain-bridging prior in a generalizable way, since accurate reconstruction during training does not guarantee tolerance to test-time prior discrepancy. The generator should exploit reliable scene information and limit its dependence on degradation-sensitive components.

To address these challenges, we propose **UG-Prior**, an Uncertainty-Guided framework for generalizable zero-shot low-light enhancement. We first enrich the prior representation space with spatially adaptive multi-scale formulation, then use uncertainty (Ning et al., 2021) to guide multi-scale fusion and luminance-derived reliability to weight chromatic cues, extracting **robust** scene information while suppressing unstable responses. To promote effective and generalizable prior utilization, we propose prior decoding regularization, which enforces intermediate generative features to reconstruct a clean intrinsic representation through an auxiliary decoder. This supervision encourages scene-relevant features and discourages the encoding of degradation-dependent cues. At inference time, we further formulate prior-conditioned enhancement as online prior-image co-optimization: the evolving image state provides feedback through the decoding proxy to refine the intrinsic condition, which in turn guides subsequent generation. Extensive experiments across multiple benchmarks demonstrate that our method generally outperforms existing zero-shot LLIE approaches in both fidelity and perceptual quality, with a 0.35 dB increase in PSNR and a 0.02 improvement in LPIPS. We establish a new state-of-the-art and achieve faithful, perceptually convincing appearance, and strongly generalizable low-light enhancement across diverse scenes and illumination conditions.

Our contributions can be summarized as follows:

- We propose **UG-Prior** to bridge normal-light and low-light domains through informative and robust scene representations. Accounting for spatially and scale-dependent unreliability in low-light observations, it adaptively integrates multi-scale structural priors and reliability-aware chromatic cues, preserving recoverable scene information while suppressing unreliable components.
- To improve generalization under diverse unseen degradation, we propose a prior decoding regularization, incorporating an auxiliary decoder to reconstruct prior latents toward clean intrinsic targets from the generative model's intermediate features. This explicitly encourages the generative representation to retain reconstruction-relevant scene information and reduce reliance on degradation-sensitive cues.

- We reformulate enhancement with a fixed prior as as an online prior-image co-optimization process at inference time. The progressively enhanced image estimate provides feedback through the decoding proxy to refine the intrinsic state, which in turn guides subsequent generation. This enables synergistic refinement of both the conditioning prior and the generated appearance for each test image.

2 Related Work

2.1 Low-Light Image Enhancement

Supervised LLIE methods typically learn enhancement mappings through Retinex decomposition (Wei et al., 2018; Yi et al., 2023), advanced architectures (Cai et al., 2023; Zhang et al., 2019b; 2021; Wu et al., 2022; Zhang et al., 2025; Yan et al., 2025), and generative modeling (Jiang et al., 2023; Yi et al., 2023; Wu et al., 2024b). To reduce reliance on paired supervision, EnlightenGAN (Jiang et al., 2021) and NeRCo (Yang et al., 2023) adopt adversarial learning, while Zero-DCE (Guo et al., 2020) and SCI (Ma et al., 2022) use reference-free objectives. PairLIE (Fu et al., 2023) alleviates noise with paired low-light inputs. LightenDiffusion (Jiang et al., 2024) combines Retinex decomposition with unpaired diffusion learning, and ZeroIDIR (Jiang et al., 2026) learns from degraded images through perturbed diffusion consistency. Since these methods still require real low-light training data, the generalization ability is still limited to the specific degradation presented in the low-light training set. Alternatively, ExCNet (Zhang et al., 2019a) and CoLIE (Chobola et al., 2024) optimize per-image mapping via handcrafted constraints.

Pretrained generative priors provide another way to enhancement without low-light training data. GDP (Fei et al., 2023) guides diffusion sampling with restoration constraints, while FourierDiff (Lv et al., 2024) and AGLLDiff (Lin et al., 2025) introduce frequency-domain and image-attribute guidance, respectively. MR. Illuminate (Cho et al., 2026) combines latent modulation with inversion-derived self-attention features. These approaches primarily constrain enhancement through observation-derived guidance or desired image properties. In contrast, our method learns an uncertainty-guided intrinsic condition from normal-light data and refines it during generation.

2.2 Illumination-Invariants and Uncertainty-Aware Representations

Physics-derived priors extract intrinsic scene properties from diverse and complex illumination appearance. CIConv (Lengyel et al., 2021) introduces color-invariant convolutions for zero-shot day-night domain adaptation. QuadPrior (Wang et al., 2024) uses physical illumination invariants to learn prior-to-image reconstruction from normal-light data, and QuadPrior++ (Huang et al., 2025a) extends this approach through multi-dimensional augmentation for distillation. However, analytical invariance does not ensure reliable estimation under real-world low-light environments. To alleviate this gap, researchers formulate enhancement as a dynamic process with uncertainty. Some work uses spatial uncertainty to adapt supervision in super-resolution (Ning et al., 2021), while GSAD (Hou et al., 2023) learns an uncertainty decoder to build a focal diffusion loss. Different from existing works, we instead use uncertainty to construct a more reliable and informative prior, enabling faithful and perceptual low-light image enhancement.

3 UG-Prior

3.1 Motivations and Overview

Unlike pixel- and distribution-level optimization, which depends on training scenes and degradations (Wang et al., 2024), zero-shot methods with physics-derived illumination-invariant priors (Wang et al., 2024; Huang et al., 2025a) avoid low-light training data, providing a cross-domain bottleneck for scene preservation and generative illumination restoration.

However, entangled scene cues and degradation create a **robustness-informativeness trade-off**: retaining too much information may amplify degradation, whereas suppressing unstable responses may discard useful details. Existing single-scale bottlenecks (Lengyel et al., 2021; Wang et al., 2024) offer limited flexibility in balancing reconstruction fidelity and cross-illumination robustness.

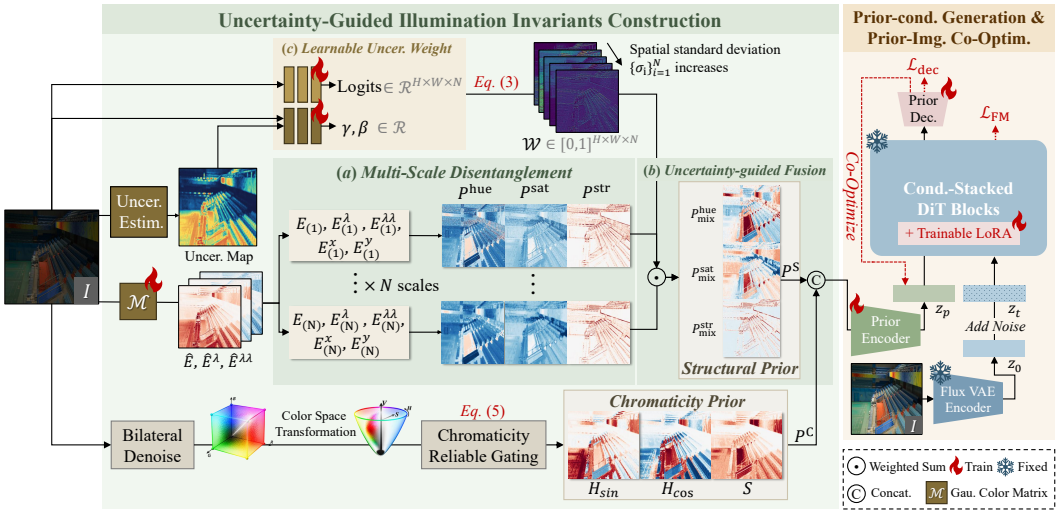


Figure 2: **Overview framework of our UG-Prior.** We propose the **Uncertainty-Guided Illumination Invariants Construction**, where we design multi-scale uncertainty-guided fusion mechanism and reliability-aware chromatic weighting, achieving an informative and robust prior. Building upon this, we propose **prior decoding regularization** that encourages generative features to recover clean intrinsic representations. During inference, **online prior-image co-optimization** will iteratively refine the prior through decoder feedback, improving guidance for subsequent generation.

Toy Experiment. We conduct a toy experiment to analyze this trade-off. We quantify *robustness* by the mean square error (MSE) between paired low-/normal-light priors, and *informativeness* by the probing reconstruction error of a lightweight decoder predicting normal-light images from their priors. [1] Lower errors indicate greater robustness and informativeness, respectively. We compare (a) globally shared single-scale priors (Lengyel et al., 2021), (b) per-image single-scale color priors (Wang et al., 2024), (c) multi-scale averaging, and (d) our uncertainty-guided multi-scale fusion.

Fig. 3 shows that single-scale priors trade robustness for informativeness. Per-image scale selection and multi-scale averaging improve this balance, while uncertainty-guided fusion further reduces both errors and extends the empirical frontier, supporting reliability-aware integration for preserving scene information and suppressing degradation.

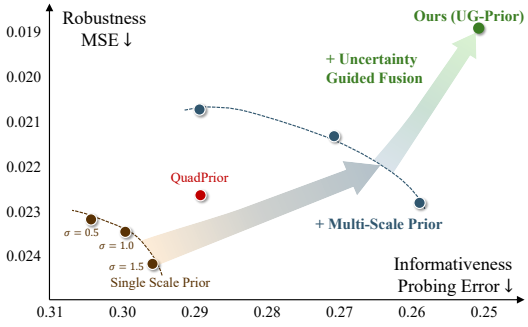


Figure 3: **Motivation experiments** on verifying that our UG-Prior effectively improves the robustness-informativeness trade-off.

Framework Overview. Building on our analysis, our UG-Prior decouples LLIE into two parts, *i.e.*, reliable prior construction and robust prior-conditioned generation. As shown in Fig. 2, our framework consists of an uncertainty estimator $\mathcal{U}$, an illumination-invariant prior constructor $\mathcal{P}_\psi$, a prior encoder E_ϕ, and a prior-conditioned generator G_θ equipped with an auxiliary intrinsic decoder D_{reg}. Given an input image $\mathbf{I}$, we first estimate its uncertainty $\mathbf{u}$ and construct a reliability-aware intrinsic condition $\mathbf{p}$:

$$\mathbf{u} = \mathcal{U}(\mathbf{I}), \qquad \mathbf{p} = E_\phi(\mathcal{P}_\psi(\mathbf{I}; \mathbf{u})), \tag{1}$$

The resulting intrinsic condition will guide the generative enhancement while preserving reliable scene information:

$$\hat{\mathbf{I}} = \mathcal{G}_\theta(\boldsymbol{\epsilon}; \mathbf{p}), \qquad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}), \tag{2}$$

where $\mathcal{G}_\theta$ denotes the prior-conditioned generation process and $\hat{I}$ is the enhanced image. We propose the prior decoding regularization during training, and formulate the online prior-image optimization during inference with the decoded output.

