LEARNING STRUCTURAL ILLUMINATION FOR UNSUPERVISED LOW-LIGHT ENHANCEMENT

Anonymous authors
Paper under double-blind review

ABSTRACT

Existing unsupervised low-light image enhancement (LLIE) methods often estimate illumination directly from the entire low-light input, without separating its spatially varying illumination pattern, termed relative illumination structure, from the absolute exposure level or preventing unreliable low signal-to-noise ratio regions from biasing the estimate. Moreover, fixed exposure targets impose a scene-agnostic enhancement criterion, limiting adaptation across diverse lighting conditions. Inspired by the spatial propagation of light, we propose a *Relative Illumination Structure Estimation* (RISE) framework that decouples relative illumination structure from absolute exposure and infers it from reliable bright regions, enabling interpretable and robust enhancement. For scene-adaptive exposure adjustment, we further propose a *Dual-Metering Exposure Reference* derived from each input, allowing RISE to adapt the enhancement strength to individual scenes and generalize across diverse lighting conditions. Extensive benchmark and real-world generalization experiments show that RISE achieves state-of-the-art performance on most datasets among unsupervised LLIE methods while producing visually natural results. Code will be made publicly available.

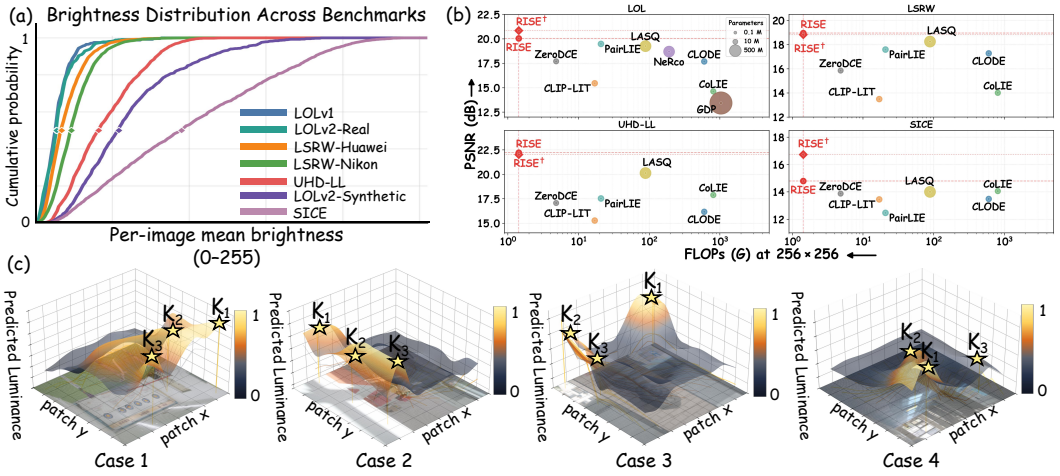


Figure 1: (a) Empirical cumulative distribution functions (CDFs) of per-image mean brightness across seven benchmarks. (b) RISE delivers a favorable PSNR–FLOPs trade-off across the evaluated benchmarks, while RISE† achieves higher PSNR than the compared prior unsupervised methods. (c) Joint visualization of relative illumination structure and absolute exposure: surface shape encodes relative illumination, elevation encodes absolute exposure, and stars mark selected KICs.

1 INTRODUCTION

Low-light image enhancement (LLIE) is a fundamental vision task that aims to recover well-exposed images from underexposed observations, benefiting both human perception and downstream understanding. However, supervised methods rely on costly, hard-to-scale, and capture-biased paired data, motivating unsupervised LLIE that learns from low-light images without paired ground truth.

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.

In recent years, deep learning has substantially advanced unsupervised LLIE. However, without paired ground-truth references, there is no direct supervisory signal for target exposure. To compensate, many unsupervised methods (Guo et al., 2020; Li et al., 2025) are trained with fixed exposure values obtained from statistical priors as surrogate supervision. Under such objectives, scene-agnostic supervision tends to bias enhanced results toward a uniform brightness level, overlooking variations among different scenes and preventing image-adaptive enhancement. Consequently, such methods fail to generalize well across the diverse brightness conditions illustrated in Fig. 1(a).

Another key challenge is the interference of low signal-to-noise ratio (SNR) regions in illumination estimation. Under severe under-exposure, pixel observations are dominated by sensor noise and contain only weak scene signals. Nevertheless, most methods (Guo et al., 2016; Zhang et al., 2019; Jiang et al., 2024) estimate illumination by treating the whole low-light image as uniformly reliable, leading to unstable enhancement. As shown in Fig. 2(a), the model may fail to brighten extremely dark regions while preventing highlight overexposure. Although recent supervised method SNR-Aware (Xu et al., 2022) exploits long-range dependencies between low-SNR and high-SNR regions to alleviate this issue, it still treats unreliable regions as useful features, leaving their interference unresolved.

Beyond these challenges, the representation of scene illumination itself is also critical. *Rather than a single quantity, illumination comprises a spatially varying structure that captures relative brightness relationships and a global scale that controls the absolute exposure level* (Zhan et al., 2021; Garon et al., 2019). Ideally, LLIE should therefore be formulated as $\mathbf{S} = g_\theta(\mathbf{I}_{\text{low}})$, $\eta = h_\phi(\mathbf{I}_{\text{low}})$, and $\hat{\mathbf{I}} = \mathcal{E}(\mathbf{I}_{\text{low}}; \mathbf{S}, \eta)$, where $\mathbf{S}$ captures the relative illumination structure and η independently determines the desired exposure level. Existing methods, however, typically predict a single enhancement representation $\mathbf{Z} = f_\theta(\mathbf{I}_{\text{low}})$ and produce the enhanced result as $\hat{\mathbf{I}} = \mathcal{E}(\mathbf{I}_{\text{low}}; \mathbf{Z})$, where $\mathbf{Z}$ may be an illumination map (Xie et al., 2024), an enhancement curve (Guo et al., 2020), or the enhanced image itself (Wang et al., 2024). Such a formulation entangles spatial illumination structure with absolute exposure, leaving their distinct roles neither explicitly identifiable nor independently controllable.

These limitations motivate us to seek illumination cues that are both structured and reliable. Such cues naturally arise from how light is distributed in real scenes. As shown in Fig. 2(b), light propagates from its source with spatially varying strength due to distance, occlusion, absorption, and other scene factors. Similar patterns can be observed in normal-light images, where brightness variations exhibit relative propagation from well-lit regions to darker areas under diverse lighting conditions, as shown in Fig. 2(c). However, directly modeling the full propagation mechanism from a single low-light image is ill-posed because the underlying scene factors are complex and unobserved. But this view suggests that illumination can be inferred through relative relationships to reliable bright regions, with the relative relationships themselves reflecting the underlying illumination structure, as illustrated in Fig. 2(d). These regions lie on strong light-propagation paths, and we refer to them as Key Illumination Cues (KICs). Based on this insight, we propose a Relative Illumination Structure Estimation (RISE) framework that learns illumination structure by modeling relative relationships to KICs and decouples it from the absolute exposure level, reducing unstable enhancement.

Meanwhile, to achieve scene-adaptive exposure adjustment, we further propose a Dual-Metering Exposure Reference (DMER) derived from each input. DMER replaces fixed exposure values and training-data statistics with an input-dependent reference that accounts for both global exposure demand and relative illumination structure. Guided by this reference, RISE adjusts exposure for individual images rather than memorizing dataset-specific brightness preferences, improving generalization across datasets and real-world scenes.

Our main contributions are summarized as follows:

- We recast unsupervised LLIE from a structured and relational illumination perspective, providing a new problem formulation for low-light image enhancement.
- We propose a Relative Illumination Structure Estimation (RISE) framework that decouples the relative illumination structure from absolute exposure and infers illumination from sparse yet reliable bright regions for interpretable and robust enhancement.
- We design Dual-Metering Exposure Reference (DMER), an input-dependent exposure reference that captures global and relative illumination cues, improving cross-dataset and real-world generalization.

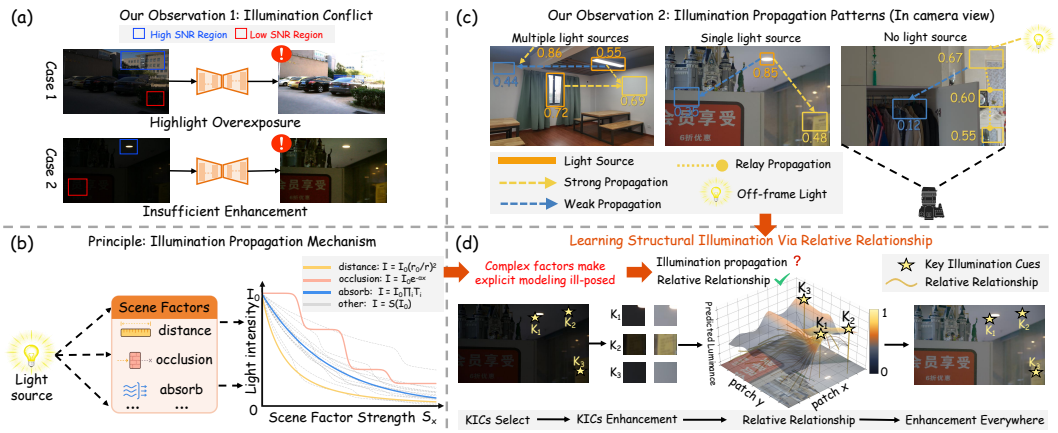


Figure 2: Motivation of RISE. (a) Existing unsupervised methods show unstable enhancement. (b) Light intensity from a source attenuates with scene factors. (c) Relative illumination propagation patterns in normal-light images (numbers denote regional brightness). (d) Our insight: RISE learns illumination structure from relative relationships to key illumination cues.

Comprehensive experiments on multiple benchmarks show that RISE achieves state-of-the-art performance on most datasets, as summarized in Fig. 1(b), and more natural enhancement, while extensive ablation studies verify the rationality and effectiveness of the proposed design.

