LEARNING STRUCTURAL ILLUMINATION FOR UNSUPERVISED LOW-LIGHT ENHANCEMENT

Tianle Du[1], **Peiyuan He**[1], **Hainuo Wang**[1], **Tianxiu Yu**[2], **Xiaojie Guo**[1,*]
[1]Tianjin University
[2]Dunhuang Academy
{dutianle, peiyuan_he, hainuo}@tju.edu.cn, xj.max.guo@gmail.com

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 among unsupervised LLIE methods while producing visually natural results.

Project page: https://dutianle.ltd/

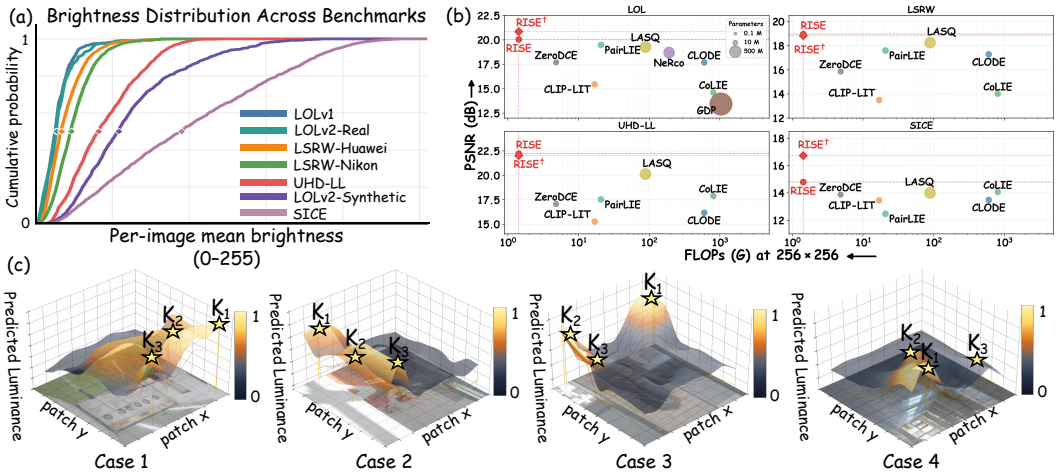

Figure 1: (a) Empirical cumulative distribution functions (CDFs) of per-image mean brightness across seven benchmarks. (b) RISE consistently delivers a superior PSNR–FLOPs trade-off across all benchmarks, achieving higher PSNR than prior SOTA methods with the fewest FLOPs. (c) Joint visualization of relative illumination structure and absolute exposure: surface shape encodes relative illumination, elevation encodes absolute exposure, and stars mark selected KICs.

*Corresponding author.

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.

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.

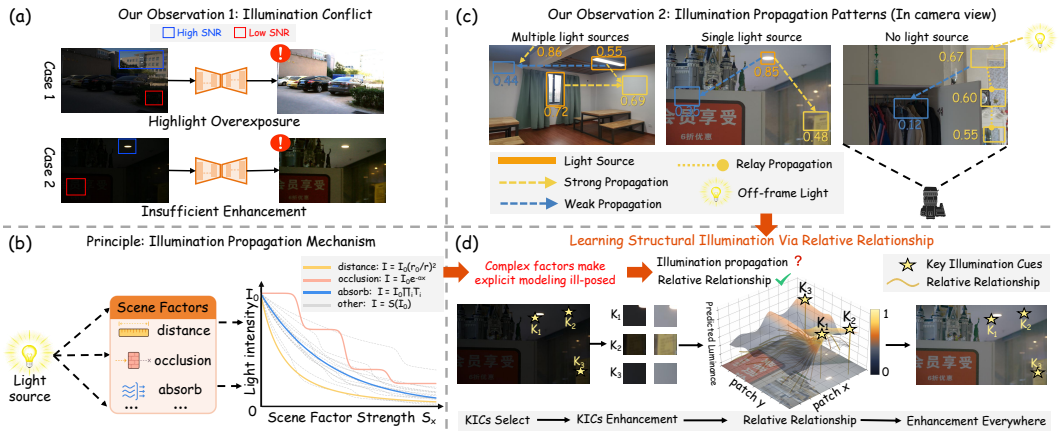

Figure 2: Motivation of RISE. (a) Existing methods show unstable enhancement. (b) Light intensity from a source attenuates with scene factors. (c) Relative illumination propagation patterns in normal-light images (numbers denote regional brightness). (d) Our insight: RISE learns illumination structure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

- 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.