[1] Evaluation uses 115 LOL test pairs; probing errors are measured after decoder convergence.

3.2 Uncertainty-Guided Illumination-Invariant Priors

We first extend single scale to multiple scales, enriching the prior representation space. We then use uncertainty (Ning et al., 2021) to guide multi-scale fusion and chromatic cues gating, achieving reliable structural and color priors.

Uncertainty estimation. We estimate a spatial confidence map $q(\mathbf{I}) \in [0,1]^{H\times W}$ from local luminance and structure-tensor statistics, and define uncertainty as $\mathbf{u} = 1 - q(\mathbf{I})$. This estimate favors strong luminance and directional structure over weak signals and isotropic variation. The formulation is detailed in Appendix B, and we provide visualization in Fig. 2, 6.

Multi-scale fusion. Following CIConv (Lengyel et al., 2021) and QuadPrior (Wang et al., 2024), we extract hue, saturation, and structural responses $\{P_i^{\text{hue}}, P_i^{\text{sat}}, P_i^{\text{str}}\}$ at Gaussian standard deviations $\sigma_1 < \cdots < \sigma_N$. Fine scales preserve detail, whereas coarse scales aggregate broader spatial support. The underlying responses and imaging assumptions are detailed in Appendix B.

Our core insight is to extract informative detail where observations are reliable and adapt them to coarse cues when they are uncertain. We learn a lightweight CNN predictor that takes $\mathbf{I}$ and $\mathbf{u}$ to produce spatial logits ℓ_i and image-conditioned parameters $\beta \in (0,1)$ and $\gamma > 0$. We define

$$\begin{aligned} c_i &= \frac{i-1}{N-1}, \qquad b_i = \beta + (1-\beta)c_i^{\gamma}, \\ g_i(s) &= q(s) + (1-q(s))b_i, \qquad w_i(s) = \frac{\exp(\ell_i(s))\, g_i(s)}{\sum_j \exp(\ell_j(s))\, g_j(s)}. \end{aligned} \tag{3}$$

where s indexes spatial locations. c_i means the ordered scale coordinates and $w_i(s)$ is a part of uncertainty-derived weighting matrix $\mathcal{W} = \{w_i\}_{i=1}^N \in [0,1]^{H\times W\times N}$. Such formulation enables uncertainty to introduce scale-dependent modulation on the routing logits. Note that at $u(s) = 0$, the weights reduce to a spatial softmax. The structural prior is

$$P^{\text{S}} = \text{IN}\left[\sum_i w_i P_i^{\text{hue}},\ \log\left(\sum_i w_i P_i^{\text{sat}} + \epsilon\right),\ \sum_i w_i P_i^{\text{str}}\right], \tag{4}$$

where IN denotes channel-wise instance normalization, $\log$ compresses the saturation response, and $\epsilon = 10^{-5}$ ensures numerical stability.

Reliability-aware chromatic encoding. Since Kubelka-Munk theory Kubelka & Munk (1999) is inherently derived based on color invariant process (Lengyel et al., 2021), we complement with explicit color priors. We extract HSV coordinates from the bilaterally filtered image $\widetilde{\mathbf{I}}$, and encode hue circularly (Yan et al., 2025) with weighting by luminance-derived reliability:

$$r_c = \frac{\widetilde{Y}^2}{\widetilde{Y}^2 + \sigma_c^2}, \qquad P^{\text{C}} = r_c\,[\sin\vartheta, \cos\vartheta, S], \tag{5}$$

where $\widetilde{Y}$ is filtered luminance, S is saturation, and $\vartheta \in [0, 2\pi)$ is the hue angle. This weighting attenuates color cues in weakly illuminated regions. Finally, the prior encoder that consists of some convolution layers and non-linear activations will map the six-channel prior $\widetilde{\mathbf{p}} = [P^{\text{S}}; P^{\text{C}}]$ to the intrinsic condition $\mathbf{p}$. We will demonstrate the necessity of color priors in Sec. 4.5.

3.3 Prior-Conditioned Generation with Decoding Regularization

Prior-conditioned generation. We learn a prior-to-image mapping with pretrained generative priors for high-quality illumination restoration. We adopt `FLUX.1-dev` (Black Forest Labs, 2024) as the generative backbone and incorporate $\mathbf{p}$ through an OminiControl-style architecture (Tan et al., 2025), where prior tokens are concatenated with noisy image tokens along the sequence dimension and interact through bidirectional attention with LoRA (Hu et al., 2022) finetuned.

Formally, let $\mathbf{z}_0$ denote the VAE latent of a normal-light training image, we construct $\mathbf{z}_t = (1-t)\mathbf{z}_0 + t\boldsymbol{\epsilon}$, with $t \sim \mathcal{U}(0,1)$ and $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, and minimize

$$\mathcal{L}_{\text{FM}} = \mathbb{E}_{\mathbf{z}_0,\mathbf{p},\boldsymbol{\epsilon},t}\left[\|v_\theta(\mathbf{z}_t, t; \mathbf{p}) - (\boldsymbol{\epsilon} - \mathbf{z}_0)\|_2^2\right]. \tag{6}$$

At inference, the learned velocity transports a Gaussian latent from $t = 1$ to $t = 0$ and project latents to pixels with the Flux decoder $\mathcal{D}_{\text{fm}}$:

$$\hat{\mathbf{z}}_0 = \mathbf{z}_1 + \int_1^0 v_\theta(\mathbf{z}_t, t; \widetilde{\mathbf{p}})\,\mathrm{d}t, \quad \mathbf{z}_1 \sim \mathcal{N}(\mathbf{0}, \mathbf{I}), \qquad \hat{\mathbf{I}} = \mathcal{D}_{\text{fm}}(\hat{\mathbf{z}}_0). \tag{7}$$

We will extend this process to an evolving intrinsic condition in Sec. 3.4.

Prior decoding regularization. Beyond the flow-matching objective, we require the final prior hidden tokens $\mathbf{h}_p$, which aggregate information through bidirectional attention, to reconstruct a clean intrinsic target with a U-Net shaped prior decoder D_{reg}:

$$\hat{\mathbf{p}} = D_{\text{reg}}(\mathbf{h}_p) + E_\phi(\widetilde{\mathbf{p}}), \qquad \mathbf{p}^+ = \mathrm{sg}\big[E_\phi\big(P_\psi(\mathbf{I}; \mathbf{u})\big)\big]. \tag{8}$$

Such regularization encourages the condition stream to retain clean scene information rather than degradation-dependent responses.

Training objective. We combine intrinsic reconstruction with spatial structural constraints:

$$\begin{aligned} \mathcal{L}_{\text{dec}} &= ||\hat{\mathbf{p}} - \mathbf{p}^+||_2^2 + \lambda_s \mathcal{L}_{\text{ssim}}(\hat{\mathbf{p}}, \mathbf{p}^+) + \lambda_g \mathcal{L}_{\text{grad}}(\hat{\mathbf{p}}, \mathbf{p}^+) + \lambda_l \mathcal{L}_{\text{lap}}(\hat{\mathbf{p}}, \mathbf{p}^+), \\ \mathcal{L} &= \mathcal{L}_{\text{FM}} + \mathcal{L}_{\text{dec}}, \end{aligned} \tag{9}$$

where $\mathcal{L}_{\text{ssim}}$ evaluates structural similarity via SSIM (Wang et al., 2004), $\mathcal{L}_{\text{grad}}$ matches horizontal and vertical gradient, and $\mathcal{L}_{\text{lap}}$ matches second-order variation using a four-neighbor Laplacian.

3.4 Online Prior-Image Co-Optimization

During flow matching denoising, the initial low-light prior can be refined with more reliable scene cues. For a schedule $1 = t_0 > \cdots > t_K = 0$, let $\mathbf{z}^{(k)}$ and $\mathbf{p}^{(k)}$ denote the image latent and intrinsic condition at step k. We initialize $\mathbf{z}^{(0)} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ and $\mathbf{p}^{(0)} = E_\phi(\widetilde{\mathbf{p}})$ in Sec. 3.2.

At step k, bidirectional attention incorporates information from $\mathbf{z}^{(k)}$ into the final prior hidden tokens $\mathbf{h}_p^{(k)}$. The regularization decoder then produces a refined proposal:

$$\hat{\mathbf{p}}^{(k)} = D_{\text{reg}}(\mathbf{h}_p^{(k)}) + \mathbf{p}^{(k)}. \tag{10}$$

This transfers image-state feedback into the intrinsic condition without per-step RGB decoding and re-encoding. Holding the proposal fixed, we anchor refinement to the observed condition:

$$\mathbf{p}^{(k+1)} = \arg\min_{\mathbf{p}^*} \frac{1 - \alpha_k}{2} \|\mathbf{p}^* - \mathbf{p}^{(0)}\|_2^2 + \frac{\alpha_k}{2} \|\mathbf{p}^* - \hat{\mathbf{p}}^{(k)}\|_2^2. \tag{11}$$

Its closed-form solution is

$$\mathbf{p}^{(k+1)} = (1 - \alpha_k)\mathbf{p}^{(0)} + \alpha_k \hat{\mathbf{p}}^{(k)}. \tag{12}$$

We use $\alpha_k = 0$ for $k < \lfloor \tau K \rfloor$ and $\alpha_k = \alpha$ thereafter, with $\tau = 0.25$ and $\alpha = 0.2$. Note that the original prior $\widetilde{\mathbf{p}}$ anchors successive refined proposals to consistently optimize the intrinsic condition.

4 Experiments

4.1 Implementation Details

We implement our UG-Prior with PyTorch on two NVIDIA H20 GPUs. We set batch size to 8 and use gradient checkpoint to save GPU memory. The training and testing data resolution is 1,024 $\times$ 1024. We set LoRA rank to 64. We train on 35,795 images randomly sampled from SA-1B (Kirillov et al., 2023) with simply synthetic perturbation following Wang et al. (2024), without using low-light training images. More details are reported in the appendix.

Evaluation Settings. We evaluate low-light enhancement on LOL (Wei et al., 2018), LSRW (Hai et al., 2023), and MIT-Adobe FiveK (Bychkovsky et al., 2011), using 115, 50, and 500 test pairs, respectively, and report PSNR (Wang & Bovik, 2009), SSIM (Wang et al., 2004), and LPIPS (Zhang et al., 2018). We also evaluate backlit enhancement on BAID (Lv et al., 2022) with 368 images and CLIPIQA (Wang et al., 2023a). We further test on 123 unpaired images from DICM (Lee et al., 2013), LIME (Guo et al., 2017), NPE (Wang et al., 2013), MEF (Ma et al., 2015), and VV (Vonikakis et al., 2018), using Q-Align (Wu et al., 2024a) and estimated noise level (NL) (Chen et al., 2015).