2 Related Work

Unsupervised LLIE. Supervised LLIE methods (Guo & Hu, 2023; Du et al., 2026) learn from paired low-/normal-light images, but their dependence on costly paired data limits scalability and motivates unsupervised enhancement. Existing unsupervised methods seek alternative supervision signals to replace paired references. Li et al. (Li et al., 2025) control exposure by enforcing image patches toward a fixed brightness value, ResQ-Net (Xie et al., 2024) uses the gray-world assumption to drive exposure adjustment, LightenDiffusion (Jiang et al., 2024) and related diffusion-based methods learn latent normal-light priors from unpaired data, CLIP-LIT (Liang et al., 2023) trains the enhancement network by maximizing similarity to positive language prompts, and PairLIE (Fu et al., 2023) mines supervision from paired low-light images of the same scene under different exposures. Although these methods have achieved promising results, their supervision still depends on fixed exposure targets, training-data statistics, or external priors. As a result, they tend to learn dataset-dependent enhancement preferences rather than scene-adaptive exposure adjustment, leading to limited generalization under diverse lighting conditions.

Illumination Modeling for LLIE. In LLIE, the illumination is explicitly or implicitly modeled. Retinex-based methods model illumination with dense maps. LIME (Guo et al., 2016) constructs an illumination map from channel-wise maxima and refines it with structure-aware smoothing. RUAS (Liu et al., 2021) searches for an illumination estimation network from a compact architecture space through architecture search. IRLE (He et al., 2026) guides illumination estimation with an inverse gain map from the input image. Another line of work models enhancement as iterative lightening curves. Zero-DCE (Guo et al., 2020) predicts pixel-wise parameters of high-order curves, while CLODE (Jung et al., 2025) formulates curve adjustment as a continuous dynamic system and solves it with neural ordinary differential equations. Implicit representations have also been explored for enhancement. LLNeRF (Wang et al., 2023a) enhances brightness in 3D neural radiance fields, NeRCo (Yang et al., 2023) uses implicit neural representations to normalize low-light inputs under different degradation levels, and NoiSER (Zhang et al., 2024) implicitly learns brightness enhancement from zero-mean Gaussian noise. Although these methods represent illumination in different forms, they do not explicitly treat scene illumination as a structured quantity decoupled from absolute exposure, nor do they account for the adverse influence of unreliable low-SNR regions on illumination representation.

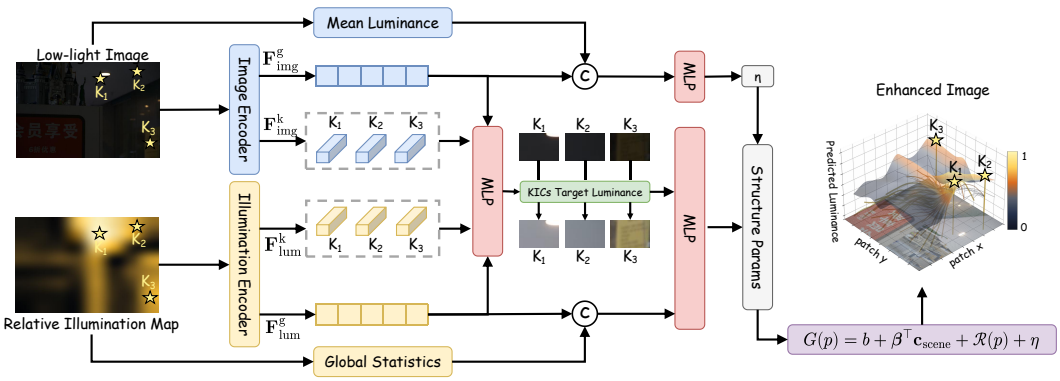


Figure 3: Overview of RISE. RISE jointly encodes image appearance and relative illumination, using spatially dispersed KICs as reliable anchors to estimate illumination structure while independently predicting absolute exposure for scene-adaptive enhancement, enabling reliable structural modeling and adaptive exposure control across diverse scenes.

3 METHOD

As schematically depicted in Fig. 3, RISE learns scene illumination structure by modeling relative relationships to KICs while separately controlling the absolute exposure level. Given a low-light input $\mathbf{I}_{\text{low}}$, RISE first derives a relative illumination map $\mathbf{Z}$, where global brightness is removed. It then selects N spatially sparse bright regions as KICs and first estimates their enhanced luminance, as their higher SNR enables more reliable luminance prediction. Based on these KICs, a relation modeling branch estimates their relative relationships to other regions, forming the relative illumination structure $\mathbf{S}$ that relates KICs to the rest of the scene. In parallel, an exposure control branch predicts an image-level exposure offset η, which determines the overall output brightness. The whole framework is trained with unsupervised objectives guided by DMER.

3.1 RELATIVE ILLUMINATION STRUCTURE ESTIMATION

Since the light source and propagation factors are not directly observable, RISE uses KICs as observable anchors for estimating the relative illumination structure. We first convert the input $\mathbf{I}_{\text{low}}$ to the luminance channel $\mathbf{Y} = 0.299\,\mathbf{I}^{\text{R}}_{\text{low}} + 0.587\,\mathbf{I}^{\text{G}}_{\text{low}} + 0.114\,\mathbf{I}^{\text{B}}_{\text{low}}$. Then, $\mathbf{Y}$ is average-pooled with an $h \times w$ window to obtain patch luminance $\mathbf{P}$, where each element corresponds to the mean luminance of a local patch. To remove global brightness and retain relative illumination, we define the relative illumination map $\mathbf{Z}$:

$$\mathbf{Z} = \log\left(\frac{\mathbf{P}}{\text{mean}(\mathbf{P}) + \epsilon}\right). \tag{1}$$

Here, ϵ ensures numerical stability, while $\log(\cdot)$ maps normalized brightness ratios to signed offsets relative to the unit-ratio reference and compresses their dynamic range.

We then select N spatially dispersed KICs from the brightest patches in $\mathbf{Z}$, denoted by $\{k_n\}_{n=1}^{N}$. An MLP predicts their target luminance values $\hat{\boldsymbol{\ell}}$, which serve as reliable anchors for relational illumination modeling. RISE builds a structured field by positioning each patch relative to the KICs in illumination and spatial domains. The spatial relation is motivated by the attenuation of light along its propagation path. However, the true propagation distance is three-dimensional and cannot be reliably recovered from a single image. We therefore use image-plane distance only as a soft attenuation surrogate rather than a physical measurement. For patch p and KIC k_n, we define:

$$\delta_n(p) = Z_{k_n} - Z_p, \qquad \rho_n(p) = \frac{1}{1 + d(p, k_n)/D}, \tag{2}$$

where $d(p, k_n)$ is the patch-grid distance and D is the grid diagonal. The kernel $\rho_n(p)$ equals one at the KIC and decays monotonically with distance, providing a bounded attenuation prior rather than an inverse-square constraint on image-plane geometry. The discrepancy $\delta_n(p)$ retains the observed

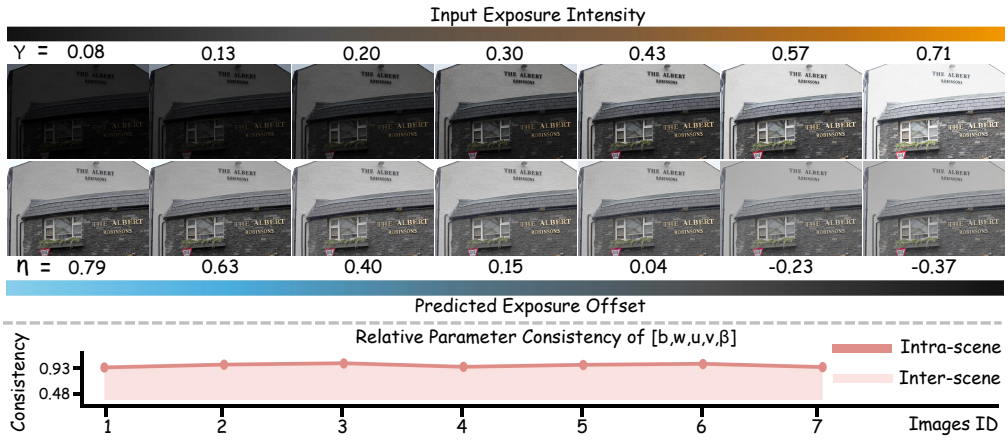


Figure 4: Validation of illumination structure and exposure decoupling on SICE. Under different exposures of the same scene, RISE maintains stable relative parameters while adjusting the absolute exposure, which demonstrates its ability to both enhance underexposed inputs and suppress overexposed scenes. In contrast, relative parameter consistency across different scenes remains low.

relative illumination change and implicitly reflects occlusion, absorption, and other effects beyond image-plane distance. Together, $\rho_n(p)$ and $\delta_n(p)$ define the relational field:

$$\mathcal{R}(p) = \sum_{n=1}^{N} \left[w_n \, \delta_n(p) + u_n \, \rho_n(p) \, \delta_n(p) + v_n \, \rho_n(p) \, Z_{k_n} \right]. \tag{3}$$

The unmodulated discrepancy $\delta_n(p)$ preserves the observed relation when image-plane distance is unreliable. Its distance-weighted form $\rho_n(p)\delta_n(p)$ combines the attenuation prior with the implicit scene factors encoded in $\delta_n(p)$, while $\rho_n(p)Z_{k_n}$ transfers the illumination strength of the KIC under the same decay. The image-adaptive coefficients w_n, u_n, and v_n balance these cues for each KIC, allowing multiple well-lit regions to jointly shape the relational field at each patch.