Comprehensive experiments on multiple benchmarks show that RISE achieves state-of-the-art performance, 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 radi-

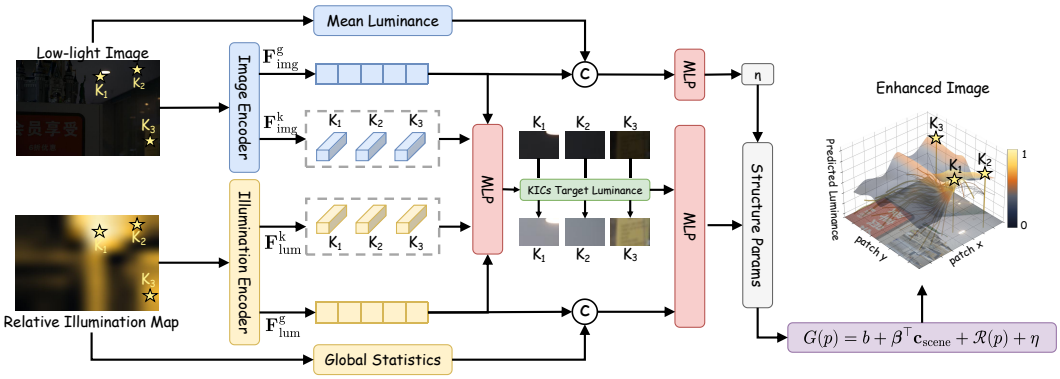

Figure 3: Overview of RISE. RISE jointly encodes image appearance and relative illumination, using spatially dispersed KICs as reliable anchors to estimate illumination structure while independently predicting absolute exposure for scene-adaptive enhancement, enabling reliable structural modeling and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

ance 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.

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

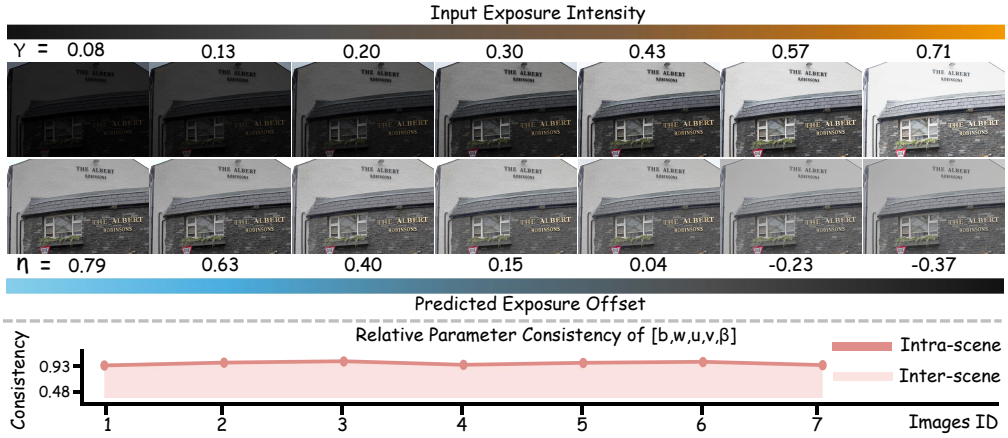

Figure 4: Validation of illumination structure and exposure decoupling on SICE. Under different exposures of the same scene, RISE maintains stable relative parameters while adjusting the absolute exposure, which demonstrates its ability to both enhance underexposed inputs and suppress overexposed scenes.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

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 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 $\boldsymbol{\beta}^\top \mathbf{c}_{\text{scene}} + b$ 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:

$$\hat{G}(p) = \underbrace{b + \boldsymbol{\beta}^\top \mathbf{c}_{\text{scene}} + \mathcal{R}(p)}_{\text{relative structure } S(p)} + \underbrace{\eta}_{\text{absolute exposure}}. \tag{4}$$

The relative parameters $\Theta = \left(b, \boldsymbol{\beta}, \{w_n, u_n, v_n\}_{n=1}^{N}\right)$ and the absolute exposure η are predicted by separate MLPs. This decoupling keeps structural parameters Θ stable across exposures of the same scene, leaving η to absorb exposure changes, as shown in Fig. 4. The enhanced result is given by

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

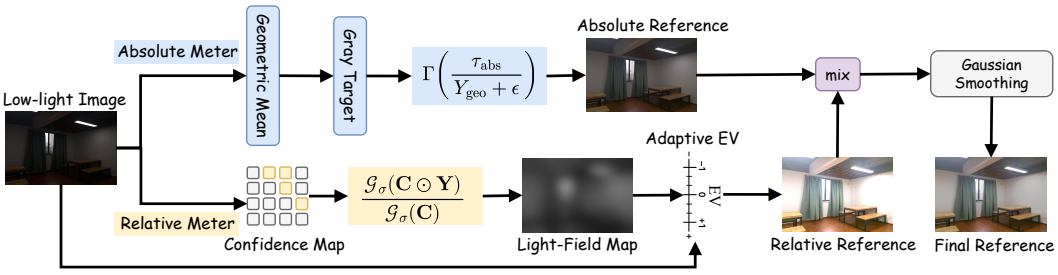

Figure 5: Overview of DMER. The absolute meter estimates global exposure, while the relative meter models spatial illumination through light-field propagation.