Table 1: **Quantitative comparison on five benchmarks.** SL, UL, and ZS denote supervised, unsupervised, and zero-shot methods, respectively; ZS requires no real low-light training data. Best and second-best results are additionally **bold** and underlined. Only UL and ZS methods are included in the ranking.

Method	Train set	LOL			LSRW			MIT-Adobe FiveK			BAID	Unpaired	
		PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	CLIPIQA↑	NL↓	Q-Align↑
Supervised (SL)													
RetinexNet	LOL	16.19	0.403	0.534	15.61	0.414	0.454	12.30	0.687	0.258	0.202	3.254	2.47
KinD	LOL	20.21	0.814	0.147	16.41	0.484	0.337	14.71	0.756	0.176	0.209	0.700	2.36
KinD++	LOL	16.64	0.662	0.410	16.08	0.402	0.371	15.76	0.650	0.319	0.242	0.431	2.45
URetinexNet	LOL	20.93	0.854	0.104	18.27	0.518	0.419	14.10	0.734	0.182	0.287	1.319	2.15
Retinexformer	LOL	28.48	0.877	0.117	14.97	0.342	0.359	13.87	0.692	0.222	0.226	1.064	3.12
Retinexformer	MIT5K	13.02	0.426	0.365	15.03	0.399	0.327	24.93	0.907	0.063	0.277	0.684	3.27
DiffLL	LOL	28.54	0.870	0.102	17.26	0.398	0.437	15.81	0.719	0.244	0.264	0.888	2.60
Diff-Retinex	LOL	26.20	0.897	0.082	15.94	0.487	0.339	16.81	0.728	0.241	0.218	0.912	2.78
CWNet	LOL+LSRW	26.90	0.918	0.09	16.82	0.556	0.336	14.26	0.781	0.172	0.250	0.559	3.55
CIDNet	LOL	20.83	0.873	0.126	17.94	0.544	0.297	19.53	0.831	0.159	0.277	0.590	3.55
CIDNet	FiveK	12.04	0.396	0.399	10.88	0.244	0.475	24.46	0.909	0.085	0.296	0.492	3.56
Unsupervised (UL)													
EnlightenGAN	Own data	18.57	0.700	0.302	17.08	0.470	0.327	13.28	0.738	0.203	0.323	0.967	2.54
PairLIE	LOL+SICE	19.70	0.774	0.235	17.60	0.512	0.329	10.55	0.642	0.273	0.341	1.641	2.50
NeRCo	LSRW	19.67	0.720	0.266	**19.45**	0.549	**0.288**	17.33	0.767	0.208	0.365	0.836	3.34
CLIP-LIT	Own data	14.82	0.524	0.371	13.48	0.405	0.353	17.00	0.781	0.159	0.426	1.864	3.40
Zero-DCE	SICE	17.64	0.572	0.316	15.85	0.453	0.317	13.53	0.725	0.201	0.398	1.569	2.78
Zero-DCE++	Own data	17.03	0.445	0.314	16.25	0.462	0.329	12.33	0.408	0.280	0.379	1.150	2.33
RUAS	MIT5K	13.62	0.462	0.346	13.02	0.359	0.379	9.53	0.610	0.301	0.372	2.051	3.02
RUAS	LOL	15.47	0.490	0.305	14.27	0.470	0.465	5.15	0.373	0.669	0.301	3.312	2.38
RUAS	FACE	15.05	0.456	0.371	14.03	0.403	0.384	5.00	0.366	0.685	0.279	3.633	1.33
SCI	MIT5K	11.67	0.395	0.361	11.79	0.317	0.401	16.29	0.795	**0.143**	0.411	0.853	2.67
SCI	LOL+LSRW	16.97	0.532	0.312	15.24	0.424	0.322	7.83	0.573	0.360	0.383	1.893	2.03
SCI	FACE	16.80	0.543	0.322	15.16	0.408	0.326	10.95	0.684	0.272	0.396	1.335	1.82
LightenDiffusion	LOL+LSRW	20.44	0.801	0.202	18.44	0.534	0.320	**21.23**	0.793	0.181	0.451	1.071	2.77
ZeroIDIR	Own data	20.87	0.818	0.179	18.41	0.540	0.309	20.13	0.798	0.169	0.471	0.491	3.37
Zero-shot (ZS)													
ExCNet	Test image	16.25	0.454	0.382	15.74	0.408	0.339	14.21	0.719	0.197	0.429	1.563	2.06
CoLIE	Test image	14.90	0.499	0.327	14.00	0.408	0.351	18.50	0.789	0.167	0.482	0.964	3.17
FourierDiff	Test image	16.95	0.604	0.293	15.64	0.466	0.327	17.81	0.793	0.168	0.403	1.221	2.77
QuadPrior	COCO	20.31	0.808	0.202	16.96	0.564	0.408	18.51	0.785	0.163	0.451	0.422	3.12
AGLLDiff	Test image	19.84	0.806	0.192	17.36	0.544	0.359	8.94	0.622	0.316	0.432	1.897	2.29
MR. Illuminate	Test image	20.94	0.812	0.180	17.43	0.533	0.301	18.96	**0.820**	0.171	0.478	0.389	2.67
UG-Prior	SA-1B subset	**21.29**	**0.838**	**0.159**	18.51	**0.575**	0.301	20.25	0.799	0.168	**0.500**	**0.376**	**3.55**



Figure 4: **Qualitative comparison on LOL. (Zoom-in for better review.)**

4.2 Quantitative Comparison

Table 1 compares our UG-Prior with supervised, unsupervised, and zero-shot methods on three paired benchmarks, one backlit benchmark, and an unpaired in-the-wild benchmark. UG-Prior establishes a **new state-of-the-art**, improving both the single dataset performance and cross-dataset generalization. Note that even the SotA supervised and unsupervised methods (*e.g.*, RetinexFormer, CIDNet and NeRCo) still struggle with performance drops under cross-dataset evaluation, our method achieves the strongest overall performance among unsupervised and zero-shot methods with superior generalization ability.

4.3 Qualitative Comparison

Our method achieves the most suitable illumination restoration with the clearest structures, less noise and color shifts, as shown in Fig. 4. Fig. 5 further compares enhancement results on challenging in-



Figure 5: **Qualitative comparison on unpaired benchmarks.**

the-wild scenes. In Fig. 5, our UG-Prior produces a cleaner night sky and fewer artifacts on the building facade, preserves stronger contrast, and produces a warmer and more plausible candlelit appearance. These examples illustrate its ability to balance visibility, structural preservation, and plausible color restoration.

4.4 Model Distillation

To reduce inference cost, we distill UG-Prior into a lightweight enhancement network. Following QuadPrior (Wang et al., 2024), we adopt the same student architecture based on U-Net (Ronneberger et al., 2015) and Restormer (Zamir et al., 2022), and use our enhanced images of LOL and LSRW **trainset** as supervision (8,000 images in total). We additionally inject the intrinsic prior $\tilde{\mathbf{p}}$ into the student architecture, transferring both the enhanced appearance and the scene guidance underlying its reconstruction. We also compare students with and without such prior injection.

Tab. 2 shows that the prior-conditioned student reduces FLOPs by approximately 3.29×10^4 and peak GPU memory by 98.3% (from 32GB to 0.538 GB) relative to the teacher, while achieving 19.47/18.31 dB PSNR on LOL/LSRW. Compared with the unconditioned student, prior injection improves PSNR by 0.66/0.30 dB and SSIM by 0.015/0.028, with only 0.025 GB additional memory. These results demonstrate the superiority of our teacher model and practical utility.

Table 2: Performance and inference cost of the teacher and distilled students. Peak GPU memory is measured on the H20 GPU with 1024 × 1024 inputs.

Model	LOL		LSRW		FLOPs	Mem. (GB)
	PSNR↑	SSIM↑	PSNR↑	SSIM↑		
Ours main model	21.29	0.838	18.51	0.575	4.6e+15	32
Student w/o prior	18.81	0.788	18.01	0.545	7.8e+10	0.513
Student w/ SFT prior	19.47	0.803	18.31	0.573	1.4e+11	0.538

4.5 Ablation Study and Visualized Analysis

We conduct thorough ablation experiments to verify: **(1)** whether the proposed uncertainty-guided fusion improves the robustness-informativeness balance of the prior, **(2)** whether the decoding regularization promotes robust prior utilization, and **(3)** whether the online refinement improves enhancement beyond a fixed condition.

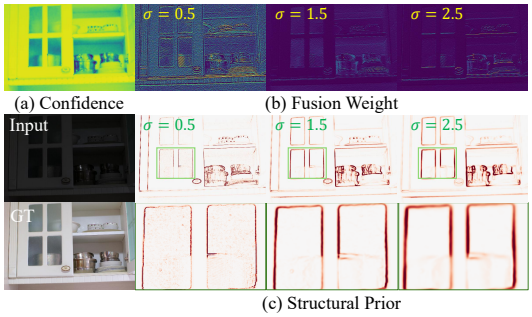


Figure 6: **Visualization of weighted fusion and the prior with different scales.**

Effects of Uncertainty Guidance. We compare (1) uniform averaging, (2) learned spatial fusion without uncertainty modulation, (3) implicit fusion without modulation in Eq. 3, and (4) the proposed uncertainty-guided fusion. As shown in Tab. 3, uncertainty-guided multi-scale prior achieves improved performance, while averaging leads to insufficient modeling ability. Implicit fusion uses uncertainty as input but discard "more confidence, more detail" constraints in Eq. 3, which is hard to optimize and leads to sub-optimal results. We also replace our designed prior with QuadPrior (Wang et al., 2024), observing obvious performance drop due to ambiguous colors and remaining noises. We provide further comparison on different uncertainty modeling in the Appendix A.2, A.3.

Visualization of Uncertainty-Guided Prior. As shown in Fig. 6, the learned weights will assign greater importance to fine-scale responses in high-confidence regions (brighter ones in Fig. 6 (a)),

and reduce the impact in low-confidence regions (dark patches in Fig. 6 (a)), which is consistent with our assumption and core modeling in Sec. 3.2.

Effects of Chromaticity Prior. We incorporate chromaticity prior to complement the color information lost in Sec. 3.2. As for ablation, we compare (1) the structural prior alone, (2) its combination with unweighted chromatic cues, (3) color prior without denoising, and (4) the full version. Tab. 3 and Fig. 8 show that discarding color prior may lead to severe color shift and performance drop, while reliability weighting effectively restore colors hidden in the dark region. The proposed bilateral denoising also reduce noises within color priors and preserve color information against high-frequency perturbation, improving perceptual quality of the final enhanced results. Fig. 13 further shows that RGB orders adopted by QuadPrior leads to obvious color shifts.