Although $\mathcal{R}(p)$ captures pairwise KIC-to-patch relations, similar local relations may arise from different scene-level illumination patterns. We therefore introduce $\mathbf{c}_{\text{scene}}$ to encode these differences by summarizing global statistics of $\mathbf{Z}$. The affine projection $C_{\text{scene}} = \boldsymbol{\beta}^\top \mathbf{c}_{\text{scene}} + b$ broadcasts this scene context to all regions. Together, the relational field and scene context define the relative illumination structure, while η independently controls the absolute exposure. Since illumination correction changes luminance by rescaling rather than shifting it, we parameterize the correction in log-gain space, where separate gain factors compose additively:

$$\hat{G}(p) = \underbrace{C_{\text{scene}} + \mathcal{R}(p)}_{\text{relative structure } S(p)} + \underbrace{\eta}_{\text{absolute exposure}}. \tag{4}$$

The relative parameters $\boldsymbol{\Theta} = \left(b, \boldsymbol{\beta}, \{w_n, u_n, v_n\}_{n=1}^{N}\right)$ and the absolute exposure η are predicted by separate MLPs. Each region has a spatially varying $S(p)$ derived from KIC based relations, while all regions share the same η. Thus, $S(p)$ represents the relative illumination structure and η determines the absolute exposure level, as further validated in Fig. 4. The enhanced result is given by

$$\hat{\mathbf{I}} = \mathbf{I}_{\text{low}} \odot \exp\left(\text{Up}(\hat{\mathbf{G}})\right). \tag{5}$$

3.2 Dual-Metering Exposure Reference

Scene-adaptive enhancement should not rely on a fixed global target, since low-light images differ not only in exposure level but also in the spatial distribution of illumination. As shown in Fig. 5, DMER estimates the enhancement reference from each input with two complementary exposure meters. The absolute meter captures the global exposure trend through a conservative gray-world

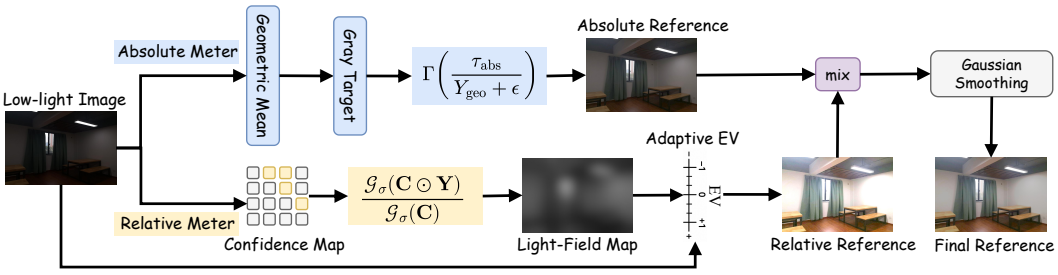


Figure 5: Overview of DMER. The absolute meter estimates global exposure, while the relative meter models spatial illumination through light-field propagation.

estimate, while the relative meter adjusts the reference according to spatial illumination differences. The absolute meter constructs its reference by rescaling the global luminance toward a gray level:

$$g_{\text{abs}} = \Gamma\left(\frac{\tau_{\text{abs}}}{Y_{\text{geo}}+\epsilon}\right), \qquad \mathbf{T}_{\text{abs}} = g_{\text{abs}}\mathbf{I}_{\text{low}}. \tag{6}$$

Here, Y_{geo} is the geometric mean luminance of $\mathbf{Y}$, and τ_{abs} is a fixed gray target across datasets in the normalized intensity range. If this gain would push bright regions beyond the valid intensity range, $\Gamma(\cdot)$ clips it to the maximum gain that avoids overexposure.

Although this absolute reference provides a stable global baseline, it is spatially uniform and cannot reflect illumination imbalance within the image. DMER therefore complements it with a relative meter that uses the estimated illumination structure to modulate the reference. Specifically, we first estimate a luminance based confidence map as $\mathbf{C} = \mathbf{Y}/(\mathbf{Y} + \text{mean}(\mathbf{Y}) + \epsilon)$, which assigns higher confidence to reliably illuminated regions. We then diffuse reliable luminance cues to construct a structure-aware light field:

$$\mathbf{L} = \frac{\mathcal{G}_\sigma(\mathbf{C}\odot\mathbf{Y})}{\mathcal{G}_\sigma(\mathbf{C})}, \tag{7}$$

where $\mathcal{G}_\sigma$ is Gaussian smoothing. This operation approximates the smooth spatial variation of illumination from reliable lit regions, allowing $\mathbf{L}$ to capture relative lighting strength across regions.

To translate $\mathbf{L}$ into an exposure adjustment, we adopt the Exposure-Value (EV) scale used in photography, where *one stop corresponds to a twofold change in exposure*. Following the conventional normal-key setting in photographic tone reproduction (Reinhard et al., 2002), we use $m = 0.18$ as the 0-stop reference. The observed luminance $\mathbf{Y}$ and the structure-aware light field $\mathbf{L}$ provide complementary exposure corrections, which are naturally additive in EV stops:

$$e_{\text{rel}} = \log_2 \frac{m}{\text{mean}(\mathbf{Y})} + \log_2 \frac{m}{\text{mean}(\mathbf{L})}. \tag{8}$$

Because the raw EV compensation can become excessive in severely underexposed scenes, a range-control function $\Phi(\cdot)$ constrains it before conversion to a linear gain, yielding the relative reference:

$$\mathbf{T}_{\text{rel}} = 2^{\Phi(e_{\text{rel}})}\,\mathbf{I}_{\text{low}}. \tag{9}$$

We blend the two references and apply Gaussian smoothing to obtain the final exposure reference:

$$\mathbf{T} = \mathcal{G}_\sigma\left((1-\alpha)\,\mathbf{T}_{\text{abs}} + \alpha\,\mathbf{T}_{\text{rel}}\right), \tag{10}$$

where $\alpha = 0.4$ balances the two references, and $\mathcal{G}_\sigma$ suppresses high-frequency fluctuations while preserving low-frequency brightness structure. Since both references are input-dependent, $\mathbf{T}$ provides scene-conditioned supervision rather than a fixed target, encouraging RISE to learn exposure mappings that generalize across illumination distributions.

3.3 Learning Objective

Exposure control loss. Let $\mathbf{G}_\mathbf{T}$ and $\boldsymbol{\ell}_\mathbf{T}$ be the patch log-gain and KIC luminance derived from $\mathbf{T}$. The exposure loss aligns the output with $\mathbf{T}$ at three levels,

$$\mathcal{L}_{\text{exp}} = \|\mathbf{Y}_{\hat{\mathbf{I}}} - \mathbf{Y}_\mathbf{T}\|_1 + \|\hat{\mathbf{G}} - \mathbf{G}_\mathbf{T}\|_1 + \|\hat{\boldsymbol{\ell}} - \boldsymbol{\ell}_\mathbf{T}\|_1. \tag{11}$$

Table 1: Quantitative comparison on LOL. FLOPs is tested on a single 256×256 image. The best and second-best results among unsupervised methods are highlighted in **orange** and yellow, respectively. SL and UL denote supervised and unsupervised learning.

Type	Method	LOL-v1			LOLv2-Real			LOLv2-Synthetic			Params (M)	FLOPs (G)
		PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓		
SL	MIRNet (Zamir et al., 2020)	20.57	0.772	0.254	21.28	0.792	0.354	21.74	0.878	0.138	31.79	760.10
SL	CIDNet (Yan et al., 2025)	23.97	0.849	0.104	23.90	0.866	0.122	25.27	0.935	0.048	1.88	7.57
SL	Multinex (Brateanu et al., 2026)	23.19	0.843	0.129	23.04	0.860	0.178	25.04	0.930	0.068	0.04	2.50
SL	M.-Nano (Brateanu et al., 2026)	19.42	0.742	0.276	19.66	0.784	0.266	21.05	0.882	0.143	0.001	0.04
UL	Zero-DCE (Guo et al., 2020)	14.86	0.559	0.335	18.06	0.574	0.313	17.76	0.816	**0.168**	0.08	4.83
UL	RUAS (Liu et al., 2021)	16.41	0.500	0.270	15.33	0.488	0.310	13.40	0.644	0.364	0.003	0.83
UL	SCI (Ma et al., 2022)	14.78	0.522	0.339	17.30	0.534	0.308	15.43	0.748	0.233	0.001	0.06
UL	PairLIE (Fu et al., 2023)	19.51	0.736	0.248	19.88	0.778	0.234	19.07	0.797	0.230	0.34	20.81
UL	CLIP-LIT (Liang et al., 2023)	12.39	0.493	0.382	15.18	0.529	0.369	16.19	0.775	0.204	0.28	16.96
UL	GDP (Fei et al., 2023)	15.82	0.541	0.339	14.40	0.494	0.362	12.12	0.497	0.356	552.81	1037.26
UL	NeRCo (Yang et al., 2023)	19.74	0.743	0.234	19.66	0.717	0.271	17.59	0.734	0.301	23.05	191.06
UL	CoLIE (Chobola et al., 2024)	13.76	0.481	0.356	15.08	0.501	0.322	14.30	0.654	0.251	0.13	806.25
UL	Li et al. (Li et al., 2025)	19.82	0.751	0.254	20.35	0.795	0.266	17.82	0.802	0.293	0.35	19.00
UL	CLODE (Jung et al., 2025)	19.60	0.718	0.287	17.87	0.681	0.335	17.21	0.783	0.313	0.22	600.96
UL	LASQ (Kong et al., 2026)	17.22	0.739	0.240	20.45	0.778	0.239	18.37	0.791	0.240	24.08	88.67
	RISE	20.21	0.772	0.219	21.33	0.780	**0.215**	18.72	**0.846**	0.169	0.35	1.43
	RISE†	**21.03**	**0.776**	**0.215**	**22.34**	**0.801**	0.243	**19.33**	0.838	0.200	0.35	1.43

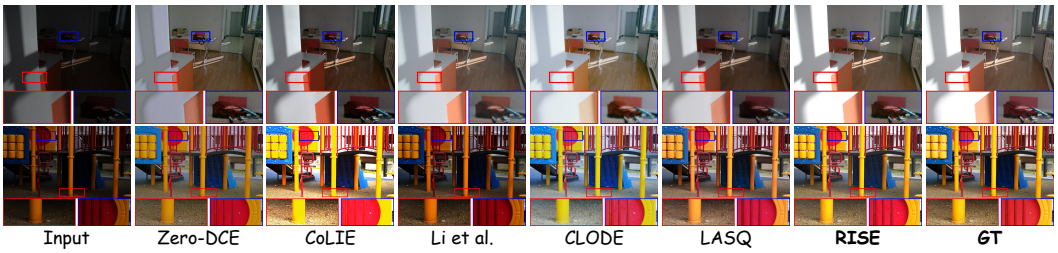


Figure 6: Visual comparisons of the enhanced results by different methods on LOL and UHD-LL.