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 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:

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

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.

Table 1: Quantitative comparison on LOL. FLOPs is tested on a single 256×256 image. The best and second-best results are highlighted in **orange** and yellow, respectively, with the best results also shown in boldface. 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

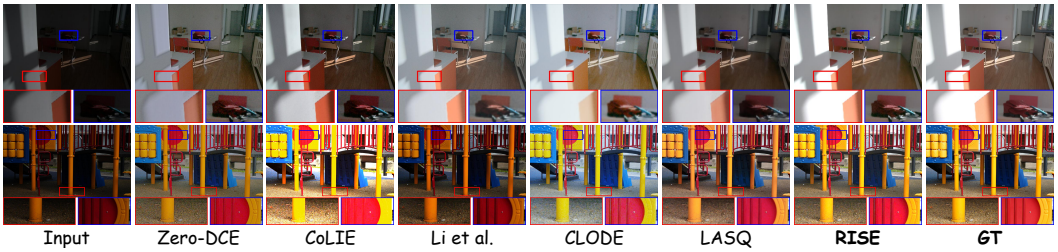

Figure 6: Visual comparisons of the enhanced results by different methods on LOL and UHD-LL.

3.3 Learning Objective

Exposure control loss. Let $\mathbf{G_T}$ 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,

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

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

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

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 train-

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)	2.7	1.90	42.3	2.4	1.78	35.7	2.2	1.57	42.3	2.2	1.71	31.0	2.6	1.67	31.4
SL	CIDNet (Yan et al., 2025)	2.9	2.05	34.0	2.8	2.12	**17.3**	2.8	2.20	**20.2**	2.9	1.87	28.2	2.8	1.85	**13.6**
UL	RUAS (Liu et al., 2021)	2.0	1.77	47.5	2.1	1.85	29.3	2.7	2.11	32.5	1.7	1.42	49.9	2.3	1.66	40.5
UL	Li et al. (Li et al., 2025)	2.7	1.89	31.9	2.4	1.82	18.5	2.5	1.85	23.7	2.9	1.92	28.7	2.6	1.75	20.8
UL	CLODE (Jung et al., 2025)	2.5	1.71	36.8	2.4	1.79	29.9	2.3	1.72	33.8	2.4	1.58	32.7	2.6	1.60	28.5
UL	LASQ (Kong et al., 2026)	3.4	2.26	30.2	3.0	2.18	17.4	3.0	2.23	25.2	3.4	2.20	23.3	3.4	2.02	17.2
	RISE	**3.7**	**2.47**	**30.0**	**3.4**	**2.54**	21.5	**3.5**	**2.44**	25.0	**3.6**	**2.45**	**21.5**	**3.5**	**2.06**	14.3

ing/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 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 official weights trained on their optimal training sets.

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 achieves the highest PSNR among unsupervised methods on all datasets. Its consistent performance without dataset-specific retraining indicates transferable exposure regularity rather than overfitting to particular brightness distributions. We further evaluate generalization on five unpaired benchmarks in Table 3, using the same checkpoint trained on LSRW-Nikon. RISE achieves the highest IQA and IAA scores across all datasets by clear margins, demonstrating perceptually natural enhancement on unseen real-world scenes. Moreover, RISE requires only 0.35M parameters and 1.43 GFLOPs, highlighting its computational efficiency.

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 enhance-

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↓
$v_n\rho_n(p)Z_{k_n}$	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
$\boldsymbol{\beta}^\top\mathbf{c}_{\text{scene}}+b$	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**

ment effects across different scenes, while other methods exhibit evident fixed enhancement styles. For more visual comparisons based on the benchmarks, please refer to supplementary material.

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 and highlight details while preserving faithful colors. More results are provided in supplementary material.

Figure 7: Generalization to real-world scenes captured with an iPhone 16 Pro. Zoom in for the best view.

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 failure cases in the supplementary material. Incorporating generative restoration remains a promising future direction.

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.