Table 3: **Ablation experiments on LOL.**

Configuration	PSNR↑	SSIM↑	LPIPS↓
Prior construction			
QuadPrior (Wang et al., 2024) as condition	19.10	0.788	0.242
Uniform averaging	20.81	0.808	0.182
Learned fusion w/o modulation	20.75	0.812	0.177
Learned implicit fusion	20.22	0.786	0.239
w/o chromatic prior	19.27	0.776	0.283
Unweighted chromatic cues	21.07	0.829	0.203
w/o bilateral denoising on P^{C}	20.99	0.811	0.194
Prior-conditioned generation			
w/o prior decoding regularization	20.89	0.802	0.188
w/o $\mathcal{L}_{\mathrm{ssim}}$, $\mathcal{L}_{\mathrm{grad}}$, $\mathcal{L}_{\mathrm{lap}}$	21.15	0.820	0.167
w/o online co-optimization	20.82	0.826	0.170
Ours	**21.29**	**0.838**	**0.159**

Effects of Prior Decoding Regularization. We discard the prior decoding regularization, observing much more noisy and obviously color-biased results than the regularized ones in Fig. 7. Besides, we visualize the enhanced result with large co-optimization scale in Fig. 7 (d). Although effectively denoised, meaningful characters also become illegible. We also verify that spatial structural supervision (*e.g.*, $\mathcal{L}_{\mathrm{ssim}}$, $\mathcal{L}_{\mathrm{grad}}$) can benefit the regularization towards more precise clean intrinsics prediction.



(a) w/o prior decode (b) w/o prior-image co-optimization (c) Ours Full Version (d) prior-image co-optim. with large scale

Figure 7: **Ablation on prior decoding regularization and online prior-image co-optimization.**

Effects of Online Prior-Image Co-Optimization. We evaluate different refinement strengths and starting steps in Fig. 11. Compared with a fixed prior as reported in Tab. 3, our proposed online refinement effectively improves both fidelity and perceptual quality. We present more visualized ablation results in Fig. 10 to demonstrate its ability to preserve faithful color during enhancement.



(a) w/o color prior (b) w/o reliable gating (c) w/o bilateral denoise (d) Ours-Full Version

Figure 8: **Ablation on color prior.**

5 Conclusion

We present UG-Prior for generalizable zero-shot low-light enhancement. The proposed uncertainty-guided multi-scale fusion improves the prior robustness–informativeness trade-off, while decoding regularization promotes robust prior utilization through clean intrinsic reconstruction. The online prior–image co-optimization further refines the condition using reconstruction feedback. Extensive experiments demonstrate state-of-the-art zero-shot performance, improving PSNR by 0.35 dB and reducing LPIPS by 0.02, with strong generalization across scenes and illumination conditions.

Limitation and Future Work. When scene information is lost due to extremely low intensity (*e.g.*, pixel value $= 0$), generative reconstruction may introduce inaccurate content, as shown in Fig. 9. In our future works, we will explore additional modalities, such as RAW (Chen et al., 2018) or event stream (Liu et al., 2026) as complementary information for faithful reconstruction.

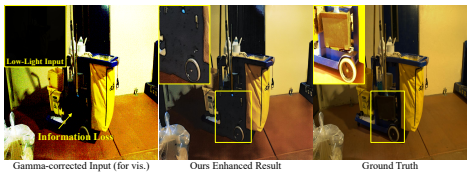


Figure 9: **Failure Case.** UG-Prior may fail when scene information is lost in extremely dark regions.

AI USE STATEMENT

In this work, we used generative AI tools only to polish the English writing and to assist in writing simple code for data preprocessing and visualization. These tools were not used to formulate research ideas, design the method or experiments, generate experimental data or results, or draw scientific conclusions. All AI-assisted text was manually reviewed and revised by the authors. All AI-assisted code was inspected, executed, and verified for correctness, and the resulting visualizations and preprocessed data were checked against the source data and expected outputs. The authors take full responsibility for the final content of this work, including all text, code, results, and claims.

REFERENCES

Black Forest Labs. Flux, 2024. URL https://github.com/black-forest-labs/flux.

Yochai Blau and Tomer Michaeli. The perception-distortion tradeoff. In *IEEE/CVF Conf. Comput. Vis. Pattern Recog.*, 2018.

Vladimir Bychkovsky, Sylvain Paris, Eric Chan, and Frédo Durand. Learning photographic global tonal adjustment with a database of input / output image pairs. In *IEEE/CVF Conf. Comput. Vis. Pattern Recog.*, 2011.

Yuanhao Cai, Hao Bian, Jing Lin, Haoqian Wang, Radu Timofte, and Yulun Zhang. Retinexformer: One-stage retinex-based transformer for low-light image enhancement. In *Int. Conf. Comput. Vis.*, 2023.

Chen Chen, Qifeng Chen, Jia Xu, and Vladlen Koltun. Learning to see in the dark. In *IEEE/CVF Conf. Comput. Vis. Pattern Recog.*, 2018.

Guangyong Chen, Fengyuan Zhu, and Pheng-Ann Heng. An efficient statistical method for image noise level estimation. In *Int. Conf. Comput. Vis.*, 2015.

Joshua Cho, Sara Aghajanzadeh, Zhen Zhu, and D. A. Forsyth. MR. Illuminate: Zero-shot low-light image enhancement with diffusion prior. In *IEEE/CVF Conf. Comput. Vis. Pattern Recog.*, 2026.

Tomáš Chobola, Yu Liu, Hanyi Zhang, Julia A. Schnabel, and Tingying Peng. Fast context-based low-light image enhancement via neural implicit representations. In *Eur. Conf. Comput. Vis.*, 2024.

Ben Fei, Zhaoyang Lyu, Liang Pan, Junzhe Zhang, Weidong Yang, Tianyue Luo, Bo Zhang, and Bo Dai. Generative diffusion prior for unified image restoration and enhancement. In *IEEE/CVF Conf. Comput. Vis. Pattern Recog.*, 2023.

Zhenqi Fu, Yan Yang, Xiaotong Tu, Yue Huang, Xinghao Ding, and Kai-Kuang Ma. Learning a simple low-light image enhancer from paired low-light instances. In *IEEE/CVF Conf. Comput. Vis. Pattern Recog.*, 2023.

Chunle Guo, Chongyi Li, Jichang Guo, Chen Change Loy, Junhui Hou, Sam Kwong, and Runmin Cong. Zero-reference deep curve estimation for low-light image enhancement. In *IEEE/CVF Conf. Comput. Vis. Pattern Recog.*, 2020.

Xiaojie Guo, Yu Li, and Haibin Ling. LIME: Low-light image enhancement via illumination map estimation. *IEEE Trans. Image Process.*, 26(2):982–993, 2017.

Jiang Hai, Zhu Xuan, Ren Yang, Yutong Hao, Fengzhu Zou, Fang Lin, and Songchen Han. R2RNet: Low-light image enhancement via real-low to real-normal network. *J. Vis. Commun. Image Represent.*, 90:103712, 2023.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Annu. Conf. Neural Inf. Process. Systems*, 2020.

Jinhui Hou, Zhiyu Zhu, Junhui Hou, Hui Liu, Huanqiang Zeng, and Hui Yuan. Global structure-aware diffusion process for low-light image enhancement. In *Annu. Conf. Neural Inf. Process. Systems*, 2023.

Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. LoRA: Low-rank adaptation of large language models. In *Int. Conf. Learn. Represent.*, 2022.

Haofeng Huang, Yifan Li, Wenjing Wang, Wenhan Yang, Ling-Yu Duan, and Jiaying Liu. Quad-Prior++: Multi-dimension augmented physical prior for zero-reference illumination enhancement. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2025a.

Yan Huang, Xiaoshan Liao, Jinxiu Liang, Yuhui Quan, Boxin Shi, and Yong Xu. Zero-shot low-light image enhancement via latent diffusion models. In *AAAI Conf. Artif. Intel.*, 2025b.

Hai Jiang, Ao Luo, Songchen Han, Haoqiang Fan, and Shuaicheng Liu. Low-light image enhancement with wavelet-based diffusion models. In *Siggraph Asia*, 2023.

Hai Jiang, Ao Luo, Xiaohong Liu, Songchen Han, and Shuaicheng Liu. LightenDiffusion: Unsupervised low-light image enhancement with latent-retinex diffusion models. In *Eur. Conf. Comput. Vis.*, 2024.

Hai Jiang, Zhen Liu, Yinjie Lei, Songchen Han, Bing Zeng, and Shuaicheng Liu. ZeroIDIR: Zero-reference illumination degradation image restoration with perturbed consistency diffusion models. In *IEEE/CVF Conf. Comput. Vis. Pattern Recog.*, 2026.

Yifan Jiang, Xinyu Gong, Ding Liu, Yu Cheng, Chen Fang, Xiaohui Shen, Jianchao Yang, Pan Zhou, and Zhangyang Wang. EnlightenGAN: Deep light enhancement without paired supervision. *IEEE Trans. Image Process.*, 30:2340–2349, 2021.

Donggoo Jung, Daehyun Kim, and Tae Hyun Kim. Continuous exposure learning for low-light image enhancement using neural ODEs. In *Int. Conf. Learn. Represent.*, 2025.

Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C. Berg, Wan-Yen Lo, Piotr Dollár, and Ross Girshick. Segment anything. *arXiv:2304.02643*, 2023.

P. Kubelka and F. Munk. Ein beitrag zur optik der farbanstriche. In *Zeitung fur Technische Physik*, volume 12, pp. 593, 1999.

Chulwoo Lee, Chul Lee, and Chang-Su Kim. Contrast enhancement based on layered difference representation of 2D histograms. *IEEE Trans. Image Process.*, 22(12):5372–5384, 2013.

Attila Lengyel, Sourav Garg, Michael Milford, and Jan C. van Gemert. Zero-shot day-night domain adaptation with a physics prior. In *Int. Conf. Comput. Vis.*, 2021.

Chongyi Li, Chunle Guo, and Chen Change Loy. Learning to enhance low-light image via zero-reference deep curve estimation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 44(8):4225–4238, 2021.

Chongyi Li, Chun-Le Guo, Man Zhou, Zhexin Liang, Shangchen Zhou, Ruicheng Feng, and Chen Change Loy. Embedding fourier for ultra-high-definition low-light image enhancement. In *Int. Conf. Learn. Represent.*, 2023.

Zhexin Liang, Chongyi Li, Shangchen Zhou, Ruicheng Feng, and Chen Change Loy. Iterative prompt learning for unsupervised backlit image enhancement. In *Int. Conf. Comput. Vis.*, 2023.