Regularizers. Scale equivariance preserves relative structure under global intensity changes. Given $(\boldsymbol{\Theta}_s, \eta_s)$ from scaled input $s\mathbf{I}_{\text{low}}$, structural parameters remain stable while η absorbs the change,

$$\mathcal{L}_{\text{sc}} = \|\boldsymbol{\Theta}_s - \boldsymbol{\Theta}\|_1 + \|\eta_s - \eta + \log s\|_1. \tag{12}$$

An idempotence regularizer $\mathcal{L}_{\text{idt}} = \|\mathbf{G}_{\text{e}}\|_1 + \|\eta_{\text{e}}\|_1 + \|\boldsymbol{\ell}_{\text{e}} - \hat{\boldsymbol{\ell}}\|_1$ discourages further enhancement on already enhanced outputs $\hat{\mathbf{I}}$, yielding the re-enhanced result $\mathbf{I}_{\text{e}}$, while an edge-aware smoothness loss $\mathcal{L}_{\text{smooth}}$ suppresses gain variations unsupported by luminance edges. The overall objective is:

$$\mathcal{L} = \lambda_{\text{exp}}\mathcal{L}_{\text{exp}} + \lambda_{\text{sc}}\mathcal{L}_{\text{sc}} + \lambda_{\text{idt}}\mathcal{L}_{\text{idt}} + \lambda_{\text{smooth}}\mathcal{L}_{\text{smooth}}. \tag{13}$$

We empirically set the loss weights $\lambda_{\text{exp}}, \lambda_{\text{sc}}, \lambda_{\text{idt}}, \lambda_{\text{smooth}} = [1.0, 0.1, 0.05, 0.01]$.

4 Experimental Validation

4.1 Datasets and Experimental Details

Datasets. We evaluate RISE on seven paired benchmarks and five unpaired datasets. LOL-v1 contains 485 training and 15 testing pairs (Wei et al., 2018). LOLv2-Real includes 689/100 training/testing pairs, while LOLv2-Synthetic includes 900/100 (Yang et al., 2020; 2021). The Huawei and Nikon subsets of LSRW contain 2,450/30 and 3,150/20 training/testing pairs, respectively (Hai et al., 2023). UHD-LL provides 2,000 training and 150 testing pairs at 4K resolution (Wang et al., 2023b). SICE (Cai et al., 2018) contains 4,413 images from 589 high-resolution multi-exposure sequences; we use all 360 Part 1 sequences for training and all 229 Part 2 sequences for testing. For

Table 2: Quantitative comparison on LSRW, UHD-LL, and SICE.

Type	Method	LSRW-Huawei			LSRW-Nikon			UHD-LL			SICE		
		PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
SL	MIRNet (Zamir et al., 2020)	20.88	0.628	0.451	17.38	0.502	0.539	20.55	0.807	0.435	-	-	-
SL	CIDNet (Yan et al., 2025)	20.71	0.618	0.359	17.14	0.503	0.300	23.80	0.875	0.209	18.60	0.611	0.407
SL	Multinex (Brateanu et al., 2026)	21.26	0.632	0.362	17.59	0.516	0.325	21.12	0.837	0.372	18.39	0.614	0.465
SL	M.-Nano (Brateanu et al., 2026)	19.73	0.604	0.357	16.87	0.506	0.354	19.28	0.807	0.419	16.93	0.572	0.443
UL	Zero-DCE (Guo et al., 2020)	16.40	0.475	**0.368**	15.03	0.418	0.244	17.07	0.639	0.514	13.88	0.542	0.436
UL	RUAS (Liu et al., 2021)	15.71	0.501	0.493	12.14	0.438	0.413	11.69	0.632	0.519	7.80	0.432	0.726
UL	SCI (Ma et al., 2022)	15.70	0.439	0.379	14.55	0.410	0.241	15.47	0.598	0.531	9.68	0.479	0.580
UL	PairLIE (Fu et al., 2023)	18.98	0.562	0.377	15.52	0.435	0.257	17.53	0.663	0.456	12.47	0.512	0.497
UL	CLIP-LIT (Liang et al., 2023)	13.56	0.424	0.405	13.37	0.382	0.276	15.27	0.590	0.551	13.45	0.530	0.416
UL	CoLIE (Chobola et al., 2024)	14.76	0.420	0.401	12.87	0.389	0.277	17.89	0.625	-	14.06	0.543	0.409
UL	CLODE (Jung et al., 2025)	18.44	0.564	0.482	15.54	0.487	0.388	16.17	0.777	0.454	13.49	0.514	0.567
UL	LASQ (Kong et al., 2026)	19.67	**0.599**	0.389	16.12	**0.491**	0.257	20.13	0.749	0.399	14.01	0.540	0.450
	RISE	**20.45**	0.590	0.398	**16.71**	**0.491**	**0.236**	**22.22**	0.771	0.384	14.80	0.554	0.466
	RISE†	20.24	0.580	0.374	**16.71**	**0.491**	**0.236**	22.03	**0.788**	**0.373**	**16.73**	**0.570**	**0.407**

Table 3: No-reference quantitative comparison on unpaired datasets. B. denotes BRISQUE.

Type	Method	DICM			LIME			MEF			NPE			VV		
		IQA↑	IAA↑	B.↓	IQA↑	IAA↑	B.↓	IQA↑	IAA↑	B.↓	IQA↑	IAA↑	B.↓	IQA↑	IAA↑	B.↓
SL	MIRNet (Zamir et al., 2020)	2.69	1.90	42.33	2.40	1.78	35.68	2.17	1.57	42.34	2.21	1.71	31.01	2.58	1.67	31.43
SL	CIDNet (Yan et al., 2025)	2.87	2.05	34.00	2.76	2.12	**17.25**	2.75	2.20	**20.20**	2.88	1.87	28.18	2.84	1.85	**13.63**
UL	RUAS (Liu et al., 2021)	2.04	1.77	47.50	2.11	1.85	29.26	2.69	2.11	32.49	1.66	1.42	49.94	2.27	1.66	40.48
UL	Li et al. (Li et al., 2025)	2.73	1.89	31.94	2.38	1.82	18.48	2.48	1.85	23.68	2.87	1.92	28.74	2.64	1.75	20.78
UL	CLODE (Jung et al., 2025)	2.48	1.71	36.80	2.43	1.79	29.93	2.30	1.72	33.84	2.42	1.58	32.70	2.56	1.60	28.47
UL	LASQ (Kong et al., 2026)	3.43	2.26	30.20	2.98	2.18	17.37	3.04	2.23	25.24	3.43	2.20	23.33	3.37	2.02	17.24
	RISE	**3.73**	**2.47**	**30.00**	**3.42**	**2.54**	21.54	**3.45**	**2.44**	24.95	**3.57**	**2.45**	**21.48**	**3.47**	**2.06**	14.33

no-reference evaluation, we use DICM (Lee et al., 2013), LIME (Guo et al., 2016), MEF (Ma et al., 2015), NPE (Wang et al., 2013), and VV (Vonikakis et al., 2018).

Metrics. For reference-based evaluation, we report PSNR, SSIM (Wang et al., 2004), and LPIPS with AlexNet (Zhang et al., 2018). For no-reference evaluation, we use the IQA and IAA scores from Q-Align (Wu et al., 2024), together with BRISQUE (Mittal et al., 2012).

Implementation Details. RISE is trained for 100 epochs on a single NVIDIA V100 GPU with a batch size of 8. We use Adam (Kingma & Ba, 2015) with an initial learning rate of 2×10^{-4}. The patch size is 75×50, and we set $N = 3$. RISE is trained only on LSRW-Nikon, whereas RISE† is trained separately on each benchmark. We use official LSRW-Nikon checkpoints when available, otherwise, we use officially released weights. The training data associated with these checkpoints are detailed in Appendix C.1.

4.2 Comparisons with State-of-the-Art Methods

Quantitative Evaluation. We compare RISE with representative supervised and unsupervised methods in Tables 1 and 2. RISE outperforms all prior unsupervised methods in PSNR on six of the seven paired benchmarks, while RISE† achieves the highest PSNR among the compared unsupervised methods on all seven paired benchmarks. On the five unpaired benchmarks in Table 3, RISE consistently achieves the highest IQA and IAA scores and competitive BRISQUE scores. Moreover, RISE requires only 0.35M parameters and 1.43 GFLOPs, demonstrating a favorable accuracy-efficiency trade-off.

Qualitative Evaluation. The visual results in Fig. 6 indicate that RISE produces more natural exposure correction with well-preserved color fidelity. Furthermore, RISE achieves adaptive enhancement effects across different scenes, while other methods exhibit evident fixed enhancement styles. Additional benchmark comparisons are provided in Appendix D.1.

Real-world Generalization. To evaluate generalization beyond benchmarks, we apply the LSRW-Nikon checkpoint to low-light photos captured with an iPhone 16 Pro, as shown in Figure 7. Without additional tuning, RISE produces natural exposure with a broader dynamic range, recovering shadow

Table 4: Ablation studies of relative structure, KIC selection, DMER, decoupling, and losses.

(a) Relative Structure				(b) KICs and DMER				(c) Decoupling and Loss			
Removed term	PSNR↑	SSIM↑	LPIPS↓	**Setting**	PSNR↑	SSIM↑	LPIPS↓	**Setting**	PSNR↑	SSIM↑	LPIPS↓
$v_n\rho_n(p)Z_{k_\eta}$	15.74	0.480	0.257	Random KICs	15.84	0.480	0.248	w/o $\mathbf{S}$	15.16	0.476	0.256
$u_n\rho_n(p)\delta_n(p)$	16.09	0.485	0.248	Darkest KICs	14.49	0.482	0.278	w/o η	15.47	0.489	0.264
$w_n\delta_n(p)$	15.46	0.477	0.242	$\mathbf{T}_{\text{abs}}$	14.92	0.471	0.269	w/o $\mathcal{L}_{\text{sc}}$	16.04	0.485	0.249
C_{scene}	15.96	0.482	0.244	$\mathbf{T}_{\text{rel}}$	15.36	0.482	0.251	w/o $\mathcal{L}_{\text{idt}}$	16.40	0.484	0.247
None (Full)	**16.71**	**0.491**	**0.236**	**Brightest+DMER**	**16.71**	**0.491**	**0.236**	**Full**	**16.71**	**0.491**	**0.236**

and highlight details while preserving faithful colors. Additional real-world results are provided in Appendix D.2.



Figure 7: Generalization to real-world scenes captured with an iPhone 16 Pro.

Table 5: Ablation studies on the number and size of KICs.

Number of KICs			**KIC Size**		
N	PSNR↑	SSIM↑	Size	PSNR↑	SSIM↑
1	16.65	0.492	50×25	16.66	0.492
5	16.81	0.492	150×100	16.59	0.481
3	16.71	0.491	75×50	16.71	0.491

Table 6: Sensitivity analysis of the blending weight α in DMER.

α	PSNR↑	SSIM↑	α	PSNR↑	SSIM↑	α	PSNR↑	SSIM↑	α	PSNR↑	SSIM↑	α	PSNR↑	SSIM↑
0.3	16.50	0.490	0.4	**16.71**	**0.491**	0.5	16.48	**0.491**	0.6	16.41	0.490	0.7	16.10	0.488

4.3 Ablation Study

Relative Structure. Table 4(a) reveals the interplay among the relational terms. Ablations show that direct discrepancies, distance-aware relations, and scene context jointly form the coherent scene-wide relative illumination structure.

KICs and DMER. Table 4(b) evaluates KIC selection and DMER, while Table 5 examines the number and size of KICs. Random or darkest KICs degrade performance, especially the latter, supporting bright regions as reliable illumination cues. In contrast, varying the number or size of KICs yields comparable performance, indicating that RISE is largely insensitive to these configurations. Using either meter alone also degrades performance, confirming their complementarity in global calibration and spatial illumination adaptation. Table 6 further shows that DMER is robust to variations in α: $\alpha = 0.4$ achieves the best PSNR, while $\alpha \in [0.3, 0.5]$ yields comparable performance. The degradation at larger α indicates that retaining a slightly larger contribution from the absolute meter better balances global exposure calibration and scene-adaptive spatial adjustment.

Decoupling and Loss. Table 4(c) ablates both our formulation of the enhancement problem and the regularization terms in the objective. Relative structure alone cannot provide sufficient exposure adjustment, whereas absolute exposure alone cannot model spatial variations in illumination. The performance drop caused by removing $\mathcal{L}_{\text{sc}}$ further supports our premise that scene illumination structure should remain stable under global intensity changes, while such changes should be absorbed by the absolute exposure term rather than structural parameters.

5 Conclusion and Limitation

This paper proposes RISE, formulating low-light enhancement as relative illumination structure estimation and absolute exposure control. RISE derives structure from reliable KIC-based spatial relations, while DMER provides scene-conditioned unsupervised guidance. Extensive experiments consistently demonstrate leading unsupervised performance and strong generalization across lighting conditions and unseen real-world scenes. However, when signals fall below the sensor noise floor, missing content cannot be recovered through exposure correction alone, as shown by the failure cases in Appendix E. Incorporating generative restoration remains a promising future direction.

AI USE STATEMENT

We used generative AI tools to improve the readability of the manuscript. The authors reviewed and verified the edited content.

ETHICS STATEMENT

RISE is designed to improve the visibility of low light images. However, its outputs are estimates derived from incomplete observations and should not be treated as faithful recovery of content that was not recorded by the camera, especially in forensic, surveillance, or other safety critical applications. This work does not involve human subject studies.

REPRODUCIBILITY STATEMENT

The main paper provides the model formulation, training objectives, implementation settings, datasets, and evaluation metrics. The appendix further specifies the training data and checkpoints used by the compared methods and provides additional analyses and results. Source code, configurations, and pretrained models will be released to facilitate reproduction of the reported results.

REFERENCES

Alexandru Brateanu, Tingting Mu, Codruta O Ancuti, and Cosmin Ancuti. Multinex: Lightweight low-light image enhancement via multi-prior retinex. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 29887–29896, 2026.

Jianrui Cai, Shuhang Gu, and Lei Zhang. Learning a deep single image contrast enhancer from multi-exposure images. *IEEE transactions on image processing*, 27(4):2049–2062, 2018.

Tomáš Chobola, Yu Liu, Hanyi Zhang, Julia A Schnabel, and Tingying Peng. Fast context-based low-light image enhancement via neural implicit representations. In *European Conference on Computer Vision*, pp. 413–430. Springer, 2024.

Tianle Du, Mingjia Li, Hainuo Wang, and Xiaojie Guo. Anchor then polish for low-light enhancement. *arXiv preprint arXiv:2603.15472*, 2026.

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 *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 9935–9946, 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 *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 22252–22261, 2023.

Mathieu Garon, Kalyan Sunkavalli, Sunil Hadap, Nathan Carr, and Jean-François Lalonde. Fast spatially-varying indoor lighting estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6908–6917, 2019.

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 *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 1780–1789, 2020.

Xiaojie Guo and Qiming Hu. Low-light image enhancement via breaking down the darkness. *International Journal of Computer Vision*, 131(1):48–66, 2023.

Xiaojie Guo, Yu Li, and Haibin Ling. Lime: Low-light image enhancement via illumination map estimation. *IEEE Transactions on image processing*, 26(2):982–993, 2016.

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. *Journal of Visual Communication and Image Representation*, 90:103712, 2023.

Peiyuan He, Hainuo Wang, Hengxing Liu, Mingjia Li, and Xiaojie Guo. Internally referenced low-light enhancement. *arXiv preprint arXiv:2605.28605*, 2026.

Hai Jiang, Ao Luo, Xiaohong Liu, Songchen Han, and Shuaicheng Liu. Lightendiffusion: Unsupervised low-light image enhancement with latent-retinex diffusion models. In *European Conference on Computer Vision*, pp. 161–179. Springer, 2024.

Donggoo Jung, Daehyun Kim, and Tae Hyun Kim. Continuous exposure learning for low-light image enhancement using neural odes. In *The Thirteenth International Conference on Learning Representations*, 2025.

Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *International Conference on Learning Representations (ICLR)*, 2015.

Derong Kong, Zhixiong Yang, Shengxi Li, Shuaifeng Zhi, Li Liu, Zhen Liu, and Jingyuan Xia. Luminance-aware statistical quantization: Unsupervised hierarchical learning for illumination enhancement. *Advances in Neural Information Processing Systems*, 38:116286–116313, 2026.

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

Huaqiu Li, Xiaowan Hu, and Haoqian Wang. Interpretable unsupervised joint denoising and enhancement for real-world low-light scenarios. *arXiv preprint arXiv:2503.14535*, 2025.

Zhexin Liang, Chongyi Li, Shangchen Zhou, Ruicheng Feng, and Chen Change Loy. Iterative prompt learning for unsupervised backlit image enhancement. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 8094–8103, 2023.

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 *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 10561–10570, 2021.

Kede Ma, Kai Zeng, and Zhou Wang. Perceptual quality assessment for multi-exposure image fusion. *IEEE Transactions on Image Processing*, 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 *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 5637–5646, 2022.

Anish Mittal, Rajiv Soundararajan, and Alan C. Bovik. Making a completely blind image quality analyzer. *IEEE Signal Processing Letters*, 20(3):209–212, 2012.

Erik Reinhard, Michael Stark, Peter Shirley, and James Ferwerda. Photographic tone reproduction for digital images. *ACM Transactions on Graphics*, 21(3):267–276, 2002. doi: 10.1145/566654.566575.

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

Haoyuan Wang, Xiaogang Xu, Ke Xu, and Rynson WH Lau. Lighting up nerf via unsupervised decomposition and enhancement. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 12632–12641, 2023a.

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

Tao Wang, Kaihao Zhang, Tianrun Shen, Wenhan Luo, Björn Stenger, and Tong Lu. Ultra-high-definition low-light image enhancement: A benchmark and transformer-based method. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pp. 2654–2662, 2023b.

Wenjing Wang, Huan Yang, Jianlong Fu, and Jiaying Liu. Zero-reference low-light enhancement via physical quadruple priors. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 26057–26066, 2024.

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

Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu. Deep retinex decomposition for low-light enhancement. In *British Machine Vision Conference (BMVC)*, 2018.

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

Chao Xie, Linfeng Fei, Huanjie Tao, Yaocong Hu, Wei Zhou, Jiun Tian Hoe, Weipeng Hu, and Yap-Peng Tan. Residual quotient learning for zero-reference low-light image enhancement. *IEEE Transactions on Image Processing*, 34:365–378, 2024.

Xiaogang Xu, Ruixing Wang, Chi-Wing Fu, and Jiaya Jia. Snr-aware low-light image enhancement. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 17714–17724, 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 *Proceedings of the computer vision and pattern recognition conference*, pp. 5678–5687, 2025.

Shuzhou Yang, Moxuan Ding, Yanmin Wu, Zihan Li, and Jian Zhang. Implicit neural representation for cooperative low-light image enhancement. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 12918–12927, 2023.

Wenhan Yang, Shiqi Wang, Yuming Fang, Yue Wang, and Jiaying Liu. From fidelity to perceptual quality: A semi-supervised approach for low-light image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3063–3072, 2020.

Wenhan Yang, Wenjing Wang, Haofeng Huang, Shiqi Wang, and Jiaying Liu. Sparse gradient regularized deep retinex network for robust low-light image enhancement. *IEEE Transactions on Image Processing*, 30:2072–2086, 2021.

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 *European conference on computer vision*, pp. 492–511. Springer, 2020.

Fangneng Zhan, Changgong Zhang, Yingchen Yu, Yuan Chang, Shijian Lu, Feiying Ma, and Xuansong Xie. Emlight: Lighting estimation via spherical distribution approximation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pp. 3287–3295, 2021.

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 (CVPR)*, pp. 586–595, 2018.

Yonghua Zhang, Jiawan Zhang, and Xiaojie Guo. Kindling the darkness: A practical low-light image enhancer. In *Proceedings of the 27th ACM international conference on multimedia*, pp. 1632–1640, 2019.

Zhao Zhang, Suiyi Zhao, Xiaojie Jin, Mingliang Xu, Yi Yang, Shuicheng Yan, and Meng Wang. Noise self-regression: A new learning paradigm to enhance low-light images without task-related data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(2):1073–1088, 2024.