Yunlong Lin, Tian Ye, Sixiang Chen, Zhenqi Fu, Yingying Wang, Wenhao Chai, Zhaohu Xing, Lei Zhu, and Xinghao Ding. AGLLDiff: Guiding diffusion models towards unsupervised training-free real-world low-light image enhancement. In *AAAI Conf. Artif. Intel.*, 2025.

Haoyue Liu, Jinghan Xu, Luxin Feng, Hanyu Zhou, Haozhi Zhao, Yi Chang, and Luxin Yan. NEC-Diff: Noise-robust event-raw complementary diffusion for seeing motion in extreme darkness. In *IEEE/CVF Conf. Comput. Vis. Pattern Recog.*, 2026.

Risheng Liu, Long Ma, Jiaao Zhang, Xin Fan, and Zhongxuan Luo. Retinex-inspired unrolling with cooperative prior architecture search for low-light image enhancement. In *IEEE/CVF Conf. Comput. Vis. Pattern Recog.*, 2021.

Xiaoqian Lv, Shengping Zhang, Qinglin Liu, Haozhe Xie, Bineng Zhong, and Huiyu Zhou. BacklitNet: A dataset and network for backlit image enhancement. *Comput. Vis. and Image Understanding*, 218:103403, 2022.

Xiaoqian Lv, Shengping Zhang, Chenyang Wang, Yichen Zheng, Bineng Zhong, Chongyi Li, and Liqiang Nie. Fourier priors-guided diffusion for zero-shot joint low-light enhancement and deblurring. In *IEEE/CVF Conf. Comput. Vis. Pattern Recog.*, 2024.

Kede Ma, Kai Zeng, and Zhou Wang. Perceptual quality assessment for multi-exposure image fusion. *IEEE Trans. Image Process.*, 24(11):3345–3356, 2015.

Long Ma, Tengyu Ma, Risheng Liu, Xin Fan, and Zhongxuan Luo. Toward fast, flexible, and robust low-light image enhancement. In *IEEE/CVF Conf. Comput. Vis. Pattern Recog.*, 2022.

Qian Ning, Weisheng Dong, Xin Li, Jinjian Wu, and Guangming Shi. Uncertainty-driven loss for single image super-resolution. In *Annu. Conf. Neural Inf. Process. Systems*, 2021.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *IEEE/CVF Conf. Comput. Vis. Pattern Recog.*, 2022.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. 2015.

Kai Shang, Mingwen Shao, Chao Wang, Yuanjian Qiao, and Yecong Wan. Training-free prior guided diffusion model for zero-reference low-light image enhancement. *Neurocomputing*, 617: 128974, 2025.

Zhenxiong Tan, Songhua Liu, Xingyi Yang, Qiaochu Xue, and Xinchao Wang. OminiControl: Minimal and universal control for diffusion transformer. In *Int. Conf. Comput. Vis.*, 2025.

Vassilios Vonikakis, Rigas Kouskouridas, and Antonios Gasteratos. On the evaluation of illumination compensation algorithms. *Multimedia Tools and Applications*, 77(8):9211–9231, 2018.

Jianyi Wang, Kelvin CK Chan, and Chen Change Loy. Exploring clip for assessing the look and feel of images. In *AAAI Conf. Artif. Intel.*, 2023a.

Shuhang Wang, Jin Zheng, Hai-Miao Hu, and Bo Li. Naturalness preserved enhancement algorithm for non-uniform illumination images. *IEEE Trans. Image Process.*, 22(9):3538–3548, 2013.

Tao Wang, Kaihao Zhang, Tianrun Shen, Wenhan Luo, Bjorn Stenger, and Tong Lu. Ultra-high-definition low-light image enhancement: A benchmark and transformer-based method. In *AAAI Conf. Artif. Intel.*, 2023b.

Wenjing Wang, Huan Yang, Jianlong Fu, and Jiaying Liu. Zero-reference low-light enhancement via physical quadruple priors. In *IEEE/CVF Conf. Comput. Vis. Pattern Recog.*, 2024.

Yufei Wang, Renjie Wan, Wenhan Yang, Haoliang Li, Lap-Pui Chau, and Alex C. Kot. Low-light image enhancement with normalizing flow. In *AAAI Conf. Artif. Intel.*, 2022.

Zhou Wang and Alan C. Bovik. Mean squared error: Love it or leave it? a new look at signal fidelity measures. *IEEE Signal Processing Magazine*, 26(1):98–117, 2009.

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

Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu. Deep retinex decomposition for low-light enhancement. In *Brit. Mach. Vis. Conf.*, 2018.

Haoning Wu, Zicheng Zhang, Weixia Zhang, Chaofeng Chen, Chunyi Li, Liang Liao, Annan Wang, Erli Zhang, Wenxiu Sun, Qiong Yan, Xiongkuo Min, Guangtai Zhai, and Weisi Lin. Q-Align: Teaching lmms for visual scoring via discrete text-defined levels. 2024a.

Wenhui Wu, Jian Weng, Pingping Zhang, Xu Wang, Wenhan Yang, and Jianmin Jiang. Uretinex-net: Retinex-based deep unfolding network for low-light image enhancement. In *IEEE/CVF Conf. Comput. Vis. Pattern Recog.*, 2022.

Yuhui Wu, Guoqing Wang, Zhiwen Wang, Yang Yang, Tianyu Li, Malu Zhang, Chongyi Li, and Heng Tao Shen. JoReS-Diff: Joint retinex and semantic priors in diffusion model for low-light image enhancement. In *ACM Int. Conf. Multimedia*, 2024b.

Xiaogang Xu, Ruixing Wang, Chi-Wing Fu, and Jiaya Jia. SNR-aware low-light image enhancement. In *IEEE/CVF Conf. Comput. Vis. Pattern Recog.*, 2022.

Qingsen Yan, Yixu Feng, Cheng Zhang, Guansong Pang, Kangbiao Shi, Peng Wu, Wei Dong, Jinqiu Sun, and Yanning Zhang. HVI: A new color space for low-light image enhancement. In *IEEE/CVF Conf. Comput. Vis. Pattern Recog.*, 2025.

Shuzhou Yang, Moxuan Ding, Yanmin Wu, Zihan Li, and Jian Zhang. Implicit neural representation for cooperative low-light image enhancement. In *Int. Conf. Comput. Vis.*, 2023.

Xunpeng Yi, Han Xu, Hao Zhang, Linfeng Tang, and Jiayi Ma. Diff-retinex: Rethinking low-light image enhancement with A generative diffusion model. In *Int. Conf. Comput. Vis.*, 2023.

Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Learning enriched features for real image restoration and enhancement. In *Eur. Conf. Comput. Vis.*, 2020.

Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In *IEEE/CVF Conf. Comput. Vis. Pattern Recog.*, 2022.

Lin Zhang, Lijun Zhang, Xinyu Liu, Ying Shen, Shaoming Zhang, and Shengjie Zhao. Zero-shot restoration of back-lit images using deep internal learning. In *ACM Int. Conf. Multimedia*, 2019a.

Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *IEEE/CVF Conf. Comput. Vis. Pattern Recog.*, pp. 586–595, 2018.

Tongshun Zhang, Pingping Liu, Yubing Lu, Mengen Cai, Zijian Zhang, Zhe Zhang, and Qiuzhan Zhou. CWNet: Causal wavelet network for low-light image enhancement. In *Int. Conf. Comput. Vis.*, 2025.

Yonghua Zhang, Jiawan Zhang, and Xiaojie Guo. Kindling the darkness: A practical low-light image enhancer. In *ACM Int. Conf. Multimedia*, 2019b.

Yonghua Zhang, Xiaojie Guo, Jiayi Ma, Wei Liu, and Jiawan Zhang. Beyond brightening low-light images. *Int. J. Comput. Vis.*, 129(4):1013–1037, 2021.

Zunjin Zhao and Daming Shi. Retinex-guided generative diffusion prior for low-light image enhancement. *Pattern Recognit.*, 172:112421, 2026.

A Additional Experiments

A.1 Ablation on Online Prior-Image Co-Optimization

Prior-image co-optimization allows the intrinsic condition to evolve beyond its initial analytical estimate, using feedback from the pretrained generative model to refine observation-derived cues. This couples physical constraints with learned image priors for adaptive enhancement. We conduct further experiments to validate its effectiveness.

Color correction. We found that the prior-image co-optimization can effectively alleviate color shifts and produce more faithful colors.



(a) Input (b) w/o Prior-Image Co-Optim. (c) Ours Full Version (d) Ground Truth

Figure 10: **Ablation on the online prior-image co-optimization.** The co-evolving paradigm can effectively correct the color bias.

Scales and schedule. We ablate on different scale and schedule setting in the prior-image co-optimization in Fig. 11. Although larger scale and longer optimization leads to worse performance, applying such co-optimization effectively reduce noises and correct color shift, as demonstrated in Tab. 3 and Fig. 10.

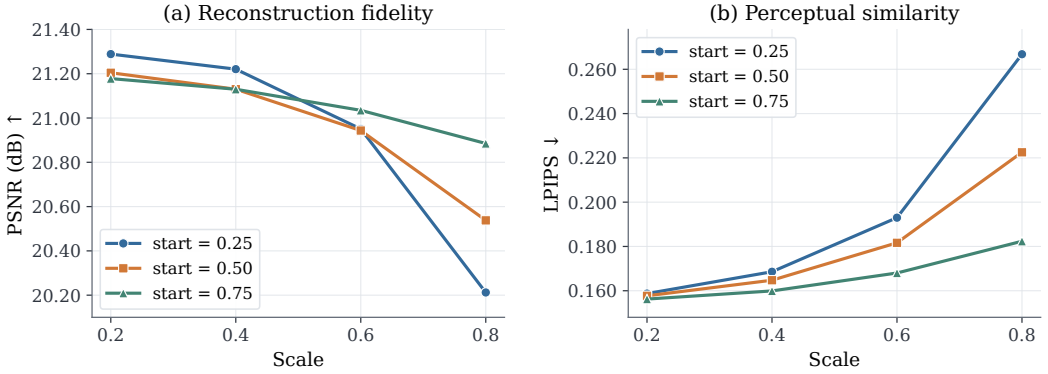


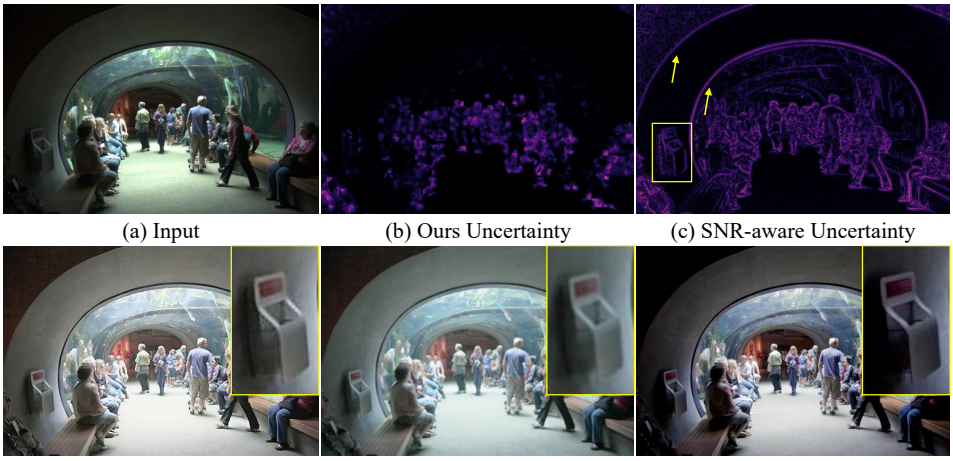
Figure 11: **Ablation on prior-image co-optimization with different scale and schedule.**