A Additional Method Analysis

A.1 Scope of the Light Propagation Approximation

The central idea of RISE is motivated by light propagation in real scenes. As illustrated in Fig. 2(b), physical light propagation is inherently conditioned on three dimensional scene geometry and is affected by depth, occlusion, surface reflectance, light source configuration, and other scene factors. Therefore, image plane distance is not equivalent to physical propagation distance, and RISE is not intended to reconstruct the exact light transport process of a three dimensional scene. Recovering all of these latent factors accurately from a single low light observation is itself severely ill posed.

One possible solution is to estimate scene depth using a strong pretrained monocular depth model. However, reliable depth estimation is particularly difficult when the input contains weak signals, severe noise, and limited texture. Depth alone also does not fully determine occlusion, reflection, material properties, or the spatial configuration of light sources. Moreover, requiring users to provide an additional depth map would limit the practical single image setting considered in this work.

We therefore use image plane distance as a tractable spatial proximity prior and relative illumination as observation dependent evidence. Together with the learned scene context, they provide a soft approximation to the projected effect of light propagation rather than an exact physical model. The image plane term constrains spatial variation, while the relative illumination and learned features compensate for propagation effects that cannot be described by distance alone. If reliable geometry, visibility, reflection, or light source information can be recovered from a single low light image in the future, these cues can naturally replace or augment the current spatial term, allowing RISE to model scene illumination more accurately.

A.2 Beyond Brightness Ordering

We clarify why RISE models continuous, spatially conditioned illumination relations instead of directly adopting the regional brightness ordering observed in a low-light input. The analysis first examines whether the input ordering reliably describes the desired normal-light appearance, and then highlights the additional information captured by our relative illumination formulation.

The relative brightness ordering in a low-light observation is not necessarily consistent with that in its normally exposed counterpart. Regional luminance is not a direct measurement of illumination, but is jointly determined by illumination, surface reflectance, sensor noise, quantization, and the camera response. In particular, weak measurements in severely underexposed regions can compress or even reverse pairwise brightness relations. Directly treating the input ordering as the target illumination structure may therefore impose incorrect ordinal constraints on the enhanced result. We analyze this phenomenon using paired low-light and ground-truth images only for evaluation. The ground truth is never used to train RISE.



Figure 8: Inconsistency of relative brightness ordering. Representative examples where the regional brightness ordering in low-light inputs differs from that in the corresponding ground-truth images, indicating that input brightness order is not a reliable target illumination constraint.

These inconsistencies indicate that input brightness relations should serve as evidence for illumination estimation rather than constraints to be directly preserved. A rank-based representation retains only whether one region is brighter than another. In contrast, the discrepancy used by RISE satisfies $\delta_n(p) = Z_{k_n} - Z_p = \log(P_{k_n}/P_p)$, preserving the magnitude of the relative brightness ratio while remaining invariant to global intensity scaling. RISE further combines this continuous relation with

spatial attenuation $\rho_n(p)$, multiple KICs, predicted KIC luminance, and scene context to construct a scene-conditioned gain field. Importantly, KICs are selected as high-SNR anchors rather than fixed ordering targets. Their relations guide the estimated correction without enforcing the input brightness ordering on the enhanced output. Consequently, RISE can distinguish illumination patterns that share the same ordinal ranking but exhibit different attenuation strengths, while correcting unreliable relations inherited from low-SNR observations.

A.3 KIC Selection in Representative Scenes

For each input, we divide the luminance image into non-overlapping patches and compute the mean luminance of each patch. KICs are selected greedily: we choose the brightest remaining patch, suppress candidates whose centers lie within a radius of two patch-grid units from the selected patch, and repeat until N anchors have been selected. This non-maximum suppression encourages spatially dispersed cues instead of redundant anchors from the same local bright region. Patch brightness is used as an observable proxy for signal reliability, not as a guarantee of high SNR.

Figure 9 shows four representative cases. The colored boxes mark the selected KICs, and the labels report their patch indices and mean luminance. The examples illustrate how the greedy spatial suppression retains multiple bright cues from different image regions for relational modeling.

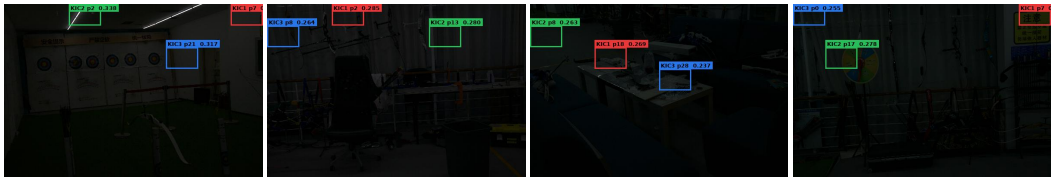


Figure 9: KIC selection in four representative low-light scenes. Colored boxes denote the selected anchors (K_1 in red, K_2 in green, and K_3 in blue); labels indicate the patch index and mean luminance. Spatial suppression discourages selecting neighboring patches as separate cues.

A.4 Stability to Challenging KICs

Selecting bright regions as KICs raises a potential concern that some selected patches may contain saturation, specular reflections, or isolated bright noise. The role of a KIC, however, is to identify a region reached by a strong light path rather than to provide an exact radiometric measurement. Although the magnitudes of saturated regions and specular highlights may be clipped or view dependent, their high observed brightness is still caused by sufficiently strong incident or reflected light reaching these regions and the camera. Their locations therefore remain informative anchors for modeling relative light propagation. RISE uses their relational features instead of directly copying their observed intensities, so exact recovery of the clipped magnitude is not required. For isolated bright noise, patch averaging reduces its influence, while multiple spatially dispersed KICs prevent a single noisy anchor from dominating the estimated field.

Figure 10 provides a direct qualitative stress case. Here, K_1 overlaps the saturated window region, while K_2 and K_3 are located near strong reflected highlights on the floor. Despite these clipped or reflection dependent measurements, the selected regions still mark strong light paths and provide effective propagation anchors. RISE consequently achieves a satisfactory enhancement result. Together with the robustness to KIC number and size reported in Table 5, this example supports the stability of the proposed multi KIC relational formulation when individual bright cues are imperfect.



Figure 10: Stability under challenging KIC selection. K_1 overlaps a saturated window region, while K_2 and K_3 lie near strong reflected highlights. Although their measured magnitudes are imperfect, these regions still indicate strong light paths and serve as effective propagation anchors, allowing RISE to produce balanced exposure and faithful color.

A.5 From Log-Gain to Image Space

We further clarify the transition from Eq. (4) to Eq. (5) in the main paper. RISE does not predict an additive offset to the image intensity. Instead, it predicts the logarithm of a multiplicative exposure gain. At the patch level, Eq. (4) decomposes this log-gain into relative illumination structure and absolute exposure:

$$\hat{G}(p) = S(p) + \eta.$$

Before it is applied to the input image, the patch-level field is upsampled to the pixel grid and mapped back to the linear-gain domain through the exponential function. For a pixel x, the resulting gain is

$$g(x) = \exp\Big(\mathrm{Up}(\hat{\mathbf{G}})(x)\Big) = \exp(\mathrm{Up}(\mathbf{S})(x) + \eta) = \exp(\mathrm{Up}(\mathbf{S})(x))\exp(\eta).$$

The enhanced image in Eq. (5) is therefore obtained by multiplying this gain with the low-light input:

$$\hat{\mathbf{I}}(x) = \mathbf{I}_{\mathrm{low}}(x) \odot g(x).$$

This derivation explains why the two components are additive in log-gain space but multiplicative in the linear-gain domain: the relative term determines the spatially varying correction, whereas η scales the exposure of the entire image. The parameterization also gives a direct interpretation of the predicted values. Specifically, $\hat{G} = 0$ yields $g = 1$ and leaves the intensity unchanged. Likewise, $\hat{G} = \log 2$ yields $g = 2$ and doubles it, while $\hat{G} = \log 0.5$ yields $g = 0.5$ and halves it. More generally, positive log-gain values enhance exposure, while negative values suppress it.

A.6 Structure–Exposure Separation

Although C_{scene} and the exposure offset η are both spatially uniform additive terms in log-gain space, they serve different roles in RISE. We distinguish their responsibilities through both architectural design and optimization constraints. Architecturally, C_{scene} calibrates the KIC-based relational field using exposure-normalized scene statistics, while η controls absolute exposure using features that retain the input brightness scale. During optimization, the scale-equivariance objective encourages the components forming C_{scene} to remain insensitive to global exposure changes and assigns those changes to η.

Scene-context calibration. The relational field $\mathcal{R}(p)$ is constructed from observed luminance discrepancies and image-plane distances, which act as tractable surrogates but cannot exactly recover the underlying three-dimensional light propagation. Consequently, similar local KIC-to-patch relations may occur under different scene-level illumination patterns. We introduce $\mathbf{c}_{\mathrm{scene}}$ to summarize the global statistics of the normalized relative map $\mathbf{Z}$ and provide the scene context needed to distinguish such locally similar relational configurations. In Eq. 4, its projection forms $C_{\mathrm{scene}} = \boldsymbol{\beta}^\top \mathbf{c}_{\mathrm{scene}} + b$. Here, $\boldsymbol{\beta}^\top \mathbf{c}_{\mathrm{scene}}$ supplies the scene-dependent calibration, while b is the intercept of the relational mapping. Because the relational basis functions $\delta_n(p)$, $\rho_n(p)\delta_n(p)$, and $\rho_n(p)Z_{k_n}$ are not mean-centered, b provides a calibrated baseline without forcing their learned combination through the origin. Thus, C_{scene} complements the approximate local relations with global relative-illumination context rather than serving as an estimate of the absolute exposure of the input.

Architectural conditioning. The exposure branch predicts η from features of the entire input image, retaining the absolute brightness information needed to estimate global exposure compensation. The relative branch receives only the normalized relative luminance map $\mathbf{Z}$ and quantities derived from it, including the KIC-to-patch relations and scene statistics. Its input therefore describes relative illumination relationships without retaining the global exposure scale. For multiplicative exposure scaling $\mathbf{I}_s = s\mathbf{I}$, $s > 0$, before clipping and omitting the numerical stabilizer in Eq. 1,

$$Z_s(p) = \log \frac{sP_p}{\mathrm{mean}(s\mathbf{P})} = \log \frac{P_p}{\mathrm{mean}(\mathbf{P})} = Z(p).$$

Inputs differing only in global exposure thus yield the same relative representation and, consequently, the same scene statistics $\mathbf{c}_{\mathrm{scene}}$. The relative branch therefore has no access to the global scale factor needed to use C_{scene} as an absolute exposure predictor. In contrast, the exposure branch retains full-image brightness cues, allowing η to respond directly to the exposure change. This difference in available information establishes the architectural division between scene-context calibration and absolute exposure compensation.

Optimization under exposure changes. The scale-equivariance loss in Eq. 12 explicitly includes b and $\boldsymbol{\beta}$ in $\boldsymbol{\Theta} = (b, \boldsymbol{\beta}, \{w_n, u_n, v_n\}_{n=1}^{N})$ and encourages

$$\boldsymbol{\Theta}(s\mathbf{I}) \approx \boldsymbol{\Theta}(\mathbf{I}), \qquad \eta(s\mathbf{I}) \approx \eta(\mathbf{I}) - \log s.$$

Because $\mathbf{c}_{\mathrm{scene}}$ is computed from the scale-invariant $\mathbf{Z}$, these constraints imply $C_{\mathrm{scene}}(s\mathbf{I}) \approx C_{\mathrm{scene}}(\mathbf{I})$, while the exposure change is assigned to η. The objective therefore encourages the relative branch to remain insensitive to global exposure changes and trains η to account for them. In the ideal case, $S_s(p) = S(p)$ and $\eta_s = \eta - \log s$, which gives $\hat{G}_s(p) = \hat{G}(p) - \log s$ and cancels the multiplicative input scaling after enhancement. The architectural conditioning and scale-equivariance objective therefore promote different transformation behaviors for C_{scene} and η.

Table 7: Parameter values across the seven exposures of the same scene in Fig. 4. $\bar{Y}$ denotes mean input luminance. We report b and representative coordinates of the structural parameter groups. Ranges are computed as maximum minus minimum from the displayed values.

Exposure	$\bar{Y}$	η	b	w_1	u_1	v_1	β_1
1	0.080	0.789	0.096	−0.305	0.238	−0.330	0.323
2	0.128	0.628	0.091	−0.297	0.243	−0.336	0.327
3	0.199	0.398	0.095	−0.290	0.247	−0.339	0.327
4	0.300	0.151	0.094	−0.273	0.263	−0.354	0.325
5	0.433	0.043	0.088	−0.286	0.257	−0.343	0.330
6	0.572	−0.224	0.085	−0.301	0.246	−0.325	0.331
7	0.705	−0.368	0.093	−0.314	0.238	−0.306	0.327
Range	—	1.157	0.011	0.041	0.025	0.048	0.008

Empirical parameter behavior. Table 7 complements Fig. 4. Because all seven inputs depict the same scene, their normalized relative maps and scene-level relative statistics remain similar despite their different absolute exposure levels. As mean input luminance increases from 0.080 to 0.705, η decreases monotonically from 0.789 to -0.368, spanning 1.157. In comparison, the parameters forming the scene-context projection are substantially less sensitive to exposure changes. Their ranges are 0.011 for b and only 0.008 for the representative coefficient β_1. The other displayed relational parameters also have small ranges of 0.041, 0.025, and 0.048 for w_1, u_1, and v_1, respectively. The strong, monotonic response of η and the relative stability of b, $\boldsymbol{\beta}$, and the relational parameters support the learned division of responsibilities. Input exposure variation is primarily absorbed by η, while the components that form C_{scene} remain comparatively insensitive to the exposure scale and provide scene-level context for the relational field.

Algebraic equivalence versus functional redundancy. For a fixed input, the two spatially uniform terms can be combined algebraically by defining $\eta' = \eta + C_{\mathrm{scene}}$, which leaves the total log-gain unchanged. This sample-wise equivalence does not imply that the two prediction pathways are functionally redundant across images. The scene-context term is a mapping from normalized structural evidence, $C_{\mathrm{scene}} = q(\mathbf{Z})$, while the exposure term is predicted as $\eta = h(\mathbf{I})$ from globally pooled appearance features and mean luminance. Removing C_{scene} therefore requires the exposure head to

learn $h'(\mathbf{I}) = h(\mathbf{I}) + q(\mathbf{Z}(\mathbf{I}))$. However, this head does not receive the normalized scene statistics explicitly and must recover both the scene-dependent relational baseline and the absolute exposure response from a compressed representation under fixed model capacity. The performance degradation in Table 4(a) after removing C_{scene} indicates that this joint mapping is not fully recovered under the current architecture and training objective. Thus, the ablation supports the usefulness of a dedicated scene-context pathway rather than contradicting the additive form of Eq. 4.

B Details of DMER

B.1 Definition of the Range-Control Function

The role of $\Phi(\cdot)$ is to convert the raw dual-meter exposure score into a stable log-domain gain. Directly using e_{rel} as the gain may produce excessive enhancement when the luminance statistics are unreliable or extremely small. Therefore, $\Phi(\cdot)$ first normalizes the input toward a reference luminance and then bounds the scene-dependent adjustment:

$$\Phi(e_{\text{rel}}) = \log_2 \frac{\tau_{\text{rel}}}{\bar{Y}_\epsilon} - \tanh(e_{\text{rel}}), \qquad \bar{Y}_\epsilon = \max(\text{mean}(\mathbf{Y}), \epsilon)\,, \tag{14}$$

where $\tau_{\text{rel}} = \sqrt{m}$ is a fixed reference luminance. Because exposure is parameterized in the log domain, we choose τ_{rel} as the geometric midpoint between the gray reference m and the normalized intensity upper bound of one. With $m = 0.18$, this gives $\tau_{\text{rel}} \approx 0.424$. Combined with the one-stop range imposed by $\tanh(\cdot)$, the resulting target luminance remains approximately within $(0.212, 0.849)$, providing a broad adjustment range while staying away from the intensity boundaries.

The first term gives the basic exposure gain required to map the input mean luminance $\bar{Y}_\epsilon$ toward τ_{rel}. The second term adaptively adjusts this target according to the dual-meter score. In particular, the negative sign assigns a more conservative target to low-key scenes instead of forcing all inputs toward the same brightness. The hyperbolic tangent smoothly limits this adaptive adjustment to one EV stop, since $\tanh(e_{\text{rel}}) \in (-1, 1)$.

Substituting $\Phi(\cdot)$ into Eq. 9, the corresponding target mean luminance is

$$\mu_{\text{rel}} = \tau_{\text{rel}} 2^{-\tanh(e_{\text{rel}})} \in \left(\frac{\tau_{\text{rel}}}{2}, 2\tau_{\text{rel}}\right). \tag{15}$$

Thus, $\Phi(\cdot)$ prevents unbounded scene-dependent exposure changes while retaining different target brightness levels for different scenes. Its dependence on $\bar{Y}_\epsilon$ is omitted from the notation in the main paper for simplicity.

Example with a dark input. Consider an input with $\bar{Y}_\epsilon = 0.05$ and $\text{mean}(\mathbf{L}) = 0.08$. With $m = 0.18$, Eq. 8 gives

$$e_{\text{rel}} = \log_2 \frac{0.18}{0.05} + \log_2 \frac{0.18}{0.08} \approx 3.02. \tag{16}$$

Equation 15 then gives $\mu_{\text{rel}} \approx 0.213$, which corresponds to an exposure gain of $2^{\Phi(e_{\text{rel}})} = 0.213/0.05 \approx 4.26$. Thus, although the bounded adjustment assigns a conservative target to this low-key input, the target remains substantially above its observed mean luminance and the resulting operation brightens the image.

B.2 Spatial Sensitivity of the Relative Meter

Although the relative exposure term in Eq. 8 uses the global mean of the light field $\mathbf{L}$, the spatial information encoded by $\mathbf{L}$ is not discarded. Each value $L(x)$ is obtained by locally aggregating luminance and confidence through normalized Gaussian propagation. Because this normalization is performed independently at every spatial location before global averaging, the two operations are not interchangeable:

$$\text{mean}\left(\frac{\mathcal{G}_\sigma(\mathbf{C} \odot \mathbf{Y})}{\mathcal{G}_\sigma(\mathbf{C})}\right) \neq \frac{\sum_x C(x)Y(x)}{\sum_x C(x)} \quad \text{in general.} \tag{17}$$

Consequently, rearranging image regions changes the local neighborhoods within the numerator and denominator, producing different local ratios and hence a different $\text{mean}(\mathbf{L})$, even when the image

content and its global luminance distribution remain unchanged. The relative meter thus aggregates spatial illumination into an image level exposure decision rather than ignoring spatial arrangement.

We verify this property through a spatial sensitivity test. For each input, we keep the content of every image patch unchanged but randomly permute the patch locations. This operation preserves the image content distribution and $\mathrm{mean}(\mathbf{Y})$, while altering only the spatial arrangement. Let $M_{\mathrm{rel}}(\mathbf{I}) = \mathrm{mean}(\mathbf{L})$ denote the illumination statistic used by the relative meter. We report its relative change after permutation as

$$\Delta_{\mathrm{meter}} = \frac{|M_{\mathrm{rel}}(\mathbf{I}^{\pi}) - M_{\mathrm{rel}}(\mathbf{I})|}{M_{\mathrm{rel}}(\mathbf{I}) + \epsilon} \times 100\%, \tag{18}$$

where $\mathbf{I}^{\pi}$ is the patch permuted input. As shown in Table 8, the meter changes substantially on all five datasets, with average changes from 11.14% to 41.77%. These changes directly affect e_{rel} because $\mathrm{mean}(\mathbf{Y})$ is invariant to the permutation. The results therefore confirm that the relative meter remains sensitive to spatial illumination organization despite its final global aggregation.

Table 8: Spatial sensitivity of the DMER relative illumination meter. The values report the dataset average relative change after permuting image patch locations while preserving patch content.