A.2 Ablation on the Choices of Uncertainty Estimation

We notice that SNR-aware Xu et al. (2022) adopts another uncertainty estimation method, which calculates the SNR map via relationship between original low-light images I_{input} and non-learning-based denoised images I_{denoise}:

$$I = \text{Lum}(I_{\text{input}}), \quad I_{\text{denoise}} = \text{Denoiser}(I),$$
$$I_{\text{noise}} = \text{abs}(I_{\text{denoise}} - I), \quad \mathbf{snr} = \frac{I_{\text{denoise}}}{I_{\text{noise}}}, \quad \mathbf{u} = \frac{1}{1 + \mathbf{snr}}, \tag{13}$$

where Lum denotes the process of extracting the luminance of images, and Denoise is the non-learning denoising method where we adopt `cv2.boxFilter()` with different window size. We



(a) Input (b) Ours Uncertainty (c) SNR-aware Uncertainty

(d) Ours Enhanced Result (e) Enhanced by SNR (f) Ground-Truth

Figure 12: **Ablative comparison between SNR maps and our uncertainty estimation.** Although they preserve structural details, SNR maps are not compatible with our uncertainty-guided fusion framework, leading to oversmoothed results without clear and sharp edges.

substitute our uncertainty estimation with $\mathbf{u}$ for ablation, where higher value in SNR maps corresponds to lower values in the uncertainty. The qualitative and quantitative results are reported in Fig. 12 and Tab. 4, respectively.

Table 4: **Ablation on uncertainty estimation methods.**

Variants	PSNR↑	SSIM↑	LPIPS↓
SNR map with window size $= 5$	21.32	0.816	0.167
SNR map with window size $= 7$	20.99	0.827	0.170
SNR map with window size $= 9$	21.27	0.830	0.165
Ours	**21.29**	**0.838**	**0.159**

Analysis. We argue that our uncertainty estimation provides a more precise modeling of noise and luminance, and SNR-aware is not fully compatible with our multi-scale fusion framework. The noise-denoise residual in SNR-aware also contains clear edges and textures, potentially underestimating structural reliability. Besides, SNR-aware ignore luminance, which has a negligible effect on the enhancer. As a result, the uncertainty-guided weighting will be confused to assign high attributes to the edge regions, leading to over-smooth results without sufficient supporting evidence.

In contrast, our uncertainty estimator uses the eigenvalues $\lambda_1 \geq \lambda_2$ of the local structure tensor:

$$\mathbf{u} = \frac{2\lambda_2/\kappa + \sigma_q^2}{\mu^2 + (\lambda_1 + \lambda_2)/\kappa + \sigma_q^2}. \tag{14}$$

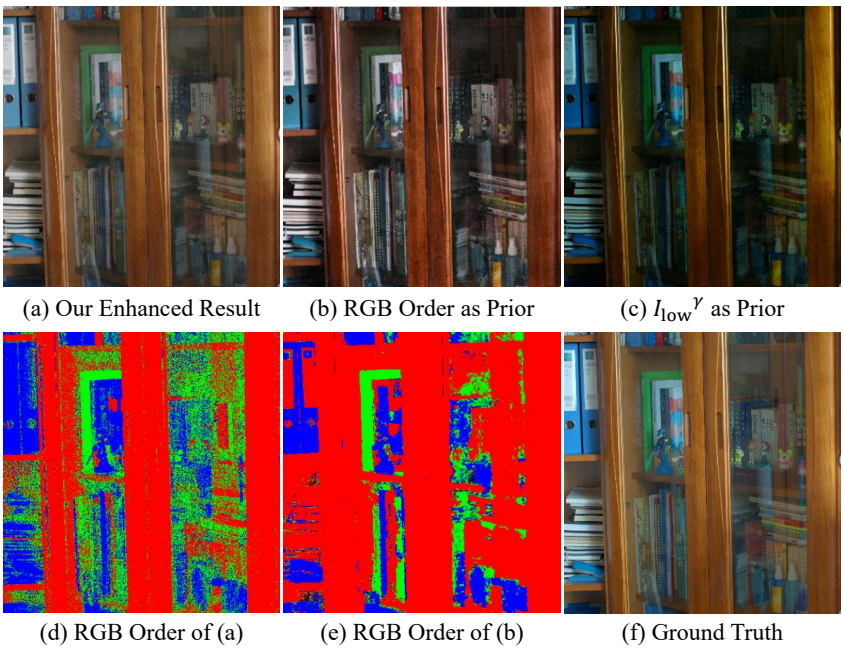
Here, $\lambda_1 - \lambda_2$ captures directional coherence, whereas $2\lambda_2$ measures isotropic gradient energy. We also detail this process in Eq. 17. Coherent edges typically yield $\lambda_1 \gg \lambda_2$ and low uncertainty, while isotropic noise tends to increase uncertainty. Therefore, our uncertainty estimation plays a more suitable role than SNR-aware modeling in our multi-scale fusion framework.

A.3 Ablation on the Choices of Uncertainty Scales

Our default configuration uses five Gaussian standard deviations, *i.e.*, $\{0.5, 1.0, 1.5, 2.0, 2.5\}$. We compare individual scales with their joint representation, and then vary the number and range of scale candidates, as shown in Tab. 5

Table 5: **Ablation experiments of prior scales on LOL.**

Variants	PSNR↑	SSIM↑	LPIPS↓
$N = 1,\ \sigma = 1.5$	20.56	0.791	0.301
$N = 3,\ \sigma \in [0.5, 2.5]$	20.69	0.801	0.312
$N = 7,\ \sigma \in [0.5, 2.5]$	19.38	0.782	0.352
$N = 5,\ \sigma \in [0.5, 1.5]$	20.29	0.817	0.291
$N = 5,\ \sigma \in [0.5, 3.5]$	21.10	0.825	0.266
Ours	**21.29**	**0.838**	**0.159**



(a) Our Enhanced Result (b) RGB Order as Prior (c) $I_{\text{low}}{}^{\gamma}$ as Prior

(d) RGB Order of (a) (e) RGB Order of (b) (f) Ground Truth

Figure 13: **Ablation on different color priors.**

A.4 Ablation on Color Prior

We conduct an ablation with another color prior modeling in Fig. 13, *i.e.*, we follow Wang et al. (2024) and adopt *RGB channel ordering* as the color prior, where each pixel is encoded as a three-channel feature with values in $\{-1, 0, 1\}$ to represent the relative ordering among the red, green, and blue channels. Although effective in QuadPrior to correct color bias, the RGB order is too ambiguous to represent precise chromatic cues. As verified in Fig. 13 (d),(e), obviously different colors in Fig. 13 (a),(b) are projected into the same RGB order. We further demonstrate the "robustness-informativeness trade-off" in the main text, using the gamma-corrected low-light input image as the condition during generative enhancement. The result is under-enhanced with dark regions and bluish books, showcasing the importance to construct a proper information bottleneck.

A.5 Ablation on Perturbed Data Training

We apply perturbed data during training following QuadPrior Wang et al. (2024). We use gamma correction with $\gamma \in [2.0, 3.5]$, Gaussian noise with standard deviation $\sigma \in [0.5, 5.0]$ and Poisson noise with $\lambda \in [0.18, 0.26]$. We found such training scheme is necessary. Training without perturbed data leads to severe color shifts and noisy results.



w/o perturbed data training Ours (Full Version)

Figure 14: **Ablation experiment on perturbed data training.**

B Details of Illumination-Invariant Prior Construction

Confidence estimation. We compute luminance as

$$Y = 0.2126I_r + 0.7152I_g + 0.0722I_b. \tag{15}$$

Let $\mathcal{A}_w$ denote local averaging with a $w \times w$ window, implemented by average pooling, where we set $w = 7$. The local mean and averaged structure tensor are

$$\mu = \mathcal{A}_w(Y), \qquad \mathbf{J} = \mathcal{A}_w(\nabla Y \nabla Y^\top). \tag{16}$$

Given the eigenvalues $\lambda_1 \geq \lambda_2 \geq 0$ of $\mathbf{J}$, we define

$$q(\mathbf{I}) = \frac{\mu^2 + (\lambda_1 - \lambda_2)/\kappa}{\mu^2 + (\lambda_1 + \lambda_2)/\kappa + \sigma_q^2}, \qquad \mathbf{u} = 1 - q(\mathbf{I}), \tag{17}$$

where $\kappa = 0.23$ scales derivative energy and $\sigma_q^2 > 0$ sets a stabilizing confidence floor in the denominator. The eigenvalue difference captures directional structure, while the sum measures total gradient energy. For fixed total energy, stronger directional structure increases confidence, whereas isotropic variation lowers it. In a locally constant region, $q = \mu^2/(\mu^2 + \sigma_q^2)$, so confidence depends on brightness rather than flatness alone.

Gaussian scale-space construction. Following CIConv Lengyel et al. (2021) and Quad-Prior Wang et al. (2024), a Gaussian color transformation maps $\mathbf{I}$ to an intensity descriptor E and spectral descriptors E_λ and $E_{\lambda\lambda}$. For each standard deviation σ_i, we compute

$$\begin{aligned} E_i &= G_{\sigma_i} * E, \\ E_{\lambda,i} &= G_{\sigma_i} * E_\lambda, \qquad E_{\lambda\lambda,i} = G_{\sigma_i} * E_{\lambda\lambda}, \\ E_{x,i} &= \partial_x E_i, \qquad E_{y,i} = \partial_y E_i, \end{aligned} \tag{18}$$

where G_{σ_i} is a spatial Gaussian kernel with variance σ_i^2, and $*$ denotes convolution. These operations provide complementary responses across fine and coarse spatial support.

Physical invariant responses. At each scale, we construct

$$\begin{aligned} P_i^{\text{hue}} &= \operatorname{atan2}(E_{\lambda\lambda,i}, E_{\lambda,i} + \epsilon), \\ P_i^{\text{sat}} &= \frac{E_{\lambda,i}^2 + E_{\lambda\lambda,i}^2}{E_i^2 + \epsilon}, \\ P_i^{\text{str}} &= \arctan\left[\left(\frac{E_{x,i}}{E_i + \epsilon}\right)^2 + \left(\frac{E_{y,i}}{E_i + \epsilon}\right)^2\right], \end{aligned} \tag{19}$$

with $\epsilon = 10^{-5}$. These responses follow physical invariants derived under simplifying assumptions of the Kubelka–Munk model Kubelka & Munk (1999); Lengyel et al. (2021); Wang et al. (2024). They reduce sensitivity to illumination changes under those assumptions; exact invariance is not guaranteed under arbitrary spatially or spectrally varying illumination, additive noise, or numerical stabilization.

Chromatic encoding. Bilateral filtering suppresses local fluctuations while retaining prominent boundaries before HSV extraction. The RGB input to the HSV chromaticity branch is filtered with a reflected 5×5 bilateral filter. The range bandwidth is $\sigma_{cw} = 0.08$, the spatial bandwidth is $\sigma_{sw} = 2.0$, and the weight for neighbor offset δ is

$$\exp\left(-\frac{\|y_p - y_{p+\delta}\|_2^2/3}{2\sigma_{cw}^2} - \frac{\|\delta\|_2^2}{2\sigma_{sw}^2}\right). \tag{20}$$

This bilateral filter is not applied to the structural-invariant branch to preserve all original information. The hue angle ϑ in Eq. 5 spans $[0, 2\pi)$: $\vartheta = \pi H/3$ for $H \in [0, 6)$, or $\vartheta = 2\pi H$ for normalized hue $H \in [0, 1)$. Its sine-cosine encoding avoids the discontinuity at the hue wraparound. The reliability parameter $\sigma_c = 0.03$ controls attenuation of chromatic cues at low luminance. We concatenate the weighted chromatic channels with the normalized structural channels and encode them through convolution and downsampling.

C Proof: Illumination Invariance of the Multi-Scale Priors

In this section, we will establish the illumination invariance of the constructed priors in Eq. 19 under idealized imaging assumptions, and clarify its limitations under practical low-light conditions.

Imaging assumptions. Following the Kubelka–Munk-based formulation adopted in Quad-Prior Wang et al. (2024); Kubelka & Munk (1999), a matte surface under equal-energy, spatially uniform illumination satisfies

$$E(\lambda, s) = aR_\infty(\lambda, s), \qquad a > 0, \tag{21}$$

where E denotes the observed spectral energy, R_∞ denotes material reflectance, and a is the illumination intensity. For a fixed scene and a linear sensor response, changing only the illumination intensity therefore gives

$$\mathbf{I}' = c\mathbf{I}, \qquad c > 0. \tag{22}$$

This relation assumes unchanged imaging geometry and excludes additive noise, clipping, and nonlinear camera processing.

Proposition. For a fixed linear color transformation and fixed Gaussian scales, the three priors in Eq. 19, with $\epsilon = 0$, are invariant to Eq. equation 22 wherever their defining expressions are well defined.

Proof. The Gaussian color transformation is linear. Hence,

$$E' = cE, \qquad E'_\lambda = cE_\lambda, \qquad E'_{\lambda\lambda} = cE_{\lambda\lambda}. \tag{23}$$

Spatial Gaussian convolution and differentiation are also linear. Since c is spatially constant, at every scale i we obtain

$$\begin{aligned} E'_i &= cE_i, \qquad E'_{\lambda,i} = cE_{\lambda,i}, \quad E'_{\lambda\lambda,i} = cE_{\lambda\lambda,i}, \\ E'_{x,i} &= cE_{x,i}, \quad E'_{y,i} = cE_{y,i}. \end{aligned} \tag{24}$$

Substitution into the prior definitions yields

$$\begin{aligned} P_i^{\text{hue}}(\mathbf{I}') &= \operatorname{atan2}(cE_{\lambda\lambda,i}, cE_{\lambda,i}) = P_i^{\text{hue}}(\mathbf{I}), \\ P_i^{\text{sat}}(\mathbf{I}') &= \frac{c^2E_{\lambda,i}^2 + c^2E_{\lambda\lambda,i}^2}{c^2E_i^2} = P_i^{\text{sat}}(\mathbf{I}), \\ P_i^{\text{str}}(\mathbf{I}') &= \arctan\left[\left(\frac{cE_{x,i}}{cE_i}\right)^2 + \left(\frac{cE_{y,i}}{cE_i}\right)^2\right] = P_i^{\text{str}}(\mathbf{I}). \end{aligned} \tag{25}$$

The hue equality follows from the invariance of $\mathrm{atan2}$ to positive common scaling. The remaining equalities follow by cancellation of the illumination factor. The result holds independently at each fixed scale, so extending the construction to multiple scales preserves this property for every constituent prior.

Effect of numerical stabilization. The implementation uses $\epsilon = 10^{-5}$ to stabilize weak-signal computations. With a fixed $\epsilon > 0$, the transformed responses become

$$\begin{aligned} P^{\mathrm{hue}}_{i,\epsilon}(\mathbf{I}') &= \mathrm{atan2}(E_{\lambda\lambda,i}, E_{\lambda,i} + \epsilon/c), \\ P^{\mathrm{sat}}_{i,\epsilon}(\mathbf{I}') &= \frac{E^2_{\lambda,i} + E^2_{\lambda\lambda,i}}{E^2_i + \epsilon/c^2}, \\ P^{\mathrm{str}}_{i,\epsilon}(\mathbf{I}') &= \arctan\left[\left(\frac{E_{x,i}}{E_i + \epsilon/c}\right)^2 + \left(\frac{E_{y,i}}{E_i + \epsilon/c}\right)^2\right]. \end{aligned} \tag{26}$$

Thus, numerical stabilization introduces illumination-dependent deviations. Approximate invariance holds when the stabilization terms are negligible relative to the relevant signal magnitudes, away from singularities and angular branch discontinuities.

Departures under low-light degradation. The proposition does not imply invariance to arbitrary spatial or spectral illumination changes. For spatially varying illumination $c(s)$, Gaussian smoothing does not generally commute with multiplication:

$$G_{\sigma_i} * (cE) \neq c\,(G_{\sigma_i} * E). \tag{27}$$

A locally constant approximation is therefore appropriate only when illumination varies sufficiently slowly over the filtering neighborhood.

Additive degradation further perturbs the normalized responses. For example, let a degraded intensity descriptor satisfy $E'_i = cE_i + n_i$, where n_i is the filtered perturbation. Ignoring numerical stabilization, its normalized derivative obeys

$$\frac{E'_{x,i}}{E'_i} - \frac{E_{x,i}}{E_i} = \frac{E_i\,\partial_x n_i - E_{x,i} n_i}{E_i(cE_i + n_i)}. \tag{28}$$

This expression shows how weak intensity and perturbation gradients can destabilize the structural response. Spectral perturbations similarly affect hue and saturation.

These observations distinguish analytical illumination invariance from reliable estimation. Our uncertainty-guided fusion addresses the latter by adapting the contribution of each scale to observation reliability. Since the routing weights depend on the input, exact invariance of the constituent priors alone does not guarantee invariance of their adaptive fusion; its cross-illumination robustness is evaluated empirically.

D Implementation Details for Reproduction

We optimize LoRA parameters with Prodigy using a learning rate of 1.0, $(\beta_1, \beta_2) = (0.9, 0.99)$, and weight decay 0.01, with decoupled weight decay, bias correction, and safeguard warm-up enabled. The prior projection layer is optimized with AdamW using a `lr` of 10^{-4} and weight decay 10^{-4}. We train our model with 24k iterations and finetune it with 1.5k iterations in the second stage.

Two-stage training. Since the decoding target is defined in a learned intrinsic space, we first establish this representation before introducing reconstruction regularization. Therefore, the training process is separated into two stages: Stage 1 optimizes the prior encoder and condition-stream LoRA with $\mathcal{L}_{\mathrm{FM}}$. Stage 2 introduces the regularization decoder and optimizes Eq. 9, updating LoRA, the confidence-curve branch, and the decoder.

Statistics of the Training Data. We visualize the distribution of 35,795 images in terms of mean brightness and contrast (pixel-intensity standard deviation). The samples cover a broad range of illumination conditions, including dark regions, high-dynamic-range scenes, and overexposed cases, demonstrating substantial variation in both global brightness and local intensity dynamics.

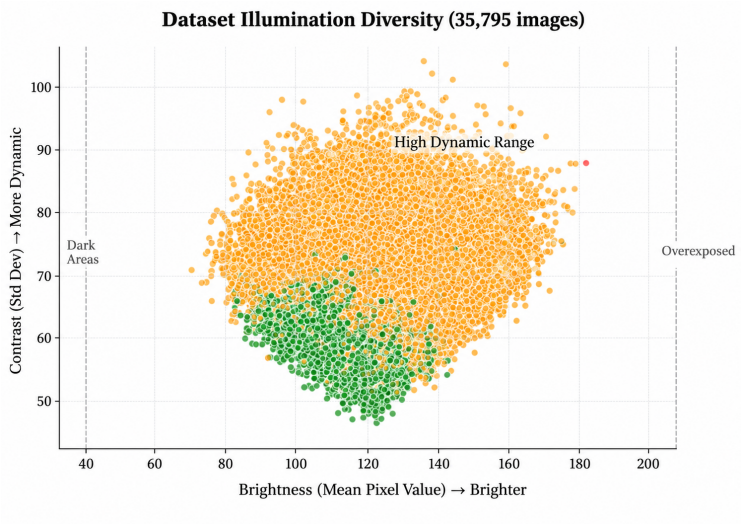


Figure 15: **Illumination diversity of the training dataset.** The 35,795 images span a wide range of brightness and contrast levels, covering dark, high-dynamic-range, and overexposed conditions.

Prior encoder. The analytical prior has six channels. Its convolutional encoder runs in FP32 and remains frozen in Stage 2 to maintain a relatively stable prior feature space for decoding regularization. Table 6 gives the architecture of prior encoder, which converts a 1024×1024 input using every convolution with a 3×3 kernel with padding 1.

Table 6: **Prior encoder architecture.**

Layer	Operation	Output shape
E_1	Conv. $6 \to 32$, stride 2; SiLU	$32 \times 512 \times 512$
E_2	Conv. $32 \to 32$, stride 2; SiLU	$32 \times 256 \times 256$
E_3	Conv. $32 \to 32$, stride 2; SiLU	$32 \times 128 \times 128$
E_4	Conv. $32 \to 16$, stride 1	$16 \times 128 \times 128$

Finally, FLUX [2]adopts 2×2 packing to convert the E_4 feature map into 4096 prior tokens of width 64.