Dataset	Mean change
LOL-v1	11.14%
LOLv2-Real	18.04%
LOLv2-Syn	41.77%
LSRW-Nikon	13.33%
LSRW-Huawei	24.00%

B.3 Direct Evaluation of the DMER Reference

To disentangle the contribution of the manually designed exposure reference from that of the learned RISE architecture, we directly use the DMER reference $\mathbf{T}$ as the enhanced output and evaluate its PSNR. As shown in Table 9, the DMER reference alone surpasses the best prior unsupervised method on LOL-v1, LOLv2-Real, and LSRW-Nikon, showing that DMER provides an effective exposure target adapted to each scene. RISE† further improves upon the reference on all five benchmarks. These results separate the complementary contributions of the two designs: DMER supplies strong supervision adapted to each image, while the learned relative illumination structure and exposure control of RISE provide additional restoration gains. Together, they confirm that both designs make substantial and distinct contributions to the final performance.

Table 9: PSNR comparison of the directly evaluated DMER reference and RISE†. Best prior UL denotes the highest PSNR among prior unsupervised methods in the main comparison tables.

Dataset	Best prior UL	DMER reference	RISE†
LOL-v1	19.820	20.299	**21.030**
LOLv2-Real	20.450	21.226	**22.340**
LOLv2-Syn	19.070	18.012	**19.330**
LSRW-Huawei	19.670	19.577	**20.240**
LSRW-Nikon	16.120	16.297	**16.710**

C Additional Experimental Details

C.1 Training Data of Compared Methods

We use the official checkpoints released by the respective authors whenever available. Table 10 summarizes the training data associated with the checkpoints used in our comparisons.

Table 10: Training data used by the compared methods. SF denotes dataset-specific training, where each model is trained on the training split of the corresponding benchmark and evaluated on its test split. Own-Data denotes data collected by the method's authors, and None denotes a training-free method.

Method	Training data
MIRNet (Zamir et al., 2020)	SF
CIDNet (Yan et al., 2025)	SF
Multinex (Brateanu et al., 2026)	SF
M.-Nano (Brateanu et al., 2026)	SF
Zero-DCE (Guo et al., 2020)	SICE
RUAS (Liu et al., 2021)	SICE
SCI (Ma et al., 2022)	MIT, LSRW
PairLIE (Fu et al., 2023)	LOL-v1, SICE
CLIP-LIT (Liang et al., 2023)	Own-Data
GDP (Fei et al., 2023)	ImageNet
NeRCo (Yang et al., 2023)	LSRW-Full
CoLIE (Chobola et al., 2024)	None
Li et al. (Li et al., 2025)	SF
CLODE (Jung et al., 2025)	LOL-v1
LASQ (Kong et al., 2026)	LSRW-Nikon
RISE	LSRW-Nikon
RISE†	SF

C.2 Quantitative Luminance Distribution Comparison

Beyond standard image restoration metrics, we further examine the luminance distributions produced by different unsupervised LLIE methods. We compute the Wasserstein Distance between the luminance distribution of each enhanced result and that of the corresponding ground truth. A smaller distance indicates that the enhanced image better matches the target exposure distribution. As shown in Table 11, RISE obtains the lowest or near-lowest distance on most benchmarks, suggesting that the proposed illumination modeling yields more faithful luminance distributions across diverse datasets.

Table 11: Quantitative comparison of luminance distributions. Values represent the Wasserstein Distance (↓) to the corresponding ground truth. The best and second-best results are highlighted in **orange** and yellow, respectively, with the best results also shown in boldface.

Method	LOL-v1	LOLv2-Real	LOLv2-Syn	LSRW-Huawei	LSRW-Nikon	UHD-LL
Zero-DCE (Guo et al., 2020)	0.1684	0.1070	0.0907	0.1223	0.0923	0.0744
RUAS (Liu et al., 2021)	0.0792	0.0890	0.1022	0.0928	0.1300	0.2075
SCI (Ma et al., 2022)	0.1816	0.1179	0.0618	0.1402	0.0572	0.1071
CLIP-LIT (Liang et al., 2023)	0.2374	0.1587	0.1325	0.1797	0.1589	0.1520
CoLIE (Chobola et al., 2024)	0.2087	0.1643	0.1961	0.1623	0.1922	0.1414
CLODE (Jung et al., 2025)	**0.0266**	0.0503	0.0636	0.0475	0.0829	0.1186
RISE	0.0769	**0.0213**	0.0348	**0.0305**	**0.0281**	**0.0271**
RISE†	0.0609	0.0422	**0.0345**	0.0351	**0.0281**	0.0525

C.3 Effect of Dataset Specific Training

Dataset specific training does not necessarily improve every metric monotonically because RISE is trained without ground truth, and neither PSNR nor Wasserstein distance is directly optimized. The single RISE model benefits from the relatively large and diverse LSRW-Nikon training set, whose illumination distribution overlaps with multiple test benchmarks, as shown in Fig. 1(a). This broad

exposure coverage also acts as an implicit regularizer and can provide better transfer than training on a narrower target distribution.

Consequently, RISE† is slightly lower than RISE by only 0.21 dB on LSRW-Huawei and 0.19 dB on UHD-LL. These differences reflect metric tradeoffs rather than consistent degradation: RISE† improves LPIPS on both datasets and also improves SSIM on UHD-LL. Wasserstein distance measures only the marginal luminance distribution and does not account for spatial structure or perceptual fidelity. Its increase on some benchmarks can therefore coexist with improvements in perceptual metrics. Moreover, LASQ, which is also trained on LSRW-Nikon, remains consistently below RISE in PSNR, indicating that the observed gains cannot be attributed to the training data alone and further supporting the effectiveness of our method.

D Additional Qualitative Results

D.1 Benchmark Results

Additional qualitative results on LOL, LSRW, UHD-LL, and SICE complement the main comparisons. These examples cover diverse illumination distributions and emphasize both the recovery of poorly illuminated content and the preservation of details in bright regions.



Figure 11: Additional visual comparisons on the LOL benchmarks.

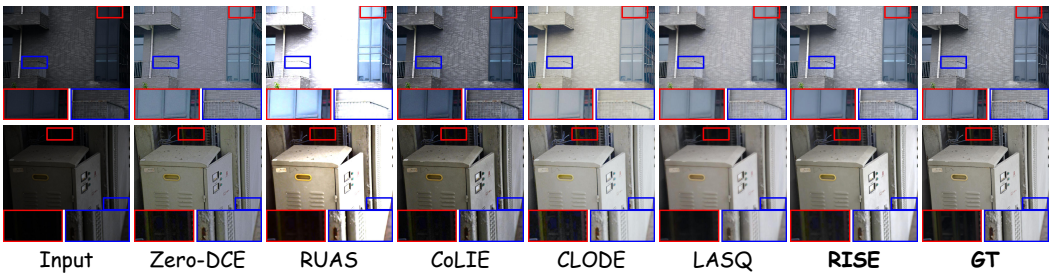


Figure 12: Additional visual comparisons on LSRW.

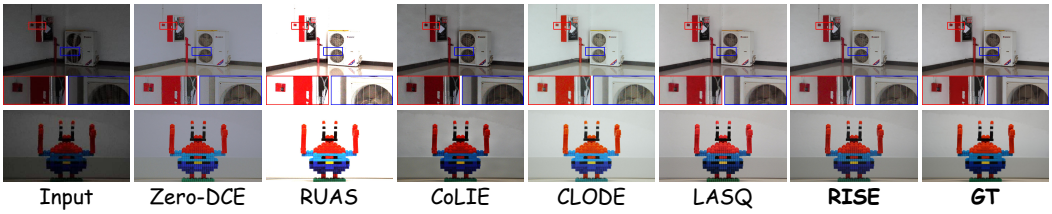


Figure 13: Additional visual comparisons on UHD-LL.



Figure 14: Additional visual comparisons on SICE.



Figure 15: Visual comparisons on No-Reference Datasets.

D.2 Real-world Generalization Results

For out-of-benchmark generalization, we apply the same LSRW-Nikon checkpoint to low-light photographs captured with an iPhone 16 Pro, without additional training or scene-specific tuning.



Figure 16: Additional comparisons on real-world low-light scenes. RISE provides balanced exposure in both outdoor and indoor scenes while avoiding severe highlight saturation. All results are produced by the LSRW-Nikon checkpoint without additional tuning.

E Failure Cases

The high SNR assumption of KICs has a clear validity boundary. The relative normalization in Eq. 1 and the subsequent brightest patch selection identify regions that are brighter than the rest of the same input, but they cannot create reliable scene signals when no effective illumination is recorded. In an extremely dark scene, the observations may fall below the sensor noise floor, so even the selected brightest patches can be dominated by sensor noise rather than illumination. The same signal loss illustrated in Fig. 17 therefore causes the KIC assumption to fail: without a valid high SNR anchor, relational propagation is guided by unreliable noise cues and cannot recover scene structures that were never captured. This failure of the KIC assumption and the unrecoverable content therefore share the same cause: input signal loss below the sensor noise floor. Addressing this regime requires integrating RISE with generative models, which is an important direction for future work.

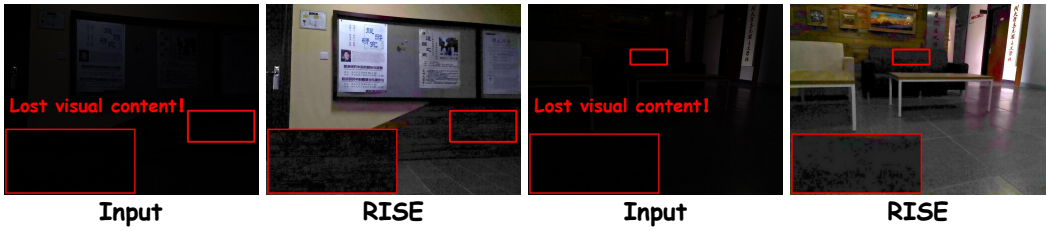


Figure 17: Failure cases under unrecoverable signal loss. When no valid high SNR region remains, the brightest patches may be dominated by sensor noise and cannot serve as reliable KICs. RISE improves overall visibility but cannot recover structures absent from the recorded input and may retain residual noise in these regions. Red boxes and enlarged crops indicate the affected areas.