Scale routing networks. The learnable weighting module which predict spatial weights in Eq. 3 applies two reflection-padded 3×3 convolutions, each with 16 output channels and SiLU activation, followed by a 1×1 convolution that maps 16 channels to five spatial logits. A softmax over the five Gaussian scales produces the spatial fusion weights. The trainable image-level curve branch which output the modulating parameters, *i.e.*, γ, β, applies reflection-padded 3×3 convolutions with channel widths $4 \to 16 \to 16$, each followed by SiLU, then adaptive average pooling to 1×1 and a 1×1 convolution.

Regularization decoder architecture. The regularization decoder D_{reg} receives the final prior-stream hidden tokens of width 3072. At 1024×1024 resolution, these tokens form a 64×64 packed-token grid, which a 1×1 projection maps to 96 channels. Each residual block consists of ChannelLayerNorm, SiLU, a 3×3 convolution, ChannelLayerNorm, SiLU, and another 3×3 convolution, together with an identity or 1×1 projected skip connection. The encoder uses residual-block widths 96, 192, 384, and 384. Each transition downsamples by antialiased bilinear interpolation by a factor of two and then applies a 3×3 convolution. A 384-channel residual bottleneck follows the fourth downsampling and operates on a 4×4 grid.

The decoder contains four bilinear-resize-plus-3×3-convolution upsampling stages. After skip concatenation, their residual-block widths are 384, 384, 192, and 96. A 3×3 output projection produces

[2]https://huggingface.co/black-forest-labs/FLUX.1-dev

64 packed residual channels. FLUX unpacking converts them to a $16 \times 128 \times 128$ residual, which is added to the unpacked input prior. The unpacked refinement head concatenates the 16-channel base prior and 16-channel residual, then applies a 3×3 convolution from 32 to 64 channels, two 64-channel residual blocks, and a 3×3 convolution from 64 to 16 channels. Both the packed output projection and final refinement projection are zero-initialized, so the decoder initially implements the identity mapping. The instantiated decoder contains 26,194,992 parameters.

Stage-2 objective. The exact decoder and total losses are

$$\begin{aligned} \mathcal{L}_{\mathrm{dec}} &= \mathcal{L}_{\mathrm{mse}} + 10\mathcal{L}_{\mathrm{ssim}} + 0.2\mathcal{L}_{\mathrm{grad}} + 0.2\mathcal{L}_{\mathrm{lap}}, \\ \mathcal{L} &= \mathcal{L}_{\mathrm{FM}} + 1.0\mathcal{L}_{\mathrm{dec}}. \end{aligned} \tag{29}$$

Thus $(\lambda_s, \lambda_g, \lambda_l) = (10, 0.2, 0.2)$ and $\lambda_p = 1.0$. SSIM averages reflected-window losses at 5×5 and 11×11 after applying detached, per-channel target-range normalization. The gradient and four-neighbor Laplacian terms use L_1 differences on the 16-channel prior latents. We use synthetic perturbed data for training following QuadPrior to reduce noise and color shifts. During stage-2, the clean intrinsic condition is determined with the unperturbed clean normal-light image.

Inference protocol and latency. Inference uses `FlowMatchEulerDiscreteScheduler` with dynamic shifting enabled: base and maximum sequence lengths are 256 and 4096, and the corresponding shifts are 0.5 and 1.15. The default replay uses 1024×1024 resolution, 28 Euler flow steps, an empty prompt, guidance scale 1.0, seed 42, and the bundled BF16 packed initial noise. Prior refinement is an anchored feed-forward update with strength $\alpha = 0.2$ beginning after 25% of the steps; gradient-based dynamic prior optimization is disabled.

On one NVIDIA H20 96 GB GPU, a 1024×1024 image takes 53.50 ± 0.13 seconds per image, where the uncertainty is the population standard deviation over three measured runs after one warm-up. Timing begins after model loading and includes prior construction, FLUX denoising, decoder refinement, VAE decoding, and image post-processing.

D.1 Details of Model Distillation

We follow the data-distillation strategy of QuadPrior Wang et al. (2024) to transfer the enhancement capability of UG-Prior to a feed-forward student. We adopt Restormer Zamir et al. (2022) as the student backbone and incorporate our illumination-invariant prior through spatial feature transform (SFT) layers, following QuadPrior++ Huang et al. (2025a). This design transfers the teacher's enhanced appearance while retaining explicit scene guidance in the student.

Teacher supervision. Let $\mathcal{T}$ denote the trained UG-Prior teacher with prior constructor $\mathcal{P}_\psi$. For each distillation input $\mathbf{I}_n$, we generate a pseudo-target $\mathbf{Y}_n = \mathcal{T}(\mathbf{I}_n)$ using the complete teacher inference procedure. The student predicts

$$\widehat{\mathbf{Y}}_n = \mathcal{S}_\omega\big(\mathbf{I}_n, \mathbf{P}_n\big), \qquad \mathbf{P}_n = \mathcal{P}_\psi(\mathbf{I}_n), \tag{30}$$

where $\mathbf{P}_n$ is the spatial prior comprising the fused structural and reliability-weighted chromatic cues. We supervise the student with the teacher-generated images using

$$\mathcal{L}_{\mathrm{distill}} = \lambda_1 \left\| \hat{\mathbf{Y}}_n - \mathrm{sg}(\mathbf{Y}_n) \right\|_1 + \lambda_{\mathrm{ssim}} \mathcal{L}_{\mathrm{ssim}} \left(\hat{\mathbf{Y}}_n, \mathrm{sg}(\mathbf{Y}_n) \right) + \lambda_{\mathrm{lpips}} \mathcal{L}_{\mathrm{lpips}} \left(\hat{\mathbf{Y}}_n, \mathrm{sg}(\mathbf{Y}_n) \right), \tag{31}$$

where $\mathcal{L}_{\mathrm{ssim}} = 1 - \mathrm{SSIM}$, and $\mathcal{L}_{\mathrm{lpips}}$ denotes the LPIPS perceptual loss. We set $\lambda_1 = 1.0$, $\lambda_{\mathrm{ssim}} = 2.0$, and $\lambda_{\mathrm{lpips}} = 5.0$. The stop-gradient operator ($\mathrm{sg}(\cdot)$) ensures that the UG-Prior teacher remains fixed throughout distillation. The teacher remains fixed throughout distillation. Unlike the intrinsic reconstruction loss used to regularize the teacher, this objective supervises the student's RGB output.

Student architecture. We adopt a lightweight residual encoder–decoder with a Restormer-style bottleneck, following the architecture used in QuadPrior++ Huang et al. (2025a). The network comprises four resolution levels with channel widths $(C, 2C, 4C, 8C)$. Convolutional residual blocks process the encoder and decoder features, while two Transformer blocks operate at the bottleneck. The six-channel prior modulates features through four spatial feature transform (SFT) modules: after

each downsampling operation and after the bottleneck. Skip connections concatenate encoder and decoder features, and a global residual connection preserves the input signal. Table 7 summarizes the architecture.

Table 7: Student architecture with the default block configuration. C denotes the base channel width. RB and TB denote residual and Transformer blocks, respectively.

Stage	Operations	Resolution	Channels
Embedding	3×3 convolution	$H \times W$	C
Encoder 1	$2\times$ RB	$H \times W$	C
Encoder 2	Downsample, SFT_1, $2\times$ RB	$\frac{H}{2} \times \frac{W}{2}$	$2C$
Encoder 3	Downsample, SFT_2, $2\times$ RB	$\frac{H}{4} \times \frac{W}{4}$	$4C$
Bottleneck	Downsample, SFT_3, $2\times$ TB, SFT_4	$\frac{H}{8} \times \frac{W}{8}$	$8C$
Decoder 3	Upsample, concatenate, 1×1 conv., $2\times$ RB	$\frac{H}{4} \times \frac{W}{4}$	$4C$
Decoder 2	Upsample, concatenate, 1×1 conv., $2\times$ RB	$\frac{H}{2} \times \frac{W}{2}$	$2C$
Decoder 1	Upsample, concatenate, $2\times$ RB	$H \times W$	$2C$
Refinement	$2\times$ RB	$H \times W$	$2C$
Output	3×3 convolution, input residual addition	$H \times W$	3

Implementation Details. We use a part of training split of LOL and LSRW, in total 8,000 images. We crop them into 256×256 patches. We optimize the distilled model with L_1 loss, SSIM loss, and LPIPS loss, weighted by 1.0, 2.0, and 5.0 respectively. The whole training is completed on a H20 GPU with `AdamW` optimizer.

D.2 Evaluation Protocol

We resize all other models' outputs to 1024 resolution, and use `pyiqa` package [3] to compute metrics for fair comparison. We use the pretrained model weights to evaluate other methods from their released checkpoints.

Table 8: GitHub repositories of the compared methods.

Method	Repository
	Supervised (SL)
RetinexNet	https://github.com/weichen582/RetinexNet
KinD	https://github.com/zhangyhuaee/KinD
KinD++	https://github.com/zhangyhuaee/KinD_plus
URetinexNet	https://github.com/AndersonYong/URetinex-Net
Retinexformer	https://github.com/caiyuanhao1998/Retinexformer
DiffLL	https://github.com/JianghaiSCU/Diffusion-Low-Light
Diff-Retinex	https://github.com/XunpengYi/Diff-Retinex
	Unsupervised (UL)
EnlightenGAN	https://github.com/VITA-Group/EnlightenGAN
PairLIE	https://github.com/zhenqifu/PairLIE
NeRCo	https://github.com/Ysz2022/NeRCo
CLIP-LIT	https://github.com/ZhexinLiang/CLIP-LIT
Zero-DCE	https://github.com/Li-Chongyi/Zero-DCE
Zero-DCE++	https://github.com/Li-Chongyi/Zero-DCE_extension
RUAS	https://github.com/KarelZhang/RUAS
SCI	https://github.com/vis-opt-group/SCI
LightenDiffusion	https://github.com/JianghaiSCU/LightenDiffusion
ZeroIDIR	https://github.com/JianghaiSCU/ZeroIDIR
	Zero-shot (ZS)
ExCNet	https://github.com/csLinZhang/ExCNet
CoLIE	https://github.com/ctom2/colie
FourierDiff	https://github.com/aipixel/FourierDiff
QuadPrior	https://github.com/daooshee/QuadPrior
AGLLDiff	https://github.com/LYL1015/AGLLDiff
MR. Illuminate	https://github.com/joshua-cho-story/MR-Illuminate

[3] https://github.com/chaofengc/IQA-PyTorch